A Study of Chemotherapy Resistance in Acute Myeloid Leukemia

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Susan Kay Rathe

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

David A. Largaespada

October, 2013

Acknowledgements

Funding: The leukemia research was funded by The Leukemia & Lymphoma Society (grant 7019-04). The MMuFLR workflow was funded by The Children's Cancer Research Fund.

University of Minnesota Resources: The BioMedical Genomics Center provided services for gene expression microarray, RNA sequencing, oligo preparation, and Sanger sequencing. The Minnesota Supercomputing Institute maintains the Galaxy Software, as well as provides data management services and training. The Masonic Cancer Center Bioinformatics Core provided guidance in the analysis of the RNA-seq data. The FDA Drug Screen was performed at the Institute for Therapeutics Discovery and Development.

Colleagues: First and foremost, I would like to recognize Dr. David Largaespada for his mentorship and the opportunity to continue research he started while a postdoc in Dr. Neal Copeland's lab. I would also like to thank the many members of the Largaespada lab who provided direct or indirect assistance to my research. Dr. Bin Yin continued Dr. Largaespada's work on the BXH-2 AML cell lines by making drug resistant derivatives and provided training in cell culture and drug assay techniques. Dr. Won-Il Kim developed the TNM mouse model and provided training in mouse related experimental techniques. Miechaleen Diers did all of the tail vein injections for the AML transplants and assisted with necropsies. Dr. Branden Moriarity designed the *Dck* knockout TALENs, the CRISPR KOs, and *DCK* overexpression vectors and did the transfections

with the assistance of Dominic Beckmann. Dr. Eric Rahrmann did the Western blots. Dr. Timothy Starr provided training in qPCR and various lab protocols. I would also like to acknowledge the assistance I received from my undergraduate volunteers. Natashay Bailey participated in the research for 3 years, Teng Zhang, Alexandra Ellingston, Christopher Stoltenberg, Ellen Melrose, and Natalie Aumann for 2 years, Chase Liska and Willemijn Veldhuijzen for 1 year, and Natasha Delgado and Nigest Balla for a semester. Their participation is outlined in the following table.

	Teng	Alexandra	Natashay	Christopher	Ellen	Natalie	Chase	Willemijn	Natasha	Nigest
Tissue culture	√	√	√	√	√	√	√	√	√	√
Drug assays	√	√	√	√	√	√	√	√	√	√
Drug screens						√		√		
Eyebleeds			√							
IP injections			√	√						
Necropsies	√	√	√	√	√	√				
Primer design			√	√	√	√				
RT-PCR	√	√	√	√	√	√	√	√		
Gel electrophoresis	√	√	√	√	√	√	√	√		
Gel extractions	√	√	√	√	√	√	√	√		
Sequencing & analysis			√	√	√	√				
qPCR	√	√	√	√	√	√		√		
RNAi Knockdowns				√						
TALEN Knockouts							√			
Overexpression							√			
RNA isolation			√	√	√	√	√			
DNA isolation			√			√	√			
Protein isolation				√						
Bradford assay				√						

I would also like to thank all of the other members of the Largaespada lab both past and present, the members of the MICaB program, my prelim committee members (Dr.

Timothy Hallstrom, Dr. Jaime Modiano, Dr. Xianzheng Zhou, Dr. Yoji Shimizu and Dr. Steve Rice), and my thesis committee (Dr. David Largaespada, Dr. Scott McIvor, Dr. Chris Pennell, and Dr. Nikunj Somia).

Dedication

I dedicate this thesis to my husband, David D. Rathe. I'm finding it difficult to find words to adequately describe the importance of his patience and encouragement during this very demanding process. He took the time to learn and understand many aspects of my research, and took sincere interest in its progress. He provided proof-reading services, and was a sounding board for problem solving. He retired and took on most of the household duties, so when I did get a break from research, I could devote the time to enjoying my family and friends. His support never wavered even when he was called back to work on a critical project for his former company. During the many months he was working over 60 hours a week, he still managed to make me feel my research was still what mattered most. He is an amazing person, and I feel so lucky to have him for a husband and best friend.

Abstract

Acute myeloid leukemia (AML) is the most deadly of the leukemias. Due to its heterogeneous genetic nature it has been difficult to find effective targeted treatments. Standard induction chemotherapy, which includes cytarabine (Ara-C) as its primary component, will in most cases result in remission, but the remission is short-lived and usually results in the presentation of an Ara-C resistant form of AML at relapse. A thorough understanding of how chemotherapy resistance develops in AML would lead to the establishment of drug profiles describing the molecular conditions under which a drug should be considered or rejected as a treatment option. With the ultimate goal of providing patient specific drug treatment options to restore or preserve chemotherapy sensitivity, the research presented here had five objectives: (1) to develop and/or evaluate mouse models of Ara-C resistance *in vitro* and *in vivo* and determine their ability to mimic the Ara-C resistance found in human disease, (2) to discover cellular mechanisms of Ara-C resistance, (3) to find effective drug partners for Ara-C in treating *de novo* AML, (4) to find drugs for treating chemotherapy refractory AML, and (5) to predict drug response based on the molecular profile (gene expression patterns and specific mutations) for each AML patient. Although these objectives are broad and aggressive, this research has resulted in some significant strides towards meeting them. Using gene expression microarray, transcriptome sequencing, and targeted mutagenesis via TAL endonuclease treatment, it was determined mutations in *Dck* were the primary factor in the development of Ara-C resistance in an *in vitro* model of Ara-C resistance. Drug screens were used to determine Ara-C resistant cells became more sensitive to

glucocorticoids, and cladribine is an affective partner to Ara-C in treating *de novo* disease. Also, an *in vivo* model of Ara-C resistance was developed by passaging AML cells through SCID/beige mice and treating the mice with low doses of Ara-C.

Table of Contents

List of Tables

List of Figures

Abbreviations

ACS: American Cancer Society

ALL: Acute lymphocytic leukemia

AML: Acute myeloid leukemia

APL: Acute promyelocytic leukemia

Ara-C: Cytarabine, cytosine arabinoside, arabinosyl cytosine

BAM: Binary alignment/map

cDNA: Complementary DNA

CDS: Coding DNA sequence

CI: Combination index

CLL: Chronic lymphocytic leukemia

CML: Chronic myeloid leukemia

CRISPRs: Clustered regularly interspaced short palindromic repeats

DCK: Deoxycytidine kinase

dCTP: Deoxycytidine triphosphate

DNA: Deoxyribonucleic acid

dNTP: Deoxyribonucleotide triphosphates

dUTP: Deoxyuridine triphosphates

ELN: European LeukemiaNet

FAB: French-American-British

FDA: Food and Drug Administration

FPKM: Fragments per kilobase of transcript per million mapped reads

GR: Glucocorticoid receptor

HiDAC: High dose Ara-C

IC: Inhibitory concentration

IP: Intraperitoneal

IGV: Integrative Genomics Viewer

IPA: Ingenuity Pathway Analysis

ITD: Internal tandem duplication

KD: Knockdown

KO: Knockout

LSC: Leukemic stem cells

MDS: Myelodysplastic syndrome

MOI: Multiplicity of infection

MPD: Myeloproliferative disease

MTS: 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium

MMuFLR: Missense Mutation and Frameshift Location Reporter

NCBI: National Center for Biotechnology Information

NGS: Next generation sequencing

NT: Nucleotide

OE: Overexpression

qPCR: Quantitative polymerase chain reaction

RNA: Ribonucleic acid

RNAi: RNA interference

RNA-seq: RNA sequencing

RT-PCR: Reverse transcriptase polymerase chain reactions

SAM: Sequence alignment/map

SCID: Severe combined immunodeficiency

shRNA: Short hairpin RNA

SNPs: Single-nucleotide polymorphisms

TALEN: Transcription activator-like effector nucleases

TNM: Tet-off, *N-ras*^{G12V}, *Mll-AF9*

TRC: The RNAi Consortium

UTR: Untranslated region

VCF: Variant call format

WBC: White blood cell

WHO: World Health Organization

Chapter 1: The Challenges of Treating Acute Myeloid Leukemia

The development of drug resistance, to the chemotherapy drug cytarabine, is a life-threatening complication which occurs in the treatment of most acute myeloid leukemia patients. This introductory chapter will provide a background on characteristics of acute myeloid leukemia, the use of cytarabine in induction chemotherapy, the methods we are using to study cytarabine resistance, and the primary goals of this research project.

Acute Myeloid Leukemia Background

Hematopoietic stem cells, residing in the bone marrow, give rise to all of the various types of blood cells, via a complex cellular program of cell replication and differentiation. There are two primary branches of hematopoietic differentiation, myeloid and lymphoid, and both pathways are characterized by a series of differentiation steps. Each step of each pathway produces a distinctly different progenitor cell, which is capable of further differentiation or self-renewal. If oncogenic mutations appear in any of these progenitors, it can result in leukemia, which is the unfettered proliferation of immature cells incapable of undergoing normal differentiation (Passegue et al., 2003).

The American Cancer Society (ACS) divides leukemias into five groups based predominantly on the maturity of the cells (acute or chronic) and the primary developmental pathway (myeloid or lymphoid). In general, the term “acute” refers to cells which more closely resemble the hematopoietic stem cells or early progenitors, while the term “chronic” refers to cells which are more differentiated. The main four types of leukemias are acute lymphocytic leukemia (ALL), acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL), and chronic myeloid leukemia (CML). Leukemias not falling into these four categories are grouped into an “other” grouping and they include hairy cell leukemia, chronic myelomonocytic leukemia, and juvenile

myelomonocytic leukemia. Each year the ACS issues a "Cancer Facts and Figures" document describing cancer incidence trends and providing estimates for the current year. The "Cancer Facts and Figures 2013" showed leukemia as the 6th leading cause of cancer death in the United States for both men and women. They estimated 48,650 people will be diagnosed with leukemia in 2013 and 23,720 will die from the disease (**Table 1.1**). Although acute myeloid leukemia makes up only 30% of the newly diagnosed leukemias, it accounts for 43.7% of the leukemia deaths.

The high mortality rate is reflective of the aggressiveness of AML and its heterogeneous nature. Although AML likely originates in the bone marrow, it rapidly expands as blast cells into the peripheral blood, where it outcompetes normal white blood cells, red blood cells and platelets, resulting in immunodeficiency-related symptoms, as well as anemia and thrombocytopenia. Blood tests can detect a rise in white blood cell count, but bone marrow samples are required to accurately determine the subtype of AML. Currently, prognosis and treatment options are dependent on the AML subtype (Estey, 2013).

Classification of AML

The first attempt to define subtypes of AML occurred in 1976, when the French-American-British (FAB) classification system was introduced (Bennett et al., 1976). FAB offered 8 subtypes, M0-M7, and involved the characterization of AML based on microscopic examination to determine how closely the AML cells resembled various myeloid progenitors (**Figure 1.1**) (van Doorn-Khosrovani, 2004).

The FAB system also employed cytogenetics to detect specific genetic alterations, and attempted to group each major chromosomal abnormality (i.e. translocation) into a specific subtype. However, defining a one-to-one correlation between translocations and FAB subtypes was only possible in the case of the t(15:17) translocation PLM-RARα found in acute promyelocytic leukemia (APL), which is classified as FAB subtype M3 (Piazza et al., 2001).

In 2001, the World Health Organization (WHO) proposed a new method of AML classification, based on factors such as genetic alterations, treatment-induced disease, or evolution from other diseases, such as myelodysplastic syndrome (MDS) or myeloproliferative disease (MPD), as outlined in **Table 1.**2 (Vardiman et al., 2002). One of the key distinctions in the WHO classification as compared to the FAB system was the separation of MDS-derived AML from *de novo* AML. MDS, formerly referred to as preleukemia, is exemplified by anemia or cytopenia, with few, if any, blast cells in the peripheral blood, and small percentages of blast cells in the bone marrow. The WHO classification changed the blast cell percentage in the bone marrow from 30% to 20% for establishing the AML threshold. The WHO classification also added a distinct sub-classification in the MDS group for 5q-syndrome, a slow growing dysplasia characterized by a 5q chromosomal deletion.

With the advent of gene expression microarray technology, an attempt was made to create AML subtypes based on gene expression patterns (Valk et al., 2004). Unsupervised clustering of gene expression patterns of 285 AML patient samples identified 16 distinct groupings. Although there was a not a definitive correspondence to the FAB or WHO classification systems, within the 16 clusters there was a cluster for

AMLs containing *CBGβ-MYH11* fusions, a cluster for *PML-RARα* fusions, and a cluster for AMLs containing *AML1-ETO* fusions. FAB subtypes M4 and M5 also fell into a single cluster.

Through the development of mouse models, AML researchers have shown the evolution of AML requires two distinct mutations, one to encourage unchecked proliferation (Class I mutations), and a second to block normal differentiation and/or promote self-renewal (Class II mutations). When the two class hypothesis was originally proposed (Gilliland and Griffin, 2002) the Class I mutations were described as mutations in a single gene, while the Class II mutations were shown as fusion events, such as the ones listed in **Table 1.2**. Examples of Class I mutations were the *FLT3* internal tandem duplication (ITD), and various point mutations in *KIT*, *KRAS*, *NRAS*, and *PTPN11*.

However, nearly half of all human AML samples display a normal karyotype (Dohner and Dohner, 2008). Since the original definition of the two class model, a number of point mutations have been identified as Class II events. These include mutations in several transcription factors: *AML1*, *C/EBPα*, and *MLL*. There are also a number of recently identified recurring mutations which have not been fully characterized, and so have not yet been assigned to either class, such as *DNMT3a* and *NPM1* (Takahashi, 2011).

If a Class I mutation, such as $K\text{-}ras^{G12D}$, is introduced to a transgenic mouse model, the result is often a MPD phenotype (Chan and Gilliland, 2004). If a Class II mutation, such as $AML1^{S291fsX100}$, is introduced to a transgenic mouse, a MDS phenotype is frequently the result (Beachy and Aplan, 2010). Combining Class I and Class II mutations in a transgenic mouse typically result in AML, such as in the mouse model that

combined a *FLT3-LM* mutation with an *AML1-ETO* fusion (Schessl et al., 2005), and in the TNM mouse model that combined a *N-ras* oncogene with an *Mll-AF9* fusion (Kim et al., 2009). Many patient samples display a similar pattern of Class I and Class II mutations in AML (Ishikawa et al., 2009), which makes the mouse models particularly useful tools for studying cellular changes associated with various mutations, and for testing various drugs and drug combinations.

Because of the abundance of AML samples available for study, the relative simplicity of acquiring AML samples from the bone marrow, and the ease at which subsets of cells can be characterized by flow cytometry, many of the advances in the understanding of the role of translocations in cancer and the theory of cancer stem cells has been progressed from the study of acute myeloid leukemia. It has been theorized there is a small population of leukemia initiating cells, commonly referred to as leukemic stem cells (LSC), present in each case of AML. Fractions of human AML cells, separated based on surface markers using flow cytometry, where introduced into severe combined immunodeficient (SCID) mice. It was shown the $CD34^+$ $CD38^-$ population was able to establish disease in the SCID mice, but not the $CD34^+$ $CD38^+$ or $CD34^-$ populations (Lapidot et al., 1994). Many attempts have been made to further characterize these LSCs, but it has proved as difficult as classifying AMLs into subtypes (Horton and Huntly, 2012). One popular theory maintains specific drugs and drug combinations can be developed which will prove toxic to LSCs, but not to normal hematopoietic cells (Jordan, 2007; Ravandi and Estrov, 2006).

Prognosis of AML

AML prognosis is dependent upon a number of factors including age, medical history, AML subtype, and cytogenetic/molecular features. In 2010, the European LeukemiaNet (ELN) proposed prognostic groups based on cytogenetic and molecular abnormalities for AML, excluding APL (**Table 1.3**) (Dohner et al., 2010).

In 2011, a German research group used the ELN system to characterize the outcome of 1557 patients undergoing standard leukemia treatments between 1996 and 2005 (Rollig et al., 2011). In the 847 patients falling into the 18-60 age group, the overall survival and probability of survival was distinctly different for each of the ELN prognostic groups. The probability for relapse within 5 years for the adverse group was near 80% while the probability of relapse for the favorable group was just below 40%. The Intermediate-II group had more favorable outcomes than the Intermediate-I group, with probabilities of relapse at approximately 75% and 50%, respectively.

The Valk et al. study, which identified 16 subtypes of AML based on the clustering gene expression, was followed by a plethora of studies with the goal of developing prognostic gene expression patterns. The most recent attempt resulted in the identification of a set of 24 genes which can be used in predicting survival (Li et al., 2013). Two sets of patients, 277 patients from the Netherlands and 548 patients from Germany, were statistically separated into two groups based on the expression levels of the 24 genes. The 5 year survival rate was twice as high in one group versus the other; ~50% vs. ~20% in the Netherlands group and ~40% vs. ~20% in the Germany group. The set of genes were composed of *ALS2CR8, ANGEL1, ARL6IP4, BSPRY, C1RL, CPT1A, DAPK1, ETFB, FGFR1, HEATR6, LAPTM4B, MAP7, NDFIP1, PBX3,*

PLA2G4A, *PLOD3*, *PTP4A3*, *SLC25A12*, *SLC2A5*, *TMEM149*, *TRIM44*, *TRPS1*, and *VAV3*.

Despite the progress being made in this area, prognostic factors are inexorably tied to treatment options. Introducing treatment options beyond the current standard of care and based on the molecular characteristics of an individual's AML could radically alter the gene expression versus prognosis correlations as they are currently defined.

Induction Chemotherapy with Cytarabine (Ara-C)

With the exception of APL (subtype M3), which is treated with retinoic acid, clinicians use the ELN risk stratification system to predict patient response and outline an appropriate treatment strategy (Estey, 2013). All patients are initially treated with induction chemotherapy to eliminate blast cells, and the AMLs characterized as "favorable" or "intermediate 1" are followed by consolidation therapy to target the leukemic stem cells. *FLT3-ITD*$^{+}$ patients are also given a FLT3 inhibitor. If the patient is strong enough, and the patient achieves a complete remission following induction therapy, bone marrow transplants are given to the "intermediate 1" group. Patients falling into the "intermediate 2" or "adverse" groups are typically shuttled to clinical trials appropriate to the genetic abnormalities of their AML.

For over 40 years cytarabine, referred to as cytosine arabinoside or Ara-C (Arabinosyl Cytosine), has been the primary component of both standard induction therapy and consolidation therapy. High dose Ara-C combined with an anthracycline (HiDAC 3+7) is given for induction chemotherapy, while lower doses of Ara-C are given for consolidation (Estey, 2013). Most AML patients will initially respond well to

chemotherapy, but then most will relapse with an Ara-C resistant form (Kubal and Lancet, 2013).

Ara-C is a cytidine analog which contains an arabinose sugar in place of the normal ribose sugar found in cytidines. It enters the cell using the same highly conserved dNTP Salvage Pathway used by the cell to recover normal cytidines from its surroundings (**Figure 1.2**) (Toy et al., 2010).

Ara-C is metabolically activated by the addition of three phosphates. The first phosphate is added by deoxycytidine kinase (DCK), which is also the rate limiting enzyme in the dNTP Salvage Pathway. Ara-C is incorporated into the DNA during the S-phase, and interferes with replication by inhibiting polymerase and topoisomerase activity. As an independent molecule, Ara-C is quickly deaminated into a nontoxic uracil derivative and has a half-life of less than an hour, but once incorporated into DNA it can still be detected after 24 hours (Grem et al., 1995). As a result, Ara-C is more toxic to fast growing cancer cells than to slow growing cells, and to be an effective treatment for AML, it must be given intravenously in a continuous manner over a period of 7 days.

Ara-C has been found to be particularly effective when combined with an anthracycline. However, this course of treatment is harsh and is not always a viable option for older patients with other medical conditions. In the hopes of finding alternative treatment options, hundreds of clinical trials are underway, using a wide variety of new drugs and drug combinations (clinicaltrials.gov).

Ara-C Resistance Research

In one of the earliest studies of Ara-C resistance, it was noted the presence of Ara-C resistance was more prevalent in ALL cells if the patient had been treated with Ara-C

prior to the acquisition of the cell sample (Preisler et al., 1984). The first study finding a connection between Ara-C resistance and loss of deoxycytidine kinase (*Dck*) function was published in 1985 (Young et al., 1985). In 1995, transfection of *Dck* cDNA into Ara-C resistant rat leukemic cells with loss of *Dck* function, restored Ara-C sensitivity (Stegmann et al., 1995). Mutations in *Dck* and *Slc29a1*, both components of the dNTP Salvage Pathway, were identified in Ara-C resistant human leukemic cell lines (Cai et al., 2008). Interestingly, the loss of *Dck* function is the only suggested mechanism of Ara-C resistance that has appeared in more than one research study, suggesting it may the most prominent mechanism of Ara-C resistance. Despite this emphasis on *Dck*, there has been no published studies describing how cells adapt to *Dck* loss, what drugs can be used to treat AML experiencing functional loss of *Dck,* or any clinical studies which include an evaluation of *Dck* functionality.

Models of Ara-C Resistance in AML

Although there have been a number of publications describing the analysis of Ara-C resistant human and mouse AML cell lines (Abe et al., 2006; Cai et al., 2008; Jin et al., 2009; Negoro et al., 2011; Veuger et al., 2000; Yin et al., 2006a), there are none describing the generation of an *in vivo* mouse model of Ara-C resistance. The Abe et al. study established Ara-C resistant K562 human AML cell lines, the Cai et al. study used Ara-C resistant human CCRF-CEM cell lines, the Jin et al. study created Ara-C resistant HF6 murine AML and K562 human AML cell lines, and the Negoro et al. study established 5 Ara-C resistant human cell lines (from HL-60, K562, CEM, THP1, and U937 cells). The research conducted for this thesis utilizes an *in vitro* model of Ara-C

resistance utilizing 2 Ara-C resistant murine BXH-2 cell lines established in the Yin et al. study, and involves the generation of an *in vivo* model of Ara-C resistance.

In vitro model of Ara-C resistance

In 1969, twelve BXH strains were developed at The Jackson Laboratory by Dr. Benjamin Taylor from crosses between C57BL/6J and C3H/HeJ strains. It was quickly discovered that the BXH-2 strain spontaneously developed AML, due to the presence of a retrovirus (Bedigian et al., 1981). AML cells from two BXH-2 mice, B117 and B140, were used to create two AML cell lines, B117P and B140P, respectively (Largaespada et al., 1995). *Nf1* transcription in the B117P cells was disrupted by proviral insertions in the *Evi-2* locus. This mimics human disease in that loss of *NF1* has been shown to be a driver of human AML in complex karyotypes (Mrozek, 2008). With the goal of creating a model of Ara-C resistant refractory AML, these two cells lines were subjected to increasing concentrations of Ara-C until two highly resistant cell lines resulted, B117H and B140H, which tolerated Ara-C concentrations 500-1000 times that of the parental lines (Yin et al., 2006a). All four cell lines (B117P, B117H, B140P, and B140H) have been characterized by flow cytometry of phenotypic surface markers (unpublished data). Interestingly, the B117 cells display phenotypic markers consistent with stem cells or early progenitors, such as CD34 and c-kit, while the B140 cells do not.

In vivo model of Ara-C resistance

Vav-tTA; TRE-*NRAS*^{G12V}; *Mll-AF9* (TNM) transgenic mice were generated to study the nature of oncogene addiction (Kim et al., 2009). This model is representative of a combination of a Class I mutation, *NRAS*^{G12V}, and a Class II mutation, *Mll-AF9*. A

bank of viable cells was made from spleen cells harvested from these mice. This is a Tet-off model, in which the oncogenic *NRAS* can be turned off by presence of doxycycline. Doxycycline can be given continuously to the mice by adding it to their drinking water, or can be given in pulses by IP injection. If doxycycline is given as an IP injection with subsequent oral delivery in drinking water, transcription of the *NRAS* oncogene is terminated within 6 hours of IP injection, while the NRAS protein persists in the cells for about 48 hours (Sachs et al, submitted). This model has a number of advantages. It is based on the most common *MLL* fusion found in human AMLs. It engrafts readily and the disease progresses rapidly in SCID/beige mice with full-blown AML (WBC > 150 K/ml) occurring at approximately 16 days. Depending on the type of experiment being conducted, the *NRAS* oncogene can be left active, or deactivated by treating the mice with doxycycline. Deactivation of the *NRAS* oncogene by treatment with doxycycline results in a nearly complete remission. If the doxycycline is subsequently withheld, the *NRAS* oncogene will be expressed again and the AML reestablished.

Methods for Evaluating Gene Expression

The turn of the current century brought about the introduction of gene expression microarrays to evaluate the levels of expression of RNA in cells. This provided a powerful tool for comparing Ara-C resistant cells to Ara-C sensitive cells. Prior to the introduction of microarrays, RNA expression was evaluated a gene at a time. Microarrays are small, thumb-nail size chips with organized sections of embedded oligonucleotides, typically 25-mers, whose sequences correspond to a specific RNA sequence. The processing of the microarray involves the isolation of RNA, the generation of a cDNA copy of the RNA, fragmentation of the cDNA, addition of

fluorescent labels to the cDNA fragments, and exposure of the labeled fragments to the microarray chip. The fragments preferentially bind to complementary sequences on the chip, and the intensity of the fluorescence is measured for each section of the chip. Software tools are used to normalize the fluorescent signals across each individual chip and to normal expression levels between chips. The final product is a list of genes and their relative expression levels for each of the RNA samples being evaluated (Harrington et al., 2000).

Although gene expression microarrays were a significant advancement, they still had limitations. They were only able to evaluate a predefined set of genes, it was not possible to design probes to distinguish between genes with similar sequences, and there was an inherent level of background signal, which could distort the signal levels, especially for genes expressed at low levels.

The year 2010 brought about the rapid deployment of high-throughput deep-sequencing devices. The University of Minnesota ran its first set of RNA-sequencing samples in November of 2010. The RNA-seq processing also starts with the isolation of RNA and the generation of a cDNA copy, followed by fragmentation. The goal of the fragmentation process is to create fragments of similar length. The fragments are run out on a gel and the location with highest level of RNA is cut out and isolated from the gel. Adaptors are added to one or both ends of the fragments, depending on whether single reads or paired-end reads are being performed. The fragments are then sequenced to a predefined length by the deep-sequencing device (Wang et al., 2009).

Theoretically, RNA-sequencing should provide the opportunity for superior gene expression analysis from that of its predecessor, gene expression microarrays, due to its

unbiased approach for selecting and evaluating cDNA fragments. In addition, RNA-sequencing has the potential to evaluate expression levels at an isoform level, detect rare and novel transcripts and splice variants, distinguish between similar variants, and identify the existence of various types of mutations and fusions. However, the analysis of the RNA-seq data is only as good as the software tools being used, and the software tools are still in their infancy. As the software tools evolve, the true power of RNA-sequencing will be realized. An in-depth comparison of the microarray and RNA-seq approaches for gene expression analysis is provided in the **Discussion** section of **Chapter 3**. A description of the analysis of RNA-sequencing data is included in **Appendix D**.

Methods for Knocking Down and Knocking Out Gene Expression

Finding candidate genes, by either microarray or RNA-seq, is only the first step in determining the gene changes involved in drug resistance. The candidate genes need to be tested to confirm their involvement. Two methods were used in this research: RNA interference to knockdown gene expression, and Transcription Activator-Like Effector Nucleases (TALENs) to achieve total knockout of genes.

Since the initial discovery of the process of RNA interference (RNAi) in *C. elegans* in 1998 (Fire et al., 1998), a plethora of techniques have been developed to take advantage of introducing small fragments of RNA designed to interfere with a specific gene's expression. Unfortunately, the effectiveness of these techniques is typically dependent of the quantity of vectors transfected into the cells. This is a problem for difficult-to-transfect cells, such as AML. Partial knockdown can be accomplished in AML cells and may suggest a trend in the change of drug response, but the results may not be statistically significant.

Techniques, which have the ability to specifically and completely knockout a gene by damaging all copies of the gene at the DNA level, such as TALENs, are far superior to RNAi, since many drug resistant phenotypes are associated with specific gene mutations that completely knockout gene function. The TAL system was first discovered in *Xanthomonas* bacteria, and consists of the FokI endonuclease and TAL effector proteins that guide FokI to a specific DNA sequence, where FokI executes a non-specific double-strand break (Cermak et al., 2011). Double-strand break repair mechanisms repair the break, but usually with the loss of one or more nucleotides. The TAL effector proteins contain a tandem repeat sequence. Toward the middle of each repeat sequence there are two amino acids that specificly bind to a single nucleotide. Researchers take advantage of this naturally occurring mechanism by engineering a pair of TAL effector proteins targeting two nucleotide (NT) sequences about 15 NTs in length and approximately 15 NTs apart, on opposite sides of the DNA strand. If the TAL effector is designed to target the first translated exon of the target gene, there is a high probability it will cause a deletions in all DNA copies of the gene, resulting in complete loss of gene function.

Pre-clinical Drug Assays for Testing Anti-Leukemia Drugs and Drug Combinations

MTS-tetrazolium based assays were optimized for this research to provide a consistent method of determining inhibitory concentrations and combination indices. Inhibitory concentrations can be established describing the concentration of drug needed to negatively influence the expansion of live cells at a specific percentage. The inhibitory concentration at 50% (IC_{50}) is a standard measure of toxicity of a drug in a specific cell line. Although the MTS assay can be used to measure inhibition of cell expansion, it

provides no indication of whether the inhibition was due to a drop in proliferation or an increase in cell death.

The combination index (CI), invented by Chou and Talalay in 1984, is used to measure the level of synergy or antagonism exhibited by a combination of two or more drugs in a specific cell line (Chou and Talalay, 1984). When combining two drugs at a constant ratio, the formula for calculating the CI_{50} is $CI_{A+B} = (D_{A|A+B}/D_A) + (D_{B|A+B}/D_B)$, where D is the drug concentration at IC_{50}, and A and B represent each of the two drugs. The nomenclature of the $D_{A|A+B}/D_A$ calculation represents the concentration of drug A in the combination assay (A+B) at the IC_{50} point divided by the IC_{50} concentration of drug A when given alone. If the drug combination is synergistic the CI will be < 1.0, if perfectly additive the CI = 1.0, and if antagonistic the CI is > 1.0.

The exact protocol for the MTS-tetrazolium assay is provided in the **Appendix A: Methods and Materials**. It is a highly reproducible colorimetric assay performed in 96-well plates. Its sensitivity and specificity has been compared favorably to other toxicological techniques for determining IC_{50} (Malich et al., 1997). A general layout for the plates is provided in **Figure 1.3**.

The drug concentrations used in each drug assay are specifically selected to optimize the number of data points within the range of 0.05-0.95 for the fraction of cells affected. After performing regression analysis of the data points, the IC_{50} is determined by locating the drug concentration associated with the point where half the cells are affected. A generic example of results from an MTS-Tetrazolium assay is provided in **Figure 1.4**.

The drug assays are conducted at least 3 times for each drug and cell line combination. For the results to be deemed acceptable for this research, there must be at least 5 data points between 0.05 and 0.95 fraction affected, and the r value for the regression analysis must be greater than 0.88.

Research Objectives

There have been over 14,000 research papers published on AML (www.ncbi.nlm.nih.gov/pubmed/) and there are over 1,600 active or completed clinical trials (clinicaltrials.gov). Despite the significant focus on AML and potential treatment options, there remains an incomplete understanding of what molecular factors determine how well patients will initially respond to induction chemotherapy, and what causes Ara-C resistance to develop.

The research conducted for this thesis had a number of objectives. The first goal was to determine the genetic alterations responsible the Ara-C resistance in the B117H and B140H cell lines described above. This was accomplished using two distinct technologies, microarray gene expression analysis (**Chapter 2**), and RNA-sequencing (**Chapter 3**). The second goal was to identify drugs which are most appropriate for treating Ara-C resistant disease, and to identify drugs synergistic with Ara-C to treat *de novo* AML (**Chapter 4**). The third goal was to establish an *in vivo* model of Ara-C resistance in mice, to mimic refractory disease for testing new drug combinations (**Chapter 5**). A summary of the results, along with conclusions, and subsequent research opportunities are also described in **Chapter 5**.

Table 1.1: Estimated New Cases of Leukemia and Deaths in 2013 (ACS)

Projections of new leukemia cases and deaths for 2013 from the American Cancer Society's Cancer Facts & Figures 2013.

	2013 - Estimated			
	New cases		Deaths	
Type of Leukemia	**Number**	**% of total**	**Number**	**% of total**
ALL - Acute lymphocytic	6,070	12.5	1,430	6.0
AML - Acute myeloid	14,590	30.0	10,370	43.7
CLL - Chronic lymphocytic	15,680	32.3	4,580	19.3
CML - Chronic myeloid	5,920	12.2	610	2.6
Other	6,350	13.1	6,730	28.4
Total	**48,610**		**23,720**	

Table 1.2: WHO Classification of Acute Myeloid Leukemia

The World Health Organization created categories of AML based on genetic alterations and therapeutic response (Vardiman et al., 2002).

Category	Includes
de novo AML with translocations	*AML1-ETO*, *CBFβ-MYH11*, *PML-RARα*, *MLL* fusions
Multilineage dysplasias	Progression from MDS/MPS, 2+ myeloid lineages
Therapy related	Alkylating agents, radiation, topoisomerase II inhibitors
AML not categorized above	

Table 1.3: ELN Proposed Prognostic Groups for AML

ELN prognostic groups for AML are based on both cytogenetic and molecular abnormalities. Adapted from Table 4 (Dohner et al., 2010).

Prognosis	Includes
Favorable	*AML1-ETO*, *CBFβ-MYH11*, Solo *NPM1* or *CEBPA* mutations
Intermediate I	Normal karyotypes, mutated *NPM1* with *FLT3-ITD*
Intermediate II	*MLLT3-MLL*, abnormal karyotypes not in other groups
Adverse	*RPN1-EV1*, *DEK-NUP214*, *MLL* fusions, del(5q), complex karyotypes (3 or more chromosomal abnormalities)

Figure 1.1: FAB Classification of Acute Myeloid Leukemia

FAB subtypes (M0-M7) for AML are determined based on how closely the cancer cells resemble various myeloid progenitors.

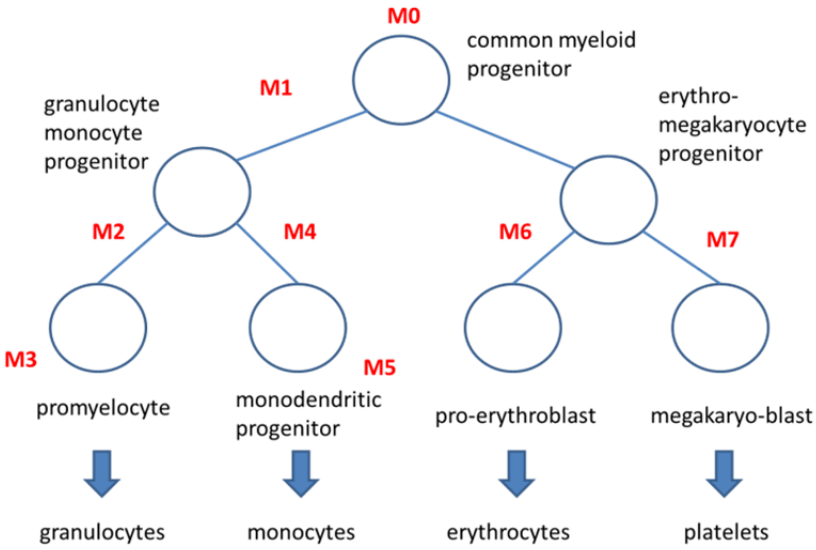

M0
common myeloid progenitor
M1
granulocyte monocyte progenitor
erythro-megakaryocyte progenitor
M2
M4
M6
M7
M3
M5
promyelocyte
monodendritic progenitor
pro-erythroblast
megakaryo-blast
granulocytes
monocytes
erythrocytes
platelets

Figure 1.2: Ara-C Is Metabolized by the dNTP Salvage Pathway

Ara-C is processed in the same manner as salvaged cytidines. It enters the cell via the ENT1 transporter (Slc29a1), and is triply phosphated before being incorporated into DNA where it interferes with DNA replication.

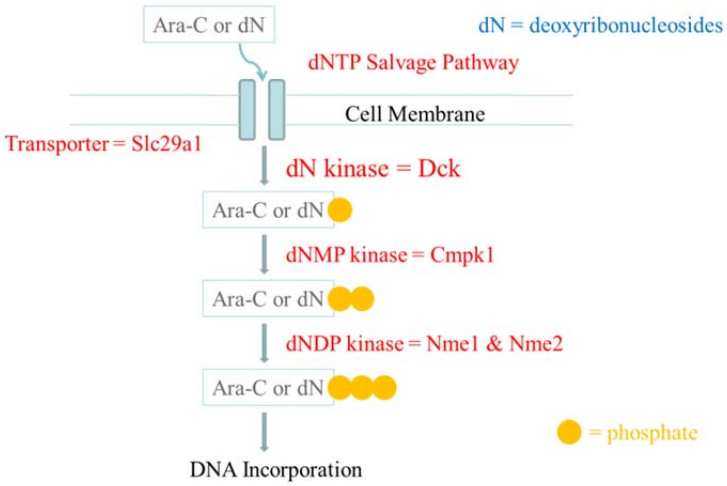

Ara-C or dN
dN = deoxyribonucleosides
dNTP Salvage Pathway
Cell Membrane
Transporter = Slc29a1
dN kinase = Dck
Ara-C or dN
dNMP kinase = Cmpk1
Ara-C or dN
dNDP kinase = Nme1 & Nme2
Ara-C or dN
= phosphate
DNA Incorporation

Figure 1.3: Plate Layout for the MTS-Tetrazolium Drug Assay

Plate layout for the MTS-Tetrazolium drug assay includes each cell sample and drug combination being done in quadruplicate, plus 10 separate drug concentrations being tested.

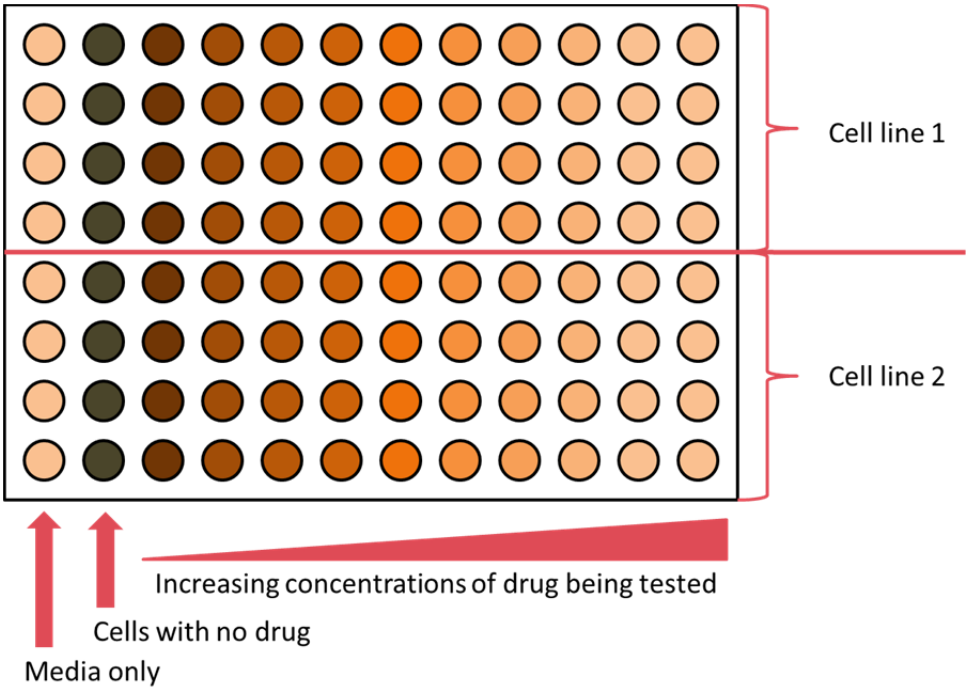
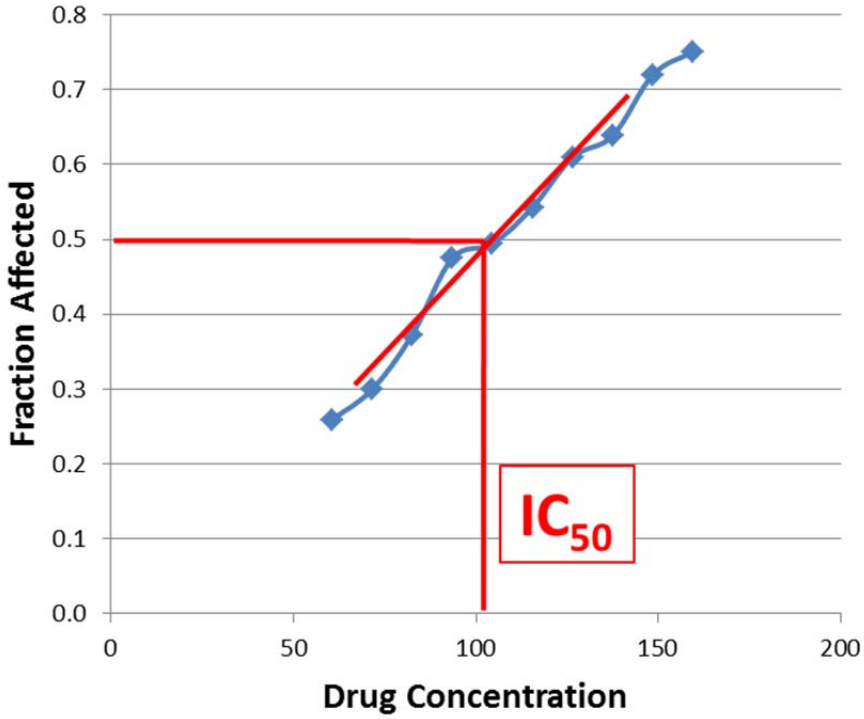

Cell line 1
Cell line 2
Increasing concentrations of drug being tested
Cells with no drug
Media only

Figure 1.4: Determining the IC_{50} from the MTS-Tetrazolium Assay Results

Regression analysis is used to determine a trend line. The IC_{50} concentration is the location on the trend line where 50% of the cells are affected.


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Fraction Affected
0
50
100
150
200
Drug Concentration
IC50

Chapter 2: Deoxycytidine Kinase Is Down-regulated in Ara-C Resistant Acute Myeloid Leukemia Murine Cell Lines

[Portions of this chapter were published in *Leukemia* (Rathe and Largaespada, 2010). *Leukemia* does not require a release form for including this material in a thesis by the primary author.]

Susan K. Rathe[1,2], David A. Largaespada PhD[1,2]

[1]Masonic Cancer Center, University of Minnesota, Minneapolis, MN, USA; [2]Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, MN, USA

Contributors

Teng Zhang and Alexandra Ellingston assisted with drug assays and qPCR experiments.

Acute myeloid leukemia (AML) is a malignant expansion of immature cells in the bone marrow. Although chemotherapeutic treatments containing cytosine arabinoside (Ara-C) typically result in dramatic remissions, resistance to chemotherapy frequently occurs. To achieve a better understanding of how Ara-C resistance develops, two highly Ara-C resistant BXH-2 derived AML cell lines (B117H and B140H) were compared to their respective parental cell lines (B117P and B140P) using gene expression microarray. The microarray identified deoxycytidine kinase (*Dck*) as the gene with the most significant change in expression levels. PCR identified the presence of an aberrant form of *Dck* in the B117H cells. Evaluation of several other chemotherapy drugs showed that the Ara-C resistant cells were also highly resistant to decitabine. The results of the drug assays are consistent with the microarray results, since both Ara-C and decitabine rely on Dck for metabolic activation.

Introduction

Patients presenting with Ara-C resistant AML have a reduction in the active metabolized form of Ara-C (Ara-CTP), but the mechanisms responsible for this reduction are still being investigated (Chou et al., 1977). In previously published work two AML cell lines (B117P and B140P), derived from BXH-2 strain mouse leukemias, were subjected to increasing concentrations of Ara-C to create two highly resistant derivatives (B117H and B140H) that tolerate Ara-C concentrations ~800 times that of their parental lines (Yin et al., 2006a). In this study, gene expression microarrays were conducted to determine the expression changes common to both sets of cell lines. The results were verified by qPCR and the specificity of drug resistance was evaluated.

Materials and Methods

Procedures for performing cell culture, RNA isolation, gene expression microarray and analysis, quantitative PCR, and drug assays for determining IC_{50} are all described in **Appendix A**. Primers are provided in **Supplementary Table A.1**.

Results

Gene expression microarrays identify genes that may be associated with Ara-C resistance

Microarray gene expression analysis to compare differences between the parental cell lines (B117P and B140P) and their highly resistant derivatives (B117H and B140H) was performed on 12 RNA isolations (3 isolations for each cell line from different passages) using Affymetrix Mouse Genome 430 2.0 Arrays containing 45,101 murine oligonucleotide sequences. The detection values were processed using Genedata Expressionist software (Genedata AG, Basel, Switzerland) using the robust multiarray average (RMA) normalization technique. The raw data are available online at http://www.ncbi.nlm.nih.gov/geo/ (accession number GSE18322). Group t-test analysis was used to compare the B117P cells to the B117H cells, and the B140P cells to the B140H cells. Only expression changes greater than 2.0 fold with p-values < 0.01 were considered. One hundred and twenty differentially expressed probe sets representing 104 genes were obtained when comparing B140H to B140P. Two hundred and eighty six differentially expressed probe sets representing 239 genes were obtained when comparing B117H to B117P. In **Table 2.1** those genes with consistent expression changes (>2.0 fold) upon selection for Ara-C resistance in both cell lines are listed.

The most significant changes occurred in the expression of *Dck*. However, there

were three Affymetrix probes designed to detect *Dck* levels and only two showed significant changes. The average detection levels for each of the *Dck* probes are presented in **Table 2.2**.

It is also possible the Ara-C resistance was caused by different mechanistic changes to each of the Ara-C resistant cell lines. The number of genes with greater than 2-fold of differentially expressed genes, when comparing each Ara-C resistant cell line to its parental cell line, is in the hundreds. **Table 2.3** and **Table 2.4** list a subset of differentially expressed genes with fold changes greater than 10 fold when comparing B117H to B117P, and B140H to B140P, respectively.

Downregulation of *Dck* is found in both Ara-C resistant cell lines

According to the microarray gene expression data the deoxycytidine kinase (*Dck*) gene mRNA level was altered to a greater extent than any other single gene. Starting levels of *Dck* expression were similar in both the B117P and B140P parental cell lines. However, the *Dck* levels in the B117H cells were near zero, while the *Dck* levels in B140H cells were decreased 8.0 fold compared to B140P. Affymetrix had three probes designed to detect *Dck*. The probe sets showing a significant decrease in *Dck* expression targeted the 3'-UTR of *Dck*. Expression changes for the *Dck* gene were confirmed by quantitative PCR (qPCR). PCR primers were designed to target both the translated portion (primer set 1) and the 3' end of the *Dck* transcript (primer set 2). *Rps9,* a gene that is expressed at moderate and consistent levels across all four cell lines, was used to normalize the results. The qPCR analysis using primer set 2 confirmed the altered expression levels found at the 3' end in the microarray (**Figure 2.1**). The qPCR using primer set 1 showed an 8 fold decrease in transcript levels in both sets of cells, while the

microarray probe 1428838_a_at showed no significant change in expression levels between all four cell lines.

Abnormal *Dck* transcript found in B117H cells

qPCR could find no evidence of expression toward the 3' end of the *Dck* transcript in the B117H cell line, indicating the presence of an abnormal *Dck* transcript. To determine if the *Dck* transcript was unusual, four sets of primers were designed to detect 800 bp overlapping sections of *Dck* (**Figure 2.2**). The amplified bands were sequenced. In addition to amplifying a segment of *Dck*, the Q3 primers also amplified a 700 bp segment of *Prpf38b*. Only the Q1 primers amplified a segment of *Dck* in the B117H cells.

Ara-C resistant cells are also resistant to decitabine

Drug assays were conducted on all four cell lines to determine if Ara-C resistance conferred drug resistance to other chemotherapy drugs (**Table 2.4**). Resistance to Ara-C did not confer resistance to other drugs, expect in the case of decitabine. There was a significant change in the response to decitabine in both sets of cells, with a greater than 900 fold increase in IC_{50} values in Ara-C resistant derivatives compared to parental cell lines.

If the cause of the Ara-C resistance is related to downregulation of *Dck*, it would seem probable the Ara-C resistance cells would be more sensitive to CPEC, an inhibitor of the *de novo* pathway for generating cytidines. However, there is only a slight drop in the IC_{50} for CPEC in the Ara-C resistant cells when compared to the parentals,

Phenotypic differences between the B117P and B140P cell lines

The parental cell lines (B117P and B140P) are markedly different phenotypically and yet their response to Ara-C is about the same with similar 50% inhibitory concentrations (IC_{50}). B117P grows 2.3 times faster than B140P (data not shown), and they have differences in gene expression patterns, as detected by microarray analysis. Although CD34 has not been found to be associated with any particular AML subtype or to be prognostically significant, B117P and B117H highly express *Cd34*, while B140P and B140H do not (a 183-fold difference). Furthermore, the level of myeloperoxidase (*Mpo*) is 34-fold greater in the B117P cells when compared to the B140P cells. In addition, the B117P expressed significantly higher levels of both *Gp49a* (27.6 fold) and *Laptm4b* (8.7 fold) than the B140P cells. The presence of *Gp49a* and *Laptm4b* are indicators of "stem-ness" in hematopoietic cells, while their absence is associated with more mature hematopoietic cells (Lee et al., 2010). Despite these differences, both Ara-C resistant cell lines showed a significant decrease in the gene expression of *Dck* upon selection for Ara-C resistance.

Discussion

The parental cell lines used in this study (B117P and B140P) were markedly different in both phenotypic markers and in response to various drugs. Despite these differences, both sets of cells have similar sensitivity to Ara-C, and both were readily transformed into Ara-C resistant cell lines (Yin et al., 2006a). RNA was isolated from all four cell lines from three separate passages and the results analyzed by gene expression microarray to search for common changes associated with Ara-C resistance.

The Affymetrix microarray chip used in this study had three probes designed to detect expression levels of *Dck*. The 1428838_a_at probe targeted the translated portion of the transcript while the other two probes targeted the 3'-UTR. The microarray analysis identified the two most significant expression changes between the Ara-C resistant cell lines and their parental lines, which were in the 3'-UTR of *Dck*. qPCR confirmed the *Dck* expression level changes in the 3'-UTR, but the qPCR of the area targeted by the 1428838_a_at detected 8-fold decreases in that area as well. The signal strength for the 1428838_a_at probe was significantly less than that observed in the other two probes, and the signal was comparable across all 4 cell lines, indicating there may be other probes attracting the *Dck* fragments. This is always a possibility when working with microarray technology. However, the downregulation of *Dck* in the Ara-C resistant cell lines was verified by qPCR and an aberrant form of *Dck* was detected in the B117H cells by RT-PCR.

Dck encodes deoxycytidine kinase, which phosphorylates deoxyribonucleosides including deoxycytidine, deoxyguanosine and deoxyadenosine. DCK performs the rate-limiting step in Ara-C metabolic activation, which is monophosphosphorylation after import of Ara-C through the SLC29A1 nucleoside transporter (Grant, 1998). Thus, down-regulation of this kinase would be expected to lead to resistance to Ara-C. Indeed, polymorphisms in the human *DCK* gene have been associated with patient's response to AML chemotherapies containing Ara-C (Shi et al., 2004), and expression of alternatively spliced isoforms of *DCK* are found in some Ara-C resistant AML patients (Veuger et al., 2000). A recent paper described alterations in *DCK* expression in a human AML cell line selected for resistance to Ara-C *in vitro* (Song et al., 2009). These investigators also

noted that primary human AML samples differ in the level of *DCK* transcript, and that these differences may be related to Ara-C sensitivity, at least as measured in an *in vitro* assay.

Both Ara-C and decitabine are deoxycytidine analogs, and are tri-phosphorylated by the same kinases (including *Dck*). However, their methods of interfering with cell proliferation are entirely different. Once triphosphorylated, Ara-C is incorporated into DNA during replication, blocking polymerase and topoisomerase activity, whereas decitabine inhibits DNA methyltransferase activity. Since the change in resistance to both Ara-C and decitabine are comparable, this supports the hypothesis that the Ara-C drug resistance of the B117H and B140H cells is driven specifically by the down-regulation of *Dck*.

If the lack of *Dck* is the primary change associated with the Ara-C resistance, it would be logical to assume that in order for the Ara-C resistant cells to survive, the cells would become dependent on the *de novo* pathway of deoxycytidine triphosphate (dCTP) generation rather than the savage pathway. In the *de novo* pathway, dCTPs are created from deoxyuridine triphosphates (dUTPs) by CTP synthetase (CTPS). CTPS is ubiquitously expressed and its role is to keep dUTPs and dCTPs in balance, which is accomplished by negative feedback in which dCTP inhibits CTPS (Chang and Carman, 2008). The drug cyclopentenyl cytosine (CPEC) inhibits CTPS (Kang et al., 1989), so if the Ara-C resistant cells lack the ability to process dCTPs via the dNTP salvage pathway, it seems likely the cells are relying on the *de novo* pathway for dCTPs during replication. If so, there should be a dramatic decrease in the IC_{50} for CPEC in the Ara-C resistant cell lines. However, there was only a slight decrease in the CPEC IC_{50}, which can be

interpreted in a number of ways: 1) Dck is not the only kinase capable of monophosphorylating deoxyribonucleotides in the dNTP salvage pathway, 2) CPEC is not completely effective at blocking CTPS activity, 3) there is another method of converting dUTPs to dCTPs other than CTPS, or 4) there is a third unknown method of providing dCTPs to the cells other than the dNTP salvage pathway or the *de novo* pathway.

Regardless, there are other changes taking place in the cells that may play a supporting role in the development of drug resistance. For example, the urea transporter *Slc14a1* is up-regulated in both sets of Ara-C resistant cells. This suggests that the Ara-C resistant cells are producing more urea than their Ara-C sensitive parental cells. Down-regulation of *Dck* would also interfere with the cells ability to get deoxycytidines via the salvage pathway. So, one possible explanation for this increase in urea is that the Ara-C resistant cells are deriving the building blocks for *de novo* synthesis of cytosine from the degradation of thymidine. Urea is a byproduct of the thymidine degradation process.

Manipulation of the Wnt pathway may also be involved in Ara-C resistance. *Dab2*, an adaptor protein that stabilizes Axin, and thus blocks Wnt signaling (Jiang et al., 2008), is down-regulated in both Ara-C resistant cell lines, suggesting an increase of Wnt/β-catenin signaling may be present in the Ara-C resistant derivatives.

Another gene of particular interest is *Ifitm1*, which was significantly down-regulated in both Ara-C resistant cell lines. *Ifitm1*, also called *Leu13*, is an interferon induced antiproliferative transmembrane protein. Lower levels of *IFITM1* expression have been shown to correlate with decreased survival in chronic myeloid leukemia (Akyerli et al., 2005). A decrease in *IFITM1* expression in esophageal squamous cell

carcinoma has been shown to correlate with increased resistance to cis-platinum (Fumoto et al., 2008).

Since the gene expression patterns of B117P and B140P cells are so different, analysis of gene expression changes for each cell line upon selection for Ara-C may provide clues for identifying additional drug resistance genes (**Tables 2.2 and 2.3**). For example, in the B140H cells there was a 36-fold down-regulation of *Spint2*, which has been characterized as a tumor suppressor gene in medulloblastoma (Kongkham et al., 2008). B117H cells have upregulated the neuregulin-4 (*Nrg4*) gene transcript 10-fold. Nrg4 can bind to and activate epidermal growth factor receptor (EGFR)-family receptors and so their activation could contribute to the Ara-C resistant phenotype in this cell line perhaps by activation of the Ras/MAPK pathway which has been shown to modulate Ara-C responsiveness (Yin et al., 2006b). *S100a10* gene expression is dramatically down-regulated in B140H cells and the product of this gene interacts with Annexin-11, which has been described as an anti-apopototic protein involved in chemotherapy resistance (Chuthapisith et al., 2009).

Further studies are underway to determine the mechanisms behind the reduction in *Dck* expression and to identify whether the expression changes of other genes identified by the microarray analysis are required by the cells to compensate for loss of *Dck* expression or participate in the development of Ara-C resistance by other mechanisms. Drugs targeting the pathways that can cooperate with *Dck* downregulation to allow Ara-C resistance may be useful for preventing the selection of Ara-C resistant clones in patients.

Table 2.1: Common gene expression changes between Ara-C resistant cell lines and their Ara-C sensitive parental lines (> 2.0 fold).

Gene expression microarray analysis comparing B117H to B117P and B140H to B140P. Genes with greater than 2-fold in both comparisons are listed. Microarray was done on RNA isolated from 3 separate passages of cells.

Probe ID	Gene ID	Description	Fold Change	
			B117H vs. B117P	B140H vs. B140P
1439012_a_at	*Dck*	deoxycytidine kinase	-273.157	-8.264
1449176_a_at	*Dck*	deoxycytidine kinase	-92.503	-7.488
1415673_at	*Psph*	phosphoserine phosphatase	-4.571	-2.606
1424254_at	*Ifitm1*	interferon induced transmembrane protein 1	-4.456	-2.788
1420498_a_at	*Dab2*	disabled homolog 2 (Drosophila)	-4.338	-3.251
1423805_at	*Dab2*	disabled homolog 2 (Drosophila)	-3.632	-2.108
1429693_at	*Dab2*	disabled homolog 2 (Drosophila)	-3.566	-2.480
1455991_at	*Ccbl2*	cysteine conjugate-beta lyase 2	-2.176	-3.308
1433521_at	*Ankrd13c*	ankyrin repeat domain 13C	2.055	2.021
1428114_at	*Slc14a1*	solute carrier family 14 (urea transporter), member 1	2.692	2.135
1433939_at	*Aff3*	AF4/FMR2 family, member 3	2.737	2.180

Table 2.2: Average probe signal strength for each of the *Dck* probes.

Average of the signal strength for each of the Affymetrix probes for *Dck* in the 4 BXH-2 cell lines.

Affy Probe ID	B117P	B117H	B140P	B140H
1428838_a_at	245.8	224.7	206.8	78.9
1439012_a_at	2793.6	10.2	2464.9	298.3
1449176_a_at	1288.1	13.8	1067.9	141.2

Table 2.3: 10-fold gene expression changes comparing Ara-C resistant B117H cells to B117P parental cells.

Gene expression microarray analysis comparing B117H to B117P. Genes with greater than 10-fold are listed. Microarray was done on RNA isolated from 3 separate passages of cells.

Probe ID	Gene ID	Description	Fold Change B117H vs. B117P
1439012_a_at	*Dck*	deoxycytidine kinase	-273.157
1426438_at	*Ddx3y*	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked	-112.117
1449176_a_at	*Dck*	deoxycytidine kinase	-92.503
1452077_at	*Ddx3y*	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked	-30.279
1426598_at	*Uty*	Ubiquitously transcribed tetratricopeptide repeat gene, Y chromosome	-15.103
1419905_s_at	*Hpgd*	hydroxyprostaglandin dehydrogenase 15 (NAD)	-13.998
1421408_at	*Igsf6*	immunoglobulin superfamily, member 6	-13.870
1421571_a_at	*LOC100045833, Ly6c1, Ly6c2*	lymphocyte antigen 6 complex, locus C	-13.366
1433930_at	*Hpse*	heparanase	-11.545
1426439_at	*Ddx3y*	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked	-11.326
1433944_at	*Hectd2*	HECT domain containing 2	10.212
1421681_at	*Nrg4*	neuregulin 4	10.217
1459860_x_at	*Trim2*	tripartite motif-containing 2	12.334
1446594_at	*Gm4371*	eukaryotic translation initiation factor 3, subunit I pseudogene	13.897
1439036_a_at	*Atp1b1*	ATPase, Na+/K+ transporting, beta 1 polypeptide	13.990
1420822_s_at	*Sgpp1*	sphingosine-1-phosphate phosphatase 1	15.374
1420821_at	*Sgpp1*	sphingosine-1-phosphate phosphatase 1	17.392
1424292_at	*Depdc1a*	DEP domain containing 1a	31.617
1426236_a_at	*Glul*	RIKEN cDNA 5830403L16 gene	43.289

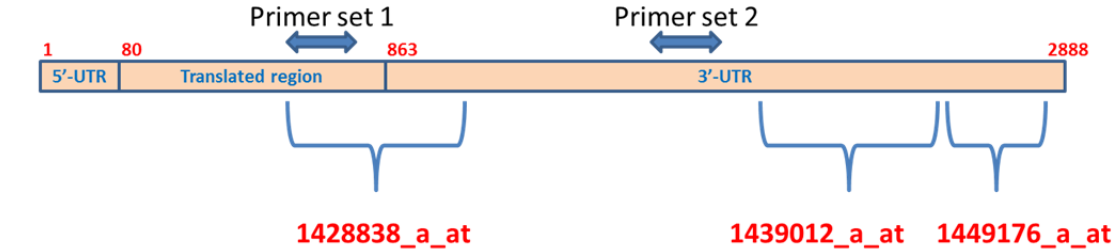
Table 2.4: 10-fold gene expression changes comparing Ara-C resistant B140H cells to B140P parental cells.

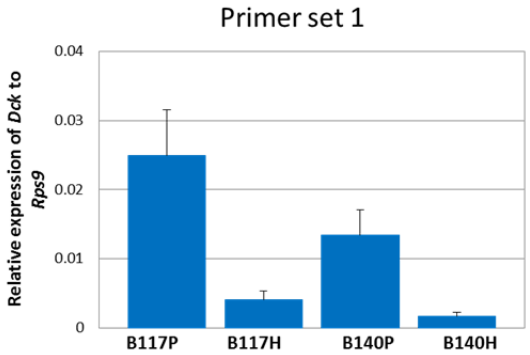
Gene expression microarray analysis comparing B140H to B140P. Genes with greater than 10-fold are listed. Microarray was done on RNA isolated from 3 separate passages of cells.

Probe ID	Gene ID	Description	Fold Change B140H vs. B140P
1456642_x_at	*S100a10*	S100 calcium binding protein A10 (calpactin)	-95.306
1416762_at	*S100a10*	S100 calcium binding protein A10 (calpactin)	-44.896
1438968_x_at	*Spint2*	serine protease inhibitor, Kunitz type 2	-36.941
1426980_s_at	*E130012A19Rik*	RIKEN cDNA E130012A19 gene	-13.315
1422776_at	*Serpinb8*	serine (or cysteine) proteinase inhibitor, clade B, member 8	-11.758
1455865_at	*Insm1*	insulinoma-associated 1	-10.926
1434353_at	*Sfmbt2*	Scm-like with four mbt domains 2	-10.866
1426523_a_at	*Gnpda2*	glucosamine-6-phosphate deaminase 2	-10.369

Figure 2.1: qPCR Measurement of *Dck* Expression in Ara-C resistant cell lines and their parental lines.

qPCR was repeated 3 times for each cell line. *Rps9* was used to normalize the data. Primers are described in **Supplementary Table C.1**.


Primer set 1
Primer set 2
1
80
863
2888
5'-UTR
Translated region
3'-UTR
1428838_a_at
1439012_a_at
1449176_a_at


Primer set 1
Relative expression of Dck to Rps9
0.04
0.03
0.02
0.01
0
B117P
B117H
B140P
B140H

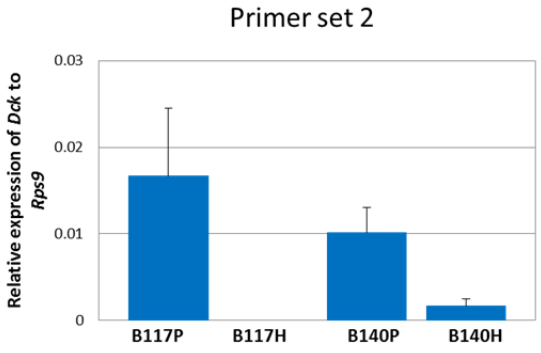
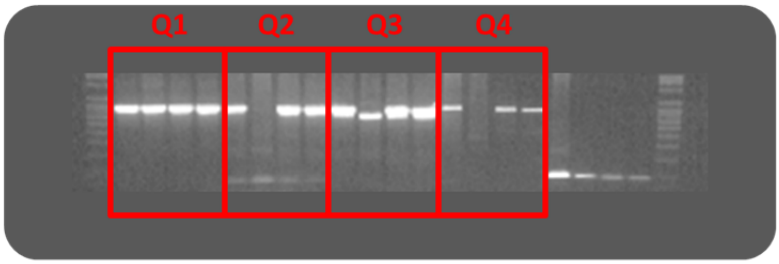

Primer set 2
Relative expression of Dck to Rps9
0.03
0.02
0.01
0
B117P
B117H
B140P
B140H

Figure 2.2: B117H cells lack normal transcripts in the area of quartiles 2, 3, and 4.

RT-PCR was performed using primers designed to amplify 800bp overlapping segments of the Dck transcript. *Rps9* was used as a loading control.



Description of PCR products of *Dck* by Quartile (number of bases)

Quartile	B117P	B117H	B140P	B140H
1	800	800	800	800
2	800	None	800	800
3	700, 800	700	700, 800	700, 800
4	800	None	800	800

Table 2.5: IC50 Values for Other Chemotherapy Drugs

MTS-Tetrozolium Assays (as described in Appendix A) were repeated at least 3 times for each drug and cell line. (Except for the fluoruracil in B117H and nelarabine in B140P, which were only repeated twice).

50% Inhibitory Concentration (IC50) in µM

	B117P		B117H		Fold	B140P		B140H		Fold
Drug Tested	**Average**	**Std. Dev.**	**Average**	**Std. Dev.**	**Change**	**Average**	**Std. Dev.**	**Average**	**Std. Dev.**	**Change**
CPEC	0.132	0.026	0.080	0.015	-1.651	0.064	0.010	0.027	0.004	-2.345
Daunorubicin	0.256	0.043	0.236	0.009	-1.085	0.057	0.022	0.074	0.016	1.302
Decitibine	0.318	0.131	301.666	54.875	948.427	0.100	0.016	110.182	18.309	1099.564
Didox	65.499	4.848	47.845	5.826	-1.369	48.100	12.407	32.254	32.254	-1.491
Etoposide	0.560	0.399	0.526	0.221	-1.064	0.245	0.096	0.272	0.059	1.113
Fluorouracil	1.201	0.079	2.266	0.193	1.887	0.847	0.259	0.487	0.054	-1.741
MK457	0.950	0.158	1.015	0.088	1.068	0.242	0.013	0.254	0.040	1.050
Nelarabine	16.809	3.303	22.097	5.735	1.315	27.289	0.440	26.401	2.795	-1.034
PD-0325901	0.013	0.004	0.010	0.004	-1.312	0.113	0.046	0.032	0.008	-3.491
PI-103	2.262	0.696	1.043	0.295	-2.168	0.721	0.271	0.630	0.149	-1.143

Chapter 3: Combining the Power of Transcriptome Sequencing and Transcription Activator-Like Effector Nucleases to Identify Mechanisms of Drug Resistance in Acute Myeloid Leukemia Murine Cells

[Manuscript is being prepared for submission.]

Susan K. Rathe[1,6], Branden S. Moriarity PhD[1,2,3], Christopher B. Stoltenberg[1], Natalie K. Aumann, Eric P. Rahrmann PhD[1,2,3], Natashay J. Bailey[1], Ellen G. Melrose[1], Dominic A. Beckmann[2,3], Chase R. Liska[1], David A. Largaespada PhD[1,2,3,4,5]

[1]Masonic Cancer Center, University of Minnesota, Minneapolis, MN, USA; [2]Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, MN, USA; [3]Center for Genome Engineering, University of Minnesota, Minneapolis, MN, USA; [4]Brain Tumor Program, University of Minnesota, Minneapolis, MN, USA; [5]Department of Pediatrics, University of Minnesota, Minneapolis, MN, USA

Contributors

Branden S. Moriarity designed and constructed the TALENs for knocking out *Dck* and the vectors for overexpressing *DCK*. He also conducted the transfections of the TALENs and overexpression vectors into the cells. He created **Figure 2.3c** and **Figure 3.4c**. He wrote the "TALEN assembly and generation of KO cells" and "Inducible *DCK* overexpression vector" sections of the **Methods and Materials** in **Appendix A**.

Christopher B. Stoltenberg assisted in the RNAi knockdown experiments and associated drug assays. He assisted with the creation of **Figure 3.3a**. He wrote the "RNAi experiments" section of the **Methods and Materials** in **Appendix A**.

Natalie K. Aumann performed the DNA isolations of the KO cells. She assisted in the design of primers and the analysis of sequencing data to verify the KOs. She also assisted in drug assays. She wrote the "Cell culture" section of the **Methods and Materials** in **Appendix A**.

Eric P. Rahrmann performed the Westerns blots. He created **Figures 3.1e**, **3.3d**, and **3.4d**. He wrote the "Western blot analysis" section of the **Methods and Materials** in **Appendix A**.

Natashay J. Bailey performed the DNA isolations of the BXH-2 cells and assisted in the analysis of the sequencing data to detect the *Dck* mutations in the B117H and B140H cells.

Ellen G. Melrose assisted in the qPCR efforts and wrote the "quantitative PCR" section of the **Methods and Materials** in **Appendix A**.

Dominic A. Beckmann designed the primers to verify the TALEN-based KOs and assisted in the DNA isolation and sequencing analysis of the KOs.

Chase R. Liska assisted in the generation of single cell clones from the TALEN-based *Dck* KOs and the *DCK* OE cells, as well as assisted in the drugs assays to measure the Ara-C IC_{50} in the KO cells and OE cells. He also assisted in the creation of **Figures 3.4a & b**.

The technical evolution from gene expression microarrays to transcriptome deep-sequencing (RNA-seq) and from RNA interference to highly specific gene knockouts using Transcription Activator-Like Effector Nucleases (TALENs) has provided a new experimental partnership for identifying and quantifying the effects of gene changes on drug resistance. In a previously published microarray analysis we identified the downregulation of *Dck* as the primary common change in two Ara-C resistant murine acute myeloid leukemia cell lines (B117H and B140H), when compared to their Ara-C sensitive parental cell lines (B117P and B140P). In this publication we describe the results of deep-sequencing (RNA-seq) of the same RNA samples, and present a Transcription Activator-Like Effector Nuclease (TALEN) based method for accomplishing complete *Dck* gene knockout (KO) in AML cells. Our analysis of the RNA-seq data led to the identification of protein modifying loss-of-function mutations in the *Dck* transcripts in both Ara-C resistant cell lines. TALEN-based KO of *Dck* in the B117P cell lines dramatically increased the IC_{50} of Ara-C near those displayed by the Ara-C resistant B117H cells. Introduction of a *DCK* overexpression vector in the *Dck* KO clones resulted in a significant, but incomplete, restoration of Ara-C sensitivity. The RNA-seq analysis also identified a number of significant expression changes that did not appear in the microarray analysis, and identified other mutations, which may contribute to Ara-C resistance. This effort demonstrates the power of using RNA-seq transcriptome analysis and TALEN-based KOs to identify and verify genes associated with drug resistance.

Introduction

The 12,000+ patients diagnosed with acute myeloid leukemia (AML) in the United States each year face a dismal prognosis. The induction chemotherapy, which will most likely result in a remission, is typically not curative. However, induction chemotherapy can significantly reduce blast cells providing the clinician with additional time to try other therapies. Unfortunately, the additional therapies are generally not effective at achieving a long-term durable remission. At relapse, most patients will no longer respond to induction therapy, since the leukemic clones surviving the initial onslaught of induction chemotherapy have an innate resistance, and have therefore become the prevalent disease cells (Estey, 2012).

Arabinoside cytarabine (Ara-C) has been the primary component of induction chemotherapy for over 40 years. Ara-C, a cytidine analog, enters the cell via the dNTP salvage pathway, where it is metabolically activated by the addition of three phosphates in the same manner as cytidines. Each phosphate is added by a different kinase. The first kinase in the dNTP salvage pathway is deoxycytidine kinase (DCK), the rate limiting enzyme in the metabolic activation of Ara-C. Numerous studies have shown *DCK* expression is frequently downregulated in cells that are unresponsive to Ara-C (Cai et al., 2008; Rathe and Largaespada, 2010; Shi et al., 2004; Song et al., 2009; Veuger et al., 2000; Veuger et al., 2002).

We theorize there are other pathways whose regulation is altered in order to compensate for DCK loss. If identified, those pathways could be potential targets for drugs, to be used in combination with Ara-C to treat AML. In a previous publication, we reported the results of a microarray gene expression analysis, which compared gene

expression of two Ara-C resistant cell lines (B117H and B140H) with their respective Ara-C sensitive parental cells lines (B117P and B140P) (Rathe and Largaespada, 2010). The B117H and B140H cells tolerated concentrations of Ara-C 500-1000 times that of their parental lines (Yin et al., 2006a). The most dramatic common change identified by the microarray study was the significant downregulation of *Dck* (Rathe and Largaespada, 2010).

Here we report the results of a subsequent RNA sequencing of the transcriptome of the same four murine AML cell lines (B117P, B117H, B140P, and B140H). RNA-seq analysis uncovered evidence to the nature of the *Dck* functional impairment in both the B117H cells and the B140H cells: a large deletion of DNA spanning the splice acceptor of the last exon of *Dck* and a frameshift mutation in the fourth exon of *Dck*, respectively. Both mutations resulted in aberrant RNA transcripts for *Dck*. RNA-seq also identified gene expression changes not previously detected by gene expression microarray.

Total KO of *Dck* using Transcription Activator-Like Effector Nucleases (TALENs) in the B117P cells confirmed the loss of *Dck* expression was nearly sufficient for the high Ara-C IC_{50} levels found in the Ara-C resistant cell lines. Introduction of an inducible *DCK* overexpression vector in the B117P *Dck* KO clones restored most of the original Ara-C sensitivity.

This research demonstrates the value of using RNA-seq methods to identify changes in cells as they become resistant to drugs and a new method for generating candidate drug resistant gene KOs in difficult-to-transfect AML cells using TALENs with single step drug selection. Further study of the cells containing the TALEN-based

knock-outs of Dck may identify upregulated genes and/or pathways compensating for the loss of Dck, and thus lead to the discovery of effective companion drugs to Ara-C..

Materials and methods

Cell culture, transcriptome deep sequencing and analysis, DNA and RNA isolation and sequencing, genomic DNA PCR, quantitative PCR, RNAi experiments, TALEN assembly and generation of KO cells, inducible DCK overexpression vector, Western blot analysis, drug assays are all described in **Appendix A**.

Raw data files and processed expression files are available online in the Gene Expression Omnibus at http://www.ncbi.nlm.nih.gov/geo/ (accession number GSE47454).

Results

RNA-sequencing identifies more gene expression changes than microarray hybridization

Samples of RNA had previously been isolated from 2 murine BXH-2 AML cell lines and their Ara-C resistant derivatives, and then evaluated by microarray.(Rathe and Largaespada, 2010) In order to find gene expression changes associated with Ara-C resistance not previously implicated by the microarray analyses, aliquots of RNA from the microarray experiment were submitted for RNA-sequencing (RNA-seq). TopHat was used to map the data to the mouse transcriptome (NCBI37/mm9), and the quality of the mapping was tested using Picard-tools. All samples had over 20 million paired reads with over 90% mapped and over 89% uniquely mapped (**Supplementary Table C.2**). Cuffdiff (Roberts et al., 2011a; Roberts et al., 2011b; Trapnell et al., 2010) was used to determine changes common to both Ara-C resistant cell lines (B117H and B140H) when

compared to their parental lines. To avoid division by zero, a minimum FPKM was established at 0.001 based on FPKM distribution patterns (**Supplementary Figure C.1**). These patterns also showed genes expressed in just one sample, a phenomenon not seen when studying microarray expression data due to the presence of background noise. Genes where both the parental and its Ara-C resistant derivative had FPKM levels less than 0.5 were excluded from the analysis, since even technical replicates display a high degree of variability at these lower expression levels (Mortazavi et al., 2008). Integrated Genomic Viewer (IGV; http://www.broadinstitute.org/igv) was then used to eliminate false positives, which included distortions due to reads mapping outside the normal transcription area, a high abundance of non-unique reads, and projected non-protein coding RNA sequences.

The previous microarray analysis identified 8 genes with expression levels with 2X or more fold changes. In comparison the RNA-seq method identified 60 genes. Seven genes appeared in both lists (**Figure 3.1a**). Genes identified by RNA-seq with a 3X or more fold change in both sets of cells (B117H vs. B117P and B140H vs. B140P) are listed in **Table 3.1**, while the greater than 2-fold and less than 3-fold change genes are included in **Supplementary Table C.3**. The only gene identified by microarray and not by RNA-seq was *Psph*, where the expression did not meet the 2-fold threshold in the RNA-seq analysis. The RNA-seq list includes *Dck*, the only gene appearing in the microarray data as being changed by more than 5-fold. The expression levels of *Dck* were verified by qPCR (**Figure 3.1b**). Of the 53 genes identified by RNA-seq but not by microarray, 3 genes did not have a probe designed on the microarray chip (*2310007A19Rik*, *2210417A02Rik*, and *AI427809*), 1 gene had expression levels so high

it probably saturated the microarray chips (*Mpo*), 10 genes had probes lacking specificity, 22 had expression levels too low for the microarray chips to distinguish significant differences (including *Dab2ip*), and 15 just missed the 2 fold cutoff in the microarray analysis, most likely due to the distortive effect of background noise. As for the last 2 genes, microarray was unable to distinguish between *Ly6c2* and *Ly6c1* expression levels due to sequence similarities between these two genes, or between *Gng5* and its family members.

We next looked at changes unique to either the B117H cells or the B140H cells when compared to their individual parental cell lines, B117P and B140P, respectively. These lists were significantly longer than those generated by microarray, so we confined our analysis to the gene expression changes of 100-fold or more (**Supplementary Tables C.4 and C.5**). Again due to the distortion caused by background noise only one gene had a 100X change within the microarray data *(Ddx3y* in the B117 cells).

RNA-sequencing identifies mutations in Dck in the B117H cells and B140H cells

Missense Mutation and Frameshift Location Reporter (MMuFLR) (Rathe et al., 2013), a Galaxy (Blankenberg et al., 2010; Giardine et al., 2005; Goecks et al., 2010) based workflow developed to look for frameshift and missense mutations, was used to identify mutations in the Ara-C resistant cell lines that were not present in their respective parental cell lines. MMuFLR identified a single thymidine insertion in exon 4 of *Dck* following the 462nd nt from the translational start site, which would result in a severely truncated protein. Sanger sequencing showed the insertion was present in all expressed *Dck* transcripts (**Figure 3.2b**) and homozygous in the genomic DNA (**Supplementary Figure C.2**). The nearly complete elimination of Dck protein was verified by Western

blot (**Figure 3.1f**). The genomic insertion would result in a severely truncated protein (**Supplementary Figure C.3**).

To confirm the expression levels of *Dck* identified by Cuffdiff, and to look for any sequence anomalies within the *Dck* transcript, IGV was used to visualize the TopHat (Langmead et al., 2009; Trapnell et al., 2009) generated mapped reads (**Figure 3.1c**). Transcripts were also assembled independently of a reference genome and then mapped back to the reference genome, and visualized using IGV (**Figure 3.1d**). The IGV views of the TopHat and Cufflinks processes elucidated the changes that took place in *Dck*.

In the B117H cells, IGV clearly showed run-through transcription into intron 6, a loss of transcription in all but a small section of the 3'-UTR and a continuation of transcription beyond the 3'-UTR. The run-through transcription into intron 6 suggested a deletion of the splice acceptor site of exon 7. Cufflinks (Roberts et al., 2011a; Roberts et al., 2011b; Trapnell et al., 2010) was used to generate transcripts by evaluating overlapping reads, but without the benefit of a reference genome. When the results were mapped back to the mm9 reference genome, the aberrant nature of the *Dck* transcript in B117H was again apparent (**Figure 3.1d**). A long template PCR of genomic DNA indicated there was a deletion of approximately 1 kb in the *Dck* locus in the B117H cells (**Supplementary Figure C.4**). This deletion was verified by amplifying segments of DNA and then sequencing the amplified segments. The actual deletion was determined to be 878 bases (**Figure 3.2a**). The deletion started 750 bases before the start of exon 7, and ended 128 bases into exon 7. The loss of the splice acceptor for exon 7 resulted in the splicing machinery selecting alternative splice sites. Sequencing of the RNA transcript verified the mapping of *Dck* done by IGV. The transcription proceeded into

intron 6 up to the deletion, continued beyond the deletion for 49 bases into exon 7, skipped 190 bases, transcribed an additional 207 bases, and then skipped another 2825 bases, where it picked up transcription again, well beyond the 3'-UTR. The protein generated from this aberrant transcript should have resulted in the translation of a protein with 20 amino acids generated from the start of the intron 6 region, rather than the 8 amino acids that would have been translated from exon 7 in a normal transcript. Dck proteins form homodimers, and although there are no specific functional domains within the C-terminus of Dck, it is highly conversed across a broad spectrum of species indicating its importance in Dck function (**Supplementary Table C.6).** Western blots (WB) of Dck in the B117P and B117H cell lines showed a Dck protein was being generated in the B117H cells (**Figure 1e**), but the WB technique was not sensitive enough to detect a size change.

Mutation analyses tools identify other mutations acquired in Ara-C resistant cells

In addition to the frameshift mutation identified in *Dck* of B140H, another frameshift was identified by MMuFLR as being introduced in *Ccdc88b* of B140H (**Supplementary Table C.7**). The *Ccdc88b* frameshift was shown to be heterozygous by Sanger sequencing. No frameshifts were detected in the B117H cells that were not also present in the B117P cells (**Supplementary Table C.8**).

MMuFLR also identified 21 mutations introduced into either the B117H cells or the B140H cells (**Table 3.2**). The potential for functional changes to comparable proteins in human cells was examined using both PolyPhen-2 (Adzhubei et al., 2010) and PROVEAN Protein (Choi et al., 2012). PolyPhen-2 identified the mutations in KDM5C, PRKACB, VPS33B, PBRM1, and ZNF668 as "probably damaging", while PROVEAN

Protein identified the mutations in PRKACB, PUS7L, RASGRP2, ATPBD4, and SMARCA4 as "deleterious".

deFuse (McPherson et al., 2011) was used to look for the introduction of any protein modifying fusions in the Ara-C resistant cell lines, when compared to their respective parental lines. No protein modifying fusions were found (data not shown).

Partial suppression of *Dck* using RNAi results in an increase of the IC_{50} for Ara-C

To test whether the downregulation of *Dck* alone can change a cell's response to Ara-C, knockdowns of *Dck* were performed in the parental cell lines, B117P and B140P, using OpenBiosystems shRNA constructs. Two TRC constructs for *Dck* were used, one targeted exon 6 (KD1) and the other targeted a sequence spanning exon 2 and 3 (KD2). The *Dck* knockdowns were verified by qPCR. Drug assays were used to determine the Ara-C IC_{50} in the knockdown cell lines. The Ara-C IC_{50}'s were higher in the cell lines with the greater downregulation of *Dck* (**Figure 3.3a**). As controls, knockdowns of *Nfkb1* and *p53* were also performed on the B117P cells. No change in IC_{50} for Ara-C was observed (data not shown).

Total KO of *Dck* using TALENs results in a significant increase of the IC_{50} for Ara-C

TALENs were used to knock-out (KO) *Dck* function by performing a targeted deletion approximately 50 nt from the translational start site (**Figure 3.2c**). Single cell clones were grown out and tested for Ara-C sensitivity. Ara-C IC_{50}'s in the *Dck* knockouts were comparable to the Ara-C IC_{50}'s in the Ara-C resistant cell lines (**Figure 3.3a**). The location of the deletion in the DNA of each of the B117P KO clones (T2A, T6B, and T11A) was determined by PCR amplification of the TALEN target area and

Sanger sequencing (**Figure 3.3b**). RT-PCR was performed on RNA-derived cDNA to look for transcript changes within the first 3 exons of *Dck*, and resulted in multiple light bands for the KO clones (**Figure 3.3c**). The top two bands of the T6B clone were sequenced. The top band revealed an alternatively spliced version of *Dck* (**Supplementary Figure C.5**), and the second band was an off target amplification of another gene. The absence of Dck protein in the KO clones was verified by Western blot (**Figure 3.3d**).

Rescue of Dck expression in Dck KO clones results in a decrease of the IC_{50} for Ara-C

A doxycycline inducible human *DCK* overexpression vector (**Figure 3.4c**) was stably integrated into to the three B117P Dck KO cell lines (T2A, T6B, and T11A) using the piggyback transposon system. The control of *DCK* expression was confirmed by qPCR (**Figure 3.4a**). In the absence of doxycycline, the cells exhibited an Ara-C IC_{50} slightly lower than the Dck KO cells (**Figure 3.4b**). Inducing DCK with doxycycline resulted in a significant reduction in the Ara-C IC_{50}. Gene expression levels were measured by qPCR and the presence of DCK protein was confirmed by Western Blot (**Figure 3.4d**). However, the DCK protein levels were much lower than the Dck levels found in the B117P cells.

Discussion

The B117H and B140H cells used in this study are highly resistant to Ara-C, tolerating concentrations of Ara-C 500-1000 times greater than the parental cell lines from which they were derived (Yin et al., 2006a). We theorized this dramatic change in drug response would allow us to focus on the most prominent changes in the cells. RNA

samples, previously analyzed using gene expression microarray technology, were examined using the Illumina HiSeq 2000 RNA-sequencing platform. Numerous software tools were used to evaluate the resulting RNA-seq data. TopHat was used to map the RNA-seq data to the mouse genome, while Cuffdiff was used to measure gene expression levels. Comparing the microarray results to RNA-seq was problematic and revealed many advantages of RNA-seq. Microarray results have an inherent background signal level, while RNA-seq does not have any technically generated background levels. To avoid division by zero in the RNA-seq data, we elected to set a minimum level of 0.001 FPKM for any genes with an FPKM less than 0.001. Less than 0.01% of the genes had expression levels greater than zero and less than 0.001. Due to the absence of background signal, RNA-seq was able to identify significant changes in genes expressed at much lower levels than could be detected by microarray. Examples of such genes were *Dab2ip* and *Hectd2*, which were downregulated and upregulated, respectively, in the Ara-C resistant cell lines. In contrast to microarray probes, which lack the ability to distinguish between genes with similar sequences, RNA-seq (through the detection of single nucleotide differences) was able to uniquely assign reads to the genes with similar sequences, as with the case of *Ly6c1* and *Ly6c2*. Furthermore, RNA-seq's unbiased approach to expression analysis has the potential to identify expression changes in genes not represented on microarray chips, and to detect expression levels at an mRNA isoform level.

RNA-sequencing has at least one other critical advantage over microarray analysis. It has the potential to identify RNA variants, such as unusual transcripts, fusions, frameshifts, and missense mutations. The RNA-seq results were instrumental in

identifying the unusual changes to the *Dck* locus in both the B117H and B140H cell lines. In the B117H cells, microarray data had previously shown part of the 3'-UTR was missing, but only in the areas where the microarray probes were designed to detect. However, the IGV visualization of the RNA-seq data specifically showed expression in the first part of intron 6, and a small section of expression in the 3'-UTR, as well as a large transcribed section beyond the exon encoding the 3'-UTR. The sequence of amino acids at the C-terminus of Dck forms an alpha-helix, which is highly conserved across various species. The end of the C-terminus is in close proximity to a number of residues (Ile24, Ala119, and Pro122) important for Dck kinase activity (Kocabas et al., 2008). The replacement of 8 amino acids by 20 amino acids in the case of the aberrant *Dck* transcript in B117H may result either in instability of the resultant structure of Dck or interference with residues important to Dck's function. The frameshift mutation identified by MMuFLR in exon 4 of *Dck* in the B140H cells would result in a severely truncated version of the Dck protein.

Although RNA-seq is clearly a technological advancement from microarrays, it is not without problems or limitations. For example, the sequencing process has difficulty determining the correct number of nucleotides when reading through a poly-A or poly-T sequence, which can lead to the identification of small indels that do not really exist, an artifact referred to as "stuttering" (Fazekas et al., 2010). MMuFLR has parameters that can be set to ignore this type of error. On the analysis side, software tools to interpret the RNA-seq data are still in their early developmental stages, as are the tables used to characterize the data, such as the tables identifying SNPs and isoforms. Normalizing RNA-seq data between samples is also providing a challenge, and many efforts are

underway to improve normalization techniques (Bullard et al., 2010; Roberts et al., 2011b). Using the FPKM normalization technique provided by Cuffdiff was adequate for comparing drug resistance derivatives to parental cell lines, as in this study, where the samples were all prepared at the same time using the same technique (Roberts et al., 2011b). Although there is no agreement on which approach is more accurate for measuring differential expression changes (microarray, RNA-seq, or qPCR), we did find RNA-seq, based on the quality measurement of the reads, perfectly and uniquely mapped reads to genes that could not be verified by qPCR due to the inability to create primers specific to the gene in question.

It was not surprising to find *Dck* was the primary common change to both Ara-C resistant cell lines and the level of Dck expression correlated to the Ara-C IC_{50} level, since Dck is the rate limiting enzyme in the dNTP salvage pathway, which is required for the metabolic activation of Ara-C. An alternatively spliced version of *DCK* was also found in a study of Ara-C resistant human acute lymphoblastic cell lines (Cai et al., 2008), and downregulation of *DCK* was discovered in Ara-C resistant human acute myeloid cell lines (Song et al., 2009). In the clinical setting, the presence of *DCK* SNPs and alternatively spliced versions of *DCK* have been correlated to patient response to chemotherapy (Shi et al., 2004; Veuger et al., 2000; Veuger et al., 2002). Since *Dck* was both downregulated and mutated in the Ara-C resistant B117H and B140H cell lines, we suspect during the process of creating the Ara-C resistant cell lines, by adding increasing concentrations of Ara-C, the cells initially responded by downregulating expression of *Dck*. Eventually the Ara-C concentrations became so high only the cells with defective *Dck* were able to survive.

Although changes in *DCK* expression have been associated with Ara-C resistance, the importance of *DCK* regulation has not been quantified. The significance of *Dck* regulation alone to Ara-C resistance was demonstrated by the knockdown and KO of *Dck*. The use of new and powerful TALEN-based techniques to specifically and completely KO *Dck* in the B117P cells proved this single modification can account for over 85% of the Ara-C resistance present in the B117H cells. Rescue of *DCK* expression significantly reduced Ara-C resistance. The failure of the *DCK* overexpression to return the B117P Dck KO cells to its original Ara-C IC_{50} level may be due to the use of a human version of *DCK*, which may vary from mouse *Dck* in its kinetics or activation/repression methods. As evidence to this, the DCK protein levels in the dox-induced *DCK* overexpression cell lines were significantly less than the Dck levels found in the B117P cells. The ability to reintroduce DCK to the cells to restore most of the Ara-C sensitivity also indicated most of the changes, which took place when Dck was lost, were not permanent. However, it is possible other gene alterations or gene expression changes partially contributed to the residual Ara-C resistance in these cell lines.

Other potential gene regulation changes were identified, which may contribute directly to Ara-C drug resistance or may be compensating changes associated with the loss of *Dck*. For example, there was significant and consistent downregulation of two tumor suppressor genes, *Dab2* and *Dab2ip*, both of which are negative regulators of Ras. Missense mutations were also identified in 3 genes found within the Ras/MAPK pathway: *Prkacb*, *Rasgrp2*, and *Map3k11*. One study showed introduction of a Ras oncogene resulted in a loss of DCK activity and Ara-C resistance in Rat-1a fibroblasts and human mammary HBL100 cells (Riva et al., 1995). Their results, indicating

overactivation of Ras impedes Dck activity, and our results, indicating loss of *Dck* downregulates Ras repressors, suggests a balance and/or interdependency between these two pathways.

This study illustrates normal Dck functionality is critical for Ara-C responsiveness in murine AML cell lines. These cell lines provide a model for understanding the clinical response to Ara-C and the development of Ara-C resistant AML. It demonstrates the many ways by which Dck function can be altered by mutation. Further analysis of transcriptome changes in the *Dck* knockout cell lines will provide a better understanding of the changes taking place in the cells to compensate for the loss of Dck. This will be crucial in identifying drug targets to prevent the expansion of AML cells with defective Dck function. The TALEN-based KO techniques described here are especially suited to test each of the gene targets identified in these subsequent research efforts.

Table 3.1: Gene expression changes common to both Ara-C resistant cell lines

Gene expression changes common to both Ara-C resistant cell lines were identified by analyzing data generated from the Illumina NGS sequencing platform with the use of TopHat for genome mapping, and Cuffdiff for gene expression comparison. Genes in bold with * were also identified by gene expression microarray with 2-fold+ changes in expression (Rathe & Largaespada, 2010). Expression levels are expressed in fragments per kilobase of transcript per million mapped reads (FPKMs). Fold change compares the Ara-C resistant gene expression to the Ara-C sensitive parental lines. Negative values indicate a decrease in expression while positive values represent an increase in expression. The table is sorted by the minimum fold change, which is the least of the two fold change values.

Gene symbol	B117P FPKMs	B117H FPKMs	B117H vs. B117P fold chg	B140P FPKMs	B140H FPKMs	B140H vs. B140P fold chg	Minimum fold change
Ly6c2	27.174	0.433	-62.787	2.345	0.061	-38.362	-38.362
Dab2ip	3.835	0.213	-17.983	0.775	0.015	-52.118	-17.983
Dck	20.102	3.312	-6.069	17.756	2.444	-7.265	-6.069
C230035I16Rik	1.699	0.396	-4.287	3.242	0.602	-5.382	-4.287
Ksr1	1.210	0.222	-5.447	1.326	0.347	-3.821	-3.821
Ifitm1	22.679	2.112	-10.736	18.128	4.779	-3.793	-3.793
2310007A19Rik(Riiad1)	0.962	0.194	-4.963	0.920	0.252	-3.651	-3.651
Cd14	37.783	11.521	-3.279	0.628	0.001	-628.324	-3.279
Mpo	1450.210	473.280	-3.064	10.416	0.913	-11.403	-3.064
Fbxo44	2.339	7.264	3.106	0.876	5.796	6.616	3.106
Plag1	0.154	0.887	5.751	0.168	0.547	3.263	3.263
Chst3	0.232	0.848	3.648	0.370	5.113	13.803	3.648
Myl10	0.001	1.014	1013.690	0.534	2.278	4.267	4.267
AI427809	0.617	3.427	5.557	0.106	0.541	5.097	5.097
Cib3	0.506	5.240	10.349	0.099	0.889	9.018	9.018
Hectd2	0.001	1.316	1315.930	0.098	1.022	10.446	10.446

Table 3.2: Protein modifying missense mutations identified by MMuFLR

Protein modifying missense mutations identified by MMuFLR and verified in IGV as being found in the B117H and B140H cells, but not found in the respective parental cell lines, B117P and B140P. Mutations projected by either damaging or deleterious in human cells by either Polyphen-2 or PROVEAN and then verified by Sanger sequencing are noted in bold.

Cell Line	Gene	Variant Position	Ref.	Var.	Prevalence	Seq. Depth	Protein Change
B117H	*Commd5*	chr15:76730961	T	G	0.37	38	L43V
	Dram1	chr10:87803336	G	A	0.53	17	V91I
	Kdm3b	chr18:34987086	T	C	0.40	30	S1157P
	Kdm5c	chrX:148705903	G	T	0.33	39	V1222L
	Pigt	chr2:164328088	C	T	0.47	108	P386S
	Prkacb	chr3:146409579	G	T	0.44	94	E209D
	Pus7l	chr15:94358455	C	T	0.37	19	R499C
	Pyroxd1	chr6:142303201	A	G	0.43	7	N179S
	Rasgrp2	chr19:6407273	G	T	0.41	99	S349I
	Susd1	chr4:59328600	T	A	0.35	72	S748T
	Vps33b	chr7:87432008	G	A	0.37	27	D414N
B140H	***Atpbd4***	chr2:114470474	C	G	0.45	33	L99V
	Chd7	chr4:8678934	A	G	0.50	6	N95D
	Map3k11	chr19:5700807	G	A	0.38	52	M684I
	Mettl16	chr11:74609506	G	A	0.34	67	R203Q
	Pbrm1	chr14:31863259	A	C	0.37	62	S375R
	Rnf149	chr1:39609321	G	A	0.49	142	C394Y
	Smarca4	chr9:21445509	G	A	0.49	164	E525K
	Spag5	chr11:78118221	C	T	0.42	73	S284L
	Usp34	chr11:23321104	C	G	0.50	38	L1812V
	Zfp668	chr7:135009900	G	A	0.43	37	R542Q

Figure 3.1: *Dck* expression patterns and protein levels

Dck expression patterns and protein levels verified by IGV, qPCR and Western blot. (a) Venn diagram depicting the overlap in the genes identified as having a greater than 2-fold expression change in both sets of cell lines (B117H vs. B117P and B140H vs. B140P) when evaluated by gene expression microarray and RNA-seq. (b) Reduced expression of *Dck* in the BXH-2 cell lines was confirmed by qPCR using primers designed to span exons 5 and 6 of *Dck*. Error bars depict standard deviation. (c) RNA-seq reads were mapped to the NCBI reference genome. Visualization of *Dck* expression using IGV. (d) Transcripts were assembled independently of a reference genome using Cufflinks, then mapped to the mm9 mouse genome using Cuffcompare, and the resulting gtf file was visualized by IGV. (e) Western blot of Dck protein levels in the BXH-2 cell lines.

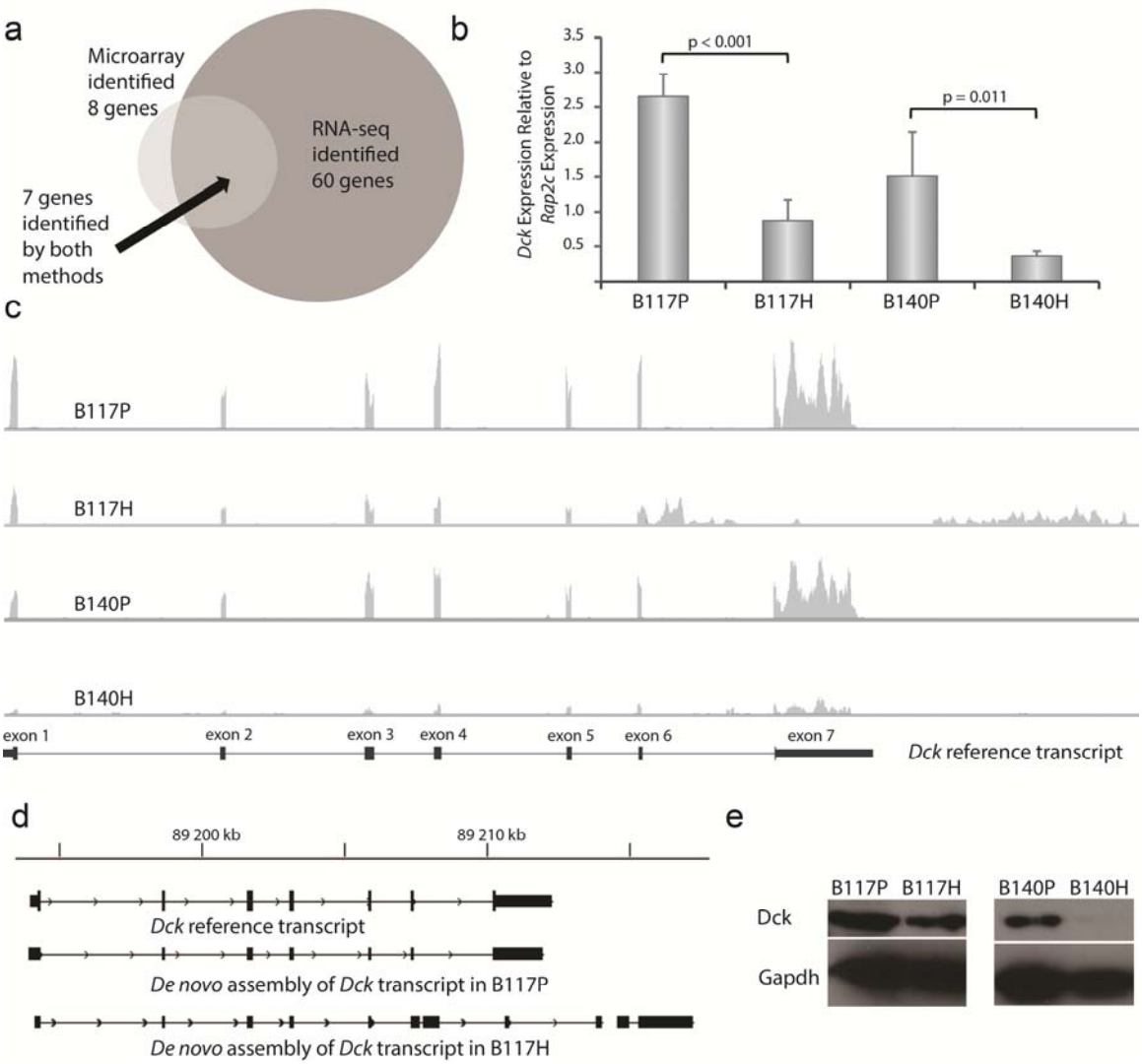

a
Microarray identified 8 genes
RNA-seq identified 60 genes
7 genes identified by both methods
b
Dck Expression Relative to Rap2c Expression
3.5
3.0
2.5
2.0
1.5
1.0
0.5
p < 0.001
p = 0.011
B117P
B117H
B140P
B140H
c
B117P
B117H
B140P
B140H
exon 1
exon 2
exon 3
exon 4
exon 5
exon 6
exon 7
Dck reference transcript
d
89 200 kb
89 210 kb
Dck reference transcript
De novo assembly of Dck transcript in B117P
De novo assembly of Dck transcript in B117H
e
B117P B117H
B140P B140H
Dck
Gapdh

Figure 3.2: Sequence abnormalities in Ara-C resistant cells

Sequence abnormalities in Ara-C resistant cells verified by Sanger sequencing. (a) Sanger sequencing of DNA in B117H cells identified an 878 nt deletion spanning the splice acceptor of intron 6 and the translated portion of exon 7. Sanger sequencing of RNA verified a transcript matching the configuration identified by TopHat and IGV. (b) Sanger sequencing of the RNA in B140H cells verified an insertion mutation in exon 4. (c) Structure of the TALENs used to knockout *Dck* in the B117P cell line.

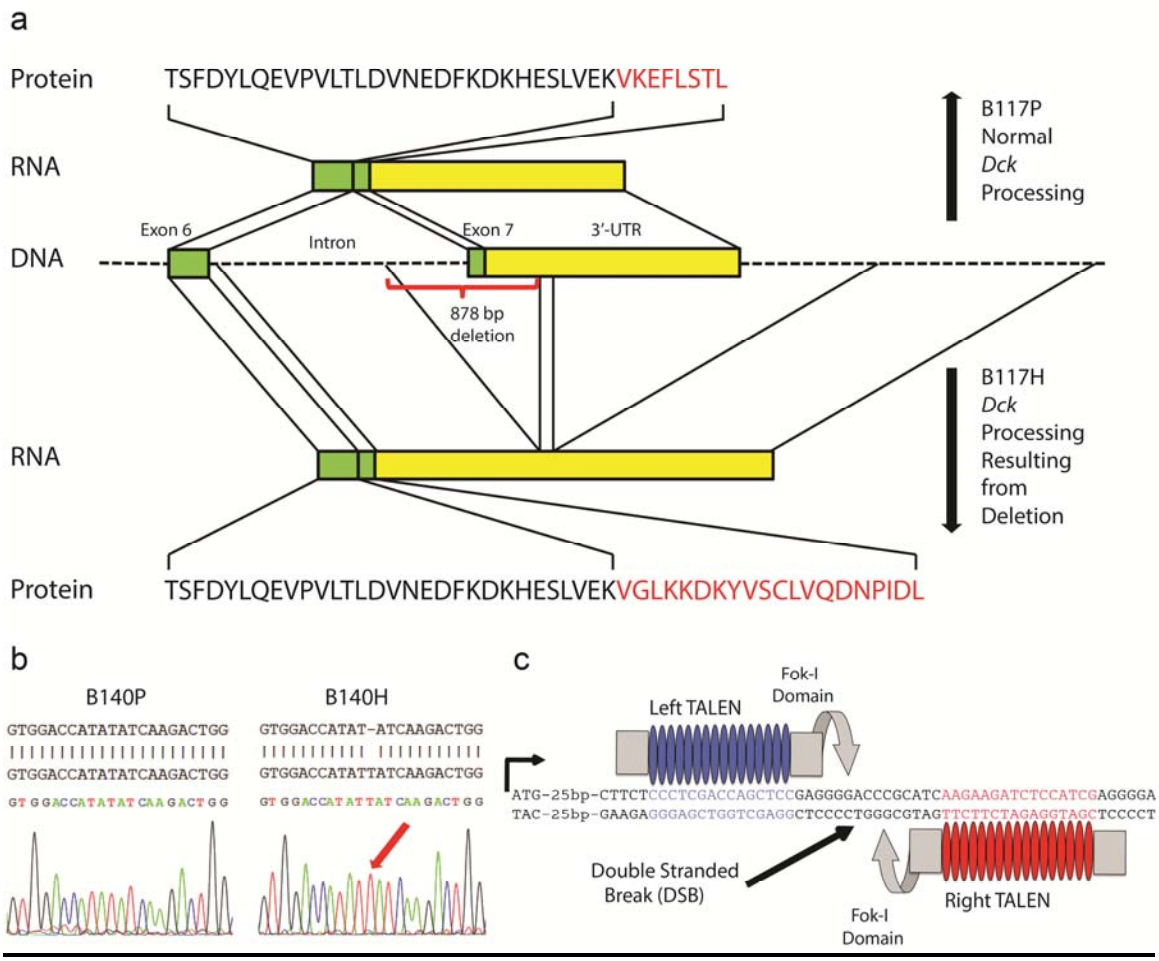

a
Protein
TSFDYLQEVPVLTLDVNEDFKDKHESLVEKVKEFLSTL
RNA
DNA
Exon 6
Intron
Exon 7
3'-UTR
878 bp
deletion
B117P
Normal
Dck
Processing
B117H
Dck
Processing
Resulting
from
Deletion
RNA
Protein
TSFDYLQEVPVLTLDVNEDFKDKHESLVEKVGLKKDKYVSCLVQDNPIDL
b
B140P
B140H
GTGGACCATATATCAAGACTGG
GTGGACCATATATCAAGACTGG
GTGGACCATAT-ATCAAGACTGG
GTGGACCATATTATCAAGACTGG
c
Fok-I
Domain
Left TALEN
ATG-25bp-CTTCTCCCTCGACCAGCTCCGAGGGGACCCGCATCAAGAAGATCTCCATCGAGGGGA
TAC-25bp-GAAGAGGGAGCTGGTCGAGGCTCCCCTGGGCGTAGTTCTTCTAGAGGTAGCTCCCCT
Double Stranded
Break (DSB)
Right TALEN
Fok-I
Domain

Figure 3.3: Reducing *Dck* expression in the B117P parental cell line

Reducing *Dck* expression in the B117P parental cell line results in increases of the IC_{50} value for Ara-C. (a) Reduction of *Dck* expression in B117P cells was accomplished marginally by RNA interference (KD1 and KD2) or completely using TALENs (KO). (b)) DNA modifications to Dck in the TALEN based KO cell lines confirmed a 1 nt deletion in the T2A cells, a 32 nt deletion in the T6B cells, and a 2 nt deletion in the T11A cells, and in each case the deletion occurred just a little over 50 nt from the translation start site. (c) RT-PCR of cDNA copy of RNA of a 595 base sequence straddling the translational start site. Primers are described in **Supplementary Table S1**. (d) Western blot verifying the absence of Dck proteins in the *Dck* KO clones.

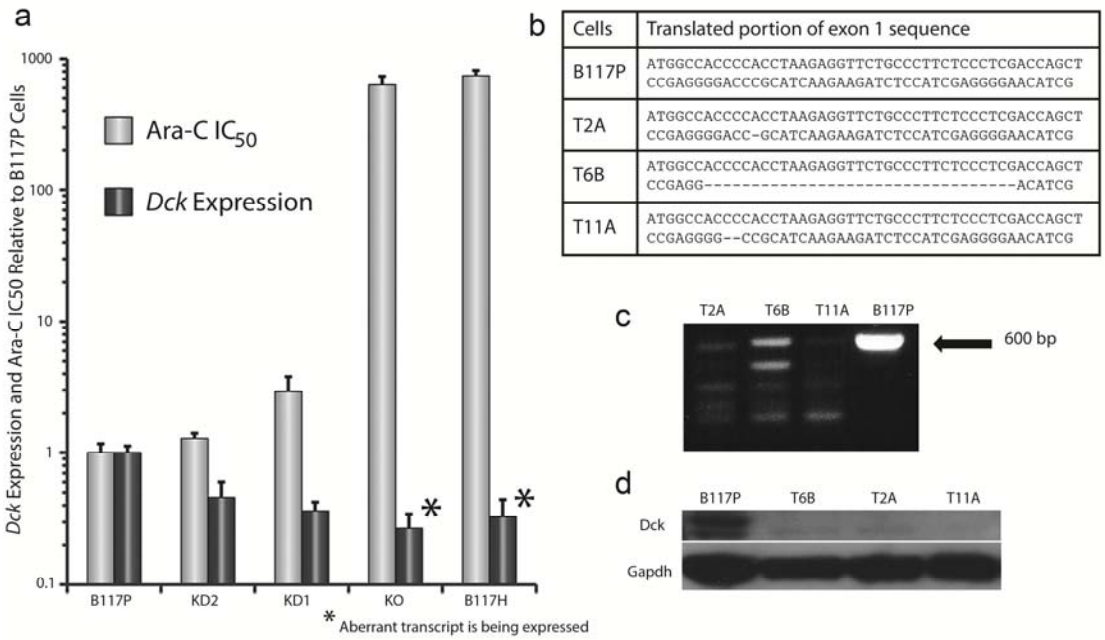
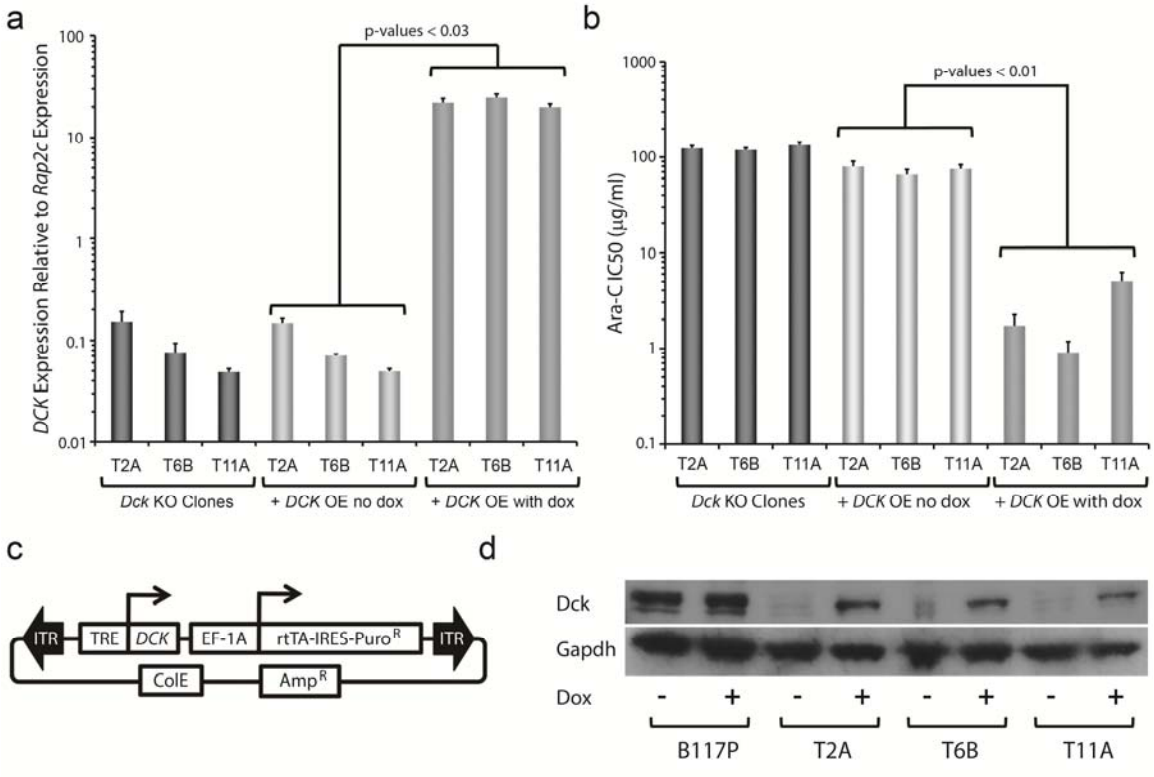

a
1000
100
10
1
0.1
Dck Expression and Ara-C IC50 Relative to B117P Cells
Ara-C IC50
Dck Expression
B117P
KD2
KD1
KO
B117H
* Aberrant transcript is being expressed
b
Cells
Translated portion of exon 1 sequence
B117P
ATGGCCACCCCACCTAAGAGGTTCTGCCCTTCTCCCTCGACCAGCT
CCGAGGGGACCCGCATCAAGAAGATCTCCATCGAGGGGAACATCG
T2A
ATGGCCACCCCACCTAAGAGGTTCTGCCCTTCTCCCTCGACCAGCT
CCGAGGGGACC-GCATCAAGAAGATCTCCATCGAGGGGAACATCG
T6B
ATGGCCACCCCACCTAAGAGGTTCTGCCCTTCTCCCTCGACCAGCT
CCGAGG----------ACATCG
T11A
ATGGCCACCCCACCTAAGAGGTTCTGCCCTTCTCCCTCGACCAGCT
CCGAGGGG--CCGCATCAAGAAGATCTCCATCGAGGGGAACATCG
c
T2A
T6B
T11A
B117P
600 bp
d
B117P
T6B
T2A
T11A
Dck
Gapdh

Figure 3.4: Rescue of *Dck* expression in the B117P KO clones

Rescue of *Dck* expression in the B117P KO clones results in a decrease of the IC_{50} value for Ara-C. (a) qPCR of *DCK* expression in *Dck* KO clones plus *DCK* overexpression with and without *DCK* expression activation by doxycycline. (b) Ara-C IC_{50} levels in *Dck* KO clones plus *DCK* overexpression with and without *DCK* expression activation by doxycycline, as determined by MTS-Tetrazolium assay. (c) Structure of the doxycycline inducible *DCK* overexpression vector. (d) Western blot of Dck/DCK proteins levels in the TALEN-based *Dck* KO clones with doxycycline induced *DCK* overexpression (OE). Protein levels shown with and without doxycycline activation of *DCK*.


a
p-values < 0.03
DCK Expression Relative to Rap2c Expression
100
10
1
0.1
0.01
T2A T6B T11A T2A T6B T11A T2A T6B T11A
Dck KO Clones
+ DCK OE no dox
+ DCK OE with dox
b
p-values < 0.01
Ara-C IC50 (µg/ml)
1000
100
10
1
0.1
T2A T6B T11A T2A T6B T11A T2A T6B T11A
Dck KO Clones
+ DCK OE no dox
+ DCK OE with dox
c
ITR
TRE
DCK
EF-1A
rtTA-IRES-Puro^R
ITR
ColE
Amp^R
d
Dck
Gapdh
Dox
- + - + - + - +
B117P
T2A
T6B
T11A

Chapter 4: Screening for Ara-C Co-resistance or Hyper-sensitivity Using Ara-C Resistant Murine AML Cell Lines

Contributors

Natashay Bailey assisted with the prednisolone IC50 drug assay.

Natalie Aumann and Willemijn Veldhuijzen assisted with the combination drug screen.

Miechaleen Diers performed all tail-vein injections and assisted with necropsies.

Acute myeloid leukemia (AML) can display *de novo* or acquired resistance to cytosine arabinoside (Ara-C), a primary component of induction chemotherapy. We hypothesized the changes required by the cells to become Ara-C resistant, might also make them more susceptible to elimination by other drugs. We also hypothesized there may be drugs which are most effective at treating *de novo* AML in combination with Ara-C. To test these hypotheses, we used two highly resistant murine AML cell lines, B117H and B140H, which were derived by introducing increasing concentrations of Ara-C to their parental cell lines, B117P and B140P, respectively. We previously demonstrated the loss of deoxycytidine kinase was one mechanism of resistance in the Ara-C resistant B117H and B140H cells, a mechanism which has also been documented in human cases of refractory disease. A medium throughput drug screen was conducted to determine if an increase in Ara-C resistance would result in increased sensitivity to other drugs. The screen evaluated 446 FDA approved drugs, and despite the diverse nature of the parental cell lines and their Ara-C resistant derivatives both Ara-C resistant cell lines showed an increase in sensitivity to prednisolone. We also conducted an assay which combined Ara-C individually with 50 other drugs. Ara-C proved most synergistic with cladribine in the parental cell lines, while the Ara-C resistant cell lines demonstrated a lack of synergy between Ara-C and cladribine.

Introduction

The standard induction therapy used by clinicians to treat patients with acute myeloid leukemia (AML) contains the chemotherapy agent cytarabine (Ara-C). Ara-C is a cytidine analog that interferes with DNA replication in fast growing cells. Ara-C is highly effective at eliminating AML blast cells. Unfortunately it is typically ineffective

in totally eradicating AML. It appears some AML cells are capable of escaping the initial assault by the chemotherapy drugs. **Chapter 3** described how an *in vitro* model of Ara-C resistance was used to identify one possible explanation for Ara-C resistance, the loss of normal deoxycytidine kinase (Dck) function. Dck is the rate-limiting enzyme in the metabolic activation of Ara-C. Through the use of *Dck* knockouts, it was shown the loss of *Dck* accounted for over 85% of the Ara-C resistance found in the B117H cells.

As cells become resistant to Ara-C, it seemed probable the cells would become more sensitive to other drugs. A standard drug screen was used to test this theory. It interrogated the response of all four cell lines to 446 FDA approved drugs. The response of the Ara-C resistant cells was compared to the response of their respective parental cells. It was found the Ara-C resistant cells became more sensitive to 3 corticosteroids with the most pronounced change in the glucocorticoid prednisolone.

A second screen was conducted to test 50 drugs for synergy with Ara-C in the B117P cells. It identified cladribine as the most synergistic drug to Ara-C in treating Ara-C sensitive cells. Subsequent assays confirmed the synergy between Ara-C and cladribine in both Ara-C sensitive parental cell lines, and lack of synergy in the Ara-C resistant cell lines.

The results from these screens on matched sets of Ara-C sensitive and Ara-C resistant murine AML cell lines are consistent with trends seen in patients. Some AML patients have been shown to respond well to treatments using glucocorticoids (Hicsonmez, 2006). Some *de novo* AML patients respond better when cladribine is included in the chemotherapy cocktail, while there is no evidence of improved outcome when cladribine is given to patients with refractory disease (Holowiecki et al., 2004;

Rubnitz et al., 2009; Rubnitz et al., 2004). Future work with this model will strive to characterize the molecular elements associated with positive drug response, for the purpose of creating clinical tools to assess a patient's probable response to various therapies and select the most effective drugs for that patient.

Methods and Material

Procedures for performing cell culture, the FDA screen, the combination screen, and drug assays for determining IC_{50} and CI_{50} are all described in **Appendix A**.

Results

Drug screen determines Ara-C resistant cells become sensitive to prednisolone

A drug screen was conducted on four BXH-2 derived cell lines, two Ara-C resistant (B117H and B140H) and the two Ara-C sensitive parental lines (B117P and B140P), from which the Ara-C resistant cells were derived. 446 FDA approved drugs were tested in each of the four cell lines at a concentration of 10 μM, and the results were expressed in the percent of cells inhibited as compared to untreated cells. Of the 446 drugs tested, 165 showed less than a 30% change in the cells' response to the drug and 218 drugs were inconclusive, since they resulted in killing all of the cells or none of the cells. 47 drugs showed a greater than 30% change in only one set of the cells, and 16 showed a greater than 30% change in drug response in both sets of cells (**Table 4.1**).

Of the 13 drugs identified by the screen as being more effective at eliminating the Ara-C resistant cell lines than their Ara-C sensitive parental lines by greater than 30%, 3 of the drugs fell into the category of corticosteroids: prednisolone, loteprednol etabonate, and corticosterone (**Table 4.2**).

Verification of prednisolone response by determining IC_{50}

Since prednisolone has been used to treat AML, an MTS tetrazolium assay was used to establish the IC_{50} for prednisolone (**Figure 4.1**). The B117P displayed a high level of prednisolone resistance (IC_{50} = 105 mM), but its Ara-C resistant derivative, B117H, had a 20-fold reduction in the prednisolone IC_{50}. The B140P cells were even more sensitive to the prednisolone than the B117H cells, and the B140P's Ara-C resistant derivative, B140H, demonstrated a 3-fold reduction in the prednisolone IC_{50}.

Dexamethasone is another glucocorticoid used in treating leukemia (clinicaltrials.gov). The IC_{50} for dexamethasone was derived using the MTS Tetrazolium Assay and provided a similar trend of IC_{50}'s in the four cell lines, but at much lower concentrations (**Table 4.3**).

Glucocorticoid sensitivity is dependent on binding to the glucocorticoid receptor

Cells respond to glucocorticoids in a number of ways, some of which are independent of glucocorticoid receptor (GR) binding (Di Santo et al., 1996). To determine if the GR is an integral part of the response to prednisolone in the BXH-2 cell lines, a combination drug assay was performed using a constant concentration of prednisolone and a variable concentration of mifepristone. Mifepristone is a steroid antagonist that binds to a number of steroid receptors including the GR (Agarwai, 1996). Mifepristone binding to the GR prevents the formation of an activated GR complex even in presence of a GR agonist, such as prednisolone (Cadepond et al., 1997). The drug assay demonstrated mifepristone was able to totally block the toxic effects of prednisolone at concentrations of 2-4 μM in all four BXH-2 cell lines (**Figure 4.2**).

Combination drug assays were also used to determine the combination index CI_{50} for the four cell lines. The resulting CI's were 5.7, 6.2, 6.0, and 6.6 for the B117P, B117H, B140P, and B140H cells, respectively, all reflective of a high degree of antagonism between prednisolone and mifepristone.

RNA-sequencing analysis identifies genes potentially envolved in glucocorticoid resistance

The high degree of variability in the four BXH-2 cell lines in their response to prednisolone and dexamethasone may be reflective of corresponding gene expression changes of similar direction and magnitude. The microarray data described in **Chapter 2** and the RNA-seq data described in **Chapter 3** were studied, looking for expression patterns mirroring the prednisolone and dexamethasone IC_{50} values. Six candidate genes were identified by the microarray, which may be associated with glucocorticoid response (**Table 4.4**).

The RNA-seq analysis, due to its unbiased approach for evaluating expression levels and its lack of distortions from background noise, identified 42 potential genes involved in prednisolone resistance (**Table 4.5**). The RNA-seq list did not include any of the same genes on the microarray candidate list. This was due to the more stringent selection criteria used in the RNA-seq analysis (30-fold vs. 10-fold between B117P and B140H, and the addition of a 2-fold change between B117P and B117H, between B140P and B140H), which was used to limit the number of potential candidates.

Combination drug screen identifies cladribine as synergistic to Ara-C in treating Ara-C sensitive AML

Synergy can be detected at a gross level by looking for graphical trends when combining a constant concentration of one drug with increasing concentrations of a second drug. **Figure 4.3** illustrates trends seen when the combination of drugs is synergistic, additive or antagonistic. The antagonistic trend was detected when prednisolone and mifepristone were tested together.

To find drugs synergistic with Ara-C, a drug assay was devised to test the response of the B117P cells when treated with 200 ng/ml of Ara-C and 7 different concentrations of the second drug being tested in 2X increments. The concentration of Ara-C used was selected to affect ~ 50% of the cells. Across the 9 plates used, the actual effect of Ara-C treatment alone was 51% (S.D. 6.7%). The additional effect provided by the second drug is present in **Table 4.6**.

Cladribine was found to have the highest level of "change". This change level reflects the steepness of the curve as the concentration of the second drug is added. To verify the synergy of the Ara-C and cladribine, a constant ratio of Ara-C and cladribine was introduced to all four cell lines using the standard MTS Tetrazolium Assay (**Table 4.7**). The constant ratio was determined by using the IC_{25} of Ara-C with the IC_{25} of cladrine. The results were compared to Ara-C or cladribine alone to determine the combination index at 50% inhibition (CI_{50}). The B117P cell line demonstrated a synergistic response when the two drugs together ($CI_{50} < 1.0$), while the B117H cell line showed an antagonistic response ($CI_{50} > 1.0$). This implies cladribine may enhance Ara-C response in cells that are already sensitive to Ara-C.

Discussion

AML is a heterogeneous disease. There is a diverse collection of mutations driving AML progression in each patient, making it problematic to establish a standard protocol for treating the disease effectively. The AML from patients with relapsed/refractory disease have additional therapy-related cellular changes. There are currently over 500 clinical trials underway for treating AML and over 1000 completed or suspended trials (clinicaltrials.gov), with approximately half the trials specific to relapsed/refracatory AML. These trials are testing a wide variety of drugs and drug combinations and yet there has been very little improvement in survival trends for AML. Most of the trials depend on the identification of the FAB subtype and a handful of known fusions. Even though over 100 of the active trials include Ara-C in the regime, none evaluate the functional status of deoxycytidine kinase or its pathway members for predicting the effectiveness of Ara-C. The same can be said for the 35 active studies testing decitabine and the 5 testing cladribine. Both decitabine and cladribine depend on deoxycytidine kinase for activation (Lotfi et al., 2003; Momparler, 2005). The use of synthetic glucocorticoids, such as prednisolone and dexamethasone, in treating AML is very rare. There are only 6 active AML trials using prednisolone and only 5 using dexamethasone. The use of glucocorticoids in treating ALL is more prevalent with over 200 clinical trials using prednisolone and over 400 using dexamethasone.

These statistics underscore the critical need for developing a better understanding of the molecular genetics driving drug response with the goal of providing patient-specific drug options. Perhaps there are some AMLs that would respond to glucocorticoid treatment if the drivers of the glucocorticoid response could be

determined. In the case of the BXH-2 cell lines, the B117P cells are resistance to glucocorticoids, while the B140P cells are significantly more sensitive (**Table 4.3**).

Synthetic glucocorticoids are given as anti-inflammatory and immunesuppresive agents in the treatment of a variety of ailments, including asthma and graft-versus-host disease (Martin et al., 2012; Rhen and Cidlowski, 2005; Schaaf and Cidlowski, 2002). Glucocorticoids affect metabolic processes and transcriptional regulation via a number of pathways, most of which are dependent upon binding to the GR (Schaaf and Cidlowski, 2002). The high number of clinical trials involving glucocorticoids to treat ALL, but not AML, is due to a mechanism of glucocorticoid-induced apoptosis, which is prevalent in ALL (Schlossmacher et al., 2011), but not in AML (1966). The development of resistance to glucocorticoids is a common problem, but only a few mechanisms of resistance have been documented to date. These mechanisms typically involve the inhibition of GR activity (Schaaf and Cidlowski, 2002).

The drug screen, reported here, evaluating 446 FDA approved drugs indicated that as cells become resistant to Ara-C they become significantly more sensitive to glucocorticoids, such as prednisolone and dexamethasone. Thus, the response to glucocorticoids in AML cells can be altered by unidentified regulatory stimulations. These results imply there may be some patients with *de novo* AML who might respond well to the inclusion of glucocorticoids in their treatment regimen, and perhaps glucocorticoids should be considered in the treatment of Ara-C resistant AML. To determine when glucocorticoids should be used in the treatment of AML, the functional changes in the cells driving both the Ara-C and the glucocorticoid response need to be identified.

As to the nature of the glucocorticoid response, there are three distinct phases that have been identified in glucocorticoid-induced apoptosis (Frankfurt and Rosen, 2004). The first phase, the "initiation stage", is triggered by glucocorticoid reception binding. The second phase, the "decision stage", occurs within the mitochondria. The last phase, the "execution stage", involves the apoptotic caspase cascade. Since all 4 of the BXH-2 cell had similar response to combination of prednisolone and mifepristone, it appears the glucocorticoid receptor is functioning normally in all 4 cell lines, and the prednisolone resistance found in the B117P is not related to the "initiation stage". One type of glucocorticoid resistance found in the "decision stage", involves the overexpression of anti-apoptotic proteins, such as the members of the Bcl-2 superfamily (Sionov, 2013). However, none of these genes were included in the list of genes which compared glucocorticoid response to gene expression levels (**Table 4.5**). The Sionov, 2013, review article also provided a comprehensive description of other genes which have been shown to be differentially expressed in glucocorticoid resistant cells, but none of these genes were shown to be differentially expressed in the glucocorticoid resistant B117P cells.

The second screen of 50 drugs used in combination with Ara-C identified cladribine as potentially the most synergistic of the drugs tested. Subsequent MTS-tetrazolium assays verified the synergy in the Ara-C sensitive B117P cell lines, and antagonism in the Ara-C resistant B117H cells (**Table 4.7**). This same pattern has been seen in clinical trials. Cladribine has been found to be an effective partner to Ara-C in treating *de novo* disease (Holowiecki et al., 2004; Rubnitz et al., 2009), but not effective in treating refractory disease (Rubnitz et al., 2004). The exact nature of the synergy between cladribine and Ara-C is not fully understood. Both drugs are activated by Dck,

and both are nucleotide analogs that interfere with DNA replication, although their mechanisms of action are different. Ara-C is incorporated into DNA (Grant, 1998), while cladribine is a strong inhibitor of ribonucleotide reductase activity, which is required for *de novo* synthesis of nucleosides (Sigal et al., 2010). Perhaps the synergy between Ara-C and cladribine is due to the combined effect of the strong interference of Ara-C to the dNTP salvage pathway and cladribine blocking the *de novo* dNTP pathway, although this could just be an additive effect. It is also interesting to note the IC_{25} for cladribine was 18-fold higher in the B117H as compared to the B117P, while the IC_{25} change in response to Ara-C was 1300-fold. This reduced level of cladribine resistance to the loss of Dck may be explained by the fact that cladribine can also be activated by deguanosine kinase (Lotfi et al., 2003) or it may be indicative of a secondary mode of action for cladribine in the cells accounting for the synergy between Ara-C and cladribine.

These drug screens support the premise that the BXH-2 cell lines are useful tools for understanding the molecular characteristics driving drug response. For example, the assay combining a fixed amount of prednisolone with variable amounts of mifepristone confirmed a high degree of antagonism between the drugs in all four cell lines confirming the glucocorticoid apoptotic activity in leukemia cells is dependent on glucocorticoid reception binding.

The RNA-seq analysis identified a number for genes that may be involved in prednisolone response. One of the genes, Cd14, is a marker of monocytes and one study showed a very different expression profile induced by prednisolone in $CD14^+$ monocytes as compared with $CD14^-CD4^+$ T cells (Toonen et al., 2011). This implies cells respond differently to prednisolone based on the cells' CD14 status. Perhaps this difference also

affects the ability of cells to undergo glucocorticoid induced apoptosis. There was a markedly different response to prednisolone in the $CD14^+$ B117P cells, which had a prednisolone IC_{50} of 100 μM, and the $CD14^-$ B140P cells, which had a prednisolone IC_{50} of 1.5 μM.

For this drug study, two drug screens were performed, one to determine if AML cells became more sensitive to other drugs when they became resistant to Ara-C (a model of refractory AML), and a second to determine what drugs were synergistic with Ara-C in treating Ara-C sensitive cells (a model of *de novo* AML). The first screen showed the BXH-2 cells became more sensitive to glucocorticoids as they became resistant to Ara-C, and the second screen identified cladribine as synergistic with Ara-C in the Ara-C sensitive B117P cells. The RNA-seq analysis provided a list of genes which may be involved in the glucocorticoid response. Uncovering the significant cellular characteristics associated with these responses, such as gene expression levels or mutated genes, could be instrumental in developing tests to determine the most effective drug combinations for treating AML patients.

Table 4.1: Summary of Drug Screen of 446 FDA Approved Drugs

446 FDA approved drugs were tested in the B117P, B117H, B140P and B140H cell lines. Each drug was tested at concentration of 10 μM. The percent of effect for each of the drugs tested were summarized and categorized.

Response of Cells to Drugs in Drug Screen	Drug Count
B117H and B140H cells became more sensitive than parentals by > 30%	13
B117H and B140H cells became more resistant than parentals by > 30%	3
B117H became more sensitive than B117P by > 30%	15
B117H became more resistant than B117P by > 30%	2
B140H became more sensitive than B140H by > 30%	29
B140H became more resistant than B140H by > 30%	1
Less than 30% change in both sets of cells	165
All 4 cell lines resistant to drugs (< 10% inhibition)	198
All 4 cell lines sensitive to drugs (> 90% inhibition)	20
Total	**446**

Table 4.2: 13 Drugs Have a Higher Effect in Ara-C Resistant Cells

13 of the drugs tested in the B117P, B117H, B140P and B140H cell lines had an increase in percent affected of >30% in the Ara-C resistant cell lines (B117H and B140H).

Compound	Description	Type	% affect of 10μM on each cell line				Least change
			B117P	B117H	B140P	B140H	
SAM001246965	AM404	Analgesic	48.3	95.5	-6.7	52.9	47.2
SAM001246897	Prednisolone	Corticosteroid	29.2	72.3	34.7	79.1	43.1
SAM001246630	Loteprednol etabonate	Corticosteroid	34.9	74.7	22.9	70.8	39.8
SAM001247051	Mesoridazine	Phenothiazine	9.1	47.8	-1.7	58.4	38.7
SAM001247098	R(+)-SCH-23390 hydrochloride	D_1 receptor antagonist	2.7	38.5	23.7	87.4	35.8
SAM001247066	CGS 12066B	5-HT_{1b} agonist	12.4	47.8	-3.0	58.2	35.4
SAM001246750	Aripiprazole	Antipsychotic	-13.9	49.6	3.6	38.4	34.8
SAM001247070	Pergolide mesylate salt	Dopaminergic agonist	-2.9	49.0	-8.6	25.4	34.0
SAM001246879	Corticosterone	Corticosteroid	24.2	57.3	6.7	53.1	33.1
SAM001246658	Naftopidil	α1-adrenergic receptor antagonist	-25.6	48.1	-2.1	30.7	32.8

Table 4.3: IC50's for Prednisolone and Dexamethasone in BXH-2 Cell Lines

The IC_{50} for prednisolone and dexamethasone was determined using the MTS Tetrazolium Assay. Assay was repeated a minimum of 3 times. r-values > 0.88.

		Cell Line			
Drug	**Statistic**	**B117P**	**B117H**	**B140P**	**B140H**
Prednisolone	Average IC50 (μM)	100.198	3.450	1.547	0.338
	Standard Deviation	14.348	0.980	0.952	0.072
Dexamethasone	Average IC50 (μM)	1.468	0.183	0.125	0.032
	Standard Deviation	0.393	0.047	0.061	0.006

Table 4.4: Potential Genes Involved in Prednisolone Resistance as Identified by Microarray Analysis

Gene involved in glucocorticoid resistance were identified from microarray data by selecting genes with both a 10-fold change in expression between the B117P and B140H and an upward or downward expression level trend across the four cell lines consistant with the trends of the prednisolone and dexamethasone IC_{50} values.

IC50 value (μM)	B117P	B117H	B140P	B140H	Ratio of B117P to B117H	Ratio of B117P to B140P	Ratio of B117P to B140H
Prednisolone	100.198	3.450	1.547	0.338	29.0	64.8	296.4
Dexamethasone	1.468	0.183	0.125	0.032	8.0	11.8	45.3
Microarray Gene Expression Levels							
Overexpressed in B117P							
BC100530, Stfa3 (1435761_at)	831.8	31.3	26.0	25.8	26.5	32.0	32.2
Hpgd (1419905_s_at)	572.1	43.7	15.0	10.0	13.1	38.1	57.3
Underexpressed in B117P							
Atp1b1 (1439036_at)	13.7	382.7	464.5	989.1	27.9	33.8	72.0
Glul (1426235_a_at)	18.5	121.9	125.4	186.1	6.6	6.8	10.1
Glul (1426236_a_at)	9.5	396.2	436.6	588.1	41.9	46.2	62.2
Il7r (1448576_at)	22.2	121.1	232.3	393.4	5.5	10.5	17.8
Pcgf6 (1424081_at)	28.1	288.0	1234.8	1269.5	10.3	44.0	45.3
Pcgf6 (1454120_a_at)	8.1	95.0	407.3	410.2	11.7	50.2	50.6

Table 4.5: Potential Genes Involved in Prednisolone Resistance as Identified by RNA-Seq Analysis

Gene candidates involved in glucocorticoid resistance were identified from RNA-seq data by identifying genes with 30-fold change in expression between the B117P and B140H, a continuous upward or downward expression level trend across the four cell lines, and at least a 2-fold change between B117P and B117H and between B140P and B140H. The FPKMs for either the B117P cells or the B140H cells were > 1.0.

RNA-Seq Gene Expression Levels (FPKMs)					Ratio of B117P to		
Overexpressed in B117P	B117P	B117H	B140P	B140H	B117H	B140P	B140H
2010300C02Rik	1.438	0.046	0.024	0.001	31.2	60.4	1437.7
Abca1	1.878	0.602	0.368	0.031	3.1	5.1	61.0
Ccbl2	15.548	6.695	3.791	0.001	2.3	4.1	15548.4
Cd14	37.783	11.521	0.628	0.001	3.3	60.1	37782.9
Cd9	30.507	4.389	0.048	0.001	7.0	640.4	30506.7
Cdc42ep4	2.778	1.039	0.107	0.001	2.7	25.8	2777.5
Cldnd2	1.075	0.496	0.105	0.001	2.2	10.3	1075.3
Csf3r	6.192	0.286	0.124	0.001	21.7	49.8	6192.2
Elane	81.118	14.846	0.385	0.154	5.5	210.7	525.3
Fkbp11	20.183	9.735	5.533	0.340	2.1	3.6	59.3
Gm5643	1.239	0.518	0.493	0.001	2.4	2.5	1239.3
Gp49a	5.403	0.289	0.043	0.001	18.7	125.6	5402.9
Hck	6.469	0.099	0.077	0.020	65.2	83.8	320.9
Hist2h3c2-ps	1.480	0.495	0.287	0.046	3.0	5.2	32.2
Id1	5.069	1.117	0.065	0.001	4.5	78.1	5069.2
Isg15	4.589	0.773	0.243	0.066	5.9	18.9	69.8
Kcng1	4.692	2.027	0.138	0.001	2.3	33.9	4692.0
Lgals3	79.326	34.940	0.040	0.001	2.3	1960.3	79325.6
Lgals3bp	24.418	11.489	0.173	0.039	2.1	140.9	629.6
Lyz1	4.368	1.712	0.446	0.106	2.6	9.8	41.2
Mocos	2.256	1.114	0.073	0.001	2.0	30.8	2256.4
Mpo	1450.210	473.280	10.416	0.913	3.1	139.2	1587.6
Ms4a3	217.785	15.342	0.115	0.001	14.2	1896.4	217785.0
Nt5e	26.361	9.577	0.029	0.011	2.8	902.3	2318.1
Oas2	2.199	0.587	0.081	0.021	3.7	27.2	104.7
Retnlb	9.076	3.842	0.090	0.001	2.4	101.0	9075.7
Sell	80.764	18.619	0.073	0.019	4.3	1111.6	4258.7
Siglecg	2.640	0.807	0.021	0.001	3.3	123.2	2640.0
Tm4sf19	1.582	0.345	0.045	0.001	4.6	35.3	1582.0
Tmem132a	2.393	0.537	0.015	0.001	4.5	159.5	2393.2
Tmem180	25.612	8.967	0.070	0.018	2.9	363.9	1394.7
Tnfsf14	1.813	0.902	0.234	0.023	2.0	7.7	78.9
Tuba8	19.869	9.444	0.349	0.092	2.1	56.9	216.0
Zfp516	2.029	0.180	0.020	0.005	11.2	101.7	394.0
Underexpressed in B117P							
Bank1	0.014	0.047	0.372	1.121	3.4	26.7	80.4
Camk2a	0.031	0.106	0.560	1.628	3.4	17.8	51.7
Dnahc6	0.001	0.004	0.044	1.309	3.9	44.2	1309.2
Gm5148	0.120	0.537	1.119	4.552	4.5	9.3	37.9
Gpr25	0.214	0.878	17.899	37.890	4.1	83.7	177.3
Hmgb1-rs17	0.033	0.185	0.384	1.273	5.6	11.6	38.6
Serinc5	0.075	0.282	0.456	5.207	3.7	6.1	69.1
Ublcp1	0.001	0.024	2.437	8.111	24.2	2437.3	8110.7

Table 4.6: Combination Assay Using Fixed Amount of Ara-C and Variable Concentrations of a Second Drug in the B117P Cell Lines

MTS Tetrazolium assay was used to measure fraction affected when 200ng/ml of Ara-C was combined with 7 different concentrations of a second drug in the B117P cell line. Fraction affected was the fraction relative to Ara-C alone. Combinations were repeated in 2 wells. The IC_{50} for the second drugs was also determined for the second drug without Ara-C, except for those highlight in AG490 and AZD6244, which used information from literature. Although 50 drugs were tested, only those with > 0.5 fraction affected at the highest concentration were included in the table. Most drugs were tested at the high concentrations (top scale), whereas aminopterin, PD-0325901, triptolide, bortezomib, and 5-fluorouridine were done at smaller concentrations (lower scale) due to their lower IC50's. Change (*) was determined adding the highest fraction affected at the 0.625 concentration level to the difference between the 0.625 and the 0.15625 (top scale), or between the 0.0625 and the 0.015625 (lower scale).

		Fraction affected at different concentrations of second drug (μM)							
		10	5	2.5	1.25	0.625	0.3125	0.15625	
Drug	**IC50(μM)**	**1**	**0.5**	**0.25**	**0.125**	**0.0625**	**0.03125**	**0.015625**	Change*
Mifepristone	5.500	0.771	0.441	0.240	0.066	0.000	0.072	0.195	-0.195
Parthenolide (PTL)	1.250	0.977	0.849	0.719	0.340	0.038	0.000	0.247	-0.171
AG490	20.000	0.795	0.651	0.410	0.135	0.000	0.000	0.148	-0.148
GSK-3B Inhibitor	0.805	0.969	0.809	0.412	0.023	0.000	0.010	0.147	-0.147
Cantharidin	1.892	0.891	0.101	0.000	0.000	0.000	0.000	0.146	-0.146
Harki1020	3.315	0.947	0.562	0.235	0.000	0.006	0.018	0.116	-0.104
IKK Inhibit VII	1.000	1.000	0.000	0.000	0.000	0.000	0.229	0.074	-0.074
MG132	1.208	1.000	0.978	0.000	0.000	0.028	0.134	0.068	-0.011
Nelarabine	12.354	0.269	0.098	0.000	0.135	0.062	0.000	0.135	-0.011
Mevastatin	2.830	0.858	0.306	0.093	0.106	0.000	0.000	0.000	0.000
Harki1001	2.235	0.783	0.284	0.000	0.000	0.000	0.000	0.000	0.000
AM404	9.812	0.897	0.729	0.438	0.208	0.108	0.174	0.213	0.004
Harki1019	5.300	0.500	0.177	0.016	0.049	0.054	0.096	0.085	0.023
RAD001	0.149	0.957	0.923	0.867	0.627	0.206	0.307	0.253	0.160
MK-0457	0.775	0.688	0.013	0.000	0.000	0.118	0.000	0.047	0.189
5-Fluoro-2'-deoxyuridine	0.149	0.507	0.456	0.413	0.425	0.416	0.418	0.393	0.438
Sunitinib	0.200	0.872	0.964	0.967	0.651	0.297	0.116	0.081	0.514
Dexamethasone	1.500	0.939	0.839	0.747	0.605	0.467	0.345	0.390	0.543
Aminopterin	0.010	0.535	0.507	0.527	0.542	0.529	0.498	0.481	0.577
GDC0941	0.650	0.944	0.840	0.750	0.636	0.517	0.523	0.369	0.665
NFkB Inhibitor	0.098	0.683	0.624	0.597	0.604	0.399	0.000	0.000	0.798
CPEC	0.157	0.717	0.751	0.735	0.736	0.776	0.744	0.709	0.842
AZD6244	10.000	0.940	0.816	0.739	0.612	0.637	0.406	0.359	0.914
PD-0325901	0.007	0.941	0.936	0.881	0.791	0.715	0.583	0.498	0.931
Triptolide (HSP inhibitor)	0.018	1.000	1.000	1.000	0.981	0.488	0.001	0.023	0.953
Bortezomib	0.006	1.000	1.000	1.000	1.000	1.000	1.000	0.994	1.006
5-Fluorouridine	0.034	0.950	0.838	0.800	0.721	0.647	0.387	0.000	1.295
Cladribine	0.800	1.000	0.957	0.828	0.805	0.718	0.433	0.095	1.341

Table 4.7: Fixed Ratio Combination Assay Verifies Synergy between Ara-C and Cladrine in B117P Cells, but Antagonism in B117H Cells

MTS Tetrazolium assay was used to determine CI_{50} when Ara-C was combined with cladribine at a fixed ratio. Assay was repeated 3 times. CI_{50} was determined using CalcuSyn, by comparing the combination assay with assays of Ara-C alone and cladribine alone, which were conducted at the same time. Each set of experiments were repeated 3 times.

	B117P	B117H
Ara-C IC_{25} (μg/ml)	0.08	110.00
Cladribine IC_{25} (μM)	0.70	13.00
Average CI_{50}	0.73	1.66
CI_{50} standard deviation	0.16	0.38

Figure 4.1: Ara-C Resistant Cells Are More Sensitive to Prednisolone

The prednisolone IC_{50} was determined using the MTS Tetrazolium Assay. Assay was repeated 3 times with comparable results.

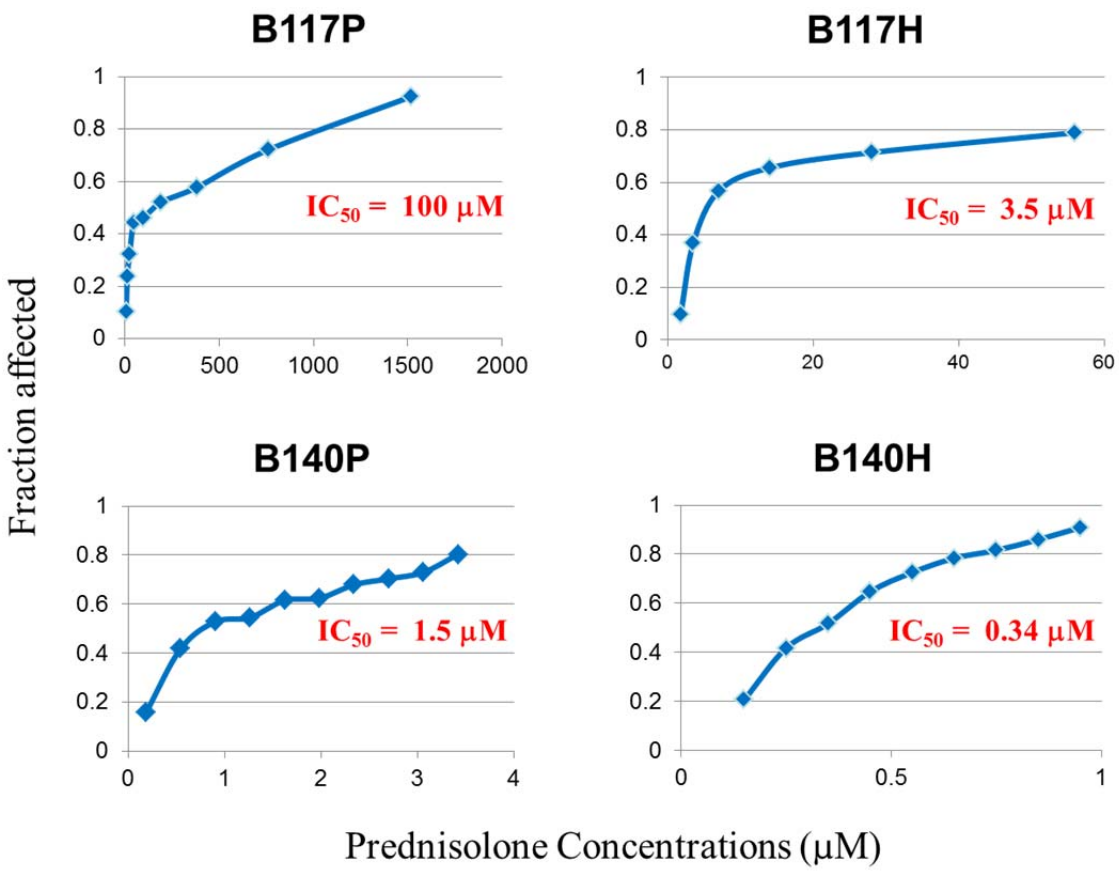

B117P
IC50 = 100 µM
B117H
IC50 = 3.5 µM
B140P
IC50 = 1.5 µM
B140H
IC50 = 0.34 µM
Fraction affected
Prednisolone Concentrations (µM)

Figure 4.2: Mifepristone Blocks the Effects of Prednisolone in All Four BXH-2 Cell Lines

MTS Tetrazolium Assay with a constant concentration of prednisolone and an increasing concentration of mifepristone. Assay was repeated 3 times with similar results.

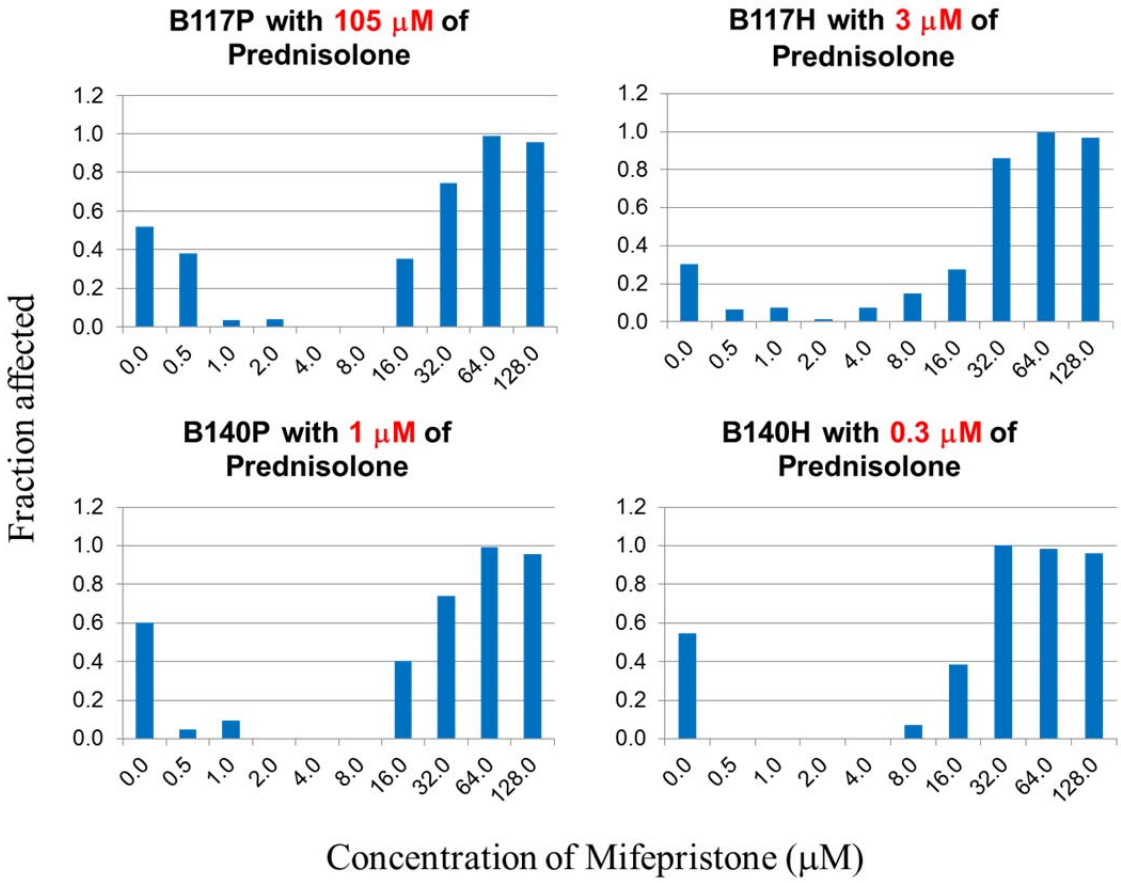
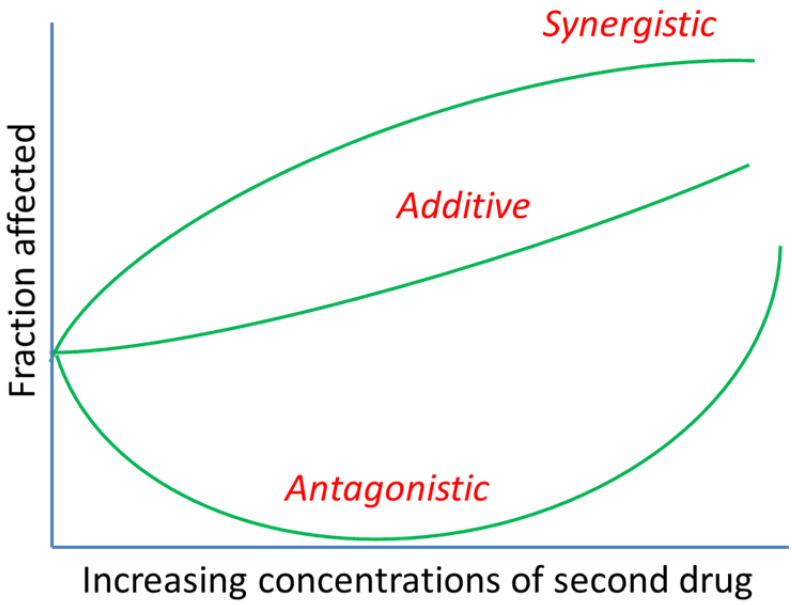

B117P with 105 μM of Prednisolone
B117H with 3 μM of Prednisolone
B140P with 1 μM of Prednisolone
B140H with 0.3 μM of Prednisolone
Fraction affected
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0.0
0.5
1.0
2.0
4.0
8.0
16.0
32.0
64.0
128.0
Concentration of Mifepristone (μM)

Figure 4.3: Graphic Trends of Combination Assay

Graphical depiction of synergistic, additive, and antagonistic trends in fraction affected when drug one of a constant concentration is tested with increasing concentrations of a second drug.


Synergistic
Additive
Antagonistic
Fraction affected
Increasing concentrations of second drug

Chapter 5: Future Research Opportunities

Contributors

Natashay Bailey and Christopher Stoltzenberg assisted with necropsies.

Summary of Research

The goal of this research was to understand how AML cells become resistant to Ara-C and to search for more effective drug treatment combinations based on cellular characteristics, such as gene expression patterns and mutations. In **Chapter 2,** the research focused on using microarray analysis of four BXH-2 murine cell lines to determine the gene expression changes associated Ara-C resistance. The primary change found by this effort was the significant downregulation of *Dck* in the two Ara-C resistant cell lines when compare their two Ara-C sensitive parental cell lines. It was also determined, by testing other chemotherapy drugs, the highly drug resistance phenotype was not applicable to any other drug except decitabine, which also depends on *Dck* for metabolic activation.

Chapter 3 described further research on these same cell lines to determine the functional changes taking place in the cells, which account for the Ara-C resistance. RNA-seq was used to locate mutations in *Dck* in both the Ara-C resistant cell lines. Knockdown and knockout of *Dck* in the Ara-C sensitive cell lines clearly showed a direct correlation between *Dck* expression and Ara-C response with loss of *Dck* accounting for 85% of the Ara-C resistance. Most of the Ara-C sensitivity was rescued by introducing *DCK* to the *Dck* knockout cell lines.

The focus of **Chapter 4** was finding drugs more effective in treating Ara-C resistant AML, as well as drugs synergistic with Ara-C for treating *de novo* AML. A drug screen of 446 FDA approved drugs, showed that the Ara-C resistant cells became more sensitive to glucocorticoids, such as prednisolone. It was also determined the sensitivity of the AML cell lines to prednisolone was dependent on glucocorticoid

receptor binding. A second drug screen evaluated the synergy of 50 different drugs with Ara-C in the Ara-C sensitive B117P cell line, and determined cladribine was the most synergistic. Combination drug assays confirmed synergy between Ara-C and cladribine in the B117P, while there was antagonism in the B117H cell lines. This result was consistent with results seen in clinical trials using the Ara-C and cladribine combination in treating *de novo* AML (Holowiecki et al., 2004; Rubnitz et al., 2009), but not refractory AML (Rubnitz et al., 2004).

Additional Research Opportunities Using the BXH-2 Model

Since the results gleaned thus far by studying the BXH-2 model of Ara-C resistance have proved consistent with findings from clinical trials, it seems evident these cells can continue to be valuable experimental tools for resolving a number of unanswered questions, such as: (1) How do Ara-C resistant cells compensate for the loss of *Dck*? (2) Do Ara-C resistant cells depend on a compensating up-regulation of the *Ras/MAPK* pathway? (3) What are the molecular processes responsible for innate glucocorticoid resistance in some Ara-C sensitive cells, why Ara-C resistant cells become more sensitive to glucocorticoids, and how can cells be induced to become glucocorticoid sensitive? (4) Are there other mechanisms of toxicity for cladribine and what accounts for the synergy existing between Ara-C and cladribine in Ara-C sensitive cells? Strategies for answering these questions are outline below.

<u>Compensating for the loss of Dck</u>

A number of changes occurred in the Ara-C resistant cells, some of which may be compensating for the loss of Dck. These include the downregulation of *Dab2ip*, *Dab2*,

Cd14, *Mpo*, *Ksr1*, and *Riiad1*. Dab2ip, Dab2, and Ksr1 are negative regulators of Ras (Calvisi et al., 2011; Therrien et al., 1996; Zhou and Hsieh, 2001), Cd14 is a monocyte marker (Griffin et al., 1981), and Mpo is involved myeloid differentiation (Koeffler et al., 1985). Very little is known about Riiad1.

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) have been designed to target three distinct sites of each of these genes. The CRISPR system was first identified in *Escherichia coli*, but has since been modified to work effectively in vertebrate cells (Mali et al., 2013; Qi et al., 2013). The system requires the preparation of CRISPR oligos, which are 23 bases long and a direct match to a protein coding segment of the target gene, preferably near the translational start site. The oligo sequence must begin with a 'G' and end with 'GG', and must be tested *in silico* for specificity to the target gene. Introduction of a vector containing the *Cas9* coding sequence, the enzyme that directs the oligo to the matching genomic sequence and executes a double strand DNA break at that site, results in double strand breaks which are inefficiently repaired by the double-strand break repair system of the cell. This results in short deletions which typically disrupt gene function. In that respect the CRISPR system is very similar to the TALEN system, but CRISPR is faster and less expensive to use. It still remains to be determined if the CRISPR system is as specific as TALENs.

The CRISPR oligos targeting *Dab2ip*, *Dab2*, *Cd14*, *Mpo*, *Ksr1*, and *Riiad1* will be introduced individually to the Ara-C sensitive B117P and B140P cells, and selected for incorporation using puromycin. A number of experiments will be conducted on the resulting knockout cell lines to determine: (1) if the Ara-C IC_{50} has increased, (2) if the *Dck* expression levels have been reduced, (3) if there is an increase in the proliferation

rate in the cell lines, and (4) how receptive the cells are to becoming Ara-C resistant. These results will help to determine the level of participation of each of the genes in the Ara-C resistant phenotype. Similar experiments will be done on the other genes identified by the RNA-seq analysis as being down-regulated or mutated in the Ara-C resistant cell lines.

RNA-seq of the TALEN based *Dck* KO cells and dox-induced *DCK* overexpression cells will be used to narrow the list of candidate genes, by identifying genes having expression levels strongly correlating to *Dck* levels, either positively or negatively. An initial RNA-seq was accomplished for the B117P T6B *Dck* KO cell line and its associated *DCK* overexpression cell line (induced by doxycycline). **Table 5.1** outlines the most prominent and reversible changes to the cells as compared to each other and the B117P starting cells.

The top 8 genes are all located in close proximity to one another on chromosome 16 of the mouse genome. They are all members of the stefin family of genes, which are only found in murine cells. They are immediately downstream from cystatin A (*Csta*), and are all protein coding genes. Their protein sequences are very similar and each most closely match cystatin A, both in murine and human cells according to Ensembl's PBLAST utility (www.ensembl.org), indicating they are *Csta* replicates in the mouse genome, which have independently evolved over time. They all have a highly conserved promoter region of 221 bp located approximately 1000 bp upstream. In the *Stfa1* gene, it is located at -1028bp from the transcriptional start site. This promoter region is not found upstream of *Csta*, and there was no significant change in *Csta* expression levels in the *Dck* KO and *Dck* KO/*DCK* OE cells, suggesting there is no relevance to this change in

human cells. As for function, Stfa1 and Stfa2 were shown to be strong inhibitors of papain-like endopeptidases, while human CSTA was not (Mihelic et al., 2006). Even though *Csta* did not undergo significant expression changes in the *Dck* KO and *Dck* KO/*DCK* overexpression cells, the stefins may be performing some other cystatin A type functions. Cystatin A has been shown to be an inhibitor of cysteine proteases (Anastasi et al., 1983; Brzin et al., 1984), a blocker of apoptosis by inhibiting cathepsin B (Takahashi et al., 2007), a modulator of migration and invasion (Chang et al., 2010), and a risk factor for breast cancer (Kuopio et al., 1998). There is nothing in the literature describing a relationship between *Dck* and any cystatin/stefin.

Another interesting pair of genes, *S100a8* and *S100a9*, appeared on the list as being highly upregulated in the *Dck* KO cells and then downregulated in the *Dck* KO/*Dck* OE cells. The *S100a8* and *S100a9* genes generate proteins which form a heterodimer called calprotectin (Hessian and Fisher, 2001). S100a8 and S100a9 have both been shown to be activators of the Ras/MAPK pathway (Kwon et al., 2013; Wu et al., 2013). Ingenuity Pathway Analysis (Ingenuity® Systems, www.ingenuity.com) of all the genes on the list indicated a close relationship of these genes to the Ras/MAPK pathway in the top 2 networks (**Table 5.2**). There is nothing in the literature describing a relationship between *Dck* and any of the genes listed in **Table 5.1**.

The expression of the genes identified in **Table 5.1** will need to be verified by qPCR, and evaluated in the other *Dck* KO B117P cell lines (T2A and T11A), as well as *Dck* KOs in the B140P cell lines. If confirmed, generation of a *DCK* knockout human AML cell line will be needed to confirm the relationship of *DCK* to these same genes in human cells. Evaluation of several human AML cell lines is underway to find cells with

moderate *DCK* expression levels, and *DCK* is also being sequenced to confirm there are no *DCK* mutations present.

Importance of *Ras* regulation in Ara-C resistance

There are several indicators in the research implicating the Ras/MAPK pathway in Ara-C resistance, including: (1) downregulation of the Ras inhibitors *Dab2ip* and *Dab2* in the Ara-C resistant cell lines (**Table 3.1** and **Supplementary Table C.3**, respectively), (2) the introduction of missense mutations into three members of the Ras/MAPK pathway (*Prkacb* and *Rasgrp2* in B117H and *Map3k11* in B140H), (3) a possible synergy between Ara-C and the MEK inhibitor PD035901 (**Table 4.6**), (4) *S100a8* and *S100a9*, activators of MAPK, are upregulated in the *Dck* KO cells and downregulated in the *Dck* KO/*DCK* OE cells (**Table 5.1**), and (5) the identification of the Ras/MAPK pathway as the primary network associated with genes whose regulation changed when *Dck* was knocked out (**Table 5.2**). Furthermore, a research study showed *Dck* can be downregulated by MAP kinase signaling in human tonsillar lymphocytes (Keszler et al., 2005). These data implicate a bidirectional relationship between Dck and the Ras/MAPK pathway. To study the role of the Ras/MAPK pathway in Ara-C resistance and its ability to compensate for the loss of *Dck*, *Raf1* will be knocked out, using either TALENs or CRISPR in the all four BXH-2 cell lines, and the *Raf1* KO cells will be evaluated for changes in response to Ara-C, changes in *Dck* expression, and changes in proliferation rates. In addition, flow cytometry of the cells will be conducted using antibodies to phospho-c-Raf, phospho-MEK1/2, and phospho-ERK1/2, to determine the activation status of the various members of the Ras/MAPK pathway. If the results are compelling, drugs targeting the Ras/MAPK pathway will be tested in

combination with Ara-C to determine their combined effectiveness in a variety of human and murine AML cell lines. The challenge of these experiments will be to find the optimal doses and the timing of the doses for the most effective outcome.

Mechanisms of glucocorticoid resistance in the B117P cell lines

A number of genes, which may be associated with innate resistance to glucocorticoids in some AML cells, have been identified in **Tables 4.4** and **4.5**. They need to be evaluated for their specific role in a cell's response to glucocorticoids. Similar techniques can be used as those defined for finding synergistic partners for Ara-C. Interestingly, *Mpo* and *Cd14* appear on both lists as being associated with Ara-C resistance and glucocorticoid resistance.

Another interesting relationship, which needs to be studied in more depth, is the one between Dck and calprotectin. Calprotectin is a tetramer, composed of two S100a8 and two S100a9 molecules, and found in high levels in neutrophils, from which it is released into the blood stream. This response is typically associated with infections/inflammation (Hessian and Fisher, 2001). Calprotectin has also been shown to have an active role in apoptosis (Zali et al., 2007). Its role in cancer has not been elucidated, although it has been shown to be elevated in colorectal carcinoma (Stulik et al., 1999), and the up-regulation of calprotectin was shown to be associated with glucocorticoid resistance in the treatment of acute lymphoid leukemia (Spijkers-Hagelstein et al., 2012). In our study the calprotectin genes, S100a8 and S100a9, were highly up-regulated in the B117P T6B *Dck* KO cells when compared to the B117P cells. The expression levels or the calprotectin genes were subsequently reduced when *DCK* was introduced into the B117P T6B *Dck* KO cells. This trend implicates a direct

relationship between Dck and calprotectin. And yet, there are no significant differences in the expression levels of the calprotectin genes in the four BXH-2 cell lines, two of which had defective Dck function. Furthermore, if the role of calprotectin in AML is similar to its role in ALL, i.e. higher levels of calprotectin result in glucocorticoid resistance, the B117P cells, which are highly resistance to glucocorticoids, should have a much higher level of calprotectin than the more glucocorticoid sensitive B117H cells. It will be interesting to see if there is a change in glucocorticoid sensitivity in direct response to the loss of *Dck*, by testing the *Dck* KO and *Dck* KO/*DCK* OE cells for changes in response to prednisolone and dexamethasone.

Determining the compensating role of calprotectin in the *Dck* KO cells will be quite complicated, especially since S100a8 and S100a9 can function independently or in the form the tetramer calprotectin. Fluorescent microscopy of the B117P cells, *Dck* KO cells, and the *Dck* KO/*DCK* OE cells, using different fluorescently labeled antibodies for the S100a8 and S100a9 proteins, can be used to determine the location of the proteins in the cell, and whether the proteins are co-localized. Immunoprecipitation of the S100a8 and S100a9 proteins from the B117P cells, *Dck* KO cells and the *Dck* KO/*DCK* OE cells, followed by mass spectrometry of the protein complexes can identify the binding partners for each protein. This will help to determine the primary function of the S100a8 and S100a9 in the *Dck* KO cells. For example, S100A8 or S100A9 may bind to TLR4 or RAGE to activate the MAPK pathway (Ichikawa et al., 2011). In addition, the relationship between DCK and calprotectin needs to be evaluated in human cell lines, as well primary AML and ALL cells. If a compensating relationship between Dck and

calprotectin is confirmed in AML cells, drug treatments can be devised to exploit this relationship.

Cladribine's role in AML toxicity

Based on the data showing cladribine is synergistic with Ara-C in Ara-C sensitive cells and yet experiences only a 10-fold increase in IC_{50} when Dck is knocked out, it appears cladribine has a second method of toxicity in AML. A second mode of toxicity for cladribine in hairy cell leukemia was suggested, involving the triggering of apoptosis by changing the transmembrane potential of the mitochondria (Marzo et al., 2001). To identify a second mode of toxicity in the AML cells, cladribine resistance cell lines will need to be developed which have a cladribine IC_{50} at least 100 times normal. Due to presence of a mutagenic retrovirus, the B117P and B140P cell lines will be ideal for this project. RNA-seq will be used to evaluate the changes required for the cells to become cladribine resistant, just as was done in defining the mechanisms of resistance in the Ara-C cell lines.

An In Vivo Model of Ara-C Resistance

Finding drug targets and drug combinations to test *in vitro* is only the first step in finding effective drug treatments for AML. These drugs need to be tested in an *in vivo* animal model of AML, which includes an Ara-C resistant derivative. To this end, a bank of mouse AML cells was established from genetically engineered mice having a Class I dox-inducible Tet-off *Nras* oncogene, and a Class II *Mll-AF9* fusion (TNM model) (Kim et al., 2009). RNA-sequencing of the TNM cells showed expression levels of *Dck* of ~20 FPKMs, consistent with the *Dck* expression levels found in the B117P and B140P cells.

IGV visualization of *Dck* found normal splicing patterns and no mutations in the TNM primary AML cells. The cells can transplanted via tail vein injection into SCID/beige mice and within 14-16 days white blood cell (WBC) counts will start to elevate. Without treatment the mice will become moribund by day 17-18 with WBC counts > 200K cells/μl. Treatment with 50 mg/kg/day of Ara-C daily over 5 days starting at day 16 after AML transplantation with reduce the WBC count significantly, usually below 50K cells/μl, but the mice will expire by day 22 due to anemia, resulting from loss of red blood cells.

To determine if Ara-C resistant cells could be developed by repeated exposure of the cells to Ara-C via passages through multiple SCID/beige mice, two strategies were implemented. In the first strategy, the mice were started with Ara-C treatment once their WBC count reached > 150K cells/μl and they received 4 injections of 60 mg/kg at 12 hour intervals. When they became moribund, their WBC count was tested and the spleen cells from the mouse with the highest WBC counts were implanted into the next set of SCID/beige mice. After 4 passages the cells were evaluated for drug resistance. 20 mice were transplanted with AML, 10 receiving cells that had never been exposed to Ara-C, and 10 receiving cells that had been passaged and exposed to Ara-C 4 times. The only difference in the two groups was the time of the initial onset of the disease. (**Table 5.3**)

Since this approach only succeeded in selecting for slower growing AML cells, rather than Ara-C resistant cells, a second approach was tried. Starting at 3 days following transplantation, the mice were given lower doses of Ara-C of 50 mg/kg every other day. These lower doses could be given indefinitely without causing anemia in the mice. Once the mice became moribund, the spleen cells from the mouse with the highest

WBC count were transplanted into the next set of mice. In each subsequent passage, the mice became moribund 3-4 days earlier than the previous passage (**Table 5.4**)

After four passages the cells were compared to cells which had never been exposed to Ara-C (**Table 5.5**). Although there was some drop in the WBC count in the mice with AML cells previously exposed to Ara-C, the mice were moribund within 24 hours of starting treatment with Ara-C, indicating the mice were indeed more resistant to Ara-C. In addition, during the necropsy of the mice, the Ara-C resistant mice presented with a dark discoloration in the bowels. Pathological examination of the tissues, as well as RNA-sequencing would be the next steps in understanding the changes that took place in the cells to make them more Ara-C resistant. It would also be interesting to compare the response of the cells to cladribine and glucocorticoids *in vivo*.

Conclusions

The BXH-2 derived cell lines (B117P and B140P) and their Ara-C resistant derivatives (B117H and B140H) have been valuable tools in understanding the molecular changes taking place in cells in response to Ara-C exposure, such as the loss of Dck function, which has also been documented in human Ara-C resistant AML. The use of RNA-seq was instrumental in describing the mutations found in *Dck* in the Ara-C resistant cell lines. Ironically, the focus of research during the past 15 years on the use of gene expression microarray analysis has actually resulted in a delay in the determination of the importance of loss of Dck function in Ara-C resistance, since mutations in *Dck*, which radically alter Dck function leading to Ara-C resistance, may not result in changes in *Dck* expression.

The drug screens and drug assays conducted on these cell lines are consistent with results seen in clinical setting, such as leukemias being responsive to Ara-C or glucocorticoids, but typically not both, and a synergy between Ara-C and cladribine in *de novo* disease, but not refractory disease. The continued interrogation of these cell lines, as well as the study of human AML cell lines and primary AML samples, may provide a better understanding of why leukemias respond differently to these drugs. This knowledge will assist clinicians in selecting the most appropriate drugs to treat patients based on the molecular profile of each individual's AML. The establishment of a molecular profile for each patient will depend on *in silico* analysis of the patient's AML using tools such as MMuFLR.

Table 5.1: Changes in the B117P T6B *Dck* KO Cells, Which Were Reverted by the Introduction of *DCK*

RNA-seq gene expression analysis using Cuffdiff and comparing the B117P sample to the B117P T6B *Dck* KO and the T6B *DCK* OE (induced with dox). Includes genes with 2-fold changes or better and total FPKM shift of greater than 5 FPKMs. Genes are sorted by decreasing fold change of the KO cells compared to parental cells. List was verified in IGV and only protein coding genes are shown. Stefin family genes include *Stfa3*, *Stfa2l1*, *Stfa2*, *Gm5416*, *BC100530*, *GM5483*, *Stfa1*, and *2010005H15Rik*, and calprotectin genes include *S100a8* and *S100a9*.

		FPKMs			Fold change	
gene_id	locus	B117P	B117P Dck_KO	B117P DCK_OE	Dck_KO vs. B117P	DCK_OE vs. Dck_KO
Stfa3	chr16:36450622-36455478	0.001	20.696	0.176	20695.800	-117.398
Stfa2l1	chr16:36156896-36162034	0.001	18.518	0.786	18517.700	-23.555
Stfa2	chr16:36404031-36408449	0.001	3.605	0.001	3604.850	-3604.850
Gm5416	chr16:36210488-36217874	0.001	1.268	0.001	1268.020	-1268.020
BC100530	chr16:36359467-36367656	0.140	25.611	0.935	182.634	-27.405
Gm5483	chr16:36184297-36188196	0.154	26.489	1.667	171.817	-15.889
Stfa1	chr16:36277233-36285457	0.606	56.992	4.290	94.057	-13.284
2010005H15Rik	chr16:36221647-36257513	0.109	5.505	0.115	50.281	-47.689
5830411N06Rik	chr7:147433199-147485690	1.612	14.898	6.195	9.243	-2.405
Scd1	chr19:44468939-44482199	0.203	1.871	0.751	9.216	-2.491
S100a9	chr3:90496554-90499613	10.106	87.544	9.172	8.663	-9.544
Lcn2	chr2:32240156-32243259	1.023	8.522	0.675	8.330	-12.630
S100a8	chr3:90472992-90473956	13.494	102.208	14.349	7.575	-7.123
Hsd11b1	chr1:195047834-195090239	0.322	2.123	0.621	6.604	-3.418
Slc22a4	chr11:53796627-53841592	0.571	2.789	0.910	4.882	-3.067
Dnmt3l	chr10:77492775-77526360	0.621	3.004	1.378	4.833	-2.179
Trat1	chr16:48734802-48772069	0.393	1.590	0.749	4.046	-2.123
Ltf	chr9:110921795-110945270	0.532	2.088	0.015	3.922	-137.298
Il7r	chr15:9435913-9459631	0.917	3.096	1.112	3.376	-2.786
Thbs1	chr2:117937657-117952869	2.053	6.170	1.842	3.006	-3.349
Ngp	chr9:110322311-110325516	1.772	4.927	0.620	2.781	-7.942
Lyz2	chr10:116714596-116719328	68.691	183.206	84.245	2.667	-2.175
Vegfa	chr17:46153942-46169326	14.748	7.021	18.350	-2.100	2.613
Ddit4	chr10:59412422-59414518	7.464	3.531	8.065	-2.114	2.284
Aldoc	chr11:78137699-78140262	2.268	0.985	10.068	-2.302	10.220
Gp9	chr6:87728129-87729756	1.534	0.632	1.786	-2.427	2.826
Asns	chr6:7625170-7643182	25.881	10.409	22.732	-2.487	2.184
Cth	chr3:157557211-157588027	3.954	1.557	4.580	-2.539	2.941
Gfi1b	chr2:28464969-28477502	16.831	6.023	12.907	-2.795	2.143
Gcet2	chr16:45610552-45622980	1.539	0.519	1.935	-2.962	3.725
Ckb	chr12:112907565-112910549	3.225	1.015	3.574	-3.179	3.522
Trib3	chr2:152163160-152169796	5.362	1.446	6.373	-3.708	4.407
Cd244	chr1:173489323-173515447	13.019	3.338	7.509	-3.900	2.250
Nupr1	chr7:133766759-133768984	11.413	2.883	18.901	-3.959	6.557
Slc6a9	chr4:117507862-117545075	9.404	2.119	8.021	-4.439	3.786
Clec11a	chr7:51559135-51562329	37.443	8.418	27.679	-4.448	3.288
Ccl3	chr11:83461344-83462880	165.965	34.841	83.274	-4.763	2.390
Cyp4f18	chr8:74512380-74533525	2.489	0.449	0.950	-5.539	2.115
Slc1a4	chr11:20202182-20232716	2.770	0.489	1.745	-5.661	3.566
Chac1	chr2:119176977-119180062	3.746	0.625	5.098	-5.991	8.152
Malat1	chr19:5795689-5802671	31.467	4.852	20.938	-6.485	4.315

Table 5.2: Ras/MAPK Pathway Associated with Genes Differentially Expressed in *Dck* KO cells

Networks identified by Ingenuity Pathway Analysis (Ingenuity® Systems, www.ingenuity.com) from the list of genes differentially expressed in the B117P T6B *Dck* KO cells when compared to the B117P cells. Genes from **Table 5.1** are in bold.

ID	Molecules in Network	Score	Focus Molecules	Top Functions
1	Alp, Alpha catenin, Ap1, **ASNS**, **CCL3L1/CCL3L3**, **CD244**, **CKB**, **CLEC11A**, **CTH**, **DDIT4**, elastase, ERK1/2 ,F Actin, Fcer1, Fibrinogen, **HSD11B1**, IL1 ,IL12 (complex),Immunoglobulin, **LCN2**, **LTF**, **Lyz1/Lyz2** ,MAP2K1/2, Mek, P38 MAPK, p70 S6k ,PDGF BB, **S100A8**, **S100A9**, **Stfa1 (includes others)**, Tgf beta, **THBS1**, **TRAT1**, **TRIB3**, **VEGFA**	45	18	Hematological System Development and Function, Inflammatory Response, Tissue Morphology
2	ABLIM, ADRB, Akt, **ALDOC**, AMPK, ANGEL1 ,caspase, CD3, chemokine, Cpe, Cyp2j9, **CYP4F2**, EPGN, ERK, estrogen receptor, FSH, **GP9**, IgG, **IL7R**, Insulin, Lh, LOC81691, Mapk, NFkB (complex), **NUPR1**, PI3K (complex), Pkc(s), Rac, Ras, **SCD**, **SLC22A4**, **SLC6A9**, **Stfa3**, Utp14b, Vegf	19	9	Hematological Disease, Hereditary Disorder, Lipid Metabolism
3	ADIRF, AGPAT6, ANKRD6, ASRGL1, BEX1, **CHAC1**, **CSTA**, CTSB, DHX34, **DNMT3L**, ENTPD6, FAM193B, GALT, **GFI1B**, GON4L, GSK3B, HDAC1, IFFO1, LPHN3, MON1A, MT1F, NFE4, **Ngp**, NTSR1, OSR1, Peg12, PHPT1, RPL36AL, RSL24D1, **SLC1A4**, SPICE1, UBC, UHRF1BP1, ZNF135, ZNF160	11	6	Gene Expression, Embryonic Development, Organismal Development
4	**5830411N06Rik**, IL17A	3	1	Cell Death and Survival, Cell-To-Cell Signaling and Interaction, Cellular Development
5	mir-223, **Stfa2/Stfa2l1**, TLR3, TLR9	2	1	Cell-To-Cell Signaling and Interaction, Inflammatory Response, Cancer
6	CDH1, IGFBP2, **MALAT1**, mir-182, miR-182-5p (and other miRNAs w/seed UUGGCAA), NCAM1, SNAI2, SRSF1	2	1	Cancer, Reproductive System Disease, Endocrine System Disorders

Table 5.3: AML Cells Exposed to High Doses of Ara-C for 4 Passages Grew Slower, but Had the Same Response the Ara-C as Cells Not Previously Exposed to Ara-C

Two cohorts of mice received AML transplants. The first 10 mice received AML cells from the spleen of a mouse which had never been exposed to Ara-C, and the second 10 mice received AML spleen cells which had been previously passaged through 4 mice and during each passage received high dose exposure to Ara-C (4 doses of 60 mg/kg) at 12 hour intervals once the mice had WBC counts > 150K/μl. WBC counts of the mice were taken every 3 days and the mice were started on Ara-C when any of the mice in the cohort exceeded a WBC of 150K/μl. Mice received injections of Ara-C (50 mg/kg) every 24 hours for 4 days. Tissues were harvested when the mice became moribund.

Mouse	ID	Previous Ara-C Exposure	WBC counts											
			Day 16	Day 17	Day 18	Day 19	Day 20	Day 21	Day 22	Day 23	Day 24	Day 25	Day 26	Day 27
SC389	00	None	122.56			27.56			22.14					
SC390	01	None	132.96			22.50			20.80					
SC391	03	None	178.10			56.48								
SC392	10	None		126.92			21.30							
SC393	30	None		115.18			14.12							
SC394	00	None		122.16			19.96							
SC395	01	None			84.98			18.54						
SC396	03	None			85.62			24.70						
SC397	10	None			85.06			87.44						
SC398	30	None			83.24			8.22						
SC399	00	4 sets	5.66			46.32			78.14			63.52		
SC400	01	4 sets	8.10			50.24			10.12			3.06		
SC401	03	4 sets	5.50			63.68			16.84			5.44		
SC402	10	4 sets		12.05			117.42			3.36				
SC403	30	4 sets		22.36			125.88			5.40		1.50		
SC404	00	4 sets		2.46			5.34			1.32			14.32	
SC405	01	4 sets			4.74			6.12			1.64		2.08	
SC406	03	4 sets			23.16			25.22			2.20		2.58	
SC407	10	4 sets			4.72			5.70			1.50		19.22	
SC408	30	4 sets			61.28									

Received Ara-C injections
Moribund

Table 5.4: AML Cells Exposed to Low Continuous Doses of Ara-C for 4 Passages Resulted in Earlier Onset of Disease and Earlier Morbidity

TNM AML cells were passaged 4 times through SCID beige mice. The mice were treated with 50 mg/kg of Ara-C every other day starting on the 3rd day after transplantation and continuing until the mice were moribund.

Mouse	Passage	Day 15	Day 17	Day 18	Day 19	Day 21	Day 22	Day 23	Day 25	Day 26	Day 29
SC422	1		3.22			13.72			78.46		378.8
SC453	2	6.52			1.54			40.24		60.92	
SC474	3		35.86			93.58	100.56				
SC529	4	3.06		9.06	38.3						

Moribund

Table 5.5: AML Cells Exposed to Low Continuous Doses of Ara-C for 4 Passages Became Less Responsive to Ara-C

Response of TNM AML cells passaged 4 times through SCID beige mice and exposed to low doses of Ara-C were compared to cells which had never been exposed to Ara-C. On day 15 following transplantation the mice were treated with 50 mg/kg of Ara-C daily for 5 day or until moribund.

Mouse	ID	Previous Ara-C Exposure	WBC Count							
			Day 15	Day 16	Day 17	Day 18	Day 19	Day 20	Day 21	Day 22
SC519	00	None	67.70			12.04				
SC520	01	None	103.13			78.46				
SC521	03	None	2.92			1.68				
SC522	10	None	80.40			14.64				
SC523	30	None	76.46			14.32				
SC524	00	4 sets	119.16		49.41					
SC525	01	4 sets	51.34		46.82					
SC526	03	4 sets	129.92		78					
SC527	10	4 sets	208.40		107.28					
SC528	30	4 sets	136.50							

Transplantion failed

Moribund

References

(1966). Treatment of acute leukaemia in adults: comparison of steroid and mercaptopurine therapy, alone and in conjunction. Second report to the Medical Research Council of the Working Party on the evaluation of different methods of therapy in leukaemia. Br Med J *1*, 1383-1389.

Abe, S., Funato, T., Takahashi, S., Yokoyama, H., Yamamoto, J., Tomiya, Y., Yamada-Fujiwara, M., Ishizawa, K., Kameoka, J., Kaku, M.*, et al.* (2006). Increased expression of insulin-like growth factor i is associated with Ara-C resistance in leukemia. Tohoku J Exp Med *209*, 217-228.

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S., and Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. Nat Methods *7*, 248-249.

Agarwai, M. K. (1996). The antiglucocorticoid action of mifepristone. Pharmacol Ther *70*, 183-213.

Akyerli, C. B., Beksac, M., Holko, M., Frevel, M., Dalva, K., Ozbek, U., Soydan, E., Ozcan, M., Ozet, G., Ilhan, O.*, et al.* (2005). Expression of IFITM1 in chronic myeloid leukemia patients. Leuk Res *29*, 283-286.

Anastasi, A., Brown, M. A., Kembhavi, A. A., Nicklin, M. J., Sayers, C. A., Sunter, D. C., and Barrett, A. J. (1983). Cystatin, a protein inhibitor of cysteine proteinases. Improved purification from egg white, characterization, and detection in chicken serum. Biochem J *211*, 129-138.

Beachy, S. H., and Aplan, P. D. (2010). Mouse models of myelodysplastic syndromes. Hematol Oncol Clin North Am *24*, 361-375.

Bedigian, H. G., Taylor, B. A., and Meier, H. (1981). Expression of murine leukemia viruses in the highly lymphomatous BXH-2 recombinant inbred mouse strain. J Virol *39*, 632-640.

Bennett, J. M., Catovsky, D., Daniel, M. T., Flandrin, G., Galton, D. A., Gralnick, H. R., and Sultan, C. (1976). Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. Br J Haematol *33*, 451-458.

Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol *Chapter 19*, Unit 19 10 11-21.

Brzin, J., Popovic, T., Turk, V., Borchart, U., and Machleidt, W. (1984). Human cystatin, a new protein inhibitor of cysteine proteinases. Biochem Biophys Res Commun *118*, 103-109.

Bullard, J. H., Purdom, E., Hansen, K. D., and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics *11*, 94.

Cadepond, F., Ulmann, A., and Baulieu, E. E. (1997). RU486 (mifepristone): mechanisms of action and clinical uses. Annu Rev Med *48*, 129-156.

Cai, J., Damaraju, V. L., Groulx, N., Mowles, D., Peng, Y., Robins, M. J., Cass, C. E., and Gros, P. (2008). Two distinct molecular mechanisms underlying cytarabine resistance in human leukemic cells. Cancer Res *68*, 2349-2357.

Calvisi, D. F., Ladu, S., Conner, E. A., Seo, D., Hsieh, J. T., Factor, V. M., and Thorgeirsson, S. S. (2011). Inactivation of Ras GTPase-activating proteins promotes unrestrained activity of wild-type Ras in human liver cancer. J Hepatol *54*, 311-319.

Carlson, D. F., Tan, W., Lillico, S. G., Stverakova, D., Proudfoot, C., Christian, M., Voytas, D. F., Long, C. R., Whitelaw, C. B., and Fahrenkrug, S. C. (2012). Efficient TALEN-mediated gene knockout in livestock. Proc Natl Acad Sci U S A *109*, 17382-17387.

Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J. A., Somia, N. V., Bogdanove, A. J., and Voytas, D. F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res *39*, e82.

Chan, I. T., and Gilliland, D. G. (2004). Oncogenic K-ras in mouse models of myeloproliferative disease and acute myeloid leukemia. Cell Cycle *3*, 536-537.

Chang, K. P., Wu, C. C., Chen, H. C., Chen, S. J., Peng, P. H., Tsang, N. M., Lee, L. Y., Liu, S. C., Liang, Y., Lee, Y. S*., et al.* (2010). Identification of candidate nasopharyngeal carcinoma serum biomarkers by cancer cell secretome and tissue transcriptome analysis: potential usage of cystatin A for predicting nodal stage and poor prognosis. Proteomics *10*, 2644-2660.

Chang, Y. F., and Carman, G. M. (2008). CTP synthetase and its role in phospholipid synthesis in the yeast Saccharomyces cerevisiae. Prog Lipid Res *47*, 333-339.

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., and Chan, A. P. (2012). Predicting the functional effect of amino acid substitutions and indels. PLoS One *7*, e46688.

Chou, T. C., Arlin, Z., Clarkson, B. D., and Phillips, F. S. (1977). Metabolism of 1-beta-D-arabinofuranosylcytosine in human leukemic cells. Cancer Res *37*, 3561-3570.

Chou, T. C., and Talalay, P. (1984). Quantitative analysis of dose-effect relationships: the combined effects of multiple drugs or enzyme inhibitors. Adv Enzyme Regul *22*, 27-55.

Chuthapisith, S., Bean, B. E., Cowley, G., Eremin, J. M., Samphao, S., Layfield, R., Kerr, I. D., Wiseman, J., El-Sheemy, M., Sreenivasan, T., and Eremin, O. (2009). Annexins in human breast cancer: Possible predictors of pathological response to neoadjuvant chemotherapy. Eur J Cancer *45*, 1274-1281.

Cingolani, P., Patel, V. M., Coon, M., Nguyen, T., Land, S. J., Ruden, D. M., and Lu, X. (2012). Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. Front Genet *3*, 35.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M.*, et al.* (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet *43*, 491-498.

Di Santo, E., Sironi, M., Mennini, T., Zinetti, M., Savoldi, G., Di Lorenzo, D., and Ghezzi, P. (1996). A glucocorticoid receptor-independent mechanism for neurosteroid inhibition of tumor necrosis factor production. Eur J Pharmacol *299*, 179-186.

Dohner, H., Estey, E. H., Amadori, S., Appelbaum, F. R., Buchner, T., Burnett, A. K., Dombret, H., Fenaux, P., Grimwade, D., Larson, R. A.*, et al.* (2010). Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. Blood *115*, 453-474.

Dohner, K., and Dohner, H. (2008). Molecular characterization of acute myeloid leukemia. Haematologica *93*, 976-982.

Estey, E. H. (2012). Acute myeloid leukemia: 2012 update on diagnosis, risk stratification, and management. Am J Hematol *87*, 89-99.

Estey, E. H. (2013). Acute myeloid leukemia: 2013 update on risk-stratification and management. Am J Hematol *88*, 318-327.

Fazekas, A., Steeves, R., and Newmaster, S. (2010). Improving sequencing quality from PCR products containing long mononucleotide repeats. Biotechniques *48*, 277-285.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.

Flicek, P., Ahmed, I., Amode, M. R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S.*, et al.* (2013). Ensembl 2013. Nucleic Acids Res *41*, D48-55.

Floquet, C., Hatin, I., Rousset, J. P., and Bidou, L. (2012). Statistical analysis of readthrough levels for nonsense mutations in mammalian cells reveals a major determinant of response to gentamicin. PLoS Genet *8*, e1002608.

Forbes, S. A., Bindal, N., Bamford, S., Cole, C., Kok, C. Y., Beare, D., Jia, M., Shepherd, R., Leung, K., Menzies, A.*, et al.* (2011). COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res *39*, D945-950.

Frankfurt, O., and Rosen, S. T. (2004). Mechanisms of glucocorticoid-induced apoptosis in hematologic malignancies: updates. Curr Opin Oncol *16*, 553-563.

Fumoto, S., Shimokuni, T., Tanimoto, K., Hiyama, K., Otani, K., Ohtaki, M., Hihara, J., Yoshida, K., Hiyama, E., Noguchi, T., and Nishiyama, M. (2008). Selection of a novel drug-response predictor in esophageal cancer: a novel screening method using microarray and identification of IFITM1 as a potent marker gene of CDDP response. Int J Oncol *32*, 413-423.

Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J.*, et al.* (2005). Galaxy: a platform for interactive large-scale genome analysis. Genome Res *15*, 1451-1455.

Gilliland, D. G., and Griffin, J. D. (2002). Role of FLT3 in leukemia. Curr Opin Hematol *9*, 274-281.

Goecks, J., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol *11*, R86.

Grant, S. (1998). Ara-C: cellular and molecular pharmacology. Adv Cancer Res *72*, 197-233.

Grem, J. L., Geoffroy, F., Politi, P. M., Cuddy, D. P., Ross, D. D., Nguyen, D., Steinberg, S. M., and Allegra, C. J. (1995). Determinants of sensitivity to 1-beta-D-arabinofuranosylcytosine in HCT 116 and NCI-H630 human colon carcinoma cells. Mol Pharmacol *48*, 305-315.

Griffin, J. D., Ritz, J., Nadler, L. M., and Schlossman, S. F. (1981). Expression of myeloid differentiation antigens on normal and malignant myeloid cells. J Clin Invest *68*, 932-941.

Guschin, D. Y., Waite, A. J., Katibah, G. E., Miller, J. C., Holmes, M. C., and Rebar, E. J. (2010). A rapid and general assay for monitoring endogenous gene modification. Methods Mol Biol *649*, 247-256.

Harrington, C. A., Rosenow, C., and Retief, J. (2000). Monitoring gene expression using DNA microarrays. Curr Opin Microbiol *3*, 285-291.

Hessian, P. A., and Fisher, L. (2001). The heterodimeric complex of MRP-8 (S100A8) and MRP-14 (S100A9). Antibody recognition, epitope definition and the implications for structure. Eur J Biochem *268*, 353-363.

Hicsonmez, G. (2006). The effect of steroid on myeloid leukemic cells: the potential of short-course high-dose methylprednisolone treatment in inducing differentiation, apoptosis and in stimulating myelopoiesis. Leuk Res *30*, 60-68.

Holowiecki, J., Grosicki, S., Robak, T., Kyrcz-Krzemien, S., Giebel, S., Hellmann, A., Skotnicki, A., Jedrzejczak, W. W., Konopka, L., Kuliczkowski, K.*, et al.* (2004). Addition of cladribine to

daunorubicin and cytarabine increases complete remission rate after a single course of induction treatment in acute myeloid leukemia. Multicenter, phase III study. Leukemia *18*, 989-997.

Horton, S. J., and Huntly, B. J. (2012). Recent advances in acute myeloid leukemia stem cell biology. Haematologica *97*, 966-974.

Ichikawa, M., Williams, R., Wang, L., Vogl, T., and Srikrishna, G. (2011). S100A8/A9 activate key genes and pathways in colon tumor progression. Mol Cancer Res *9*, 133-148.

Ishikawa, Y., Kiyoi, H., Tsujimura, A., Miyawaki, S., Miyazaki, Y., Kuriyama, K., Tomonaga, M., and Naoe, T. (2009). Comprehensive analysis of cooperative gene mutations between class I and class II in de novo acute myeloid leukemia. Eur J Haematol *83*, 90-98.

Jiang, Y., Prunier, C., and Howe, P. H. (2008). The inhibitory effects of Disabled-2 (Dab2) on Wnt signaling are mediated through Axin. Oncogene *27*, 1865-1875.

Jin, G., Matsushita, H., Asai, S., Tsukamoto, H., Ono, R., Nosaka, T., Yahata, T., Takahashi, S., and Miyachi, H. (2009). FLT3-ITD induces ara-C resistance in myeloid leukemic cells through the repression of the ENT1 expression. Biochem Biophys Res Commun *390*, 1001-1006.

Jordan, C. T. (2007). The leukemic stem cell. Best Pract Res Clin Haematol *20*, 13-18.

Kang, G. J., Cooney, D. A., Moyer, J. D., Kelley, J. A., Kim, H. Y., Marquez, V. E., and Johns, D. G. (1989). Cyclopentenylcytosine triphosphate. Formation and inhibition of CTP synthetase. J Biol Chem *264*, 713-718.

Keszler, G., Virga, S., Spasokoukotskaja, T., Bauer, P. I., Sasvari-Szekely, M., and Staub, M. (2005). Activation of deoxycytidine kinase by deoxyadenosine: implications in deoxyadenosine-mediated cytotoxicity. Arch Biochem Biophys *436*, 69-77.

Kim, W. I., Matise, I., Diers, M. D., and Largaespada, D. A. (2009). RAS oncogene suppression induces apoptosis followed by more differentiated and less myelosuppressive disease upon relapse of acute myeloid leukemia. Blood *113*, 1086-1096.

Kocabas, N. A., Aksoy, P., Pelleymounter, L. L., Moon, I., Ryu, J. S., Gilbert, J. A., Salavaggione, O. E., Eckloff, B. W., Wieben, E. D., Yee, V.*, et al.* (2008). Gemcitabine pharmacogenomics: deoxycytidine kinase and cytidylate kinase gene resequencing and functional genomics. Drug Metab Dispos *36*, 1951-1959.

Koeffler, H. P., Ranyard, J., and Pertcheck, M. (1985). Myeloperoxidase: its structure and expression during myeloid differentiation. Blood *65*, 484-491.

Kongkham, P. N., Northcott, P. A., Ra, Y. S., Nakahara, Y., Mainprize, T. G., Croul, S. E., Smith, C. A., Taylor, M. D., and Rutka, J. T. (2008). An epigenetic genome-wide screen identifies SPINT2 as a novel tumor suppressor gene in pediatric medulloblastoma. Cancer Res *68*, 9945-9953.

Kubal, T., and Lancet, J. E. (2013). The thorny issue of relapsed acute myeloid leukemia. Curr Opin Hematol *20*, 100-106.

Kuopio, T., Kankaanranta, A., Jalava, P., Kronqvist, P., Kotkansalo, T., Weber, E., and Collan, Y. (1998). Cysteine proteinase inhibitor cystatin A in breast cancer. Cancer Res *58*, 432-436.

Kwon, C. H., Moon, H. J., Park, H. J., Choi, J. H., and Park do, Y. (2013). S100A8 and S100A9 promotes invasion and migration through p38 mitogen-activated protein kinase-dependent NF-kappaB activation in gastric cancer cells. Mol Cells *35*, 226-234.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Lapidot, T., Sirard, C., Vormoor, J., Murdoch, B., Hoang, T., Caceres-Cortes, J., Minden, M., Paterson, B., Caligiuri, M. A., and Dick, J. E. (1994). A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature *367*, 645-648.

Largaespada, D. A., Shaughnessy, J. D., Jr., Jenkins, N. A., and Copeland, N. G. (1995). Retroviral integration at the Evi-2 locus in BXH-2 myeloid leukemia cell lines disrupts Nf1 expression without changes in steady-state Ras-GTP levels. J Virol *69*, 5095-5102.

Lee, H. M., Zhang, H., Schulz, V., Tuck, D. P., and Forget, B. G. (2010). Downstream targets of HOXB4 in a cell line model of primitive hematopoietic progenitor cells. Blood *116*, 720-730.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Li, Z., Herold, T., He, C., Valk, P. J., Chen, P., Jurinovic, V., Mansmann, U., Radmacher, M. D., Maharry, K. S., Sun, M.*, et al.* (2013). Identification of a 24-gene prognostic signature that improves the European LeukemiaNet risk classification of acute myeloid leukemia: an international collaborative study. J Clin Oncol *31*, 1172-1181.

Lotfi, K., Juliusson, G., and Albertioni, F. (2003). Pharmacological basis for cladribine resistance. Leuk Lymphoma *44*, 1705-1712.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823-826.

Malich, G., Markovic, B., and Winder, C. (1997). The sensitivity and specificity of the MTS tetrazolium assay for detecting the in vitro cytotoxicity of 20 chemicals using human cell lines. Toxicology *124*, 179-192.

Martin, P. J., Rizzo, J. D., Wingard, J. R., Ballen, K., Curtin, P. T., Cutler, C., Litzow, M. R., Nieto, Y., Savani, B. N., Schriber, J. R.*, et al.* (2012). First- and second-line systemic treatment of acute graft-versus-host disease: recommendations of the American Society of Blood and Marrow Transplantation. Biol Blood Marrow Transplant *18*, 1150-1163.

Marzo, I., Perez-Galan, P., Giraldo, P., Rubio-Felix, D., Anel, A., and Naval, J. (2001). Cladribine induces apoptosis in human leukaemia cells by caspase-dependent and -independent pathways acting on mitochondria. Biochem J *359*, 537-546.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res *20*, 1297-1303.

McPherson, A., Hormozdiari, F., Zayed, A., Giuliany, R., Ha, G., Sun, M. G., Griffith, M., Heravi Moussavi, A., Senz, J., Melnyk, N.*, et al.* (2011). deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. PLoS Comput Biol *7*, e1001138.

Meinke, A., Henics, T., Hanner, M., Minh, D. B., and Nagy, E. (2005). Antigenome technology: a novel approach for the selection of bacterial vaccine candidate antigens. Vaccine *23*, 2035-2041.

Mihelic, M., Teuscher, C., Turk, V., and Turk, D. (2006). Mouse stefins A1 and A2 (Stfa1 and Stfa2) differentiate between papain-like endo- and exopeptidases. FEBS Lett *580*, 4195-4199.

Momparler, R. L. (2005). Pharmacology of 5-Aza-2'-deoxycytidine (decitabine). Semin Hematol *42*, S9-16.

Moriarity, B. S., Rahrmann, E. P., Keng, V. W., Manlove, L. S., Beckmann, D. A., Wolf, N. K., Khurshid, T., Bell, J. B., and Largaespada, D. A. (2013). Modular assembly of transposon integratable multigene vectors using RecWay assembly. Nucleic Acids Res *41*, e92.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods *5*, 621-628.

Mrozek, K. (2008). Cytogenetic, molecular genetic, and clinical characteristics of acute myeloid leukemia with a complex karyotype. Semin Oncol *35*, 365-377.

Negoro, E., Yamauchi, T., Urasaki, Y., Nishi, R., Hori, H., and Ueda, T. (2011). Characterization of cytarabine-resistant leukemic cell lines established from five different blood cell lineages using gene expression and proteomic analyses. Int J Oncol *38*, 911-919.

Passegue, E., Jamieson, C. H., Ailles, L. E., and Weissman, I. L. (2003). Normal and leukemic hematopoiesis: are leukemias a stem cell disorder or a reacquisition of stem cell characteristics? Proc Natl Acad Sci U S A *100 Suppl 1*, 11842-11849.

Piazza, F., Gurrieri, C., and Pandolfi, P. P. (2001). The theory of APL. Oncogene *20*, 7216-7222.

Preisler, H. D., Epstein, J., Raza, A., Azarnia, N., Browman, G., Booker, L., Goldberg, J., Gottlieb, A., Brennan, J., Grunwald, H., and et al. (1984). Inhibition of DNA synthesis by cytosine arabinoside: relation to response of acute non-lymphocytic leukemia to remission induction therapy and to stage of the disease. Eur J Cancer Clin Oncol *20*, 1061-1068.

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., and Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173-1183.

Rathe, S. K., Johnson, J. E., Silverstein, K. A., Erdmann, J. J., Watson, A. L., Popescu, F. E., Ohlfest, J. R., and Largaespada, D. A. (2013). MMuFLR: missense mutation and frameshift location reporter. Bioinformatics.

Rathe, S. K., and Largaespada, D. A. (2010). Deoxycytidine kinase is downregulated in Ara-C-resistant acute myeloid leukemia murine cell lines. Leukemia *24*, 1513-1515.

Ravandi, F., and Estrov, Z. (2006). Eradication of leukemia stem cells as a new goal of therapy in leukemia. Clin Cancer Res *12*, 340-344.

Rhen, T., and Cidlowski, J. A. (2005). Antiinflammatory action of glucocorticoids--new mechanisms for old drugs. N Engl J Med *353*, 1711-1723.

Riva, C., el Khyari, S., Rustum, Y., and Barra, Y. (1995). Resistance to cytosine arabinoside in cells transfected with activated Ha-ras oncogene. Anticancer Res *15*, 1297-1302.

Roberts, A., Pimentel, H., Trapnell, C., and Pachter, L. (2011a). Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics *27*, 2325-2329.

Roberts, A., Trapnell, C., Donaghey, J., Rinn, J. L., and Pachter, L. (2011b). Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol *12*, R22.

Rollig, C., Bornhauser, M., Thiede, C., Taube, F., Kramer, M., Mohr, B., Aulitzky, W., Bodenstein, H., Tischler, H. J., Stuhlmann, R.*, et al.* (2011). Long-term prognosis of acute myeloid leukemia according to the new genetic risk classification of the European LeukemiaNet recommendations: evaluation of the proposed reporting system. J Clin Oncol *29*, 2758-2765.

Rubnitz, J. E., Crews, K. R., Pounds, S., Yang, S., Campana, D., Gandhi, V. V., Raimondi, S. C., Downing, J. R., Razzouk, B. I., Pui, C. H., and Ribeiro, R. C. (2009). Combination of cladribine and cytarabine is effective for childhood acute myeloid leukemia: results of the St Jude AML97 trial. Leukemia *23*, 1410-1416.

Rubnitz, J. E., Razzouk, B. I., Srivastava, D. K., Pui, C. H., Ribeiro, R. C., and Santana, V. M. (2004). Phase II trial of cladribine and cytarabine in relapsed or refractory myeloid malignancies. Leuk Res *28*, 349-352.

Schaaf, M. J., and Cidlowski, J. A. (2002). Molecular mechanisms of glucocorticoid action and resistance. J Steroid Biochem Mol Biol *83*, 37-48.

Schessl, C., Rawat, V. P., Cusan, M., Deshpande, A., Kohl, T. M., Rosten, P. M., Spiekermann, K., Humphries, R. K., Schnittger, S., Kern, W.*, et al.* (2005). The AML1-ETO fusion gene and the FLT3 length mutation collaborate in inducing acute leukemia in mice. J Clin Invest *115*, 2159-2168.

Schlossmacher, G., Stevens, A., and White, A. (2011). Glucocorticoid receptor-mediated apoptosis: mechanisms of resistance in cancer cells. J Endocrinol *211*, 17-25.

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Res *29*, 308-311.

Shi, J. Y., Shi, Z. Z., Zhang, S. J., Zhu, Y. M., Gu, B. W., Li, G., Bai, X. T., Gao, X. D., Hu, J., Jin, W.*, et al.* (2004). Association between single nucleotide polymorphisms in deoxycytidine kinase and treatment response among acute myeloid leukaemia patients. Pharmacogenetics *14*, 759-768.

Sigal, D. S., Miller, H. J., Schram, E. D., and Saven, A. (2010). Beyond hairy cell: the activity of cladribine in other hematologic malignancies. Blood *116*, 2884-2896.

Sionov, R. V. (2013). MicroRNAs and Glucocorticoid-Induced Apoptosis in Lymphoid Malignancies. ISRN Hematol *2013*, 348212.

Song, J. H., Kim, S. H., Kweon, S. H., Lee, T. H., Kim, H. J., and Kim, T. S. (2009). Defective expression of deoxycytidine kinase in cytarabine-resistant acute myeloid leukemia cells. Int J Oncol *34*, 1165-1171.

Spijkers-Hagelstein, J. A., Schneider, P., Hulleman, E., de Boer, J., Williams, O., Pieters, R., and Stam, R. W. (2012). Elevated S100A8/S100A9 expression causes glucocorticoid resistance in MLL-rearranged infant acute lymphoblastic leukemia. Leukemia *26*, 1255-1265.

Stegmann, A. P., Honders, W. H., Willemze, R., Ruiz van Haperen, V. W., and Landegent, J. E. (1995). Transfection of wild-type deoxycytidine kinase (dck) cDNA into an AraC- and DAC-resistant rat leukemic cell line of clonal origin fully restores drug sensitivity. Blood *85*, 1188-1194.

Stulik, J., Osterreicher, J., Koupilova, K., Knizek, Macela, A., Bures, J., Jandik, P., Langr, F., Dedic, K., and Jungblut, P. R. (1999). The analysis of S100A9 and S100A8 expression in matched sets of macroscopically normal colon mucosa and colorectal carcinoma: the S100A9 and S100A8 positive cells underlie and invade tumor mass. Electrophoresis *20*, 1047-1054.

Takahashi, H., Komatsu, N., Ibe, M., Ishida-Yamamoto, A., Hashimoto, Y., and Iizuka, H. (2007). Cystatin A suppresses ultraviolet B-induced apoptosis of keratinocytes. J Dermatol Sci *46*, 179-187.

Takahashi, S. (2011). Current findings for recurring mutations in acute myeloid leukemia. J Hematol Oncol *4*, 36.

Therrien, M., Michaud, N. R., Rubin, G. M., and Morrison, D. K. (1996). KSR modulates signal propagation within the MAPK cascade. Genes Dev *10*, 2684-2695.

Toonen, E. J., Fleuren, W. W., Nassander, U., van Lierop, M. J., Bauerschmidt, S., Dokter, W. H., and Alkema, W. (2011). Prednisolone-induced changes in gene-expression profiles in healthy volunteers. Pharmacogenomics *12*, 985-998.

Toy, G., Austin, W. R., Liao, H. I., Cheng, D., Singh, A., Campbell, D. O., Ishikawa, T. O., Lehmann, L. W., Satyamurthy, N., Phelps, M. E.*, et al.* (2010). Requirement for deoxycytidine kinase in T and B lymphocyte development. Proc Natl Acad Sci U S A *107*, 5551-5556.

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol *28*, 511-515.

Valk, P. J., Verhaak, R. G., Beijen, M. A., Erpelinck, C. A., Barjesteh van Waalwijk van Doorn-Khosrovani, S., Boer, J. M., Beverloo, H. B., Moorhouse, M. J., van der Spek, P. J., Lowenberg, B., and Delwel, R. (2004). Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med *350*, 1617-1628.

van Doorn-Khosrovani, S. B. (2004). Acute myeloid leukaemia: the search for novel prognostic markers: S.l. : Barjesteh van Waalwijk van Doorn).

Vardiman, J. W., Harris, N. L., and Brunning, R. D. (2002). The World Health Organization (WHO) classification of the myeloid neoplasms. Blood *100*, 2292-2302.

Veuger, M. J., Honders, M. W., Landegent, J. E., Willemze, R., and Barge, R. M. (2000). High incidence of alternatively spliced forms of deoxycytidine kinase in patients with resistant acute myeloid leukemia. Blood *96*, 1517-1524.

Veuger, M. J., Honders, M. W., Willemze, R., and Barge, R. M. (2002). Deoxycytidine kinase expression and activity in patients with resistant versus sensitive acute myeloid leukemia. Eur J Haematol *69*, 171-178.

Vita, R., Zarebski, L., Greenbaum, J. A., Emami, H., Hoof, I., Salimi, N., Damle, R., Sette, A., and Peters, B. (2010). The immune epitope database 2.0. Nucleic Acids Res *38*, D854-862.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet *10*, 57-63.

Wu, R., Duan, L., Ye, L., Wang, H., Yang, X., Zhang, Y., Chen, X., Weng, Y., Luo, J., Tang, M.*, et al.* (2013). S100A9 promotes the proliferation and invasion of HepG2 hepatocellular carcinoma cells via the activation of the MAPK signaling pathway. Int J Oncol *42*, 1001-1010.

Xu, X., Zhu, K., Liu, F., Wang, Y., Shen, J., Jin, J., Wang, Z., Chen, L., Li, J., and Xu, M. (2013). Identification of somatic mutations in human prostate cancer by RNA-Seq. Gene *519*, 343-347.

Yin, B., Kogan, S. C., Dickins, R. A., Lowe, S. W., and Largaespada, D. A. (2006a). Trp53 loss during in vitro selection contributes to acquired Ara-C resistance in acute myeloid leukemia. Exp Hematol *34*, 631-641.

Yin, B., Morgan, K., Hasz, D. E., Mao, Z., and Largaespada, D. A. (2006b). Nfl gene inactivation in acute myeloid leukemia cells confers cytarabine resistance through MAPK and mTOR pathways. Leukemia *20*, 151-154.

Young, I., Young, G. J., Wiley, J. S., and van der Weyden, M. B. (1985). Nucleoside transport and cytosine arabinoside (araC) metabolism in human T lymphoblasts resistant to araC, thymidine and 6-methylmercaptopurine riboside. Eur J Cancer Clin Oncol *21*, 1077-1082.

Zali, H., Marashi, S. A., Rezaei-Tavirani, M., Toossi, P., Rahmati-Roodsari, M., and Shokrgozar, M. A. (2007). On the mechanism of apoptosis-inducing activity of human calprotectin: zinc sequestration, induction of a signaling pathway, or something else? Med Hypotheses *68*, 1012-1015.

Zhou, J., and Hsieh, J. T. (2001). The inhibitory role of DOC-2/DAB2 in growth factor receptor-mediated signal cascade. DOC-2/DAB2-mediated inhibition of ERK phosphorylation via binding to Grb2. J Biol Chem *276*, 27793-27798.

Appendix A: Methods and Materials

Cell Culture

The B117P, B117H, B140P, and B140H (murine BXH-2) cell lines were maintained by incubating at 37°C in 10% CO_2 in ASM (Largaespada et al., 1995) in the absence of Ara-C. Cells were passaged three times each week. Cells were replaced from frozen stocks every 3-4 months, to minimize genetic drift. Knockdown cells (KDs) were maintained in the same manner, but in the presence of selective doses of puromycin (1.5 μg/ml for the B117P KDs and 0.6 μg/ml for the B140P KDs). Ara-C was acquired from Bedford Laboratories (Bedford, OH).

Proliferation Assay

Each BXH-2 cell line was plated at an initial cell concentration 2.5 x 10^5 cells/ml. Cells were incubated at 37°, 10% CO_2 in ASM with 10% FBS. After 24 hours the cell concentrations were determined using Countess® Automated Cell Counter (Life Technologies Corporation, Carlsbad, CA). Each cell line was then split to 2.5 x 10^5 cells /ml. Cell counts were calculated for 7 days and the doubling times and standard deviations determined.

Gene Expression Microarray and Analysis

RNA was isolated from 3 separate passages of the BXH-2 cell lines (B117P, B117H, B140P and B140H). The RNA was amplified and labeled using the Ambion® MessageAmp™ III RNA Amplification Kit (Life Technologies Corporation, Carlsbad, CA) and gene expression process was conducted using the GeneChip Mouse Genome

430 2.0 Array (Affymetrix, Santa Clara, CA). The resulting data was normalized using the RMA algorithm of Genedata Expressionist software (Genedata AG, Basel, Switzerland).

Transcriptome Deep Sequencing and Analysis

Aliquots of RNA derived from the cells designated "passage B" of B117P, B117H, B140P, and B140H (GSM457359, GSM457362, GSM457365, and GSM457368, respectively) for the previously published microarray experiment(Rathe and Largaespada, 2010) were submitted for transcriptome sequencing. The RNA was quality tested using a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA) and all four samples had RNA integrity numbers (RINs) of 10.0. cDNA was created by reverse transcription of oligo-dT purified polyadenylated RNA. The cDNA was fragmented, blunt-ended, and then ligated to barcoded adaptors. Lastly, the library was size selected, and the selection process was validated and quantified by capillary electrophoresis and qPCR, respectively. Sequencing was accomplished on the HiSeq 2000 (Illumina Inc., San Diego, CA), with the goal of generating a minimum of 20 million paired-end 76 bp reads. The resulting data was loaded into Galaxy.(Blankenberg et al., 2010; Giardine et al., 2005; Goecks et al., 2010) TopHat 2.0.5(Langmead et al., 2009; Trapnell et al., 2009) was used to map the paired reads to the NCBI37/mm9 assembly of the mouse genome. The mean inner distance was established using the insertion size metrics feature of Picard-tools (http://picard.sourceforge.net). Other than stipulating the use of NCBI mouse genes for the gene model annotation, the default parameters (as defined by the University of Minnesota's Galaxy implementation) were used. The resulting TopHat data served as input to other analytical tools, which compared data from B117H to B117P, and B140H

to B140P. Visualization of the mapped reads was accomplished using the Integrative Genomic Viewer (http://www.broadinstitute.org/igv). Gene expression analysis of the RNA-seq data was conducted using Cufflinks tools.(Roberts et al., 2011a; Roberts et al., 2011b; Trapnell et al., 2010) Cuffdiff mapped reads to the NCBI37/mm9 mouse genome assembly and presented the data in terms of fragments per kilobase of transcript per million mapped reads (FPKMs). Cuffdiff was executed from Galaxy using default parameters. Transcripts were also assembled using Cufflinks, but without stipulating a reference transcriptome. The resulting transcripts were then mapped back to the NCBI37/mm9 reference genome using Cuffcompare. Cufflinks and Cuffcompare were executed from Galaxy using default parameters. Fusion analysis was conducted using deFuse(McPherson et al., 2011) from the Galaxy platform using default parameters. Frameshift and missense mutations were identified by MMuFLR: Missense Mutation and Frameshift Location Reporter. (Rathe, Johnson et al., submitted to *Bioinformatics*). The potential effects of the missense mutations on protein function were evaluated using PolyPhen-2(Adzhubei et al., 2010) and PROVEAN Protein.(Choi et al., 2012)

DNA and RNA Isolation and Sequencing

RNA isolations were performed using the RNeasy® Midi Kit (QIAGEN, Venlo, Netherlands), following the protocol for isolating cytoplasmic RNA. For each sample, 10^7 cells were processed and the centrifugation steps were performed at 2850 g. DNA was eliminated using the RNase-Free DNase Set (QIAGEN) at the recommended step in the RNeasy® protocol. RNA concentration was determined using a NanoDrop™ 1000 Spectrometer (Thermo Fisher Scientific Inc., Waltham, MA). The RNA samples were then stored at -80°C.

Genomic DNA isolations from the BXH-2 cell lines (B117P, B117H, B140P, B140H) were performed using a DNeasy Blood & Tissue Kit (QIAGEN). The resulting samples were quantified using a NanoDrop™ 1000 Spectrometer (Thermo Fisher Scientific Inc.). DNA samples were stored at -20°C.

cDNA was prepared from RNA using the Invitrogen™ Superscript® III First-Strand Synthesis System (Life Technologies). DNA (or cDNA) was amplified using Taq DNA Polymerase (QIAGEN), and separated by gel electrophoresis on a 2% agarose gel. The DNA was extracted from the resulting bands using the UltraClean® GelSpin® DNA Extraction Kit (MO BIO Laboratories, Inc., Carlsbad, CA). Classic Sanger sequencing was done using the ABI PRISM® 3730xl DNA Analyzer (Life Technologies Corporation).

Genomic DNA PCR and Quantitative PCR (qPCR)

The DNA samples from the B117P and B117H cell lines were amplified using the Expand Long Template PCR System (Roche Applied Systems, Indianapolis, IN). Primers are described in **Supplementary Table C.1**.

Quantitative PCR (qPCR) was performed using SYBR® Green PCR Master Mix (Life Technologies Corporation) on a Mastercycler® ep realplex device (Eppendorf, Hamburg, Germany). Primers are described in **Supplementary Table C.1**.

RNAi Experiments

The Open Biosystems' shRNA TRC constructs for *Dck*, 25382 (KD1) and 25383 (KD2) (Thermo Fisher Scientific Inc.), were provided in *E. coli* and plated on

carbenicillin media to isolate single clones. Next, the plasmids containing the shRNA constructs were isolated from the *E. coli* using the Invitrogen™ PureLink® Quick Plasmid Miniprep Kit (Life Technologies Corporation). The plasmids were then transfected into Open Biosystems' packaging cells, TLA-HEK293T, using the Open Biosystems' Trans-Lentiviral™ Packaging System (Thermo Fisher Scientific Inc.). The TLA-HEK239T cells were maintained in the recommended growth media. Viral particles were then collected and concentrated using PEG-it™ Virus Precipitation Solution (System Biosciences, Mountain View, CA). The viral particles were transduced into the B117P and B140P cell lines by adding virus (MOI of 100) and 8 μg/ml of polybrene to the cells and incubating for 2 hours at 37°C followed by spinoculation (30 min, 300 g).

TALEN Assembly and Generation of KO Cells

Candidate *Dck* TALENs were designed using TALE-NT (https://boglab.plp.iastate. edu/node/ add/talen). From the list of candidate TALENs generated using TALE-NT, three were chosen and constructed based on optimal RVD content, spacer length, and binding site length based on previous TALEN publications and our own experiences with TALENs. TALENs were assembled using Golden Gate cloning as previously described.(Cermak et al., 2011) The truncated GoldyTALEN backbone used has also been previously described.(Carlson et al., 2012) Assembled TALENs were validated by transient transfection into NIH 3T3 cells using the NEON® Transfection System (Life Technologies Corporation, Carlsbad, CA), following manufacturer's instructions. Three days after TALEN transfection DNA was collected

using standard Phenol:Chloroform extraction and CEL-I assays were performed as previously described.(Guschin et al., 2010)

Validated *Dck* TALENs were then electroporated into B117P cells using the NEON® Transfection System (Life Technologies Corporation), following manufacturer's instructions. Three days after transfection cells were selected for *Dck* Knockout (KO) using 50 μg/ml of Ara-C for 5 days. From this pool of TALEN modified cells single cell clones were isolated by limiting dilution cloning. Clones were analyzed for *Dck* KO by direct PCR and sequencing of the TALEN targeted region of *Dck*, using primers described in **Supplementary Table C.1.**

Inducible DCK Overexpression Vector

Full-length human *DCK* cDNA was purchased from GeneCopoeia (Rockville, MD) in a ready entry ORFEXPRESS™ Gateway® Plus Shuttle (Cat#GC-C0081). The *DCK* cDNA was then transferred to PB-TRE-DEST-EF1A-rtTA-IRES-Puro(Moriarity et al., 2013) via the Invitrogen™ Gateway® LR Clonase® reaction (Life Technologies Corporation), following manufacturer's instructions. Two micrograms of PB-TRE-*Dck*-EF1A-rtTA-Puro was electroporated with 2 μg of Super *piggyBac* transposase (System Biosciences, Mountain View, CA) into B117P using the NEON® Transfection System (Life Technologies Corporation), following manufacturer's instructions. Two days after transfection cells were selected with 1.5 μg/ml puromycin for 5 days to generate stable cell lines. Overexpression was activated by adding 1.0 μg/ml of doxycycline.

Western Blot Analysis

Protein lysate was isolated from cells using RIPA lite buffer (50nM Tris-HCL pH7.6, 150mM NaCl, 1% NP40, 5mM NaF, 1mM EDTA) supplemented with protease inhibitors (Roche Applied Systems) and phosphatase inhibitors (Sigma-Aldrich, St. Louis, MO). 200μg of protein lysate was separated on a NuPAGE® Novex® 10% Bis-Tris Gel (Life Technologies Corporation) and transferred to PVDF membrane. The membrane was blocked with 5% milk in 1XTBST (4hrs RT) and incubated overnight at 4°C with the primary antibodies anti-DCK (1:500, Proteintech, Chicago, IL) and anti-Gapdh (1:10,000, Cell Signaling Technology, Danvers, MA). Goat anti-Rabbit HRP conjugated secondary antibodies were utilized at 1:5000 dilution (Santa Cruz Biotechnologies, Dallas, TX). Membranes were developed with WesternBright Ecl kit (BioExpress, Kaysville, UT).

Drug Assays

Drug assays were performed using the CellTiter 96® Aqueous Non-Radioactive Cell Proliferation Assay (Promega, Madison, WI), as described previously.(Rathe and Largaespada, 2010) For each cell line the drug was tested using 10 different concentrations, and each drug concentration was tested in quadruplicate. The drug concentrations were selected to maximize the number of data points between IC_5 and IC_{95}. For the results to be acceptable there needed to be data points on both sides of the IC_{50} and the *r*-value needed to be greater than 0.85. Inhibitory concentrations (IC) values and *r*-values were calculated using CalcuSyn 2.0 (Biosoft, Cambridge, UK).

Screen of 446 FDA Approved Drugs

A screen using a NIH Clinical Collection of 446 FDA approved drugs (http://www.nihclinicalcollection.com/) at 10μM was conducted for each of the four BXH-2 cell lines (B117P, B117H, B140P and B140H). Cell viability was measured using Alamar Blue and fluorescence readout on a Molecular Devices Spectromax 2e microplate reader.

Combination Screen of Ara-C with 50 Individual Drugs

A drug screen was conducted to test 50 different drugs in combination with Ara-C in the B117P cell line using the MTS-tetrazolium assay. The drugs tested are listed in **Supplementary Table C.9.** The concentration of Ara-C used was 200ng/ml. It was combined with 7 different concentrations at 2X increments of each of the 50 drugs. Each drug concentration was conducted in duplicate.

Appendix B: MMuFLR: Missense Mutation and Frameshift Location Reporter

[This appendix was published in its entirety in *Bioinformatics (Rathe et al., 2013).* Bioinformatics does not require a release form for including this material in a thesis by the primary author.]

Susan K. Rathe[1,*], James E. Johnson[2,*] , Kevin A.T. Silverstein[2], Jesse J. Erdmann[2], Adrienne L. Watson[1,3,4], Flavia E. Popescu[5], John R. Ohlfest[5] and David A. Largaespada[1,3,5]

[1]Masonic Cancer Center, University of Minnesota, Minneapolis, MN 55455, USA, [2]Supercomputing Institute for Advanced Computation Research, University of Minnesota, Minneapolis, MN 55455, USA, [3]Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, MN 55455, USA, [4]Brain Tumor Program, University of Minnesota, Minneapolis, MN 55455, USA, [5]Department of Pediatrics, University of Minnesota, Minneapolis, MN 55455, USA

Contributors

James E. Johnson was co-author and did most of the programming for MMuFLR. He wrote the WORKFLOW section of the manuscript.

Kevin A.T. Silverstein assisted in the design of MMuFLR.

Jesse J. Erdmann assisted in the early programming efforts for MMuFLR.

Adrienne L. Watson performed the PCRs.

Flavia E. Popescu prepared the biological samples and conducted the RNA isolations.

John R. Ohlfest was co-principal investigator.

Motivation: Cancer researchers seeking immunotherapy targets in cancer cells need tools to locate highly expressed proteins unique to cancer cells. Missense Mutation and Frameshift Location Reporter (MMuFLR), a Galaxy-based workflow, analyzes Next Generation Sequencing (NGS) paired read RNA-seq output to reliably identify small frameshift mutations, as well as missense mutations, in highly expressed protein-coding genes. MMuFLR ignores known SNPs, low quality reads, and poly-A/T sequences. For each frameshift and missense mutation identified MMuFLR provides the location and sequence of the amino acid substitutions in the novel protein candidates for direct input into epitope evaluation tools.

Availability: http://toolshed.g2.bx.psu.edu/

Introduction

Next Generation Sequencing (NGS) holds promise for individualized targeted treatment, particularly in the realm of cancer immunotherapy (Meinke et al., 2005). Identifying aberrant transcripts from a patient's tumor cells may provide candidate target epitopes which can be used to generate patient specific vaccines. Given the quantity of data generated by NGS, it is important to have an efficient and repeatable process to deconstruct the data and present the most relevant vaccine targets, which includes missense and small frameshift mutations located in protein coding regions of highly expressed genes. We offer Missense Mutation and Frameshift Location Reporter (MMuFLR) as a process to analyze NGS RNA-seq paired read output to identify a limited set of peptides which may be targets for immunotherapy. MMuFLR is implemented as a workflow in Galaxy (Blankenberg et al., 2010). Galaxy is a server based workflow system designed to run a series of applications, presented to the user as tools, in an easy to use, repeatable manner. MMuFLR uses a few custom developed Galaxy tools as well as those in the standard distribution of Galaxy. In this pilot project MMuFLR's ability to identify potential vaccine targets was tested in RNA-seq data sets derived from two benign human meningiomas. However, the use of MMuFLR is not limited to cancer immunotherapy. It can also be used to locate prominent mutations responsible for cancer development and for other maladies, such as genetic diseases, developmental defects, and drug resistance.

Workflow

2.1 *Map read sequences to reference genome* MMuFLR starts with the NGS RNA-seq paired read output in the form of two fastq-formatted files. Each read pair is individually mapped to the reference genome using TopHat (Langmead et al., 2009) resulting in a bam formatted file. The standard galaxy distribution includes TopHat in its toolset. The custom galaxy tool “Filter SAM or BAM" invokes the SAMtools (Li et al., 2009) view command to filter TopHat bam output by the MAPQ score.

2.2 *Refine the mapping* TopHat, as is typical with short read mappers, may find multiple mappings for individual read pairs which are equally valid. The mapped reads can be analyzed in aggregate to improve upon the initial mapping of individual reads. The MMuFLR workflow uses software from the Genome Analysis Toolkit (GATK) (DePristo et al., 2011; McKenna et al., 2010) along with Picard-tools (http://picard.sourceforge.net) to perform this collective analysis and realign the individual reads. GATK requires bam input to be labeled with read group information and be sorted in reference order. Several Picard-tools (“Add or Replace Groups”, “Paired Read Mate Fixer”, “Mark Duplicate Reads”, and “Reorder SAM/BAM”) add the read groups, filter out unmapped pairs, remove exact duplicate reads, and sort the resulting bam output into reference order. Two GATK tools, “Realigner Target Creator” and “Indel Realigner”, realign the reads based on aggregate information from all reads. Duplicate reads are again identified and removed with “Mark Duplicate Reads.”

2.3 *Multiple alignment* The realigned reads are then assembled into a multiple alignment using the Galaxy tool “MPileup” which invokes SAMtools mpileup. The

custom galaxy tool, "Pileup to VCF", filters the pileup results and converts them to the standard Variant Call Format (VCF). Filter parameters to the "Pileup to VCF" tool include Minimum Base Quality, Minimum Coverage Depth, and Minimum Frequency of a Specific Allele.

2.4 *Eliminate common variants* The resulting VCF file will likely contain many commonly known variants which would not be useful in determining candidate targets. MMuFLR uses the “annotate” command of SnpSift (Cingolani et al., 2012) to merge variants from the samples being evaluated with known variants from the dbSNP database (Sherry et al., 2001), as well as a user provided list. Then the “filter” command of SnpSift is used to eliminate those annotated variant entries.

2.5 *Determine variant effects* MMuFLR next categorizes the effect of each variant using SnpEff (Cingolani et al., 2012). The “filter” command of SnpSift is used to separate the missense and frameshift variants into individual files for reporting.

2.6 *Report variants* Finally, a new Galaxy tool, "SnpEff Ensembl CDS", inserts the variants into transcripts retrieved from Ensembl (Flicek et al., 2013) and reports the amino acid sequence of the abnormal peptides. Included in the report are the sequencing depth of reads at the mutation site and the prevalence of the mutation. A parameter can be set to ignore poly-A/T sequences of specified lengths.

Frameshift and missense mutations identified by MMuFLR and verified using Sanger sequencing. Included in the table are the location of the variant, the reference sequence being replaced, the fraction of reads with the mutation (prevalence) and the number of reads (depth).

MMuFLR was executed for two human meningioma RNA-seq data sets, MG1 and MG2. MMuFLR settings restricted the selection of candidate genes with minimum coverage of 5 reads, with a prevalence greater than 0.32, and a Phred quality score of 20 or more. MMuFLR also ignored poly A/T mononucleotide sequences of 5 or more. MMuFLR identified 16 frameshifts and 576 missense mutations in the MG1 sample, and 10 frameshifts and 509 missense mutations in the MG2 sample. The candidates were evaluated for their presence in other samples. Sanger sequencing of 9 prominent candidates, not found in other cancer samples, verified the existence of the mutations within the tumor samples, with the MMuFLR prevalence accurately reflecting the relative abundance of the mutation within the total transcripts present (Table 1). Since there were no normal samples available for comparison, these mutations may represent rare germline variants.

Conclusions

One challenge for developing patient specific cancer vaccines is identifying novel peptides specific to the tumor cells. MMuFLR is an innovative Galaxy-based software tool designed to locate highly abundant transcripts with frameshift or missense mutations within protein coding sequences. MMuFLR ignores potentially false candidates such as frameshifts located within poly A/T sequences of mononucleotides, which is believed to result from a loss of fidelity during replication due to dissociation within the reverse transcriptase binding site in a phenomenon referred to as “stuttering” (Fazekas, Steeves, and Newmaster, 2010). Because there is always a lag between the identification of SNPs and their incorporation into the dbSNP database, MMuFLR provides the ability to establish a supplementary SNP table in which the researcher can record and thereby

ignore SNPs identified during their research efforts. MMuFLR has been successfully tested with both human and mouse samples, but should be easily adaptable to other organisms.

Supplementary information

Processing of biological samples

Two meningioma samples (MG1 and MG2) were provided by the Biological Materials Procurement Network at the University of Minnesota. Samples were flash frozen. RNA was isolated from the cells using RNeasy® Midi Kit (QIAGEN, Venlo, Netherlands), and DNA was eliminated using RNase-Free DNase Set (QIAGEN, Venlo, Netherlands). RNA integrity numbers (RINs) were determined using a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). The RIN for the MG1 and MG2 samples were 8.6 and 7.2, respectively. RNA-seq was performed on a HiSeq 2000 (Illumina Inc., San Diego, CA) and generated 24 million paired reads of 100 cycles for each sample. cDNA was prepared from RNA using the First Strand Transcriptor cDNA Synthesis Kit (Roche Applied Systems, Indianapolis, IN). cDNA was amplified using GoTaq Green PCR Master Mix (Promega, Madison, WI), and separated by gel electrophoresis on a 1% agarose gel. The DNA was extracted from the resulting bands using QIAquick Gel Extraction Kit (QIAGEN, Venlo, Netherlands). Classic Sanger sequencing was performed on an ABI PRISM® 3730xl DNA Analyzer (Life Technologies Corporation, Carlsbad, CA).

MMuFLR features and use

MMuFLR is a biologist friendly Galaxy workflow designed to assist researchers in the identification of legitimate frameshift and missense mutations, while minimizing the number of "false positives". The generation of hundreds of "false positives" by existing bioinformatics tools frequently deters researchers from taking advantage of the wealth of valuable information provided by next generation sequencing. MMuFLR uses a series of filters to eliminate "false positives", as well as known SNPs. These filters include: 1) the mutation must occur within a coding region, 2) a setting for the minimum number of reads spanning a mutation location, 3) a setting for a minimum number of reads containing the mutation, 4) a setting for a minimum ratio of reads with the mutation relative to the total number of reads, 5) the ability to ignore poly A/T sequences which result from polymerase "stuttering", 6) the ability to ignore reads with low phred scores, 7) the elimination of PCR duplicates, 8) the ability to ignore known SNPs, and 9) the ability to ignore a list of SNPs provided by the researcher.

The minimum phred and poly A/T logic were introduced into two new algorithms created specifically for the MMuFLR workflow, "Pileup to VCF" and "SnpEff Ensembl CDS", respectively. The combination of just these two features reduced the number of mutation candidates in the MG1 and MG2 samples by more than 60%. The "SnpEff Ensembl CDS" algorithm also provides other labor saving features to the researcher. It automatically identifies the protein sequence changes resulting from the mutation, thereby eliminating the need to do this very necessary next step using another software tool. It provides the novel protein sequences in a form which can be cut and pasted directly into an epitope finder, such as the Immune Epitope Database and Analysis

Resource (www.iedb.org) (Vita et al., 2010). It describes missense mutations in a standard format which can be input directly into tools to predict the effect of the mutation, such as Polyphen-2 (Adzhubei et al., 2010) and PROVEAN Protein (Choi et al., 2012), or into tools which describe known somatic mutations, such as COSMIC (Forbes et al., 2011). Further, it identifies the sequence of the stop codon, including the nucleotide preceding and the nucleotide following the stop codon. The configuration of the stop codon and surrounding bases affects the likelihood of translation beyond the stop codon (Floquet et al., 2012), a feature which can be quite valuable to some researchers.

Ideally, researchers using MMuFLR should have paired samples, such as a cancer sample paired with a normal tissue sample, or a metastatic cancer sample paired with a primary cancer sample, or a drug resistant sample paired with its original drug sensitive sample. There are two approaches which can be used to compare samples, such as in the case where a researcher would like to eliminate germline mutations (those found in the normal tissue sample) from cancer samples, to leave only the somatic mutations. In the first method MMuFLR can be run individually for both of the paired samples and the resulting TSV files can be interrogated by the "Compare two datasets" utility of Galaxy to eliminate the mutations found in one sample from those found in the other. In the second method, MMuFLR can be run first for the normal sample, and the resulting VCF file from the first SnpSift Filter step can be used as the exception file (VCF file of additional known variants) in the second run of MMuFLR for the cancer sample. The FASTq file being input to MMuFLR for the normal sample could have been generated from exome or genome sequencing, not just RNA-sequencing.

It is still necessary to review visually all the candidate mutations presented by MMuFLR. This will insure there is no other explanation for the presence of the candidate, such as the absence of expression in the paired sample or the prominence of the mutation appearing at the end of reads. We utilized the Integrative Genomic Viewer (http://www.broadinstitute.org/igv) to verify the candidate mutations. Following this in silico verification step, all of candidates selected for verification by RT-PCR and Sanger sequencing were confirmed as true mutations. The MMuFLR Supplementary Tables contain the MMuFLR reports prior to in silico verification.

In the MMuFLR runs to analyze the MG1 and MG2 samples, we set the minimum number of reads covering the mutation location to 5, the minimum number of reads with the mutation to 2, the minimum prevalence to 0.32, the minimum phred level to 20, and the poly A/T threshold to 5. Another research group recently published their analysis of RNA-seq data to locate frameshift and missense mutations in prostate cancer samples as compared to benign matching tissue samples (Xu et al., 2013). They used some of the same filtering strategies available in MMuFLR. They looked for mutations in coding regions, a minimum of 5 reads containing the variant, a phred quality of more than 20, and they ignored known SNPs. From the description of their process, it appears they utilized a number of Unix based tools, and wrote custom programs for some of the filtering steps. MMuFLR makes a similar, but more robust, analytic process available to any biologist with access to Galaxy.

Accessing and executing MMuFLR

The MMuFLR workflow," MMuFLR_Human_v1.2", and the two new Galaxy tools, "Pileup to VCF" and "SnpEff Ensembl CDS", can be acquired from the Galaxy toolshed at Penn State University:

http://toolshed.g2.bx.psu.edu/view/jjohnson/mmuflr

http://toolshed.g2.bx.psu.edu/view/jjohnson/pileup_to_vcf

http://toolshed.g2.bx.psu.edu/view/jjohnson/snpeff_cds_report

Instructions for deploying a workflow from the Galaxy toolshed are documented in the Galaxy wiki: http://wiki.galaxyproject.org/ToolShedWorkflowSharing along with a screencast tutorial demonstrating how to run a workflow: http://screencast.g2.bx.psu.edu/mt-workflowImport.

MMuFLR uses a variety of preexisting Galaxy tools and for the MMuFLR workflow to execute properly most of the parameters should not be changed. The parameters which can be safely adjusted are described here.

The first 4 steps of the MMuFLR workflow request the identification of 4 input files.

1. the "Forward reads fastq" file for the sample being analyzed
2. the "Reverse reads fastq" file for the sample being analyzed
3. a "VCF file of known variants (dbSNP)"

4. a “VCF file additional known variants” as provided by the researcher

The last input file contains known variants which will be excluded from the MMuFLR analysis. This can be a VCF file generated by a previous run MMuFLR. For example, MMuFLR can be run first on a normal tissue sample, using workflow: “MMuFLR_Human_germline_v1.2”, and the resulting VCF file can be used as input to a run of MMuFLR to identify mutations in a matching tumor sample which are not found in the normal tissue.

The “Mean Inner Distance between Mate Pairs” and the “Std. Dev for Distance between Mate Pairs” can be set in the TopHat for Illumina step.

Pileup to VCF allows the selections of “Minimum Base Quality”, “Minimum Coverage Depth”, and “Minimum Frequency of a Specific Allele”. (Note: the Minimum base quality in the MPileup step should not be changed)

SnpEff Ensemble CDS allows the selection of the type of report (Filter SnpEffect) which can be FRAME_SHIFT or NON_SYNONYMOUS_CODING. It also provides the ability to “Ignore variants that are part of a Poly-A of at least this length”.

Table C.1. Frameshift and missense mutations identified by MMuFLR

Frameshift Mutations

Sample	Gene	Variant Position	Ref	Var	Prev	Depth
MG1	ZNF812	chr19:9800968	CT	T	0.50	20
MG2	CPNE1	chr20:34215234	G	TG	0.45	98
	MRPL16	chr11:59577372	C	GC	0.37	27
	ZP3	chr7:76071183	T	TG	0.50	14
	DNAJB7	chr22:41257834	A	TA	1.00	6

Missense Mutations

Sample	Gene	Variant Position	Ref	Var	Prev	Depth
MG1	COL6A2	chr21:47545737	A	G	0.50	899
MG2	APP	chr21:27484329	T	A	0.46	432
	FRMPD2	chr10:49389016	A	G	0.48	227
	CCDC74A	chr2:132288362	T	C	0.97	29

Appendix C: Supplementary Tables and Figures

Supplementary Table C.1: Primers

Primers designed to find mutations and to verify expression levels.

		Primers	
Experiment	**Target**	**Forward**	**Reverse**
qPCR	*Dck* (Exon 5-6)	CTGGCTCCTTCATCGGACT	CCAGGCTTTCGTGTTTGTCT
qPCR	*Dck* (3'-UTR)	TGAGGATAGAACATGCCAAGG	CGCCTAGGCTTCTCAGTGTC
qPCR - Reference gene	*Rps9* (Exon 2-3)	TCTATTCACCATGCCCGTGT	GGCGAACAATGAAGGATGG
RNA sequencing	*Dck* Quartile 1	GCAGGTCAGGATCTGGCTTAG	CTATGTAAGCATGTCTCTGGAGTGG
RNA sequencing	*Dck* Quartile 2	TGAACGATCTGTGTATAGTGACAGG	CAACCTGCTGAGGAAGACAAA
RNA sequencing	*Dck* Quartile 3	GCAGGAAATGTAGAGGTAGATGG	CACTGCACACCTTAACTAAACAAAC
RNA sequencing	*Dck* Quartile 4	AAATTAATGAAGACAGAGGCTAGAAAG	TTATTAGCGTCTTTTCAATTCTACAAA
qPCR - Reference gene	*Rap2c* (Exon 2-3)	CCCGAGAAGCAAGATCAGTG	GTGCAAATTTCCTCTGAGTTTGTAA
Expand Long Template PCR	*Dck* deletion in B117H	GACAAACACGAAAGCCTGGT	CAACCAGCTGCATGAGAGAA
RNA sequencing	*Dck* deletion in B117H	ATCCTGGACTCTAGAGGAGCATT	GGGTGTGGAGGAAACTGTGT
DNA sequencing	*Dck* deletion in B117H	CTGTCTGGCTGTTTGCTATAAAGA	TTAACCCACTCAGGATGACAAGT
RNA sequencing	*Ccdc88b* deletion in B140H	TCATGGACCAGTACCGTGTG	GGACTAAATGACGCAGAGCTTC
RNA sequencing	*Dck* insertion in B140H	TTTGAAAGATCTGTGTATAGTGACAGG	TTTCTCGGGAGTAGCTCGAA
DNA sequencing	*Dck* insertion in B140H	GATTGAGAAGCGCATTTACCC	ACCTCGGGAGTAGCTCGAAG
RNA sequencing	*Dck* start site (KO's)	AGGAACCACAGACATGC	CTCTCTCAATGGCAAGCTCA
DNA sequencing	*Dck* start site (KO's)	GAGGCGTATAGGGTGGGAGT	GGGGACACCTTACCGATGT

Supplementary Table C.2: RNA-seq stats

The number of paired-end 76 bp reads generated by Illumina NGS sequencing for the murine cell lines B117P, B117H, B140P, and B140H, and the ability of the reads to be mapped uniquely by TopHat.

	TopHat Stats PE				
Cell line	Total Read pairs	Mapped Read pairs	% Mapped	Unique Read pair	% Unique
B117P	24,914,387	22,998,275	92.3	22,640,193	90.9
B117H	21,620,354	19,656,507	90.9	19,321,761	89.4
B140P	20,570,801	19,114,959	92.9	18,838,702	91.6
B140H	27,977,177	25,952,317	92.8	25,585,528	91.5

Supplementary Table C.3: Common expression changes

Gene expression changes common to both Ara-C resistant cell lines when compared to their parental cell lines were identified by analyzing data generated from the Illumina NGS sequencing platform with the use of TopHat for genome mapping, and Cuffdiff for gene expression comparison. Included in this table are genes with a minimum fold change between 2 and 3. Genes with a greater than 3 fold change are included in **Table 3.1**. Genes in bold with * were also identified by gene expression microarray with 2-fold+ changes in expression (Rathe & Largaespada, 2010). (1) Expression levels are expressed in fragments per kilobase of transcript per million mapped reads (FPKMs). (2) Fold change compares the Ara-C resistant gene expression to the Ara-C sensitive parental lines. Negative values indicate a decrease in expression while positive values represent an increase in expression. (3) The table is sorted by the minimum fold change, which is the least of the two fold change values.

Gene symbol	B117P FPKMs[1]	B117H FPKMs[1]	B117H vs. B117P fold chg[2]	B140P FPKMs[1]	B140H FPKMs[1]	B140H vs. B140P fold chg[2]	Minimum fold change[3]
Sh3tc1	2.773	0.872	-3.181	2.337	0.788	-2.964	-2.964
Mex3a	2.270	0.639	-3.552	4.567	1.716	-2.661	-2.661
Osgin1	0.593	0.213	-2.791	1.185	0.476	-2.491	-2.491
Dab2*	68.710	12.284	-5.594	13.793	5.717	-2.413	-2.413
Satb2	3.286	0.946	-3.475	2.237	0.933	-2.398	-2.398
Cxcl10	2.890	0.884	-3.270	1.140	0.476	-2.394	-2.394
Ier3	8.200	0.505	-16.223	1.582	0.672	-2.353	-2.353
2810410L24Rik	1.654	0.709	-2.332	1.089	0.151	-7.203	-2.332
Ccbl2*	15.548	6.695	-2.322	3.791	0.001	-3790.700	-2.322
Igfbp4	101.925	13.223	-7.708	105.091	45.829	-2.293	-2.293
Smpdl3b	7.676	1.762	-4.356	0.703	0.309	-2.275	-2.275
Fam160a2	0.834	0.371	-2.247	1.920	0.790	-2.432	-2.247
1700109H08Rik	7.218	3.223	-2.240	2.172	0.919	-2.363	-2.240
Serpinb6b	5.757	2.589	-2.224	8.700	1.926	-4.517	-2.224
Gpr183	8.932	4.158	-2.148	1.829	0.783	-2.335	-2.148
Sash1	4.016	1.902	-2.112	2.270	0.116	-19.538	-2.112
Plp1	0.540	0.242	-2.231	0.543	0.261	-2.079	-2.079
Fkbp11	20.183	9.735	-2.073	5.533	0.340	-16.252	-2.073
Klhl23	5.132	2.108	-2.435	0.629	0.305	-2.059	-2.059
Cds1	5.988	2.947	-2.032	3.019	0.900	-3.354	-2.032
Gjb3	0.686	0.001	-685.977	0.832	0.410	-2.030	-2.030
2210417A02Rik	2.075	1.032	-2.010	2.551	0.909	-2.805	-2.010
Gng5	71.113	142.262	2.001	78.906	184.445	2.338	2.001
Acadm	12.394	24.907	2.010	16.925	36.069	2.131	2.010
Chpf	0.059	1.118	19.060	1.098	2.229	2.030	2.030
Fpgt	1.267	3.218	2.541	1.454	2.976	2.047	2.047
Kmo	4.019	8.514	2.118	2.236	4.611	2.062	2.062
Bcl10	10.187	21.039	2.065	8.611	20.871	2.424	2.065
Ssx2ip	5.298	11.077	2.091	4.246	9.297	2.189	2.091
Traf5	1.411	3.134	2.222	0.598	1.263	2.113	2.113
Gpr25	0.214	0.878	4.105	17.899	37.890	2.117	2.117
Depdc1a	0.001	3.868	3868.250	3.159	6.693	2.119	2.119
Znhit6	4.859	10.512	2.163	6.821	14.549	2.133	2.133
Ankrd13c*	6.405	13.990	2.184	8.611	19.645	2.282	2.184
Dact3	0.515	1.192	2.312	0.635	1.400	2.204	2.204
2410004B18Rik	3.861	8.836	2.289	5.156	13.242	2.568	2.289
Trpa1	4.909	13.575	2.765	2.901	6.775	2.335	2.335
Ccdc112	0.879	2.097	2.387	0.151	0.781	5.171	2.387
2610524H06Rik	12.971	31.344	2.417	7.959	31.839	4.001	2.417
Pigk	3.112	7.577	2.435	3.750	9.630	2.568	2.435
Rtn4rl1	2.940	7.237	2.462	1.586	4.844	3.054	2.462

Slc14a1*	0.267	1.766	6.604	0.289	0.745	2.579	2.579
Aff3*	13.743	37.371	2.719	0.120	0.870	7.259	2.719
Il1r2	0.179	0.521	2.905	0.625	2.572	4.115	2.905

Supplementary Table C.4: B117H expression changes

Gene expression changes between the Ara-C resistant B117H cells compared to its parental cell line, B117P, were identified by analyzing data generated from the Illumina NGS sequencing platform with the use of TopHat for genome mapping, and Cuffdiff for gene expression comparison. Genes in bold with * were also identified by gene expression microarray with 100-fold+ changes in expression. . (1) Expression levels are expressed in fragments per kilobase of transcript per million mapped reads (FPKMs). (2) Fold change compares the Ara-C resistant gene expression to the Ara-C sensitive parental lines. Negative values indicate a decrease in expression while positive values represent an increase in expression.

Gene symbol	Description	B117P FPKMs[1]	B117H FPKMs[1]	B117H vs. B117P fold chg[2]
5730528L13Rik	RIKEN cDNA 5730528L13 gene	4.437	0.001	-4437.220
Uty	ubiquitously transcribed tetratricopeptide repeat gene, Y chromosome	1.986	0.001	-1985.690
Ccl5	chemokine (C-C motif) ligand 5	1.630	0.001	-1629.900
Klk8	kallikrein related-peptidase 8	1.167	0.001	-1167.290
Zfyve9	zinc finger, FYVE domain containing 9	0.831	0.001	-830.902
Zfp157	zinc finger protein 157	0.798	0.001	-797.719
Gjb3	gap junction protein, beta 3	0.686	0.001	-685.977
Upk1b	uroplakin 1B	0.637	0.001	-636.522
Stfa1	stefin A1	0.606	0.001	-605.925
8430408G22Rik	RIKEN cDNA 8430408G22 gene	0.550	0.001	-549.757
Ddx3y*	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked	7.466	0.043	-172.670
Tnxb	tenascin XB	0.011	2.441	222.762
Glul	glutamate-ammonia ligase (glutamine synthetase)	0.033	15.557	474.688
2610109H07Rik	RIKEN cDNA 2610109H07 gene	0.001	0.510	509.710
Kcnmb4	potassium large conductance calcium-activated channel, subfamily M, beta member 4	0.001	0.610	609.674
Blk	B lymphoid kinase	0.001	0.623	622.966
Nxf3	nuclear RNA export factor 3	0.001	0.701	701.321
1700019E19Rik (Ift43)	intraflagellar transport 43 homolog	0.001	0.713	712.901
Thsd4	thrombospondin, type I, domain containing 4	0.001	0.731	730.994
Tspan12	tetraspanin 12	0.001	0.743	742.820
Il17rb	interleukin 17 receptor B	0.001	0.756	756.133
Myl10	myosin, light chain 10, regulatory	0.001	1.014	1013.690
Hectd2	HECT domain containing 2	0.001	1.316	1315.930
Prr5	proline rich 5 (renal)	0.001	1.826	1826.190
Lrrc49	Leucine rich repeat containing 49	0.001	1.962	1962.050
Ggt7	gamma-glutamyltransferase 7	0.001	3.278	3278.190
Depdc1a	DEP domain containing 1a	0.001	3.868	3868.250
Zfp13	zinc finger protein 13	0.001	7.147	7147.450

Supplementary Table C.5: B140H expression changes

Gene expression changes between the Ara-C resistant B140H cell compared to its parental cell line, B140P, were identified by analyzing data generated from the Illumina NGS sequencing platform with the use of TopHat for genome mapping, and Cuffdiff for gene expression comparison. (1) Expression levels are expressed in fragments per kilobase of transcript per million mapped reads (FPKMs). (2) Fold change compares the Ara-C resistant gene expression to the Ara-C sensitive parental lines. Negative values indicate a decrease in expression while positive values represent an increase in expression.

Gene symbol	Description	B140P FPKMs[1]	B140H FPKMs[1]	B140H vs. B140P fold chg[2]
Cenpb	centromere protein B	11.488	0.001	-11487.700
Pepd	peptidase D	5.479	0.001	-5478.620
Ccbl2	cysteine conjugate-beta lyase 2	3.791	0.001	-3790.700
Ephb2	Eph receptor B2	2.352	0.001	-2351.650
Pou2af1	POU domain, class 2, associating factor 1	2.215	0.001	-2215.060
Lmo1	LIM domain only 1	2.152	0.001	-2151.920
Hgsnat	heparan-alpha-glucosaminide N-acetyltransferase	2.057	0.001	-2057.290
Tarsl2	threonyl-tRNA synthetase-like 2	1.925	0.001	-1924.690
Cdh17	cadherin 17	1.770	0.001	-1769.610
Zbtb10	zinc finger and BTB domain containing 10	1.225	0.001	-1224.630
Csf2	colony stimulating factor 2 (granulocyte-macrophage)	0.841	0.001	-841.152
Col11a1	collagen, type XI, alpha 1	0.817	0.001	-817.413
Rab39	RAB39, member RAS oncogene family	0.798	0.001	-798.492
Psd3	pleckstrin and Sec7 domain containing 3	0.796	0.001	-796.090
Lipg	lipase, endothelial	0.772	0.001	-771.597
Faah	fatty acid amide hydrolase	0.738	0.001	-738.177
Nkain1	Na+/K+ transporting ATPase interacting 1	0.729	0.001	-728.751
Batf	basic leucine zipper transcription factor, ATF-like	0.720	0.001	-720.441
Cacnb2	calcium channel, voltage-dependent, beta 2 subunit	0.657	0.001	-656.769
Spef1	sperm flagellar 1	0.638	0.001	-638.260
Cd14	CD14 antigen	0.628	0.001	-628.324
Rabl2	RAB, member of RAS oncogene family-like 2	0.566	0.001	-566.329
Ppfibp1	PTPRF interacting protein, binding protein 1 (liprin beta 1)	4.666	0.008	-562.055
Apcdd1	adenomatosis polyposis coli down-regulated 1	0.510	0.001	-510.038
Fbxl5	F-box and leucine-rich repeat protein 5	5.110	0.017	-296.832
S100a10	S100 calcium binding protein A10 (calpactin)	105.583	0.401	-263.358
Tbc1d16	TBC1 domain family, member 16	1.454	0.006	-255.435
Spint2	serine protease inhibitor, Kunitz type 2	27.160	0.173	-157.145
Elovl6	ELOVL family member 6, elongation of long chain fatty acids (yeast)	2.432	0.020	-120.742
Cd2	CD2 antigen	0.001	0.595	595.349
Gm1631	predicted gene 1631	0.001	0.735	735.306
2200002J24Rik	RIKEN cDNA 2200002J24 gene	0.001	0.787	786.621

Supplementary Table C.6: Dck C-terminus protein sequence

The C-terminus of Dck is highly conserved across a number of species. (Source: Ensembl)

Organism	Amino acid sequence at end of Dck protein
B117H	ESLVEKVGLKKDKYVSCLVQDNPIDL
Mouse	ESLVEKVKEFLSTL
Human	ESLVEKVKEFLSTL
Dog	DGLIEKVKEFLSTL
Wallaby	ENLVEKVKEFLSTL
Platypus	ANLIEKVKEFLSTL
Dolphin	HDSLEKVKEFLSTL
Chicken	DHMTEKVKEFLSTL
Anole lizard	ENMIEKVKEFLSTL
Zebrafish	ADMIEKVKEFLSTL
Coelacanth	ADMVEKVREFLSSL

MALEK (highly conserved aa in alpha helices)

Supplementary Table C.7: B140H missense mutations

MMuFLR identified potential frameshift mutations in the B140P and B140H cell lines. The mutations were screened *in silico* using IGV. The two mutations confirmed by IGV as being newly introduced to the B140H cells, *Ccdc88b* and *Dck,* were confirmed as heterozygous and homozygous, respectively, by Sanger sequencing.

B140P				B140H			
Gene	Variant Position	Ref.	Var.	Gene	Variant Position	Ref.	Var.
Purb	chr11:6368602	CT	T	*Purb*	chr11:6368602	CT	T
C2cd3	chr7:107603881	A	AG	*C2cd3*	chr7:107603881	A	AG
Cyhr1	chr15:76477929	A	GA	*Cyhr1*	chr15:76477929	A	GA
Sap25	chr5:138082875	GT	G	*Sap25*	chr5:138082875	GT	G
Ptpmt1	chr2:90757866	C	TC	*Ptpmt1*	chr2:90757866	C	TC
Ip6k2	chr9:108700131	A	AC	*Ip6k2*	chr9:108700131	A	AC
Fam116b	chr15:89017849	CT	T	*Fam116b*	chr15:89017849	CT	T
Lst1	chr17:35325324	GA	A	**not expressed in B140H**			
expressed at lower levels in B140P				*Baiap2*	chr11:119860508	TC	T
mutation not found in B140P				*Ccdc88b*	chr19:6921808	CT	T
mutation not found in B140P				*Dck*	chr5:89203167	A	AT

Supplementary Table C.8: B117H missense mutations

MMuFLR identified potential frameshift mutations in the B117P and B117H cell lines. The mutations were screened *in silico* using IGV. None were found to be newly introduced to the B117H cells.

B117P				B117H			
Gene	Variant Position	Ref.	Var.	Gene	Variant Position	Ref.	Var.
Sap25	chr5:138082875	GT	G	*Sap25*	chr5:138082875	GT	G
Cyhr1	chr15:76477929	A	GA	*Cyhr1*	chr15:76477929	A	GA
C2cd3	chr7:107603881	A	AG	*C2cd3*	chr7:107603881	A	AG
Trerf1	chr17:47490865	AC	A	*Trerf1*	chr17:47490865	AC	A
Ptpmt1	chr2:90757866	C	TC	*Ptpmt1*	chr2:90757866	C	TC
Purb	chr11:6368602	CT	T	*Purb*	chr11:6368602	CT	T
Fam116b	chr15:89017849	CT	T	**found in B117H at lower levels**			
found in B117P at lower levels				*Ip6k2*	chr9:108700131	A	AC
found in B117P at lower levels				*Zfp319*	chr8:97847627	CAG	G

Supplementary Table C.9: Drugs used in combination screen

Source of the drugs used in the Ara-C combination screen.

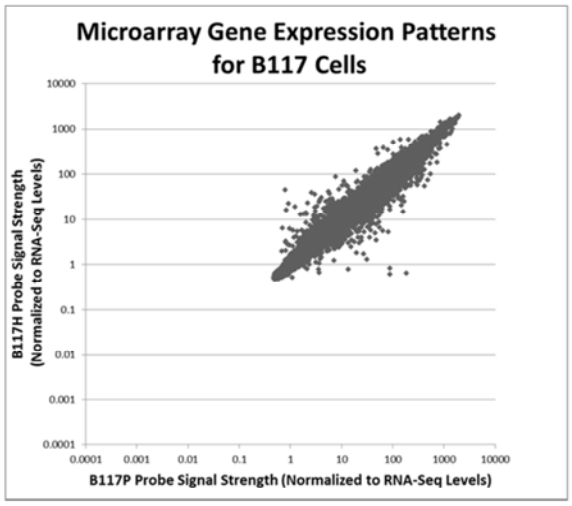
Drug	Source	Catalog #
5-Azacytidine	VWR International	TCA2033-100MG
5-Fluoro-2'-deoxyuridine	Sigma-Aldrich	F0503
5-Fluorouridine	Sigma-Aldrich	F5130
AG490 (JAK2 inh)	Sigma-Aldrich	T3434-5MG
AM 404	Tocris Bioscience	1116
Aminopterin	VWR International	10062310
AZD6244	Selleck Chemicals	S1008
Bay 11-7082	VWR International	80016-546
Bifemelane	Tocris Bioscience	0767
Bortezomib	LC Laboratories	B-1408
Cantharidin	Sigma-Aldrich	
Cladribine	Sigma-Aldrich	C4438
CPEC	NCI	NSC 375575
Dexamethasone	Fisher Scientific	ICN19456125
Dideoxycytidine	Sigma-Aldrich	D5782
Didox	Cayman Chemical	10009081
Erlotinib	ChemiTek	CT-EL002
etoposide	Sigma-Aldrich	E1383
GDC0941	Dr. Kevin Shannon (UCSF)	
GSK-3B Inh.	Calbiochem	361540
Harki1001	Dr. Dan Harki (UMN)	
Harki1002	Dr. Dan Harki (UMN)	
Harki1003	Dr. Dan Harki (UMN)	
Harki1004	Dr. Dan Harki (UMN)	
Harki1019	Dr. Dan Harki (UMN)	
Harki1020	Dr. Dan Harki (UMN)	
IKK Inhibit VII	Calbiochem	401486
Indomethacin	Fisher Scientific	ICN19021701
LC-1	Dr. Dan Harki (UMN)	
M1	Potter lab	
Mevastatin	Fisher Scientific	15997105
MG-132	Calbiochem	474790
Mifepristone	Sigma-Aldrich	M8046-100MG
MK-0457	Merck & Co.	new
Nelarabine	Selleck Chemicals	S1213
NFkB Inhibitor	Calbiochem	481406
Oxythiamine HCl	MP Biomedicals	100426
Parthenolide	Dr. Dan Harki (UMN)	
PD-0325901	Dr. Kevin Shannon (UCSF)	
PI-103	Dr. Kevin Shannon (UCSF)	
Prazosin	Fisher Scientific	ICN15378225
Prednisolone	Sigma-Aldrich	P4153-1G

RAD001	Selleck Chemicals	S1120
Rimcazole	Tocris Bioscience	1497
Ritonavir	Dr. David Potter (UMN)	
SCH23390	Sigma-Aldrich	D054-5MG
Sunitinib	Dr. Wendy Stock (Univ. of Chicago)	
Tetrahydrouridine	EMD Chemicals	584222-10MG
Trimidox	Cayman Chemical	10009083
Triptolide (HSP Inh)	A.G. Scientific	PG490

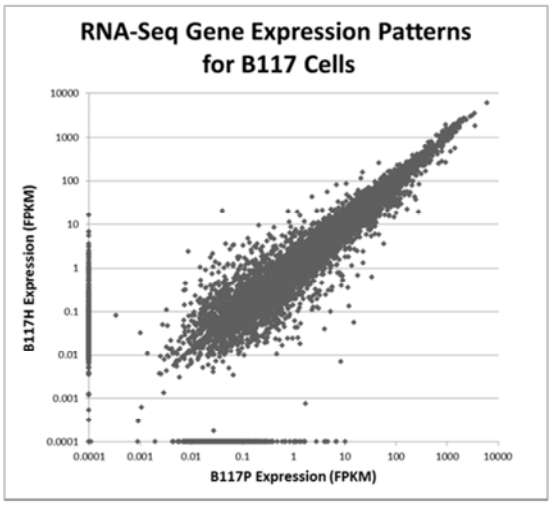
Supplementary Figure C.1: RNA-seq and microarray expression patterns

Distribution of expression comparing Ara-C resistant samples to parental samples using both microarray and RNA-seq data. a) Dot-blot of microarray gene expression comparing B117H to B117P. b) Dot-blot of RNA-seq gene expression comparing B117H to B117P. c) Dot-blot of microarray gene expression comparing B140H to B140P. d) Dot-blot of RNA-seq gene expression comparing B140H to B140P.

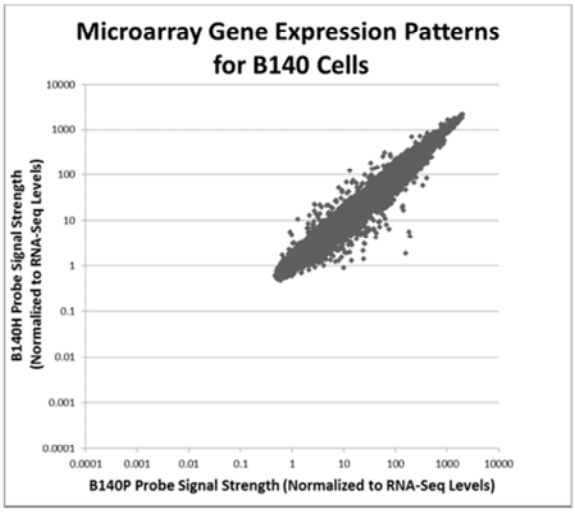
a)


Microarray Gene Expression Patterns for B117 Cells
B117H Probe Signal Strength (Normalized to RNA-Seq Levels)
B117P Probe Signal Strength (Normalized to RNA-Seq Levels)
10000
1000
100
10
1
0.1
0.01
0.001
0.0001
0.0001 0.001 0.01 0.1 1 10 100 1000 10000

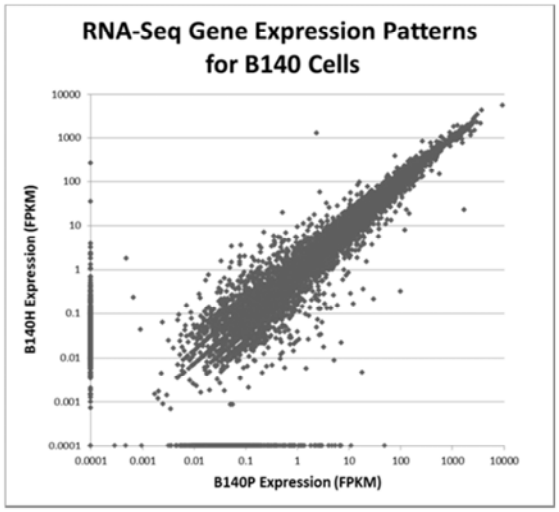
b)


RNA-Seq Gene Expression Patterns for B117 Cells
B117H Expression (FPKM)
B117P Expression (FPKM)
10000
1000
100
10
1
0.1
0.01
0.001
0.0001
0.0001 0.001 0.01 0.1 1 10 100 1000 10000

c)


Microarray Gene Expression Patterns for B140 Cells
B140H Probe Signal Strength (Normalized to RNA-Seq Levels)
B140P Probe Signal Strength (Normalized to RNA-Seq Levels)
10000
1000
100
10
1
0.1
0.01
0.001
0.0001
0.0001 0.001 0.01 0.1 1 10 100 1000 10000

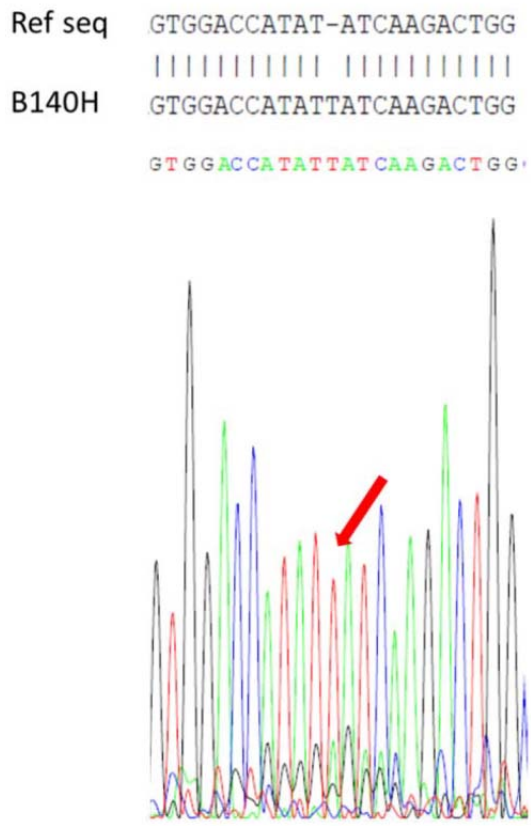
d)


RNA-Seq Gene Expression Patterns for B140 Cells
B140H Expression (FPKM)
B140P Expression (FPKM)
10000
1000
100
10
1
0.1
0.01
0.001
0.0001
0.0001 0.001 0.01 0.1 1 10 100 1000 10000

Supplementary Figure C.2: Dck insertion mutation in B140H

Insertion mutation in exon 4 of *Dck* in B140H is also present in the genomic DNA. (a) Sanger sequencing was used to interrogate the DNA sequence using primers provided in **Supplementary Table 1**. Intensity of the signal for the insertion indicates the mutation is homozygous. (b) Projected change in protein sequence for Dck in the B140H cells.

a



Normal Dck protein sequence:

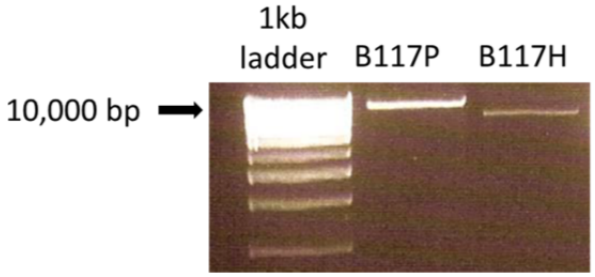
MATPPKRFCPSPSTSSEGTRIKKISIEGNIAAGKSTFVNILKQASEDWEVVPEPVARWCNVQSTQEEFEE
LTTSQKSGGNVLQMMYEKPERWSFTFQSYACLSRIRAQLASLNGKLKDAEKPVLFFERSVYSDRYIFASN
LYESDCMNETEWTIYQDWHDWMNSQFGQSLELDGIIYLRATPEKCLNRIYLRGRNEEQGIPLEYLEKLHY
KHESWLLHRTLKTSFDYLQEVPVLTLDVNEDFKDKHESLVEKVKEFLSTL

Projected Dck protein sequence in B140H:

MATPPKRFCPSPSTSSEGTRIKKISIEGNIAAGKSTFVNILKQASEDWEVVPEPVARWCNVQSTQEEFEE
LTTSQKSGGNVLQMMYEKPERWSFTFQSYACLSRIRAQLASLNGKLKDAEKPVLFFERSVYSDRYIFASN
LYESDCMNETEWTILSRLARLDEQPVWPKP

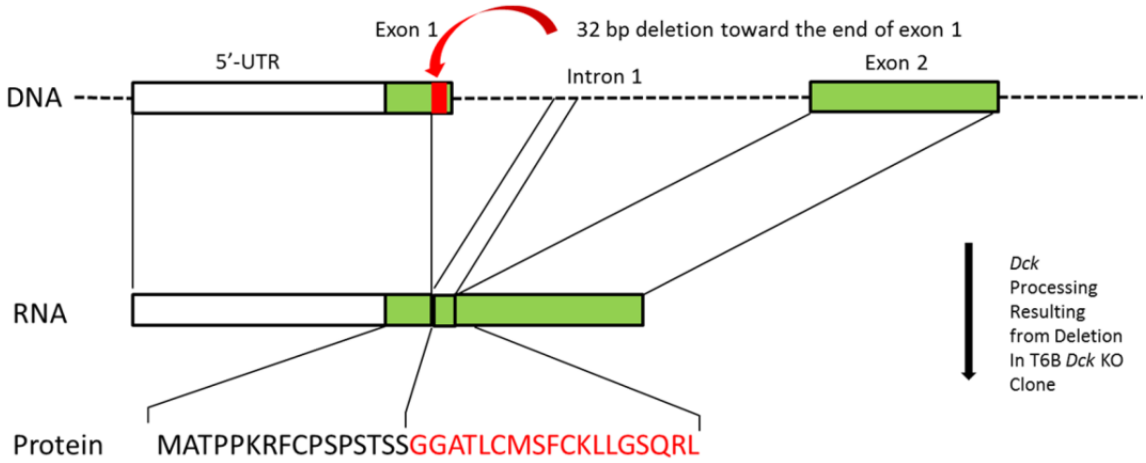
Supplementary Figure C.3: Extended PCR of 3’ end of B117H

Extended PCR of a C-terminus transcript segment of *Dck* in B117P and B117H DNA shows an approximate deletion of 1000 bp in B117H cells.


1kb
ladder
B117P
B117H
10,000 bp

Supplementary Figure C.4: RNA sequence of B117P T6B *Dck* KO clone

KO of *Dck* in the B117P T6B clone deleted a 32 bp segment toward the end of exon 1 which resulted in an abnormally spliced transcript containing 34 nucleotides from intron 1, starting at the 872nd nt of intron 1. The predicted result would be a protein consisting of 34 amino acids. The novel segment of protein is GGATLCMSFCKLLGSQRL.


Exon 1
32 bp deletion toward the end of exon 1
5′-UTR
Intron 1
Exon 2
DNA
RNA
Dck
Processing
Resulting
from Deletion
In T6B Dck KO
Clone
Protein
MATPPKRFCPSPSTSSGGATLCMSFCKLLGSQRL

Appendix D: Analysis of RNA-Sequencing Data

RNA-sequencing (RNA-seq) is a powerful technology for interrogating the transcriptome of cells. In addition to providing an unbiased approach for determining gene expression levels, RNA-seq provides detailed sequencing information which, with the power of super computers, can be evaluated to locate mutations, fusions, and unusually spliced genes. Since the advent of RNA-seq, dozens of software tools have been developed by various organizations. The Minnesota Supercomputing Institute (MSI) is committed to providing a state-of-the-art collection of tools for the biologists at the University of Minnesota to take advantage of the power of RNA-seq. To assist the MSI with their goals, I confined my analysis to the use of the tools acquired by MSI and developed by MSI, and provided practical testing of the MSI tools and detailed feedback.

MSI provides the tools within an umbrella software tool called “Galaxy”. Galaxy was originally developed at Penn State University. Its goal was to provide a biologist-friendly interface to various bioinformatics tools, with RNA-seq tools being just a subset of what is available.

The first step for acquiring RNA-seq data is the preparation of the biological samples. To generate high-quality RNA, the RNA should be isolated immediately from the tissue samples, or the tissues should be stored in a manner to preserve the RNA quality. RNAlater® (Life Technologies) works well for this. To perform the RNA isolation, it is important to use a method to produce high-quality RNA while eliminating DNA from the finished product.

Before submitting the RNA for sequencing, the biologist needs to decide on single or paired reads, the number of reads, and the length of the reads (cycles). When asked, I recommend paired reads, 80 million reads (40 million paired reads), and 50 cycles. At the University of Minnesota, the RNA is submitted to the BMGC where the quality of the RNA is tested. If the RIN value is greater than 7, the RNA is prepared for sequencing. This process involves amplification, fragmentation, selection of fragments of a similar length, and attachment of adaptors.

The output from the actual sequencing is a fastq file. The fastq file contains the actually sequences, a quality code for each of the bases read, and header information which is needed to uniquely identify each read and match paired reads. Prior to using the data, it is recommended the data be checked for quality using a tool called FastQC. If the "per base sequence quality" is poor at either end of the reads, the reads can be trimmed by a tool called FASTQ Trimmer. The final step of the quality control process is execution of the tool FASTQ Groomer, which formats the data for use by Galaxy.

To evaluate fusions and mutations, the FASTQ groomer files can be input directly into deFuse (to find fusions), or MMuFLR (to find missense and frameshift mutations). To evaluate expression data, the data must first be mapped to a reference transcriptome. Galaxy uses a tool called TopHat to do the mapping. TopHat does the mapping in two steps in order to make the most efficient use of computer processing. First, it executes the tool Bowtie to map the reads to a reference genome. Most of the reads can be mapped in this manner. However, there will be reads that actually span two exons. TopHat uses a more sophisticated algorithm to compare these reads to a reference transcriptome. One of the parameters to TopHat is the mean inner distance. This is the

fragment length minus the length of the paired reads. If this information is not known, it can be determined by running TopHat with defaults and then running NGS: Picard. Then TopHat can be run again with the actual mean inner distance supplied. The quality of the mapping can be determined by running the TopHat Stats PE tool.

The output from TopHat is a bam file and its associated index file. These files can be used as input to the Integrative Genome Viewer (IGV). IGV is a powerful visualization tool, which allows the user the ability to see all the reads mapped to an individual gene. It uses different color codes to identify reads mapped to multiple places and reads mapped outside the normal mean inner distance (which happens in the case of fusions). It also shows how reads span introns, which is helpful for evaluating alternative splicing events.

The output from TopHat can also be used as input to Cuffdiff, a tool that compares expression levels between multiple samples, or multiple groups of samples. It measures expression levels using the FPKM calculation. Because the RNA-seq process does not result in any technically-generated background noise, the user needs to decide on a minimum FPKM value for doing fold change comparisons. Low FPKMs (less than 1) can also be highly unreliable when evaluating fold changes. It is recommended IGV be used to evaluate the genes on a list of differentially expressed genes, to determine the validity of the gene expression values. For example, some of the reads may be mapped to multiple locations, or may be mapped outside the normal transcript area.

The tools currently available to locate undocumented or aberrantly spliced genes are totally inadequate. However, proving the existence of an abnormally spliced gene

can be verified by running a tool called Cufflinks, which can be run on a sample without the selection of a reference genome. Cufflinks will build transcripts from scratch by overlapping the reads. The output from Cufflinks can then be compared to a reference genome using CuffCompare.

The tool selected for Galaxy to locate fusions is called deFuse. DeFuse provides a sequence which includes a 100+ bases before and after the fusion point, with the fusion point identified by a bar (|). It identifies the two genes involved and whether the sequence occurred in a coding region, an intron, a UTR, upstream or downstream. It also provides a probability factor. A probability factor of greater than 0.5 indicates the fusion is most likely real. So far, deFuse has proven to be quite accurate in locating valid fusions, however it does not always accurately predict the location of the fusion point or the direction of the transcript. It also does not describe which exons are being transcribed from each gene, or the projected protein sequence resulting from the fusion. I'm currently working with MSI to develop a tool to provide this type of analysis of the deFuse output.