

Functional Drivers of Therapeutic Response in
Diffuse Large B Cell Lymphoma

by

Nicholas Samuel Davis

University Program in Genetics and Genomics
Duke University

Date: _____

Approved:

Sandeep S. Davé, Supervisor

Douglas Marchuk

Kris Wood

Jeffrey Marks

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the
University Program in Genetics and Genomics
in the Graduate School of Duke University

2018

ABSTRACT

Functional Drivers of Therapeutic Response in
Diffuse Large B Cell Lymphoma

by

Nicholas Samuel Davis

University Program in Genetics and Genomics
Duke University

Date: _____

Approved:

Sandeep S. Davé, Supervisor

Douglas Marchuk

Kris Wood

Jeffrey Marks

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in the
University Program in Genetics and Genomics
in the Graduate School of Duke University

2018

Copyright by
Nicholas Samuel Davis
2018

Abstract

Diffuse large B cell lymphoma (DLBCL) is the most common form of non-Hodgkin's lymphoma and a leading cause of death among B cell lymphomas. The disease displays a remarkable amount of genetic and clinical heterogeneity, hampering efforts at designing effective therapeutic agents. The last major change in frontline therapy for DLBCL came in 1997 with the addition of rituximab into widespread clinical use. Much effort has been put forth to identify new therapeutic agents in DLBCL, but these have been relatively unsuccessful as phenotypic heterogeneity renders them useful in only small percentage of patients, while responding patients develop resistance to single agents. To find broadly effective, safe therapeutics, we must first perform deep genomic characterization of the disease to understand its true heterogeneity. We can then study those genes and pathways with genomic alteration and evaluate their effects on the disease. Using this information, we can develop relevant drug screening tools that leverage genomic technologies to better understand therapeutic interactions with the disease and use them to predict possible therapeutic synergy. In this dissertation, I utilize this three-pronged approach to identify novel functional variation in DLBCL and use it to predict and verify synergistic therapeutic combinations for therapy.

First, I focus on the genomic variation across DLBCL patients. Since DLBCL displays vast genomic heterogeneity, we assemble the largest sequencing study of the

disease to date, consisting of 1,001 cases. Here, we applied exome and transcriptome sequencing to these cases with paired clinical data. We identified 150 putative driver genes that were recurrently mutated with a mean of 7.75 driver mutations per sample. Using genomic alteration types, we then classified these genes as oncogenes or tumor suppressors. We then used gene expression data to classify these tumors based on the cell of origin into activated B cell like (ABC) and germinal center B cell like (GCB) subtypes, which aligned well with standardized clinical methods. We found differential mutations in 20 genes based on these subtypes. We also found significant overlaps of driver genes with both co-occurrence and mutual exclusion, suggesting subnetworks of mutation. To identify functional variation, we then use a CRISPR screen to identify putative oncogenes and tumor suppressors genome-wide. Using this data in combination with matched clinical data, we then create the Genomic Risk Model for predicting single-patient outcomes based on genomic profile. This model outperforms current models and successfully predicts outcomes in most patients.

I then focus on placing two focal adhesion genes identified as recurrently altered in DLBCL into the functional context of the disease. The first gene is *RHOA*, a small GTPase that is found to be recurrently mutated in DLBCL in a hotspot-specific manner. Overexpression of both wildtype *RHOA* and the enriched R5Q mutant form both significantly increase proliferation in DLBCL cell lines. We also find that *RHOA* loss causes a loss of fitness in DLBCL lines, causing the cells to arrest in the G2/M phase of

the cell cycle and altering cellular morphology. I then develop a mouse model that knocks out *Rhoa* in either the full B cell lineage or germinal center B (GCB) cells specifically. We observe a massive loss of B cells across the B lineage driven by *Rhoa* deletion. In the GCB restricted knockout, we find that GCB cell numbers are reduced, the dark zone to light zone regulation in the germinal center is altered, as well as actin dysregulation in these cells. The next gene we modeled was focal adhesion kinase (*FAK*). Though *FAK* is not recurrently mutated in DLBCL, it is overexpressed in GCB cells and many cancers, and it is a master regulator of the focal adhesion pathway, which is overrepresented in mutation rates. Chemical inhibition and genetic knockdown of *FAK* in GCB cells causes cell death and a marked reduction in B cell receptor (BCR) signaling effectors. Further work in BCR signal transduction places *FAK* near the intracellular interface with the BCR, as the first effector molecules in the pathway have vastly reduced activity with *FAK* inhibition. Mouse models of *Fak* knockout also show a reduction in GCB cells, dysregulation of germinal centers in secondary lymph organs, and a reduction in serum levels of secreted immunoglobulins.

Lastly, we sought to understand the role of single agent therapeutics on gene expression to identify a method to use this data to inform combination therapy predictions. Using a panel of 152 FDA-approved drugs and 6 DLBCL cell lines, we screened all lines for drug efficacy. This revealed 3 classes of drugs: pan-effective, selective, and resistant. The selective drugs displayed pathway specific resistance as well

as a subtype specific sensitivity for certain drugs. RNA sequencing to quantify gene expression was then performed for all drug-cell line pairs. Overall gene expression patterns show that drugs targeting similar primary targets or targets in the same pathway induced tightly correlated gene expression patterns. We also found that changes in expression of genes like *MYC* were correlated with sensitivity, giving possible proxies for sensitivity and mechanisms of resistance. Furthermore, we discovered that baseline expression of the target gene was correlated with a higher dose requirement to achieve similar viability changes. Gene expression changes of the target gene were also found to be indicative of either sensitivity or resistance in some cases. We then developed a model for using gene expression to predict dual drug synergy that we are calling combination reversal of disease gene expression (cRDGE). This model accurately predicted both single agent effectiveness as well as combination synergy in previous datasets. We then validated this method across several drug combinations and found synergy between the tested combinations *in vitro*. We then tested one combination, panobinostat and ruxolitinib, and found it to be highly synergistic in both the cell lines tested within the study as well as a panel of cell lines representing a wide variety of B cell malignancies. We then show how ruxolitinib alone does not reduce STAT signaling in these cells at a lower dose, but sensitization of these cells with panobinostat greatly reduces STAT activity with the combination. Using xenograft

models, we then tested this combination *in vivo*, finding synergy of the drug combination and low hematopoietic toxicity.

Broadly, this dissertation contributes novel findings to the fields of B cell biology, lymphoma genomics, and therapeutic screening. Cancer is an incredibly challenging entity that requires an approach that integrates interrogation of genomic alterations, exploration of functional alterations, and development of new tools to identify therapeutics. Leveraging these tools to understand the molecular basis of lymphoma, its functional variation, and therapeutic interactions, we can more accurately diagnose, give prognoses, and treat patients with the disease.

To my grandparents Ed and Hazel Johnson and Sam and Betty Davis.

“If I have seen farther, it is by standing on the shoulders of giants.”

-Sir Isaac Newton

Contents

Abstract	i
List of Tables.....	xi
List of Figures	xii
1. Introduction	1
1.1 Molecular Alterations and Therapies in Cancer	2
1.1.1 Cancer as a genetic disease	4
1.1.2 Genetic alterations in cancer.....	6
1.1.3 Genetic alterations in DLBCL.....	10
1.1.4 Targeting genetic alterations for therapies.....	12
1.2 B Cell Origins and Their Relationship to Diffuse Large B Cell Lymphoma	16
1.2.1 Cell-of-origin in cancer.....	16
1.2.2 B cell lineage and germinal center biology.....	18
1.3 Focal Adhesion in B Cells.....	23
1.3.1 Role of the focal adhesion in immune function.....	23
1.3.2 The role of <i>RHOA</i> in normal and malignant cells	25
1.3.3 The role of focal adhesion kinase in normal and malignant cells	28
1.4 Cancer Genomics and Bioinformatics	30
1.4.1 Exome sequencing for discovery of somatic mutations	30
1.4.2 RNA sequencing for expression profiling.....	33
1.5 Overview of Results	35

2. Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma.....	39
2.1 Introduction.....	39
2.2 Methods	40
2.2.1 DLBCL patients and cell line samples.....	40
2.2.2 Sample processing for exome sequencing.....	40
2.2.3 CRISPR screen	42
2.2.4 RNA sequencing	44
2.2.5 Gene expression signatures	45
2.2.6 Genomic risk model.....	46
2.3 Results	47
2.3.1 Discovery of genetic drivers in DLBCL	47
2.3.2 Cell-of-origin effects in DLBCL.....	50
2.3.3 The connectivity of DLBCL driver genes	52
2.3.4 Functional genomics through CRISPR screening.....	54
2.3.5 Gene expression signature associations with genetic alterations and outcome	59
2.3.6 Clinical characteristics of DLBCL driver genes	65
2.3.7 Comparison of clinical versus genomic risk	73
2.4 Discussion.....	76
3. The Role of <i>RHOA</i> in Diffuse Large B Cell Lymphoma	79
3.1 Introduction.....	79
3.2 Methods	80
3.2.1 <i>RHOA</i> knockdown	80

3.2.2 Western blot.....	80
3.2.3 Cell cycle analysis	81
3.2.4 Mouse strains and breeding	81
3.2.5 Sheep red blood cell (SRBC) immunization	82
3.2.6 Flow cytometry of murine cells.....	82
3.3 Results	85
3.3.1 <i>In vitro</i> models of RHOA gain and loss	85
3.3.2 <i>In Vivo</i> mouse models of <i>Rhoa</i> loss	87
3.4 Discussion.....	90
4. Focal adhesion kinase in normal and malignant B cells.....	91
4.1 Introduction.....	91
4.2 Methods	92
4.2.1 <i>FAK</i> gene expression and <i>TP53</i> Status	92
4.2.2 Dose response curves	93
4.2.3 Western blot.....	94
4.2.4 Knockdown of <i>FAK</i>	95
4.2.5 Immunofluorescence	96
4.2.6 Mouse model of <i>Fak</i> loss in GC B cells.....	96
4.3 Results	98
4.3.1 Identification of <i>FAK</i> in B cell receptor signaling	98
4.3.2 <i>In Vivo</i> mouse models of <i>Fak</i> loss.....	103
4.4 Discussion.....	106

5. Gene Expression Perturbations upon Drug Treatment Enable Identification of Resistance Mechanisms and Novel Combination Therapies in Lymphoma.....	109
5.1 Introduction.....	109
5.2 Methods	112
5.2.1 Cell lines and culture.....	112
5.2.2 Compounds.....	112
5.2.3 Dose response curves	114
5.2.4 Drug treatment and RNA-seq library construction	114
5.2.5 Combination predictions	116
5.2.6 Combination screening	117
5.3 Results	118
5.3.1 Single agent screening reveals sensitive and resistant drug classes in DLBCL	118
5.3.2 Gene expression changes reveal molecular mechanisms of sensitivity and resistance in DLBCL	120
5.3.3 Predicting Synergistic Combinations Using Expression and Activity Data	122
5.3.4 HDAC and JAK/STAT combination therapy	126
5.4 Discussion.....	131
6. Conclusions.....	135
References	143
Biography.....	156

List of Tables

Table 1: Gene Set Signatures.....	61
Table 2: Antibodies used in RHOA study	84
Table 3: Antibodies used in FAK study	94
Table 4: Compounds.....	113

List of Figures

Figure 1: B cell development	19
Figure 2: Overview of DNA sequencing..	32
Figure 3: Overview of RNA-sequencing.....	35
Figure 4: The landscape of genetic drivers in 1,001 DLBCLs.....	49
Figure 5: ABC/GCB-based classification of DLBCL.	51
Figure 6: Comparison of RNA-seq subtype score.	52
Figure 7: Genetic alterations that are enriched in ABC versus GCB DLBCL	53
Figure 8: Connectivity network of genetic alterations in DLBCL	54
Figure 9: Schematic of CRISPR screen performed for six cell lines in triplicate	56
Figure 10: Distribution of CRISPR gene scores genome wide	57
Figure 11: CRISPR scores for 35 DLBCL oncogenes with functional classification.....	58
Figure 12: Schematic of integrative analysis of gene expression, genetic alterations, and outcome	60
Figure 13: Integrative analysis of gene expression, genetic alterations, and outcome.....	64
Figure 14: Common models of patient stratification	66
Figure 15: Genetic alterations and survival.....	68
Figure 16: Schematic depicting multivariate survival modeling.	69
Figure 17: Genomic risk model	71
Figure 18: Survival predictions within model types.....	72
Figure 19: Modeling risk in known high risk groups.	73
Figure 20: Comparison of clinical risk model with genomic risk model.	74

Figure 21: Genomic risk model identifies high risk groups in all clinical response groups.	75
Figure 22: R5Q mutations in DLBCL.....	85
Figure 23: <i>RHOA</i> knockout disrupts cell cycle.	87
Figure 24: <i>Rhoa</i> knockout in B lineage.....	88
Figure 25: <i>In vivo</i> knockout of <i>Rhoa</i> in GCB cells.....	89
Figure 26: Network of mutations found in DLBCL.	91
Figure 27: Relative expression of <i>FAK</i> (PTK2) in B cell lineage.....	98
Figure 28: Dose response to <i>FAK</i> inhibition..	99
Figure 29: Effects of BCR activation on <i>FAK</i>	100
Figure 30: <i>FAK</i> inhibition reduces downstream BCR signaling.	101
Figure 31: <i>FAK</i> knockdown reduces BCR signaling.	102
Figure 32: Mouse breeding schematic for <i>Fak</i> knockdown in GCB cells.....	104
Figure 33: <i>In vivo</i> knockout of <i>Fak</i> in GCB cells.	105
Figure 34: Proposed model of <i>FAK</i> involvement in BCR signaling.	107
Figure 35: Study design expression-based predictions of synergy.....	111
Figure 36: Single agent effects reveal classes of effective drugs in DLBCL..	120
Figure 37: Compound gene expression signatures reveal classes of drugs.....	121
Figure 38: Gene expression reveals mechanisms of resistance and sensitivity.....	122
Figure 39: Schematic for cRDGE model.....	123
Figure 40: Validation of cRDGE model in DREAM data.	124
Figure 41: Figure 42: Validation of synergy of cRDGE predicted combinations.	125
Figure 43: Validation of panobinostat and ruxolitinib synergy <i>in vitro</i>	127

Figure 44: Sensitizing STAT inhibition with panobinostat and ruxolitinib synergy.....	128
Figure 45: Panobinostat sensitivity across B cell malignancies	129
Figure 46: Panobinostat with ruxolitinib synergy	130
Figure 47: Synergy in xenograft models.	131

1. Introduction

Over the past century, isolated anti-microbial agents have vastly reduced mortality from exogenous disease, but in turn cancer has risen as a leading cause of death worldwide. It is estimated that one in two men and one in three women will face some type of neoplastic event in their lifetime. Even in cases where these events are not immediately life threatening, quality of life can be greatly improved with safe, effective therapies for these diseases. Large-scale efforts have been put forth to find new therapies, and while some progress has been made, millions still succumb to the disease annually. Surgical excision and radiation therapy are appropriate in certain types of cancers, but improvements in these techniques have only improved outcomes at the margins, and often are augmented with chemotherapy. Chemotherapeutics are the most commonly used therapy in cancer treatment, yet resistance and secondary neoplasms have proven to be quite common, difficult problems to overcome across cancer types. Massive efforts, both public and private, have been ongoing to find therapeutics that more effective than current standards of care. While some of these efforts have been fruitful, there are still millions of patients annually that would benefit from more diverse and effective identification and treatment of cancer.

Lymphomas are a group of cancers arising from lymphocytes that display a wide heterogeneity in clinical outcome, genetic features, and cell-of-origin. These hematological malignancies can arise from most cell types during lymphoid

differentiation, including cells from early lymphoid progenitor cells, B cells, T cells, follicular cells, and NK cells. The site of primary disease can be anywhere the normal cells are found, primarily the lymph organs, but tumors can even be found in secondary sites, such as the liver or gut and even immunoprivileged sites, such as the CNS and testes. Clinical outcomes are also quite disparate across lymphoma types, ranging from higher 5-year survival rates of over 80% for classical Hodgkin's lymphoma and primary mediastinal B cell lymphoma to less than 20% for hepatosplenic T cell lymphoma [1]. This diverse heterogeneity is also reflected in molecular markers for disease, with histological variation across disease subtypes. The vast heterogeneity in both clinical outcomes and underlying biology present a complex problem of understanding the disease, but recent tools for genomic interrogation present powerful new avenues to interrogate the biology of lymphoma. The work of this dissertation is to better understand the genomic basis of the disease, define the role of key players in the underlying biology of normal and malignant B cells, and to leverage this information to find therapeutic strategies for fighting this disease.

1.1 Molecular Alterations and Therapies in Cancer

Though cancer is only defined by unchecked growth of abnormal cells, cancer is a complex disease that exists on a spectrum of genetic alteration, signaling alterations, metabolism, and clinical outcomes. Advances in technology for individual assays and in the throughput of these assays has allowed for orders-of-magnitude more data to be

collected for each of these, yet the integration of these data types is still difficult, with connectivity of this data remaining largely obscured. A single biopsy sample removed from a patient often has millions of data points created for it, yet reducing these to personalized, actionable data is still the exception to the rule. Though millions of single nucleotide variants have been cataloged, few have actionable targets, such as vemurafenib for BRAF V600E/K variants{?}. Targeted cancers often acquire resistance to single agent therapies, thus new therapies to related pathways or other processes must be developed. Similarly, expression differences have been invaluable in understanding the underlying biology of cancer, yet few drugs targeting irregularly expressed genes, such as Herceptin targeting HER2 amplified breast cancer, have been developed. In order to continue on this front of targeted therapies based on biological variation in these cancer types, a more complete understanding of the biological variation and normal variation underpinning the affected cells must be identified.

Lymphoma, specifically DLBCL, presents an even more complex challenge. While some cancers are predominated by genetic lesions in specific genes, DLBCL has a long-tail distribution of mutations, with almost all commonly mutated genes falling under 10% of cases. With so many paths to becoming dysregulated, it becomes imperative to develop therapies that focus on the disease as a whole instead of individual mutated driver gene inhibitors. In order to do this, I have taken an approach

that identifies both recurrently mutated genes as well as functional characterization of all players genome wide that are responsible for oncogenic processes.

1.1.1 Cancer as a genetic disease

Cancer is unique among common diseases as it is generally a sporadic disease caused by genetic alterations of small numbers of cells. Unlike most exogenous disease, which involves an infectious agent or other harmful exposure, cancer arises from one's own cells. This process begins with a single somatic cell acquiring a first-hit mutation. This mutation is often proliferation or genome stability related, and leads to a clonal expansion of this cell. As these cells are proliferating, more mutations will occur. This process creates subpopulations of cells that can have a wide variety of mutations between them. These subpopulations undergo selection, with cells with pro-growth and proliferation mutations outcompeting other subclones and becoming the dominant clone within the cancer. The dominant subclones begin to outgrow and become pathogenic, and mutations continue accruing, leading to subpopulations that may be drug resistant.

There are many mechanisms by which mutations occur in somatic cells, with different mechanism leading to distinct molecular patterns and outcomes.

Environmental exposure to specific mutagens drives mutation rates and certain mutations in some cancers. Melanoma and gastric cancers display a high degree of mutations correlated with their exposure to ultraviolet radiation and environmental mutagens respectively, while soft tissue sarcomas and astrocytomas are somewhat

insulated from these environmental factors and show lower mutation rates [3]. The most common mechanism in age related cancer is fidelity errors upon DNA replication in the normal cell cycle. As cells copy their genome in the process of cell division, small numbers of errors occur during each of these cycles. As cells continue to divide, they accrue larger numbers of mutations, leading to higher rates of cancer in older populations. In certain tissue types, there are specific mechanisms for mutation as well. In germinal center B cells, activation-induced cytidine deaminase (AID) is employed to create mutations in specific areas of immunoglobulin genes, leading to more specific antibodies during the process of somatic hypermutation. While this process is generally restricted to immunoglobulin regions, off target effects have been identified targeting the same motif [4]. Mutation rate also increases when early loss of DNA repair genes, such as TP53 or MSH2, are present [5]. Defining these mutational patterns is critical to the understanding of the underlying pathogenesis of the disease, as well as what populations might be affected.

These mutational mechanisms create cancer cells with a wide variance in mutational burden. Cancer types such as medulloblastoma and acute myeloid leukemia have less than 0.5 somatic mutations per megabase, while melanoma and squamous cell carcinoma have more than 10 somatic mutations per megabase [6]. While most of these are passenger mutations, most driver genes have a mutation rate of less than 2-20% in most cancer types. An analysis of lung adenocarcinoma showed that 15% of tumors

showed no known mutation in a cancer hallmark pathway. This data suggests a much deeper analysis of many more tumor-normal samples must be completed in order to find cancer causing variants. For DLBCL, it is estimated that approximately 1000 samples would be necessary to identify all mutations in greater than 3% of cases with 90% power in 90% of genes [6]. A study of this size is an order of magnitude above current standards of sample sets, suggesting the need for large single cancer type studies to identify most driver mutations within a single cancer type.

1.1.2 Genetic alterations in cancer

Genetic alterations in cancer fall into several broad classes, each having overlapping effects with other classes but are functionally created by different mechanisms. The most common alteration is a single nucleotide variant (SNV). These variants occur across the genome, but most efforts are focused on identified functional regions such as protein coding regions and promoters. Within protein coding regions, SNVs fall into two categories: synonymous mutation and nonsynonymous mutation. Synonymous mutations code for the same amino acid. They are generally benign, but they can alter protein amounts due to codon bias in translation. Nonsynonymous mutations change the codon to a different amino acid or stop codon and can have a variety of effects based on mutation type and position in the gene. These missense mutations change a single amino acid and can alter the conformation and thus function of the protein. While many impair the ability of the protein to function, some increase the activity of a protein or

make it constitutively active. These activating mutations usually target a single codon and can be found in hotspots, with higher mutation rates at that position than the rest of the gene body.

The second class of genomic abnormalities are indels, or small insertions or deletions of 1 to 10 nucleotides. If the number of nucleotides is divisible by 3, the protein is still read in frame. If the number of nucleotides is not divisible by 3, this creates a frameshift. Frameshifts often create a stop codon immediately downstream of the indel, truncating the protein. Even if the protein is not truncated, the rest of the sequence is read out of frame with distal amino acids affected, often rendering the protein inactive or the target of the unfolded protein degradation.

Larger structural alterations are very common in cancer compared to somatic cells and come in a variety of classes and functions. Large insertions and deletions consist of more than 10 nucleotides and can disrupt normal gene function. Genetic rearrangements change the orientation of genes within a chromosome, and this may change protein or transcription levels within a cell. Copy number variation (CNV) is very common within cancer cells. Deletions reduce or ablate proteins coded within the affected region, while copy number gains usually increase transcription and protein levels of affected genes. CNVs can exist for stretches of single genes, whole chromosome arms, or entire chromosomes. These larger changes affect all genes along the duplication or deletion. The last class are translocations, which involve large parts of one

chromosome fusing to another chromosome. Hematopoietic malignancies have high rates of translocations. More than 90% of follicular lymphoma cases show a t(14;18) translocation, connecting IgH and BCL-2 [7]. The canonical Philadelphia chromosome is created by a BCR-ABL1 translocation t(9;22) in over 95% of CML [8]. All of these genetic alterations can be identified through high-throughput massively parallel sequencing, but the task then becomes assigning function to each of the genes in the specific context of that cancer type.

Once genes are identified to have any these recurrent genetic alterations, the next step is to assign function of this gene within the cancer cell type of interest. Functional genes in cancer biology fall into two broad categories: oncogenes and tumor suppressors. Oncogenes are genes that promote tumorigenesis when activated or overexpressed. Genetic alterations in these genes are generally copy number gains, overexpression, translocation to highly active promoters, and hotspot mutations (overrepresentation of mutations at a single amino acid position leading to more active protein). Tumor suppressors are usually genes that function to rein in cell growth and promote apoptosis in normal cells. These genes are often marked by genetic alterations including missense/nonsense mutations across the gene body, deletion, and dominant negative genetic alterations. Together these mutations make up a class of mutations called driver mutations. These mutations confer a selective advantage upon the cell, leading to increased growth and proliferation, and creating a clonal expansion of these

cells within the tissue of origin. Other genes that are mutated but have no fitness effect on the cell are considered passenger mutations. These often are identified as false positive in smaller sequencing studies, but with higher number of samples the false positive rate drops for discovery. Some genes, such as those with open chromatin in the cell type of interest, see a higher mutation rate above that of background passenger mutations, so functional validation of these genes is required to understand the specific function of the gene in this context.

Genetic alterations cluster in functional categories that represent the hallmarks of cancer [9]. Activating mutations or overexpression of genes that increase cell proliferation, telomerase activity, angiogenesis, tumorigenic inflammation, invasion, and cellular metabolism all lead to a more cancer-like state of the cell. These include genes like *EGFR*, *MYC*, *CDKs*, *VEGF*, telomerases, and glycolysis genes. These are commonly overexpressed or activated in tumor samples, and they are considered to be oncogenic. Other genes are deleted or inactivated by other mechanism that are involved in growth suppression, apoptosis, and genome stability. These include genes such as *TP53*, *FOXO1*, and integrin-related genes. Other genetic players have also been identified that act through mechanisms of immune evasion, such as *PDL-1*. These genes do not have a direct effect on growth or proliferation, but they do allow for the cancerous cells to evade the normal mechanisms of immune surveillance and elimination *in vivo*.

The functional characterization of these genetic alterations is critical to understanding the underlying biology of the cancer and the multitude of mechanisms that have emerged to lead to a similar disease of unchecked cell outgrowth. Different genetic alteration can lead to distinct subtypes of disease, even within the same cell of origin, with individual outcomes being driven by this genetic pattern. Learning to leverage this information can allow us not only to look back and find patterns that emerge to explain pathogenesis and outcome, but also to develop novel strategies for treating these patients based on their specific genetic alterations.

1.1.3 Genetic alterations in DLBCL

Diffuse large B cell lymphoma is the most common type of non-Hodgkin's lymphoma, and many studies have been performed to identify the underlying genetic alterations that give rise to the disease. Approximately 50% of patients succumb to the disease, yet many others achieve long-term survival. Early studies in this effort used microarray based quantification of gene expression to try to understand the wide gulf in outcomes of patients diagnosed with DLBCL [10]. These early studies described distinct subtypes that were largely based on the cell-of-origin of the subtype of the disease, with clinical outcomes associated with these subtypes [11-13]. The germinal center B cell like (GCB) subtype resembles gene expression patterns similar to those of normal B cells isolated from the germinal center of peripheral lymph organs, such as the spleen and lymph nodes. The activated B cell like (ABC) subtype displays gene expression profiles more

closely related to the of normal post-germinal center active B cells. There also exists an intermediate population that does not clearly fall into either group, with expression profiles resembling both types to some degree. These subtypes were then shown to have disparate clinical outcomes, with ABC patients succumbing to the disease at a higher rate than those in the GCB subtype. Other gene expression-based studies have identified possible subtypes of clinical importance. Lenz et al. describes a germinal center like classification while also including MHC class II, lymph node, and proliferation signatures associated with clinical outcomes [14].

Several studies have also investigated somatic mutations in DLBCL by tumor-normal paired exome sequencing [15,16]; and I will be focusing on the largest study in this field from our lab[17]. The striking finding of all of these studies is the remarkable degree of genetic heterogeneity in DLBCL. Over 300 somatic variants were implicated in the disease, yet almost all are mutated in less than 10% of patients. This long-tailed distribution of mutations suggests many possible paths to dysregulation and lymphomagenesis. Classical cancer drivers, such as *TP53*, were identified as recurrently mutated, as well as lymphoma-specific genes, such as *CARD11*. While some mutations show an even distribution across subtypes, other genes were mutated in a subtype specific manner. *MYD88* and *CD79B* were specifically mutated in the ABC subtype, while *GNA13* and *BCL2* were mutated in the GCB subtype. These subtype-specific data give clues as to the underlying biology and process of lymphomagenesis. *GNA13*

mutations in GCB cells have been shown to increase proliferation and alter migration in the normal germinal cells, thus leading to lymphomagenesis [18]. Functionally, the recurrently mutated genes fall into categories that resemble the hallmarks of cancer. Interestingly, one of the most altered pathways in DLBCL is chromatin modification, with common mutations in *MLL2* and *SETD2*.

In addition to identifying recurrently mutated genes, these studies also shed light on the high level of heterogeneity within the disease, requiring larger study sizes to accurately identify most of the players in the disease. These studies are underpowered to make prognostic claims about driver mutations, as the vast majority of genes are mutated in less than 10% of samples. Larger studies with associated clinical data would be powered to make these kinds of assessments.

1.1.4 Targeting genetic alterations for therapies

Early therapeutic agents for cancer were discovered in an entirely phenotypic manner. The effects of these drugs were invaluable, but the exact cellular mechanisms of the drugs were unknown. As technology advanced, cell line-based screens of compounds were used to define the effects of specific compounds on cellular viability. These screens identified cytotoxic compounds that could be used at specific doses to push cancer cells toward cell death, while limiting damage to normal healthy cells. The revolution in molecular biology led the identification of transforming genes that allow for normal cells to become cancerous, and many of these became attractive targets in therapeutic

discovery. This changed the direction and specificity of possible drug development from a drug-to-phenotype method to a target-based approach that uses genomics to inform targets. These targets can then be used to develop drugs based on the protein structure, and thus activity, involved. Once targets are identified by genetic study, protein biology and medicinal chemistry can be used to create molecules that are specifically targeted to these types of targets. This model of tailoring drugs to oncogenic targets was effective in its goal of creating powerful targeted inhibitors, yet resistance to these drugs became a major problem. This suggests new mechanisms have emerged that are compensating for the inhibition of the initial target. Genetic identification of targets became much easier with the emergence of high-throughput genomic technologies. Deep sequencing allows for the identification of driver mutations across cancer types. Both mutations, found through whole exome or genome sequencing, and expression changes, found through RNA quantification, can inform targets for drug development.

A popular theoretical framework for drugs that would be specific to cancer while minimizing off target effects is creating drugs that target mutant form of proteins. Vemurafenib was developed using this model. In melanoma, about 60% of patients carry a V600E point mutation that drives tumorigenesis, growth, and proliferation, making it a very attractive target in an aggressive disease [19]. The drug acts by interrupting the B-Raf/MEK interaction in cells that have the V600E mutation, and thus inducing apoptosis [20]. Wild-type *RAF* remains unaffected and can be paradoxically

activated, making it pinpoint targeted to affected cells [21]. Initial response to the drug was promising, but resistance to the drug quickly emerged, with around 40% of resistance can be explained by *PDGFRB* overexpression, *NRAS* mutations, and hepatocyte growth factor secretion by normal stroma [22-24]. Drugs based on predominant mutations can be useful clinically, but new strategies for design, including dual drugging, are called for in these cancer types to overcome resistance.

Other types of genetic lesions can be targeted if the product is targetable and unique to the cancer. Chronic myeloid leukemia and a small number other cancer contain a translocation of chromosome 9 and 22 that creates what is known as the Philadelphia chromosome. This creates a linkage between the *BCR* and *ABL1* genes, forming a chimeric protein BCR-ABL. Since this is a cancer specific protein that drives the tumor, this is an attractive target for therapy[25]. Imatinib was developed as a targeted inhibitor for this BCR-ABL that is quite selective for the chimeric protein, though it has some off-target effects[26]. Imatinib is still used as frontline therapy in CML and other Ph+ cancers. Though these types of translocations that create novel targets are rare in cancer many cancer types, this drug illustrates the power of using genetic lesions to inform targeted therapies.

Gene duplication and gene expression can also be used to inform therapeutic target in cancer. Around 20-30% of early stage breast cancers have an amplification and resulting overexpression of the gene *HER2*, and *HER2* overexpression results in more

aggressive disease without specific treatment [27]. This overexpression leads to downstream growth factor signaling that is necessary and sufficient for tumor formation in HER2+ breast cancer models [28]. This makes HER2 an attractive target in these breast cancer types, and a monoclonal antibody against it was introduced called trastuzumab. This antibody binds to surface HER2 and prevents cleavage of the protein into the more active p95-HER2 form, thus downregulating growth signaling [29]. Other targets have also been found that leverage this overexpression phenotype by other cancers to reduce growth factor signaling.

Genomics has become an invaluable tool in drug discovery for cancer. Genomics allows us to find the small differences in the genome or transcriptome that drive these malignancies, and drugs can be developed to combat these changes. Sequence variation, copy number change, overexpression and even translocation can be used to develop therapies to these targets. Other uses for genomics still present themselves in this field, however. Strategies for combining genomics with large-scale screening present an opportunity to understand not just the mechanism, but the entire cell state before and after treatment. Combining these powerful tools presents an opportunity to understand the effects of these drugs in a holistic manner by assaying the entire cell at once.

1.2 B Cell Origins and Their Relationship to Diffuse Large B Cell Lymphoma

1.2.1 Cell-of-origin in cancer

Cancer displays an extreme amount of clinical, molecular, and histological heterogeneity. From the earliest identifications of cancer, location of tumors has been noted to be related to the outcome of the patients. As diagnostic and molecular techniques advanced, typing and subtyping of cancer became the crucial method for understanding cancer. Single organs and tissues may develop diverging cancer subtypes with radically different treatment options and patient outcomes. These subtypes are molecularly different, but evidence suggests that the cell-of-origin of these cancers drives much of the biology of the emerging tumor.

Any cell with proliferative capacity is a possible cell or origin for an emerging cancer. As mutations accumulate, this proliferation becomes unregulated and neoplastic events occur. To define the cell-of-origin of tumors, many approaches use the normal histology of the cellular hierarchy of the affected organ, yet histologically similar samples may also have diverse outcomes. Histologically identical glial-cell tumors arising in different parts of the CNS have different patterns of mutation and other chromosomal abnormalities, suggesting they originated in separate subpopulations in these distinct regions [30].

Cells maintain high fidelity copies of the entire genome, but each uses a subset of the genes to be transcribed and used. Different tissues and cells display a substantial

variation in epigenetic factors, such as methylation, histone modification, and chromatin accessibility [31,32]. Cell-of-origin has also been shown to drive chromatin organizational shape and the mutational pattern of the emerging disease [33]. The innate biology of the cell that the cancer derives from drives the location and type of mutation present within the emerging tumor.

Understanding the biology of the cell-of-origin of the tumor is also critical to the development of therapies. Certain types of breast cancer retain their innate nuclear receptor signaling. These cancers have estrogen receptor and/or progesterone receptor expressed and respond in a proliferative fashion when exposed to even normal biological levels of estrogen [34]. The development of selective estrogen receptor modulators, like tamoxifen, have greatly improved outcomes in patients with this disease [35]. Similarly in prostate cancer, the use of androgen deprivation therapy leverages the ability to block the normal androgen associated growth pathways to slow down prostate cancer cells [36].

Defining the cell-of-origin of cancer subtypes is a powerful tool in studying the disease and for development of treatments. The remarkable concordance of cell lineage and the phenotype of the resulting cancer suggests that lineage restricted mechanisms can play a critical role in the pathogenesis of the disease. This is especially true in lymphoid malignancies due to their tightly specified lineage restriction and high proliferation rate at these late differentiation points. Utilizing these known biological

processes that the underlying cells rely on, possible therapeutics can be devised that exploit these characteristics.

1.2.2 B cell lineage and germinal center biology

The cells-of-origin for lymphomas are lymphocytes, and malignancies arise from many stages of differentiation, which are summarized in figure 1. Hematopoietic stem cells give rise to the first lineage committed cells, in either the myeloid or lymphoid lineage. The myeloid lineage differentiates into erythrocytes, granular leukocytes (eosinophils, basophils, and neutrophils), monocytes, and megakaryocytes. Hematopoietic stem cells can also differentiate into common lymphoid progenitor cells, and these cells largely differentiate into the T and B cell lineages.

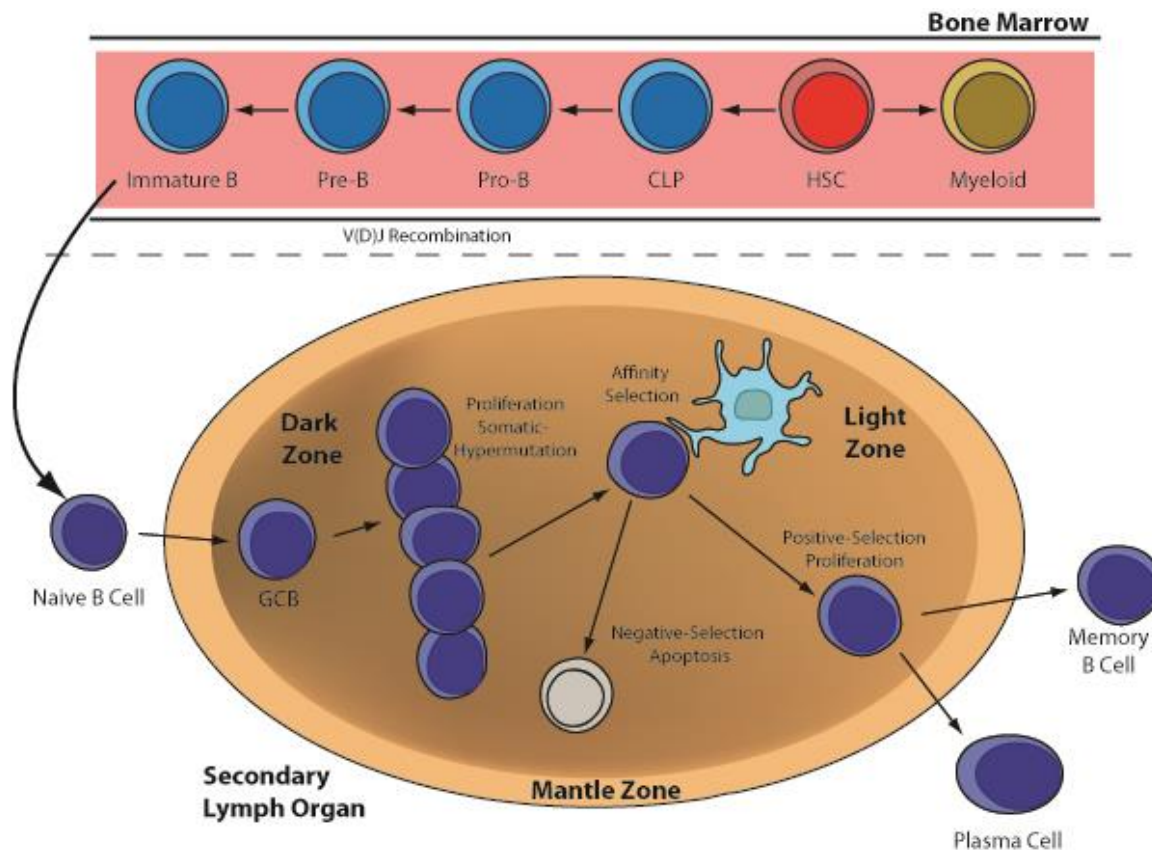


Figure 1: B cell development. Arrows indicate differentiation or the labeled process. Committed B lineage cells are in blue.

In the B cell lineage, the common lymphoid progenitor cell remains in the bone marrow and becomes a pro-B cell. Here, the antigen-independent process of V(D)J recombination begins. This process involves the recombination of immunoglobulin (Ig) genes within individual B cells that results in a nearly random recombination of the 2 constant, 44 variable, 27 diversity, and 6 joining regions of the heavy chain, as well as those of the light chains. This high number of possible recombination products leads to highly variable Ig products with high affinity to possible non-self proteins. This near-

unique recombination can also be leveraged to understand the clonality of the cells present in a tumor by sequencing the Ig regions present. The pro-B cells then differentiate into pre-B cells and then immature B cells within the bone marrow. These cells express a pre-B cell receptor that later becomes a mature BCR. As these cells exit the bone marrow into the periphery, they become naïve B cells and home to secondary lymph organs such as the spleen and lymph nodes. As these cells come into contact with antigen through contact with antigen-presenting cells, they activate and transition to marginal zone B cells and follicular B cells.

These cells then home to the germinal centers of secondary lymph organs, where they become germinal center B cells (GCBs). They first enter the dark zone of the germinal center, where they undergo massive clonal expansions. Dark zone expanding GCBs are some of the most rapidly dividing cells in the body. These cells also undergo somatic hypermutation, a process that further mutates the regions of the Ig genes that come into contact with antigen. This mutation is driven by activation-induced cytidine deaminase (AID) that deaminates cytidine at WRCY motifs. These cells then home to the light zone of the germinal center, where they undergo affinity selection. This process involves B cells and their receptor coming into contact with antigen-presenting dendritic cells. These cells are all poised to undergo apoptosis without external signaling, which protects from unchecked outgrowth of these rapidly dividing cells. GCB cells with B cell receptors (BCR) that are not high affinity for antigens presented by dendritic cells

rapidly undergo poised apoptosis. Cells that do have high affinity interactions with the antigen will have pro-growth signaling activated, and they will return to the dark zone to undergo more rounds of clonal expansion and hypermutation. After a number of rounds in this dark zone-light zone cycle, the cells will then undergo class switch recombination where they can transition from expressing the early IgM/IgD to later immunoglobulins, such as IgG, IgE, or IgA. These cells will then exit the germinal center and become late effector cells.

During acute infection, a large number of plasma cells will be produced. These cells produce a large quantity of secreted Ig and are the main effector cells of the B cell lineage. The other cell type produced are memory B cells. Unlike other B cells, these cells are extremely long-lived and do not divide rapidly. These cells then home to the bone marrow where they remain until they come into contact with antigen and become activated. These cells can remain in the body for decades and will rapidly switch to dividing into plasma cells to produce the adaptive immune response upon being challenged by the same antigen again.

Understanding the normal biology of B cell development and germinal center biology specifically gives insight into the biology of the tumors that emerge from these normal cells and how lymphomagenesis is driven. Somatic mutations in GCB cells can be driven by many normal processes in the B cell developmental process. Clonal expansion in the dark zone of the germinal center leads to rapid cell division over many

separate occasions. This cell division, like all divisions, creates small variation in the genome that will be carried on with every subsequent division. Additionally, somatic hypermutation can create mutations outside of the targeted Ig regions, and these off-target effects can be detected in disease associated genes from patient samples [37]. Since these cells are poised to undergo apoptosis without BCR signaling, it follows that genes in the B cell receptor signaling pathway or essential apoptotic genes would be overrepresented in mutations.

The underlying normal biology of the germinal center B cells can also be used to develop therapeutics for germinal center derived disease. Classically, front-line therapies for GCB lymphomas include glucocorticoids like prednisone. Glucocorticoids have long been known to have immunosuppressive effects on normal lymphoid cells through several mechanisms [38], and these have been leveraged in lymphoma treatment to be effective therapeutics for decades [39]. Newer therapies have also used the normal B cell surface expressed markers to effectively treat DLBCL. Rituximab, developed in 1997, targets CD20 on the cell surface of early to GC B cells. Once bound, rituximab causes cell death in these cells by induction of apoptosis, antibody-dependent cellular cytotoxicity, and downregulation of B cell receptor signaling [40,41]. Newer developments in immune therapies have also been targeted at DLBCL. Chimeric antigen receptor T cells (CAR-T) have been optimized for CD19, which is expressed on B cells and most arising

lymphomas, including DLBCL [42]. These chimeric T cells are autologous transplants of the patients own cells, and are highly effective in B cell malignancies [43].

1.3 Focal Adhesion in B Cells

1.3.1 Role of the focal adhesion in immune function

Focal adhesion genes play a pivotal role in the biology of all immune cells from hematopoietic stem cells to the late effector cells. Hematopoietic stem cell homing to bone marrow from the periphery is cytokine mediated, but adhesion to the bone marrow niche is maintained by expressed focal adhesion molecules such as VLA4, VLA5, LFA1, and selectins [44]. Unlike most stromal cells, later cells in the hematopoietic lineage have few long lasting focal adhesions, especially in the periphery and circulation. Where these cells do come into contact with other cells, interactions are usually with other immune cells or immune adjacent cells (e.g. dendritic cells). These interactions usually involve effector molecules, such as the TCR, BCR, or MHC, coming into contact with the appropriate cellular receptors. These interactions are then strengthened by integrin binding and its corresponding signaling. B cells express a wide variety of integrins, and the expression repertoire at each stage is critical to its function. As B cells enter the splenic marginal zone, the integrins LFA-1 and $\alpha 4\beta 1$ integrins keep these cells isolated to this zone. These integrins are then downregulated when stimulated by lipopolysaccharide, releasing these cells [45]. Further into the germinal

center, several integrins interplay with the B cell to antigen-presenting dendritic cell, and inhibition of these integrins reduces the fitness of the GC response [46].

B cell receptor signaling is particularly influenced by focal adhesion associated integrin signaling. In resting B cells, the B cell receptor complex and associated integrins are spread across the membrane of the B cell. Activation of the B cell receptor by crosslinking to antigen leads to a downstream signaling cascade that activates associated integrins by inside-out signaling [47,48]. This activation then causes a cytoskeletal reorganization that leads to clustering of the BCR into an immunological synapse mediated by other focal adhesion associated genes, including Rho GTPases [49]. The synapse is then surrounded by a sheath of integrins involved in binding to antigen-presenting cells. The integrin LFA-1 interacts with ICAM1 during the B cell-dendritic cell interaction has a key role during affinity selection, and this interaction can prevent apoptosis in these B cells [50,51]. Inside-out integrin signaling in response to BCR activation is a complex process involving many parts of the focal adhesion machinery within the cell. This process is of particular importance considering membrane-bound antigen activation is the predominant pathway by which B cells encounter antigen *in vivo* [52]. Alterations in these mechanisms may have a wide variety of effects in B cells, depending on the alteration. Activation of these pathways by genetic alterations may lead to an unregulated growth in these cells, leading to a malignant version of these

cells, while inhibition of these pathways may have therapeutic value in targeting one of the main avenues of B cell growth and proliferation.

1.3.2 The role of *RHOA* in normal and malignant cells

The Ras homology family member A (RHOA) is a small GTPase that is part of a larger family of isoforms of Rho, related to similar Ras-like proteins RAC1 and CDC42. All eukaryotes have at least one Rho-GTPase, with over twenty discovered in mammals. [53] The Rho family is biochemically distinct from the other Ras-like family members due to a “Rho insert” domain between the fifth β -sheet and fourth α -helix domains [54]. Though these family members have remarkable sequence similarity, RhoA remains distinct from other Rho GTPases as it is associated with the endoplasmic reticulum in the cytosol [55].

RHOA is often activated by extrinsic signals through cell surface receptors. Rho-family activity is known to follow activation of G-protein coupled receptors, integrins, tyrosine kinase receptors, and, importantly for the context of lymphoid cells, cytokine signaling [56-58]. RHOA is specifically activated by GNA13 [59], and loss of GNA13 have been shown to be lymphomagenic in mice [18]. Functionally, RHOA exists in 2 states: an active GTP-bound state, and an inactive GDP-bound state. RHOA binds these two molecules with high affinity yet has been shown to rapidly cycle between these states. This cycling from GTP to GDP, or vice versa, can act as a molecular switch in cells. Two classes of regulatory molecules control the state of Rho-GTPases, GTPase-activating proteins (GAPs) and guanine exchange factors (GEFs). GAPs act to enhance the normal

GTPase activity of RHOA, yielding and inactive state, while GEFs help to catalyze the exchange of inactive GDP, yielding and active form of RHOA. The first discovered GEF, DBL, was identified as a transforming protein in DLBCL [60,61].

Once activated, RHOA signals downstream through numerous downstream effector proteins. The most well studied of these is the RHOA-ROCK pathway. GTP-bound RHOA activates ROCK (RHOA Kinase), which stimulates LIM kinase, which stimulates cofilins. [62] Cofilins then actively remodels actin cytoskeleton within the cell. This downstream remodeling of actin has been implicated in a number of processes. RHOA is directly involved in cell polarity, protrusion formation, and cell spreading via actin remodeling [63-65]. The cytoskeletal tension and induced cell shape regulated by RHOA has also been shown to regulate lineage commitment of stem cells into osteoblasts and adipocytes [66].

Cytokinesis is one of the most well characterized process regulated by RHOA. Animal cells divide by a highly regulated cytokinetic process, which begins with the assembly of an actomyosin ring. As the ring begins to contract, a cleavage furrow forms between segregated chromosomes, continuing to grow as the ring contracts. The actomyosin ring is then disassembled as cytofission completes and the cells split [67,68]. RHOA, through its regulation of actin, has been shown to be essential to cytokinesis. Blocking of RHOA block furrow ingression through both chemical inhibition [69] and RNAi [70]. Both the contraction of the furrow and the accumulation of the cortical actin

itself requires RHOA. The RHOA GEF, ECT2, has been shown to also be required for this process [71].

RhoA has been implicated as important functionally across cancer types and in many cancer-related processes, though the function of the gene seems to be context dependent. In colorectal cancer, *RHOA* acts as a tumor suppressor and is downregulated in a c-MYC/SP1 dependent manner [72]. Here, its inactivation enhances Wnt signaling and promotes growth [73]. *RHOA* as a tumor suppressor has also been shown in diffuse gastric cancers [74]. Nearly one-quarter of these diffuse gastric cancers contain *RHOA* mutations; these mutations cluster but are not typical hotspot mutations. Mutant *RHOA* (but not wild-type *RHOA*) knockdown was shown to inhibit cell growth, and reintroduction of mutant *RHOA* rescued this phenotype [75]. This process is context-dependent to this cell type, as the mechanism was shown to be driven by escape of the normal anchorage-dependent cell death [74]. Among lymphomas, *RHOA* has been shown to be mutated in Burkitt lymphoma, diffuse large B cell lymphoma, and angioimmunoblastic T cell lymphoma (AITL), but mutational context and types of mutations vary by disease [37,76-79]. In Burkitt's lymphoma, *RHOA* and *GNA13* mutations are thought to work as effectors in the same pathway, and inactivating mutations of these increase proliferation *in vivo*, suggesting a tumor suppressor role [76]. In AITL, *RHOA* mutations are dominated by a hotspot mutation at G17V in two-thirds of patients [78]. These *RHOA* mutants have been shown to act in a dominant negative fashion and

lose the GTP binding properties of the GTPase domain [77,79]. Since DLBCL also has inactivating GNA13 mutations, it is inferred that the GNA13-RHOA axis may function as in Burkitt lymphoma, but this remains untested.

1.3.3 The role of focal adhesion kinase in normal and malignant cells

Focal adhesion kinase (FAK) is a critical signaling molecule involved in a several normal cellular processes as well as tumorigenic/pro-oncogenic signaling across neoplasms. Its namesake role in focal adhesion construction, maintenance, and deconstruction is well defined, but it has also been shown to play a role in many other signaling pathways. In cancer, FAK has been implicated in metastasis and cellular torsion effects, but the scope has been limited to solid tumor biology. The role of FAK in cells with more transient cell-cell interactions, such as B cells and their arising neoplasms, is less well understood.

FAK was originally identified as a target of the Src oncogene and was found to be associated with focal adhesion sites along the cell surface [80]. FAK is activated at the cell surface by integrins when they come in contact with extracellular matrix or other growth factors [81]. FAK then signals through paxillin and directly phosphorylates α -actinin to form long stress fibers at the focal adhesion sites [82,83]. Interestingly, FAK has also been shown to regulate the RHOA-ROCK pathway through regulation of GAPs and GEFs, pushing the pathway to a more active or inactive state in a context dependent manner [84]. Downstream of these primary signaling proteins, FAK activation has been shown to activate the MAPK cascade through ERK2 [83]. All of these activated processes

are thought to be related to cell motility, cell polarization, focal adhesion deconstruction, and to a lesser extent, cell growth.

FAK's role in these common and essential processes lead to its study in a variety of malignancies. *FAK* has been associated with metastasis across solid tumors, including pancreatic, prostate, ovarian, colorectal and gastric cancers [85-89]. Due to the normal function of FAK in actin dynamics and focal adhesion disassembly, research has focused on invasion and metastasis as the main role of FAK in cancer. Cellular mobility and invasion through extracellular matrix and other cells is a complex process that is tightly coordinated. Short-lived focal adhesions are assembled and disassembled on the cell surface on the polarized migrating cell. FAK is a master regulator of not only the integrin-linked focal adhesions, but also the peripheral actin structure and cadherin-dependent adhesions [90,91]. FAK also has a structural, non-kinase role in platelet-derived growth factor and epidermal growth factor signaling, being necessary for motility in these contexts [92]. This adapter functionality also forms complexes of SRC, FAK, calpain-2 and ERK2 [93]. All members of this complex have been associated with cellular invasion and mobility, while some have also been associated with survival [94-97].

FAK has been shown to promote cell survival by a variety of mechanisms in both malignant and non-malignant contexts. Expression and overexpression of *FAK* often leads to cell survival in contexts that would normally induce programmed cell death. FAK can bind and sequester receptor-interacting protein, part of the death receptor

complex, and block apoptosis directly [98]. In normal stromal cells, cells that lose cell-cell contacts and are in suspension undergo a cell death program called anoikis. The process can be suppressed in anchorage-dependent cells by overexpression of *FAK* [99].

Aside from the strong evidence of *FAK*'s role in cancer through mechanistic study, genomic studies of *FAK* expression have shown a strong link between *FAK* copy number and expression levels with tumorigenesis. Early work in this field showed that *FAK* locus amplification was present across a wide variety of solid tumors. Locus amplification was found in colon, lung and breast cancer cell lines as well as primary tumors of squamous cell carcinoma. This increase in *FAK* copy number was also observed in models of colon cancer progression, as the copy number of the *FAK* locus increased as the cells converted from adenoma to carcinoma [100]. This phenotype was also observed in skin carcinogenesis, with *Fak*^{+/-} mice showing impaired papilloma formation compared to wild-type controls, and *FAK* transcript levels increasing during conversion to more malignant cell stages [101].

1.4 Cancer Genomics and Bioinformatics

1.4.1 Exome sequencing for discovery of somatic mutations

Once it became clear that mutations in the genome are the major driver in cancer formation, sequencing the genes of cancer patients became the obvious next step in understanding the disease. Early work into previously known cancer-associated genes, such as *TP53*, led to the sequencing of these genes in patients and translating them into

functional models and patient outcomes [102,103]. Though these studies gave some insight into the biology of these cancers, they only explained a fraction of the changes within the cancer cells. Sequencing of the entire genome would be optimal for identification of variants, but this would be far too costly to perform at scale. As technology advanced in massively parallel next generation sequencing, sequencing of small percentages of the genome became feasible. This led to the routine sequencing of the protein-coding region of the genome, the exome [104]. Exome sequencing allows for the simultaneous sequencing of approximately 20,000 genes, allowing for the discovery of mutations in an unbiased manner, while enriching for the most functional regions of the genome.

As shown in figure 2, our lab uses this exome-based strategy to identify recurrently mutated genes in lymphoma. Samples include primary lymphoma tissue taken from biopsies of patients with DLBCL and matched normal samples. This pairing allows for identification of somatic mutations, those mutations that are only found in the tumor tissue, as those variants seen in the normal tissue is not likely to be causative for cancer. For our purposes, normal tissue is peripheral blood, normal bone marrow, or buccal swabs.

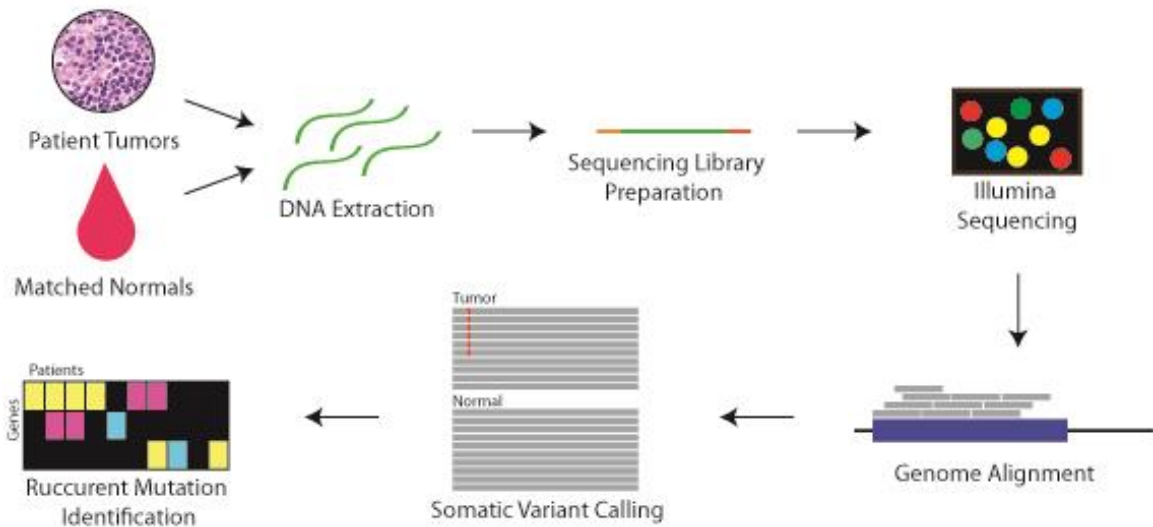


Figure 2: Overview of DNA sequencing. Patient tumors and matched normal represent primary inputs. All downstream steps are performed in parallel for all input samples.

DNA is extracted from both the tumor and normal tissue by deparaffinization of formalin fixed paraffin embedded (FFPE) tissues and extraction by column-based purification. This DNA is fragmented to approximately 250 basepairs and sequencing adapters are ligated to create whole genome DNA libraries. The library is then enriched for protein coding regions using baits that are complementary to these regions using a hybridization based capture [105]. These libraries are then sent for sequencing on the Illumina platform to a depth of about 50x.

Sequencing of this number of samples creates an extremely large amount of data that must be processed in a unified manner to detect variants. This process was automated in our lab in order effectively manage hundreds of samples simultaneously. The data is received in the form of fastq files that are preprocessed with GATK [106].

Reads are then aligned with Burrows-Wheeler Aligner [107], with coverage stats calculated with BEDTools [108]. Indels were called with GATK. Joint variant calling was performed using MuTect [109] and Annovar [110]. Further analyses were performed as needed for the goals of the project and will be described in each section.

1.4.2 RNA sequencing for expression profiling

Gene expression profiling is a powerful tool in genomics to understand the state of cells as they exist. All cells contain all genes necessary for any type of cell within the organism, but gene expression changes are private to each cell type. Gene expression profiling proves to a powerful tool in understanding not only the dosage of each gene expressed within the cell, but also the state of the cell temporally. The most straightforward way of assessing the expressed repertoire would be protein profiling. The technology for protein biology remains far behind DNA sequencing in both throughput and scope. To approximate protein expression itself, relative RNA quantification is used in its stead. Early work in this field involved quantification of individual genes using a qPCR based method [111]. While this method is limited to the scope of known functional genes, complex gene expression profiles can be produced containing entire networks of genes [112]. The next technological innovation in this space was gene expression microarrays [113], allowing for genome wide hybridization-based quantification of transcripts. This technology allowed for genome-wide exploration of expression data in a high-throughput manner. This was applied to cancer across disease

types, revealing different subtypes and predictions for outcomes [114,115]. DLBCL specifically was major target of expression profiling, with this profiling revealing molecular subtypes with disparate outcomes clinically [10,116].

The most powerful tool currently available for gene expression quantification remains RNA-sequencing (RNA-seq). This method uses reverse transcription to make cDNA that is sequenced in a similar method to massively parallel DNA sequencing [117]. These methods have transformed how we understand the underlying biology of everything from development to cancer biology. Additionally, RNA-seq removes the technical restrictions of microarrays and allows for the identification of many types of RNA specific events. Using a count based, instead of probe based, approach allows for greater dynamic range in expression quantification. Because RNA-seq does not use predefined probes, novel transcripts, including non-coding RNA species, can be identified. Alternative splicing events can also be called in RNA-seq data based on reads spanning splice junctions.

In our lab, we have utilized RNA-seq technology in a number of ways, but the process remains largely the same (figure 3). First RNA is extracted from the appropriate tissue, usually tumor sample or cell line RNA. Next ribosomal RNA (rRNA) is removed by either rRNA depletion in FFPE samples or mRNA selection in cell lines. Next, cDNA is synthesized by reverse transcription and fragmented to around 250 basepairs. Barcoded sequencing adapters are added in order to pool samples for sequencing. The

samples are then sent for deep sequencing on the Illumina platform. Sequencing reads are aligned to the transcriptome using TopHat [118], then Cufflinks is used for transcript assembly and quantification [119]. Gene expression normalization and gene signature analysis was performed as needed for each study and are described in that section.

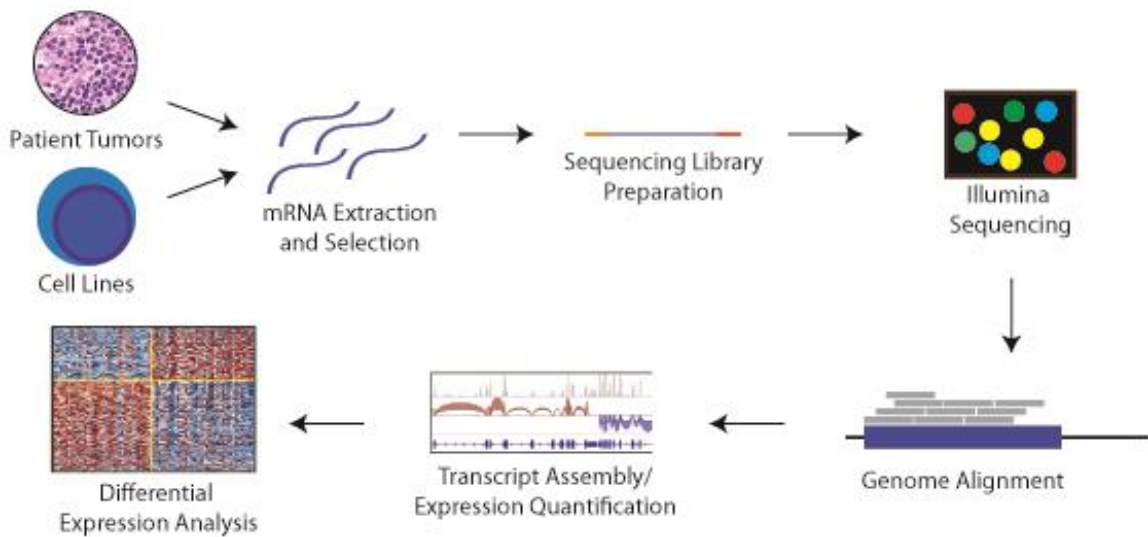


Figure 3: Overview of RNA-sequencing. Patient tumors and cell lines are the primary samples used. All downstream processing is performed in parallel for all samples.

1.5 Overview of Results

In this dissertation, I present a tripartite approach to understanding diffuse large B cell lymphoma (DLBCL) by defining the genetic and functional drivers, elucidating the role of two focal adhesion related genes in DLBCL, and developing a method for predicting synergistic response to drug perturbation based on expression analysis.

In chapter 2, I present the largest exome and transcriptome study to date in DLBCL. With over 1000 tumor samples paired with 400 normals, this study is uniquely

powered to detect genomic alterations. We identified 150 putative driver genes that were recurrently mutated in DLBCL. These genes included putative oncogenes, such as *BCL2* and *CARD11*, as well as tumor suppressors, such as *SPEN* and *CDKN2A*. We then identified subtypes of DLBCL (either ABC or GCB) based on gene expression data. We found that 20 genes were differentially mutated by subtype including *EZH2*, *SGK1*, *GNA13*, *SOCS1*, *STAT6*, and *TNFRSF14*, which were more frequently mutated in GCB DLBCLs, while *ETV6*, *MYD88*, *PIM1*, and *TBL1XR1* were more frequently mutated in ABC DLBCLs. We then performed a CRISPR screen on 6 DLBCL cell lines to identify functional genes across the genome. We identified oncogenes (including *MYC*, *RHOA*, *MTOR* and *BCL2*) as well as tumor suppressors (including *TP53*, *MGA*, and *PTEN*) genome wide. 35 driver genes identified by genetic alteration were also found in this screen, with 9 of these being in a cell type specific manner. Transcriptional profiling also found that upregulated genes were enriched in CRISPR-based oncogenes. We then developed a model based on expression, mutational profile, and clinical characteristics that outperforms traditional patient stratifications to predict patient outcomes.

The focal adhesion pathway is enriched in both mutations and expression changes, and in chapter 3 and 4 I look at the role of 2 members of this pathway (*RHOA* and *FAK* respectively) in normal and malignant B cell biology. We found that knockout of *RHOA* in DLBCL cell lines causes a major loss in fitness (CRISPR screen in chapter 2), and that knockdown of *RHOA* slows down cell cycle progression as cells begin to

accumulate in the G2/M phase. Mouse models of *Rhoa* loss in the B cell lineage confirmed this result, with major loss in B cell numbers within all secondary lymph organs. Specific knockout of *Rhoa* in the germinal center B cell population shows a similar reduction in B cells, dark zone/light zone dysregulation, as well as a decrease in filamentous actin within these cells. The role of *FAK* (chapter 4) is less clear in DLBCL, as it is not recurrently mutated, and B cells do not maintain focal adhesions like most solid tissues. Chemical inhibition of FAK shows that FAK is necessary for growth of B cells that maintain B cell receptor (BCR) signaling. Further work shows that downstream targets of BCR signaling are blocked by both chemical inhibition and genetic knockdown of FAK. Levels of phospho-FAK also rise upon activation of the BCR. Activation of SYK, a BCR linked protein, was shown to drop upon FAK inhibition. This data together suggests that FAK is acting in the B cell receptor pathway, at the interface of the BCR and membrane, and that the FAK-LYN interaction is the possible mechanism for this signaling.

In chapter 5, we sought to understand the role of single agent therapeutics on gene expression to identify a way to use this data to inform combination therapies. Using a panel of 152 FDA-approved drugs and 6 DLBCL cell lines, we screened all lines for drug efficacy. This revealed 3 classes of drugs: pan-effective, selective and resistant. The selective drugs displayed pathway specific resistance (e.g. mTOR inhibitors) as well as a subtype specific sensitivity for certain drugs. Gene expression was then performed

for all drug-cell line pairs. Overall gene expression patterns show that drugs targeting similar targets or targets in the same pathway induced tightly correlated gene expression patterns. We also found that changes in expression of genes like *MYC* were correlated with sensitivity, giving possible proxies for sensitivity. We also found that baseline expression of the target gene was correlated with a high dose requirement to achieve viability changes. Gene expression changes of the target gene were also seen to be indicative of either sensitivity or resistance in some cases. Using expression of both normal and malignant cells, we developed a model for using gene expression to predict dual drug synergy that we are calling combination reversal of disease gene expression (cRDGE). This model accurately predicted both single agent effectiveness as well as combination synergy in previous datasets. We then validated this method across several drug combinations and found synergy between the tested combinations *in vitro*. We then tested one combination, panobinostat and ruxolitinib and that they were not only synergistic but also sensitized the reduction of STAT signaling in these cell lines. Using xenograft models, we then tested this combination *in vivo*, finding synergy of the drug combination and low hematopoietic toxicity.

2. Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma

2.1 Introduction

Diffuse large B cell lymphoma (DLBCL) is the most common hematologic malignancy, with an annual incidence of over 100,000 cases worldwide. Although more than half of these patients may achieve long-term remission, the majority of the remaining patients succumb to DLBCL. Application of next-generation sequencing has revealed a striking degree of molecular and clinical heterogeneity in DLBCL [120].

The heterogeneity of DLBCL has imposed several major limitations on previous genetic studies [15,16,120,121]. First, sample-size limitations of earlier studies have hampered the discovery of low-frequency but pathogenetically important mutations. Systematic pan-cancer analysis has indicated that 400 tumor-normal DLBCL pairs would be needed to comprehensively discover all the genes mutated at a frequency of 5% or higher with greater than 90% power [6]. Second, defining the link between genetic mutations and clinical outcome has been challenging. Third, even as the genetic alterations are being defined, the functional impact of these mutations and their role in promoting the growth of lymphoma cells remains unclear, even though these are critical considerations for exploring therapeutics.

2.2 Methods

2.2.1 DLBCL patients and cell line samples

Diffuse Large B-Cell Lymphoma (DLBCL) tumors (N=1001) and paired-normal tissue (N=400), along with patient clinical information, were obtained. All cases were required to be *de novo* and treated with a rituximab-containing standard regimen. Clinical data including initial response to therapy, overall survival, gender, age, stage, performance status and number of extranodal sites were collected on nearly all cases.

OCI-LY3 were grown in RPMI 1640 with 20% HI-Fetal bovine serum and 1% Penicillin-Streptomycin (5,000 U/mL). Pfeiffer, SU-DHL-4, BJAB, TMD8, and HBL-1 were grown in RPMI 1640 with 10% HI-Fetal bovine serum and 1% Penicillin-Streptomycin (5,000 U/mL). All cell lines were grown at 37 C. OCI-LY3, TMD8, SU-DHL-4, Pfeiffer, and HBL-1 are derived from males, and BJAB is derived from a female. Cell lines were authenticated through STR sequencing (Duke University Core Facility).

2.2.2 Sample processing for exome sequencing

Genomic DNA was sheared to 250 bp and used in a custom, in-house exome library preparation process based on the Agilent XT2 pooled exome capture method. To increase the throughput of exome library preparation, a custom 96-well barcode system was designed based on the Illumina sequencing system whereby a separate sequencing read is used to identify the barcode of a particular library. Custom barcodes of length 8, compatible with the 3-read Illumina hiseq v2 platform.

Pre-capture libraries were prepared with standard library preparation protocols using the KAPA Hyper kit (Kapa Biosystems), and then pooled at equal volume and sequenced on the Illumina platform at low depth to determine exact relative abundances. Based on these abundances, libraries were balanced optimally for exome using the Agilent All Exon V5 human bait set. Library sequencing was performed to an average of 75X coverage on Illumina Hiseq 2500 platform.

Reads in fastq format were pre-processed with GATK version 3.2 to remove Illumina adapter sequences (analysis type `-T ClipReads`, `-XF illumina.adapters.fa`) and Phred-scaled base qualities of 10 and below (`-QT 10`). After GATK processing, reads were mapped to hg19 using Burrows-Wheeler Aligner (BWA) version 0.7.7 with the mem algorithm. Reads were sorted with Novoalign V2.08.03 `novosort`. SAMtools `flagstat` was used to compute the number and percent of reads that mapped to the genome. Both depth and breadth of coverage for each exome were computed using BEDTools. PCR/optical duplicates were marked by Picard. Base quality recalibration and indel realignment was performed using GATK.

Joint variant calling was performed on all 1001 DLBCL samples together using Samtools `mpileup` v. 0.1.19 and somatic variant calling was performed using MuTect version 1.1.4 for the tumor/normal pairs. This resulted in a total of ~1.1 million variants. The variant functions and population statistics were annotated using Annovar.

2.2.3 CRISPR screen

To generate GeCKO.v2 library virus, 293T cells were seeded in 15 cm dishes at 12 million cells per dish. The following day, cells were transfected using TransIT-LT1 according to manufacturer's instructions with 10 ug lentivector, 10 ug psPAX2 and 1 ug pCMV-VS. 16 hours post transfection, media was changed. Virus was harvested at 48 and 72 hours post transfection, pooled, filtered, aliquoted and stored at -80 until needed for screening.

DLBCL cells stably expressing Cas9 were generated by transduction of cells with lentiCas9-blast virus at multiplicity of infection (MOI) < 1. Forty-eight hours post transduction, cells were selected with 10 ug/ml blasticidin for four or five days and expanded. Cas9-expressing cells were transduced with the human GeCKO. GeCKO library were transduced into cells and screened separately in triplicate. 24-48 hours post transduction, cells were selected with puromycin for 72 hours. Some cells were harvested after puromycin selection to serve as the initial time point sample as well as after 14 population doublings. 1000-fold representation of the library was maintained at each passage.

Genomic DNA was harvested from each time point sample. The sgRNA cassettes from each sample were amplified and barcoded for multiplexing with two rounds of PCR with Takara Ex Taq DNA polymerase as previously described. PCR products were gel purified and isopropanol precipitated for sequencing on the HiSeq2500.

Guide RNA (sgRNA) amplicon sequencing reads for each library were downloaded from as de-multiplexed FASTQ files. Raw reads were then trimmed using CutAdapt v1.8.3 to remove 5' barcoded adapter sequences specific to each library and 3' PCR primer sequences. Successfully trimmed reads were aligned to the GeCKO v2 sgRNA library using MAGeCK-VISPR v0.5.3 to produce a raw count table of sgRNAs across cell line replicates. The resulting count table was then normalized to account for sampling differences using the variance-stabilizing transformation implemented in the MAGeCK count module.

CRISPR gene scores were used to quantify the magnitude and directionality of fitness effects from normalized sgRNA counts as in Aguirre et al., 2016. Briefly, under a scenario in which there are m total genes being screened in cell line j , the CRISPR gene score for gene i targeted by n sgRNAs is given by the following equation:

$$crisprgenescore_{ij} = \frac{\sum_{k=1}^n sgRNAscore_{ijk}}{n}$$

Where $sgRNAscore_{ijk}$ is given as the z-score transformation of the log-fold change between early and late samples in cell line j for a sgRNA k targeting gene i :

$$sgRNAscore_{ijk} = \log_2 \left(\frac{Latecount_{ijk}}{earlycount_{ijk}} \right)$$

$$sgRNAscore_{ijk} = \log_2 \left(\frac{Latecount_{ijk}}{earlycount_{ijk}} \right)$$

2.2.4 RNA sequencing

We removed rRNA from total RNA by hybridization to homologous rRNA DNA oligonucleotides. The rRNA and DNA oligos that formed double stranded DNA/RNA complexes were then digested with RNase H for 30 minutes at 45 C. The remaining DNA oligos were then digested with DNase I for 30 minutes at 37 C.

RNA libraries were prepared using the Stranded mRNA-seq kit (Kapa Biosystems) per manufacturers guidelines. Briefly, cDNA was generated using random hexamer primers followed by A-tailing. The P5 and P7 universal sequencing adaptors containing specific 8mer index sequences were ligated to the A-tailed cDNA. For this, we generated adaptors containing the P5 and P7 sequencing primers with 96 specific index sequences. The resulting libraries were PCR amplified and pools of 96 were quantified using the Bioanalyzer

To improve the sensitivity and specificity of RNAseq in FFPE cases, RNAseq libraries were hybridized to Human All Exon V6 + 3' UTR capture baits (Agilent, Cat. #5190-9306). Libraries of high quality were subjected to high throughput sequencing using the Illumina HiSeq 2500 platform per manufacturer's instructions using HiSeq V4 125 PE, to generate an average of 10 million reads per sample.

The sequencing reads were aligned using TopHat. The Cufflinks software package was then used for reference-based transcript assembly and quantification of gene and transcript level expression. Using the Cufflinks package, we quantified the

number of reads mapping to each exonic locus as exon fragments per kilobase of exon per million fragments mapped (FPKM).

Gene expression measurements were FPKM normalized. To enable broad gene expression signature discovery and validation, we omitted 150 samples with expression of fewer than 12,000 genes owing potentially to either technical or biological artefacts. The remaining 625 cases were designated as the core set for the integrative analysis. Additionally, quantile normalization was performed to remove the effects due to technical variation, and the data was log₂ normalized.

2.2.5 Gene expression signatures

We define gene signatures using a collection of widely used annotated gene set databases (Kegg, Reactome, Biocarta, Gene Ontology, MSigDB), as well as several lymphoma-specific gene signature sources (Lenz et al. 2008, Monti et al. 2005, and SigDB from Staudt et al.).

We clustered the significantly correlated gene sets using affinity propagation clustering (Frey and Dueck, 2007). Affinity propagation is a clustering algorithm that simultaneously identifies clusters and exemplars (cluster representatives) in the data without the user specifying the number of clusters (k) to be identified. Using affinity propagation on the 1228 significantly correlated gene sets we obtained 31 clusters.

Exemplar names were manually curated to best represent the gene signatures that were part of the cluster.

2.2.6 Genomic risk model

We developed a supervised learning approach for defining the association of survival with combinations of the following genetic and molecular features: (a) known expression subgroups (ABC, GCB subgroups, *MYC*, and *BCL2* high expression), (b) the mutations/copy number events in the 150 DLBCL-associated genes identified in this study. We use a comprehensive approach by first defining combinatorial features derived from genetic and molecular features.

A schematic of our predictive modeling workflow is shown in figure 12. The first step of our method involves enumerating all possible combinations that involved up to 4 separate genetic and molecular features (degree \leq 3) and affected a minimum of 20 patients (coverage \geq 20). This approach identified 313 combinatorial features, which includes 95 singletons, 166 pairs, 52 triplets. These combinatorial features formed the input for the survival predictive model.

We used elastic net regression with a Cox proportional hazards model. Elastic net is a regularized regression framework that is a linear combination of the L1 (Lasso regression) and L2 (Ridge regression) penalties. Elastic net combines the advantages of these two methods, and allows for the flexibility of choosing the combination in a dataset-specific manner by tuning parameters using cross-validation within the training data. The performance of the model was rigorously evaluated using 5-fold cross-validation which was repeated 100 times.

2.3 Results

2.3.1 Discovery of genetic drivers in DLBCL

We started with a discovery phase of identifying genes mutated in DLBCL through whole-exome sequencing of 1,001 DLBCL and 400 paired germline DNAs. The clinical characteristics of these patients are described fully in Table S1 of the publication based on this work [122]. The sample size of 400 tumor-normal pairs provided nearly 100% power for the discovery of genes mutated in 10% of patients and over 90% power to detect genes mutated in 5% of DLBCL patients [6]. To increase the sensitivity and robustness of our discovery of driver-gene mutations, we added 102 previously described tumor-germline pairs [15,16,120,121] to our set of 400 paired cases. These 502 paired tumor-germline cases represented a primary discovery set for identifying genetic drivers in DLBCL, with the remaining cases serving as a validation set to ensure the reproducibility of the findings and to characterize the frequency of genetic events and clinical associations in these driver genes. In addition to single nucleotide variants (SNVs) and insertions/deletions (indels), we also characterized copy number alterations from whole-exome sequencing data [123]. We verified the accuracy of our variant identification through Sanger sequencing of 1,130 events, representing 61 genes with 90% concordance.

The distribution of somatic alterations in tumor-normal pairs were used to identify [124] 150 putative driver genes that were recurrently mutated in DLBCL. We then

defined genetic alterations for all 1,001 DLBCLs corresponding to these 150 genes, identifying variants that meet the criteria for driver mutations as described [124,125]. The mean number of mutations per DLBCL case was 7.75. The individual variants identified by patient are, with copy number and gene-level alterations summarized in Table S1 of the publication based on this work [122].

The top 60 most frequently mutated genes are shown in the mutational heatmap in figure 4A. Genes frequently exhibited a pattern of either predominant missense and/or copy number gains consistent with an oncogene (e.g., *BCL2*, *CARD11*, and *IRF4*) or truncating mutations and/or copy number losses consistent with a tumor suppressor gene (e.g., *SPEN*, *CDKN2A*, and *TNFAIP3*).

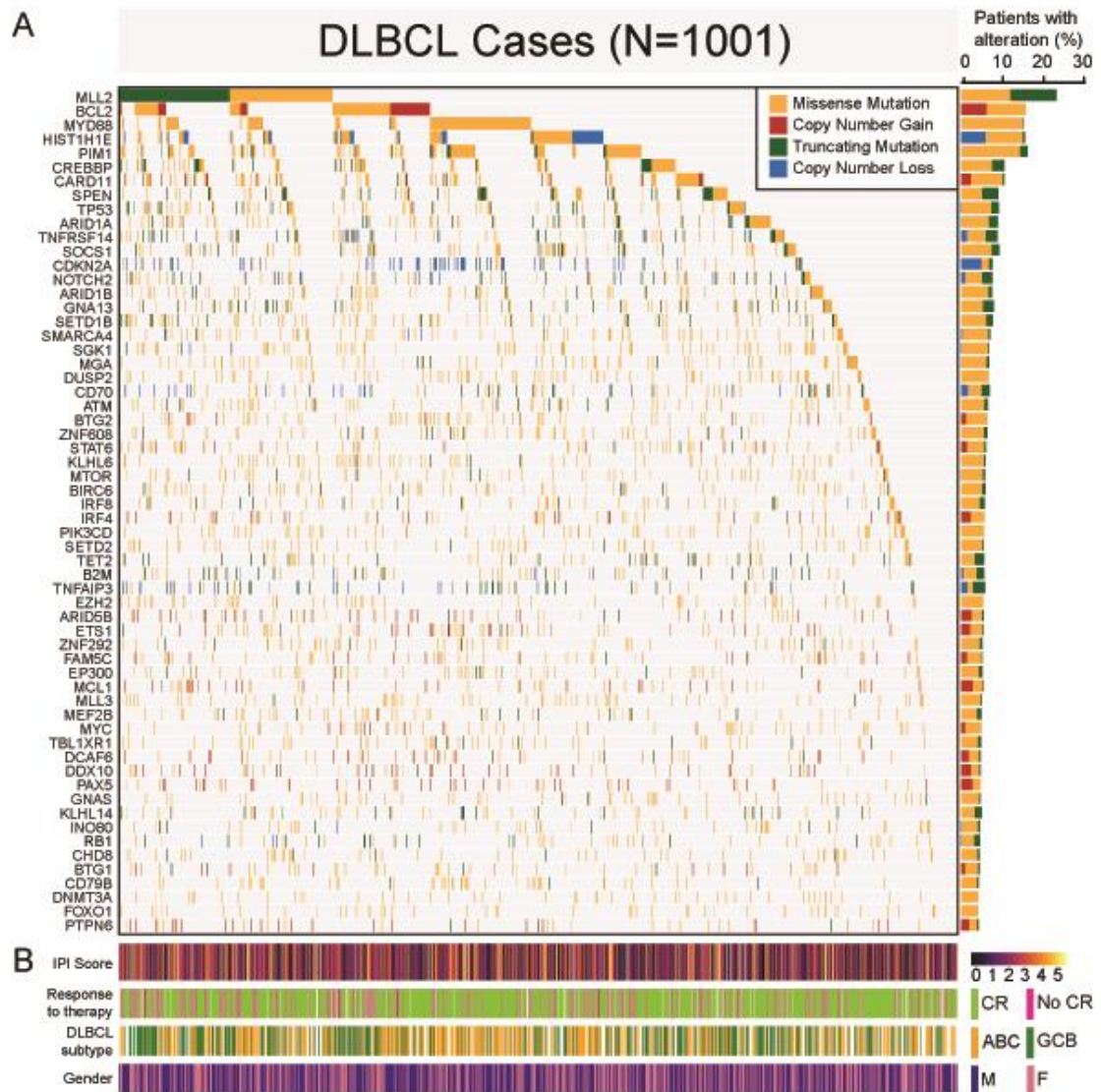


Figure 4: The Landscape of Genetic Drivers in 1,001 DLBCLs. (A) The mutational heatmap indicates the most recurrently altered genes in 1,001 DLBCL cases with frequency >5%, color coded by four genetic alteration types: missense mutation (yellow), copy number gain (red), truncating mutation (green), and copy number loss (blue). To the right of the mutational heatmap, the stacked bargraph indicates the gene-level alteration type breakdown using the same four-color scheme. (B) Clinical features of the corresponding 1,001 patients are indicated below the mutational heatmap, including the international prognostic index (IPI) score, response to therapy (complete response or not), activated B cell-like (ABC) versus germinal center B cell-like (GCB) DLBCL subtype, and gender.

Figure 4B summarizes the international prognostic index (IPI) [126] score, initial response to therapy (complete response or not), and the cell-of-origin subsets of DLBCL, activated B cell-like (ABC), and germinal center B cell-like (GCB) DLBCL. When the cases were ordered in descending order by gene-mutation frequency, there were no obvious patterns detectable in the corresponding clinical data. The large sample size in our study enabled not only the identification of driver genes in DLBCLs but also provided clarity on the types of mutations in known driver genes.

2.3.2 Cell-of-origin effects in DLBCL

We defined the lineage-based subgroups of DLBCL using RNA-seq (RNA sequencing) to generate expression data in 775 patients for whom adequate material was available.

Using gene expression to distinguish cell-of-origin [13], we identified 313 ABC DLBCLs and 331 GCB DLBCLs, while the rest were unclassified DLBCLs (figure 5).

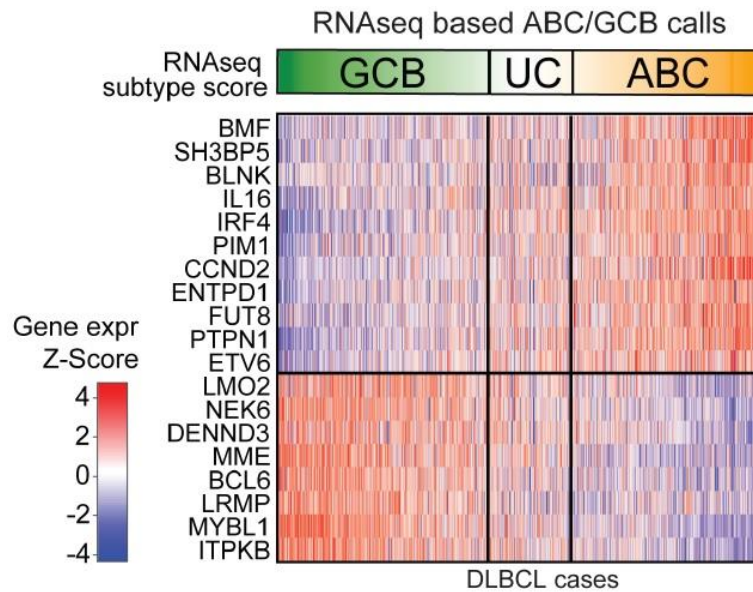


Figure 5: ABC/GCB-Based Classification of DLBCL. RNA-seq gene expression classifier distinguishes germinal center B cell-like DLBCL (GCB), unclassified DLBCL (UC), and activated B cell-like DLBCL (ABC). Color indicates Z-score

We validated the efficacy of the RNA-seq-based classifier through two complementary approaches. First, we performed Nanostring assays [127] in 200 of these cases to determine cell-of-origin. We found excellent agreement between the expression data and the cell-of-origin calls generated by RNA-seq and Nanostring (Figure 6A, $R^2 = 0.87$, $p < 10^{-6}$). We also compared the RNA-seq classifier with the immunohistochemistry-based Hans algorithm [128] commonly used to distinguish cell-of-origin. In 654 cases (Figure 6B, $p < 10^{-6}$) that we tested using immunohistochemistry for IRF4, BCL6, and CD10, we found that the RNA-seq classifier score was significantly different between the two Hans algorithm classes. As expected, ABC DLBCLs had worse

overall survival compared to GCB DLBCLs in our study ($p = 0.004$, logrank test). These data indicate that RNA-seq can be reliably used to distinguish the DLBCL cell-of-origin subsets.

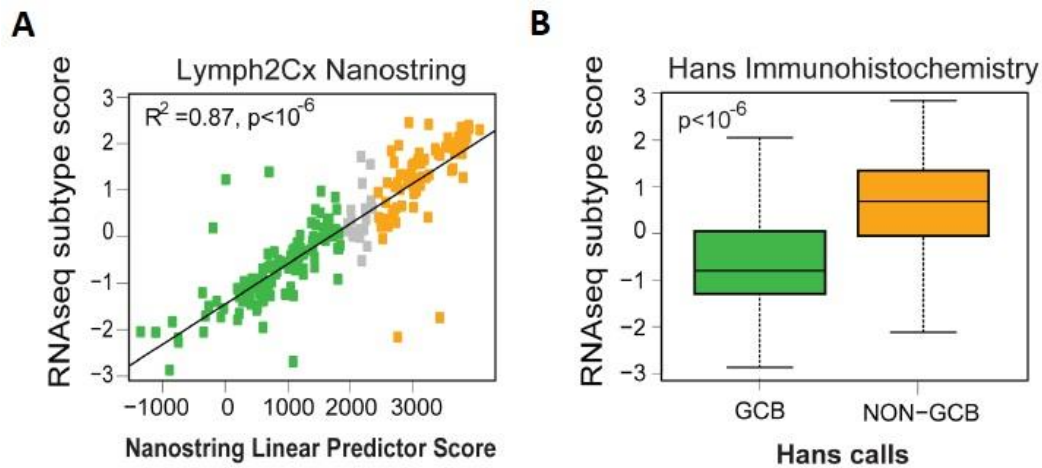


Figure 6: Comparison of RNA-seq subtype score. (A) Our scores versus the Nanostring linear predictor score (using Pearson’s correlation $R^2 = 0.87, p < 10^{-6}$) and by **(B)** immunohistochemistry Hans GCB versus non-GCB classification (Wilcoxon test $p < 10^{-6}$).

2.3.3 The connectivity of DLBCL driver genes

We further examined the differences in mutational frequencies among the ABC and GCB DLBCL subgroups. We found that these two subgroups shared the vast majority of driver genes at statistically indistinguishable frequencies. However, 20 genes were differentially mutated among the two groups (Figure 7) including *EZH2*, *SGK1*, *GNA13*, *SOCS1*, *STAT6*, and *TNFRSF14*, which were more frequently mutated in GCB DLBCLs, while *ETV6*, *MYD88*, *PIM1*, and *TBL1XR1* were more frequently mutated in ABC DLBCLs.

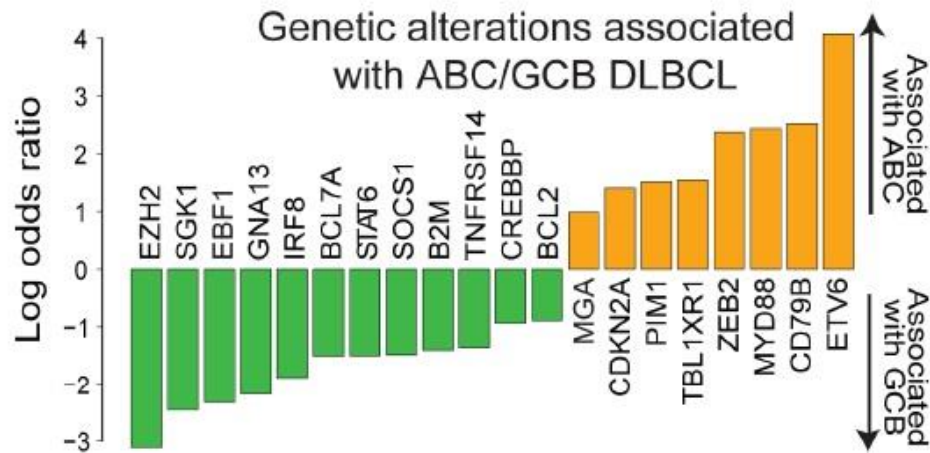


Figure 7: Genetic alterations that are enriched in ABC versus GCB DLBCL. Green indicates GCB association. Orange indicated ABC association. (Fisher’s test FDR < 0.1)

We examined pairwise overlaps using Fisher’s test and mutual exclusion [129] among the 150 DLBCL driver genes. 61 genes had statistically significant relationships with other driver genes ($p < 0.01$), depicted in the network in figure 8. We noted a number of interesting features, including subnetworks associated with cell-of-origin in DLBCL that generally exhibited mutually exclusive mutation patterns, as well as a larger network of genes shared between the subgroups. For instance, *MLL2* mutations occurred in a largely exclusive fashion with *MYC*, while *TP53* mutations occurred in a mutually exclusive fashion with *KLHL6*, implicating biologically relevant models for DLBCL.

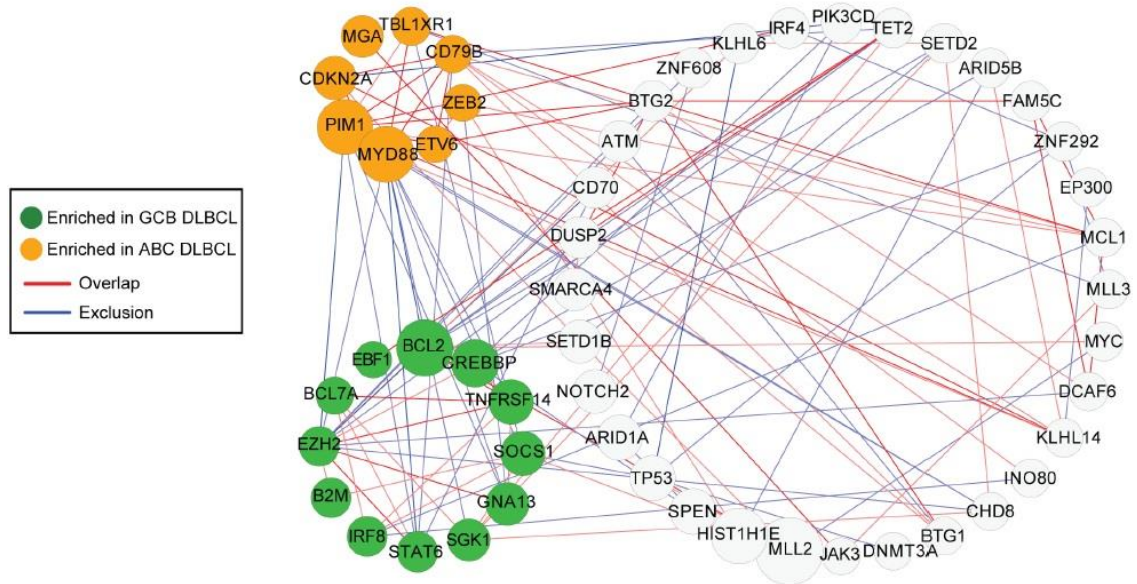


Figure 8: Connectivity network of genetic alterations in DLBCL. Network of genes with statistically significant ($p < 0.01$) overlap or co-occurrence (red edges) and mutational exclusion (blue edges). Stronger associations are indicated with thicker and darker colored lines between nodes ($p < 0.001$). Genes significantly enriched in GCB or ABC DLBCL are colored in green and orange, respectively, or gray otherwise.

2.3.4 Functional genomics through CRISPR screening

CRISPR screening has emerged as a powerful method for identifying critical functional dependencies *in vitro* [130,131]. We selected a total of six cell lines including three ABC DLBCLs (LY3, TMD8, and HBL1), two GCB DLBCLs (SUDHL4 and Pfeiffer), and one Burkitt lymphoma cell line (BJAB) that phenotypically resembles GCB DLBCLs.

We utilized the GeCKO v2 genome-wide human single guide (sgRNA) library [131], containing over 120,000 sgRNAs targeting 19,050 protein coding genes, to transduce three replicate populations from each cell line with the complete sgRNA library. The

schematic for our CRISPR screen is shown in Figure 9A. Each gene was targeted by six distinct sgRNAs in Cas9-expressing DLBCL cells with cell growth allowed to proceed for 14 population doublings. At the end of these experiments, we expected that cells expressing sgRNAs that target oncogenes promoting cell growth and proliferation would be selectively depleted from the cell population, whereas cells expressing sgRNAs that target tumor-suppressor genes inhibiting cell growth and proliferation would be selectively enriched in the cell population.

We sequenced replicate populations at two time points (Figure 9B; Early: after transduction, Late: after 14 population doublings) to observe changes in sgRNA frequency over time. The resulting changes were then used to infer the magnitude and directionality of fitness effects associated with targeted gene knockout in DLBCL cell lines, with the ultimate goal of broadly identifying genes essential to the development and/or maintenance of DLBCL. High-throughput sequencing of sgRNA libraries amplified from plasmid sequences was used to determine sgRNA abundance for populations at each time-point. Summary statistics of sgRNA sequencing library quality are detailed in and Figure 9B.

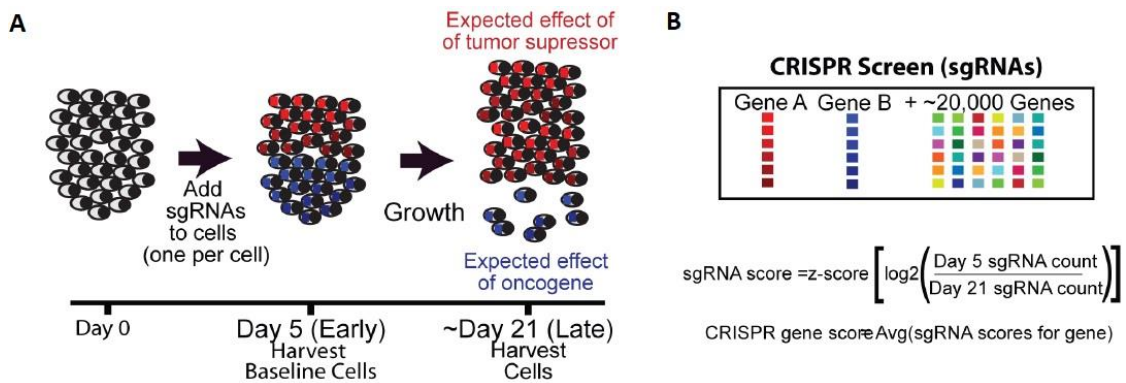


Figure 9: Schematic of CRISPR screen performed for six cell lines in triplicate. (A) Example screen schematic with 1 oncogene (blue) and one tumor suppressor (red). **(B)** Schematic of our screen with scoring sgRNA calculation.

We identified 1,956 “essential genes” whose silencing resulted in significantly decreased cell fitness in at least one cell line. We ranked all the genes based on their CRISPR score, a measure of the degree of alteration in each gene as a function of altered abundance of the guide RNAs targeting that gene (Figure 10). We found that the genetic drivers identified through whole-exome sequencing were disproportionately at the extremes of the distribution ($p = 3 \times 10^{-5}$). The genes that were depleted in the screen and therefore implicated as oncogenes critical to cell growth across the majority of DLBCLs included *MYC*, *RHOA*, *SF3B1*, *MTOR*, and *BCL2*. Among genes that were enriched in the screen and therefore implicated functionally as tumor-suppressor genes, we noted *TP53*, *MGA*, *PTEN*, and *NCOR1*. Separately, we examined the association of genomic copy number and the guide RNAs that were significantly altered in our screen (Figure 10) to rule out copy number alterations as a source of artifactual results [132].

None of the essential driver genes were found within genomic regions identified by our analysis as potentially influenced by copy number amplification. Thus, we concluded that the decreased fitness resulting from essential driver gene knockout reflects the functional importance of these genes in the context of DLBCL.

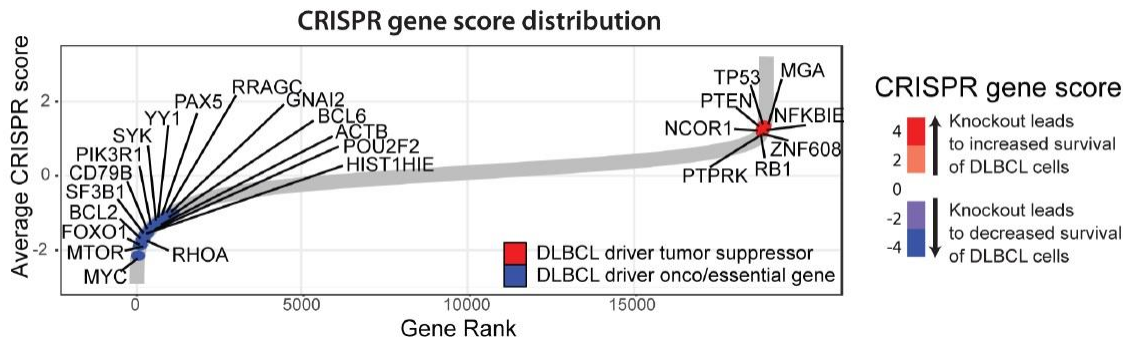


Figure 10: Distribution of CRISPR gene scores genome wide. A ranked list of CRISPR scores for the 19,032 genes targeted in the screen. Illustrative driver genes are shown in blue (likely oncogenes) or red (likely tumor suppressor genes).

In all, there were 35 driver genes whose knockout resulted in decreased viability of DLBCL cells, identifying them as functional oncogenes (Figure 11). Nine of these driver genes were found to be important in a subtype-specific fashion. Knockout of *EBF1*, *IRF4*, *CARD11*, *MYD88*, and *IKBKB* was selectively lethal in ABC DLBCL, whereas knockout of *ZBTB7A*, *XPO1*, *TGFBR2*, and *PTPN6* was selectively lethal in GCB DLBCL. These data provide critical clues to the functional dependence of DLBCL on particular oncogenes and suggest that small molecule inhibitors of these oncogenes may provide therapeutic benefit in affected DLBCL patients. All of these genes were well expressed in DLBCLs. Of the 35 CRISPR driver gene hits, 9 genes are direct targets of the therapeutic

drug targets either in human clinical trials or already in use for another indication.

Importantly, 36% of the DLBCL patients have genetic alterations in these nine drug targets and would potentially benefit from targeted therapy.

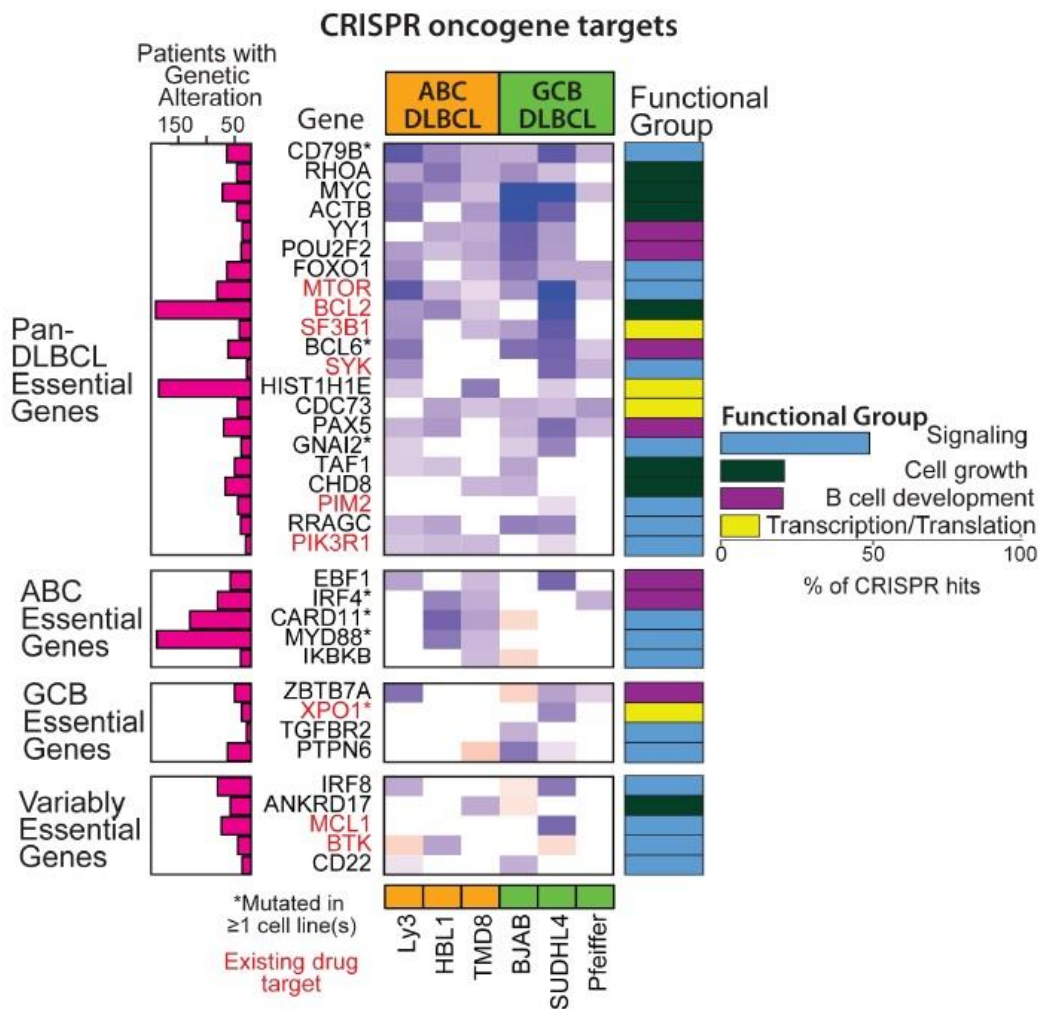


Figure 11: CRISPR scores for 35 DLBCL oncogenes with functional classification. CRISPR gene scores are shown in the middle heatmap. Gene names in red have existing drug targets. Cell lines listed on the bottom are grouped by ABC (orange) and GCB (green) subtype. Genes are separated by class of essential genes (left). Number of patients with genetic alterations are shown in the bar graphs (left, n=1,001). Gene functional classification is shown by color scale (right).

Gene set enrichment analysis of the significant driver genes (online supplement) identified four functional groups, including signaling (e.g., *MTOR*, *PIK3R1*, *PIM2*, *BTK*), cell growth (e.g., *MYC*, *CHD8*, *BCL2*), B cell development (e.g., *EBF1*, *IRF4*, *PAX5*, *POU2F2*, *YY1*), and transcription/translation (e.g., *SF3B1*, *XPO1*, *HIST1H1E*). Importantly, we observed that the lethal effects of knocking down these genes were not limited to cell lines that harbored alterations in these genes, suggesting that these driver genes represent favored pathways for promoting proliferation in DLBCL.

2.3.5 Gene expression signature associations with genetic alterations and outcome

Gene expression profiling has long been used to define subtypes and understand the heterogeneity within DLBCL. Microarray-based gene expression studies have identified cell-of-origin subgroup [11], as well as gene expression signatures derived from host inflammatory response [133] and stromal tissues [134], as well as other biological processes. However, the connections between these gene expression signatures and genetic alterations is largely unknown.

We first sought to comprehensively define the relevant gene expression signatures across our DLBCL cohort by performing RNA-seq on all tumors for which RNA was available (N = 775). From these cases, we defined a core set of 624 samples for integrative analysis.

Figure 12 provides an overview of the integrative expression analysis. We began the analysis with a comprehensive universe of nearly 9,500 annotated gene sets from widely used gene set databases (Kegg, Reactome, MSigDB), as well as several lymphoma-specific gene signature sources. We first identified annotated gene sets that showed a pattern of high correlation among the constituent genes in our dataset, indicating that these gene sets represented processes that are co-regulated, as well as variable, across DLBCL tumors. We found that 1,228 gene sets exhibited such patterns of co-regulation.

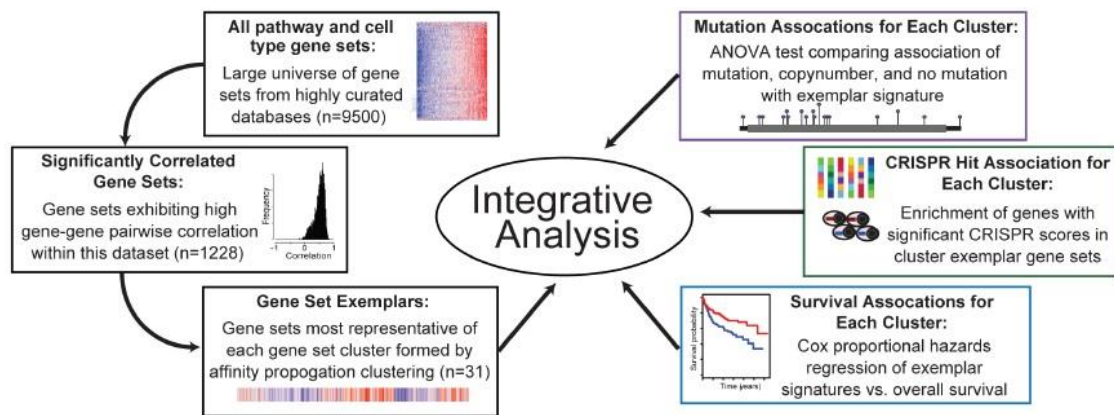


Figure 12: Schematic of integrative analysis of gene expression, genetic alterations, and outcome

In addition to the observed high correlation among their constituent genes, the vast majority of these 1,228 gene sets were strongly correlated with each other. The high inter-gene set correlation indicated a high degree of redundancy between these gene sets, likely reflecting their origin from shared biological processes. To reduce this

observed redundancy and focus on distinct gene expression sets that described the underlying biology, we applied affinity propagation clustering [135] to define the exemplars that essentially replicated the patterns of the members of the cluster. Application of affinity propagation clustering resulted in 31 unique clusters with cluster assignments for all gene sets. This analysis identified 31 non-redundant gene clusters that each comprised highly correlated gene sets (Table 1), as well as exemplar gene sets that represent each cluster (Figure 13A). Our results recapitulated a number of described patterns underlying DLBCL gene expression. The cell-of-origin-derived ABC-high and GCB-high signatures were generally expressed in a mutually exclusive fashion. In addition, the signature clustering analysis also revealed two broad classes of gene sets. The first class included signatures representing many immune-cell types (regulatory T cells, myeloid cells, NK cells) and stromal connective tissue (epithelial and endothelial cells). This group included the Monti Host Response signature [133], as well as the Lenz Stromal 1 and Stromal 2 signature [134]. The second class included signatures related to tumor cell processes including proliferation, transcription, translation, DNA replication, and cell cycle, as well as the Monti BCR and Oxidative Phosphorylation signatures [133].

Table 1: Gene set signatures. Cluster names with CRISPR essential gene enrichment odds ratios (CRISPR-O), CRISPR essential gene enrichment p-value (CRISPR P), Survival association hazard ratio (Survival-HR), and survival association (Survival-P)

Cluster	Cluster name	CRISPR-O	CRISPR P	Survival-HR	Survival-P
---------	--------------	----------	----------	-------------	------------

1	ABC-specific	1.38219815	0.52672239	1.16820126	0.21072216
2	GCB-specific (Lenz germinal center)	2.91937305	0.3119789	0.69132382	0.00033239
3	Regulatory T cells	0	0.17555093	0.51623892	0.00013796
4	Complement / Adipocytes	0	1	1.07312582	0.56649465
5	Myeloid / macrophages	0	0.00219748	0.91195806	0.59317068
6	Inflammatory response (Monti HR)	0	0.07902316	0.69420974	0.04589052
7	Immunoglobulin binding	0	0.18443309	0.72006563	0.10741642
8	T/NK Cells	0	0.64278568	0.71161091	0.00425825
9	Interferon / T cell	0.55664864	0.49887959	0.79546478	0.22473228
10	Epithelial cells	0.17147553	0.00174876	0.54503646	0.00144463
11	Stromal connective tissue (Lenz stromal 1)	0.13064916	0.01181791	0.63299189	0.00149268
12	Extracellular matrix / lymph vessels	0.53033955	0.25728244	0.41024656	0.00036766
13	Gas transport	1.4590553	0.50867543	1.1360205	0.35642502
14	Endothelial cells (Lenz stromal 2)	0	0.62744489	0.84621364	0.22503583
15	Antigen presentation and processing	0	0.41101613	0.62523258	0.00014971
16	MHC Class II	0	1	0.76539771	0.00123987
17	Dendritic cells	0	1	0.68961442	0.00150371
18	Glutamate receptor	0	0.41664398	1.35499659	0.00910192
19	B cells	2.38838214	0.36157247	1.04710794	0.71637911
20	Cholesterol biosynthesis	2.19098982	0.24695932	1.17313242	0.26446524
21	Mitochondrial fatty acid beta oxidation	0	1	1.18982912	0.10097877
22	Glycolysis	5.64422113	0.02278426	1.05367675	0.67577188
23	Oxidative Phosphorylation (Monti OxPhos)	5.19896661	4.7427E-09	1.37660014	0.03804499
24	DNA damage	7.8843663	1.079E-07	1.19163237	0.23663759
25	MYC / nucleotide biosynthesis	11.892739	4.0809E-13	1.48158	0.00402426
26	Translation Initiation	26.6883356	5.8566E-51	1.07431114	0.55585112

27	RNA processing and ribosome	12.9032758	3.5264E-50	1.36253468	0.04617716
28	Cell cycle	6.80393561	6.4267E-14	1.46562186	0.02745283
29	Chromatin organization (Monti BCR)	3.89504944	8.3537E-12	1.4414844	0.06878094
30	Transcription	1.38386041	0.54072876	1.01263803	0.92691725
31	DNA replication	7.12875868	1.4744E-19	1.31842499	0.03107574

For each of these gene set exemplars, we identified the genetic alterations associated with high and low expression. Figure 13B shows a heatmap of signatures versus mutated genes (mutations and copy number alterations). These associations demonstrate a number of interesting patterns including the association of *RHOA* and *MYC* with proliferation-related signatures. The expression of many signaling pathways such as PI3 kinase did not reliably associate with specific mutations in *PIK3R1*, *PIK3CD*, or *PTEN*, suggesting that expression patterns from non-malignant cells may confound the effective measurement of these associations. There was no association between the expression of stromal signatures and overall mutational burden ($p = 0.4$).

We further investigated the relationship of our CRISPR hits and these gene sets. We found that gene sets related to cancer-related processes (e.g., oxidative phosphorylation, DNA replication, cell-cycle progression, RNA processing) demonstrated significant enrichment of the CRISPR-identified driver genes. Strikingly, no enrichment of CRISPR hits were found in the immune and stromal gene sets. Finally, we tested each of the cluster exemplar signatures for associations with overall patient survival. Figure 13C annotates the significant associations for all gene expression

clusters. As expected, the GCB DLBCL signature was associated with better overall survival. Interestingly, the signatures in the stromal and immune response group (e.g., regulatory T cells) tended to associate with better survival, while the signatures in the proliferation and B cell group tended to associate with worse survival (e.g., *MYC*/Nucleotide Biosynthesis).

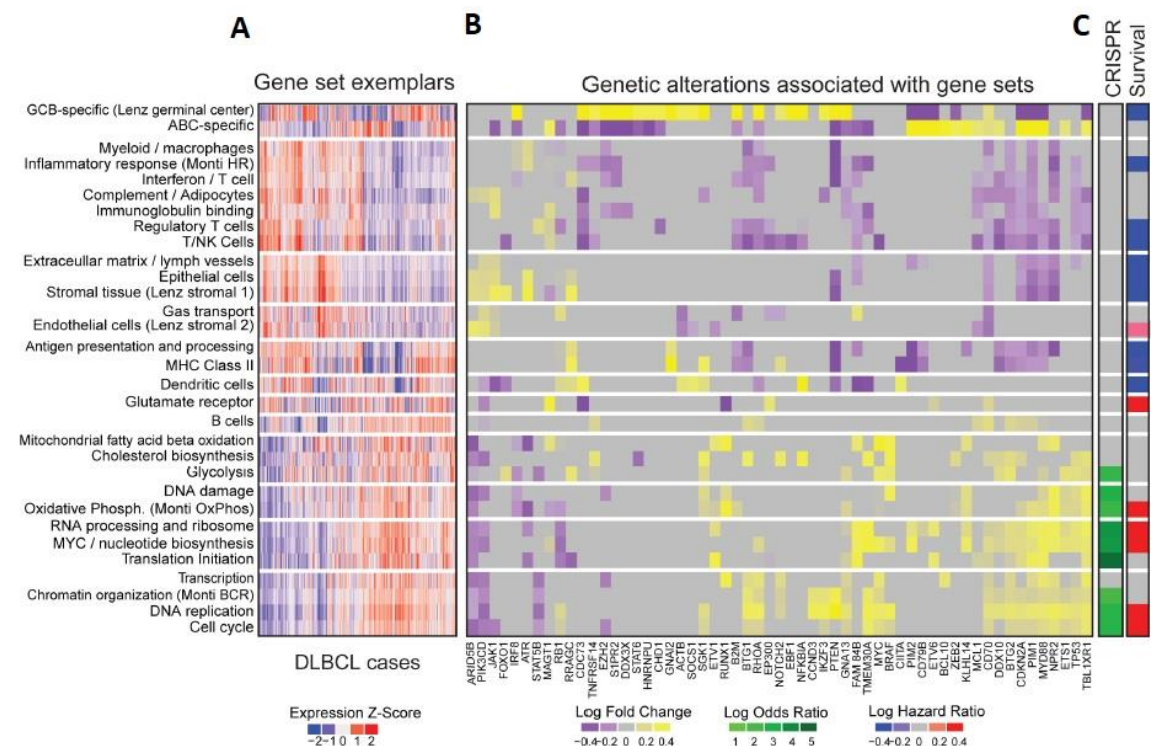


Figure 13: Integrative Analysis of Gene Expression, Genetic Alterations, and Outcome. (A) Heatmap of expression of significantly correlated gene set exemplars across DLBCL samples (N = 624). (B) Heatmap of fold-change associations of significant genetic alterations with gene set exemplars. Fold-changes for significant associations (ANOVA test, $p < 0.05$) are shown in yellow-purple color scales, while insignificant associations are colored gray. (C) Left: Enrichment of CRISPR hits within the cluster exemplar gene sets (Fisher's test, $p < 0.05$). Right: Survival association with cluster exemplar signature (logrank test, $p < 0.05$).

2.3.6 Clinical characteristics of DLBCL driver genes

Overall survival following diagnosis and initial treatment is shown in Figure 14A. The clinical risk groups of the IPI was highly prognostic in our patients ($p < 10^{-6}$), with clear distinctions in survival for the low-, intermediate-, and high-risk groups. The cell-of-origin distinction was found to be prognostic in our data (Figure 14B, 14C).

High expression of *MYC* and *BCL2* has been associated with worse prognosis [136,137]. High expressors were defined from RNA-seq data as *MYC*-high and *BCL2*-high and were significantly associated with survival. Double expressors were defined as samples that have both high expression of *MYC* and high expression of *BCL2* (Figure 14D). We assessed the independence of ABC/GCB subgrouping and double expressors (high *MYC* and *BCL2*) and found that ABC/GCB status does not further stratify the double expressors.

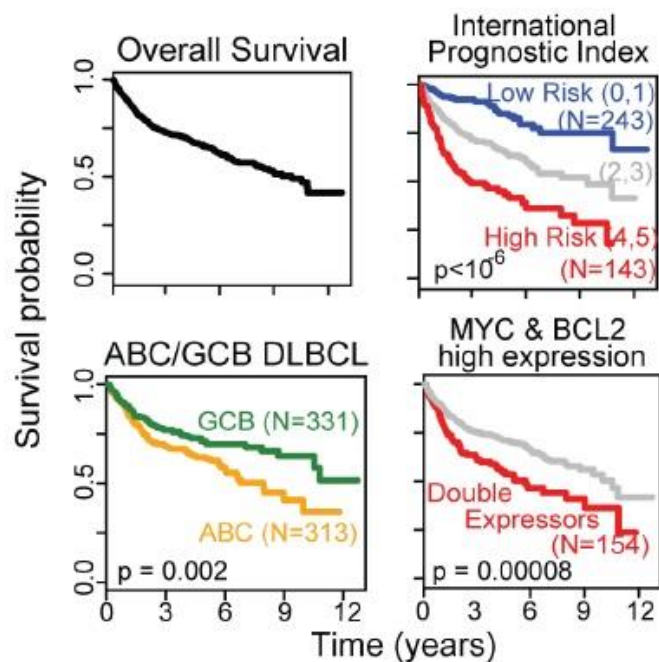


Figure 14: Common models of patient stratification. (A) Overall survival of 1,001 DLBCL cases, (B) cases stratified by IPI groups, (C) ABC/GCB DLBCL, and (D) MYC and BCL2 expression.

We separately examined the association of translocations in *MYC* and *BCL2* with mutations and expression by performing fluorescent in situ hybridization (FISH) using both genes to define translocations involving the immunoglobulin locus and *MYC* and *BCL2* (performed separately). *MYC* translocations are associated with mutations and high expression of *MYC*, while translocations in *BCL2* were associated with mutations and amplifications (Online supplement S4).

The potential association of mutations with clinical outcome has largely been unexplored. We first examined the association of specific mutations and survival among

all DLBCLs. We found that *MYC* mutations were strongly associated with poorer survival, as were mutations in *CD79B* and *ZFAT*. Mutations in *NF1* and *SGK1* were associated with more favorable survival. We further examined the association of different mutations and survival within ABC and GCB DLBCL. Genetic alterations in *KLHL14*, *BTG1*, *PAX5*, and *CDKN2A* were associated with significantly poorer survival in ABC DLBCL. Alterations in *CREBBP* were associated with favorable survival in ABC DLBCLs. In the GCB DLBCL group, genetic alterations in *NFKBIA*, *NCOR1* were associated with poorer prognosis, while alterations in *EZH2*, *MYD88*, and *ARID5B* were all associated with significantly better prognosis (Figure 15).

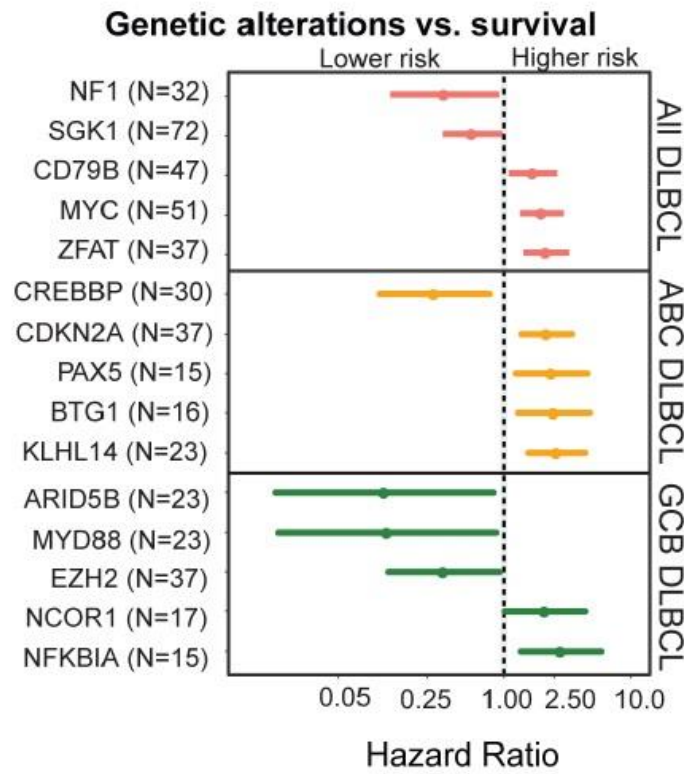


Figure 15: Genetic alterations and survival. Genetic alterations (left) with their associated hazard ratio (middle) and subtype specificity (right). Dashed line represents delineation of high vs. low risk for all hazard ratios ($p < 0.05$).

We have developed an interactive webtool (<https://dlbcl.davelab.org>) for survival analysis using clinical and genomic features. Our large dataset provided the opportunity to directly examine the combinatorial and potential confounding effects of *MYC*, *BCL2*, and cell-of-origin, along with other genetic alterations.

We developed a multivariate supervised learning approach for defining the association of survival with combinations of genetic markers (150 genetic driver genes) and gene expression markers (cell-of-origin, *MYC*, and *BCL2*). Our method is

summarized in Figure 16. We first enumerated all 313 possible combinations, which involved up to three separate genetic and gene expression markers and affected a minimum of 20 patients, as inputs for predictive model for survival [138].

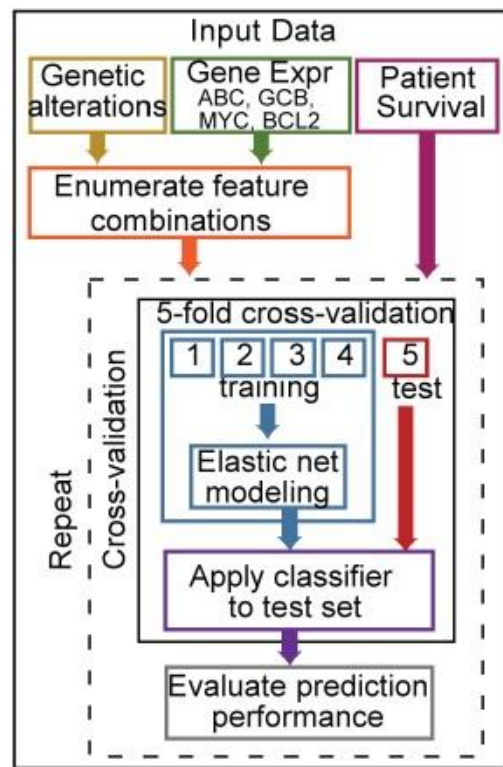


Figure 16: Schematic depicting multivariate survival modeling.

The individual components of the model are depicted in descending order of their association with poor survival in Figure 17. *MYC* genetic alterations combined with *MYC* expression defined the subset with the least favorable prognosis in DLBCLs, while GCB DLBCLs with *CD70* alterations had the most favorable prognosis. Each component in the model was defined so as to be prognostically independent from all others to

collectively identify subgroups that cover the entire set of DLBCL cases. These features were combined into a Cox proportional hazards model to generate a prediction score, and the DLBCL cases were divided into three subgroups based on their predicted risk. The full list of these combinatorial features along with sample-level information and model coefficients is provided in the online supplement.

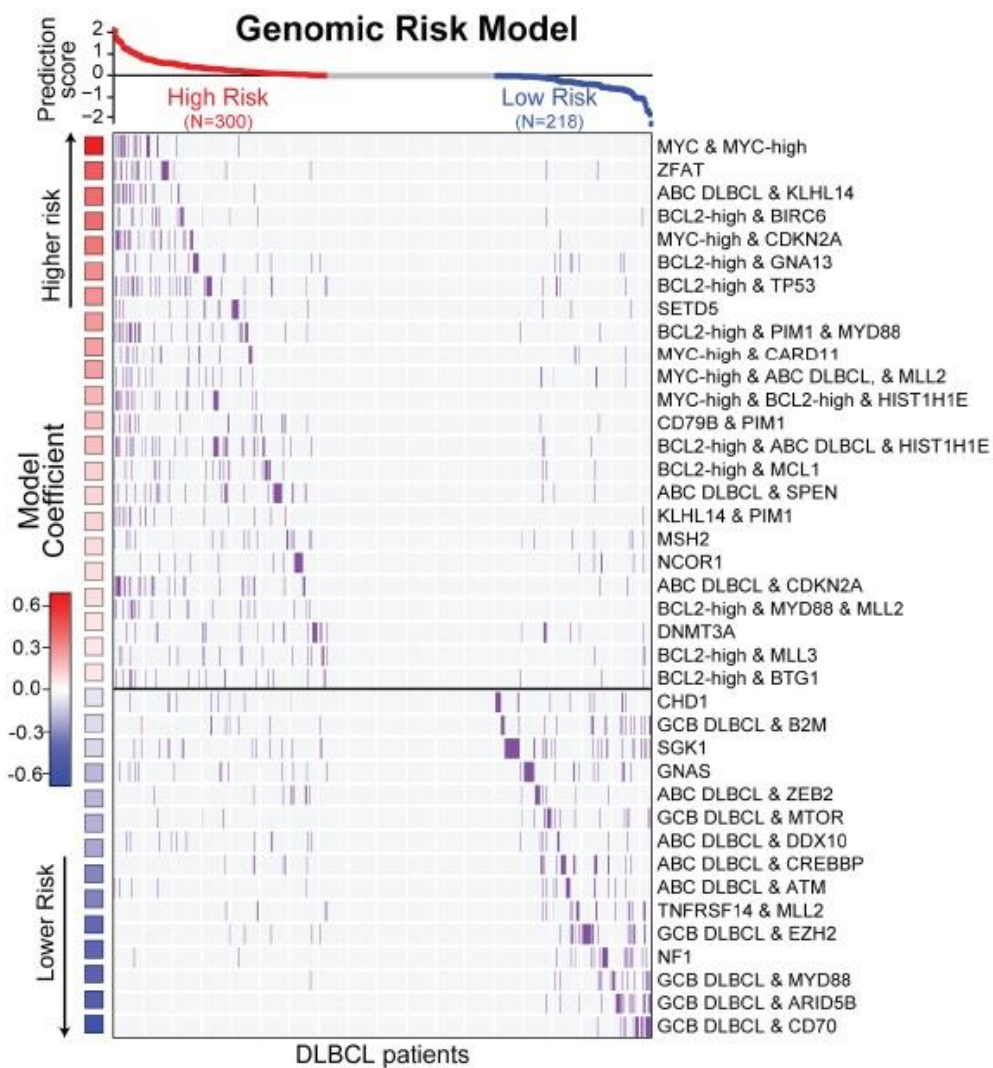


Figure 17: Genomic risk model. Combinations of genetic and expression features that comprise the genomic risk model in DLBCL patients. Model coefficient represented by color scale (left). Higher risk (red) and lower risk (blue) shown in prediction score. Features associations (right) are shown within patients (middle).

We validated our predictive modeling approach using an independent test set (20% of the data). The genomic model was highly significant in distinguishing patients

with high versus low risk of death in the test set ($p = 8 \times 10^{-5}$; Figure 18 left). In addition, we also used 5-fold cross-validation repeated 100 times to gain a robust estimation of model performance, finding that it was highly statistically significant in distinguishing the survival groups (median logrank $p = 8 \times 10^{-6}$). The integrative model strongly outperformed the other models based on genetic alterations (DNA only) or expression (RNA only) alone (Figure 18 right). The genomic risk model, which used combinatorial features, also strongly outperformed a model with the same input features (DNA+RNA) but with no combinations used, underscoring the importance of examining interactions between different genomic features for evaluating a patient's risk.

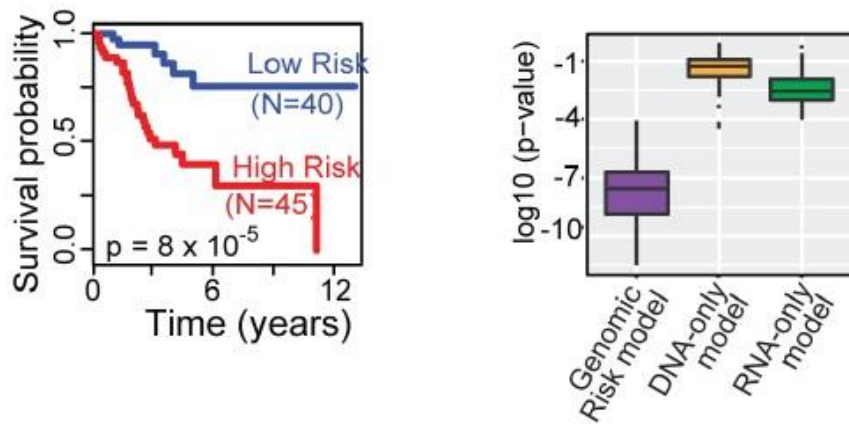


Figure 18: Survival predictions within model types. (A) Survival of the predicted risk groups in the test set (logrank test, $p = 8 \times 10^{-5}$). (B) Cross-validation performance of the genomic risk model compared to that with only genetic alterations (DNA-only), and gene expression (RNA-only).

We further examined the performance of the genomic model in the known risk groups that are known to influence survival: cell-of-origin, double expression of *MYC*

and *BCL2*, and IPI. Our genomic model was able to discern patients with significantly distinct outcomes within each group (Figure 19, $p < 10^{-3}$ in all cases). These data indicate that our genomic predictor provides robust prognostic information that is not currently captured by widely used clinical and expression-based models.

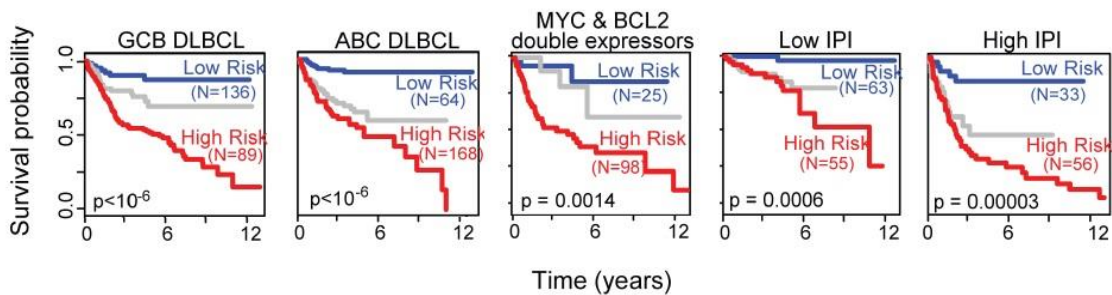


Figure 19: Modeling risk in known high risk groups. The genomic risk model significantly stratifies survival within known risk groups (logrank test).

2.3.7 Comparison of clinical versus genomic risk

The hazard ratios of the different survival models in DLBCL are compared in Figure 20A for cell-of-origin and *MYC* and *BCL2* separately and together, as well as the clinical risk (IPI) and genomic risk models. The genomic risk model outperformed all of these existing risk predictors in DLBCL. Next, we examined the time-dependent effects of the genomic model and IPI for DLBCL patients who survived for at least 1, 3, or 5 years (Figure 20B). We found that while IPI was highly prognostic in early years, its effect waned as a function of time. In contrast, the genomic risk model had significant prognostic value for predicting longer-term mortality. These data indicate that the

effects of IPI are strongest in predicting early mortality, whereas genomic risk is effective in predicting both early and late mortality.

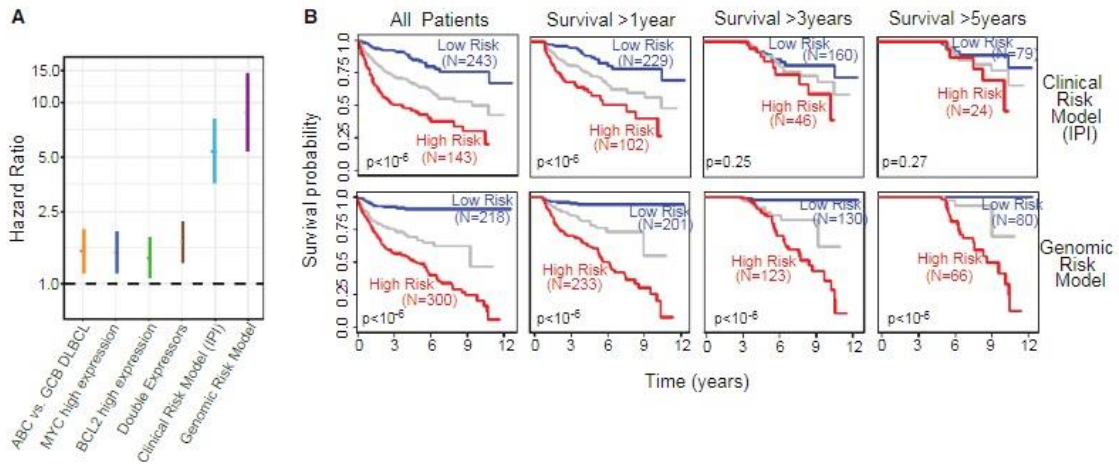


Figure 20: Comparison of clinical risk model with genomic risk model. (A) Comparison of hazard ratios (high- versus low-risk groups) and 95% confidence intervals for various DLBCL risk models, including our genomic risk model. **(B)** The matrix of Kaplan-Meier survival plots indicates risk stratification by clinical risk model (IPI) versus genomic risk model for all patients and patients stratified by minimum overall survival of 1, 3, and 5 years (logrank test).

We separately examined the survival effects of response to initial therapy. We found that response to initial therapy was significantly associated with survival, with median survival for complete responses being >12 years, while the median survival for partial (2 years) and no responses (1 year) were drastically lower (Figure 21A). We also observed a pattern of continued mortality in patients achieving a complete remission, suggesting that for a large proportion of patients, achieving a complete remission does not necessarily indicate a cure. Both IPI and genomic risk models are significantly

associated with response to initial therapy (Figure 21B), with IPI being more predictive. The genomic risk model has a higher percentage of high-risk patients in the complete response group compared to IPI. We hypothesized that while the complete responders have a generally favorable prognosis as a group, this group still harbors a significant fraction of patients with early mortality that can be identified by the genomic risk model. To test our hypothesis, we examined survival for the predicted risk models stratified by clinical responses (Figure 21C). The genomic risk model was highly prognostic within each of the clinical response groups, indicating that the genomic profile of a patient strongly influences their prognostic risk, regardless of their initial response to therapy.

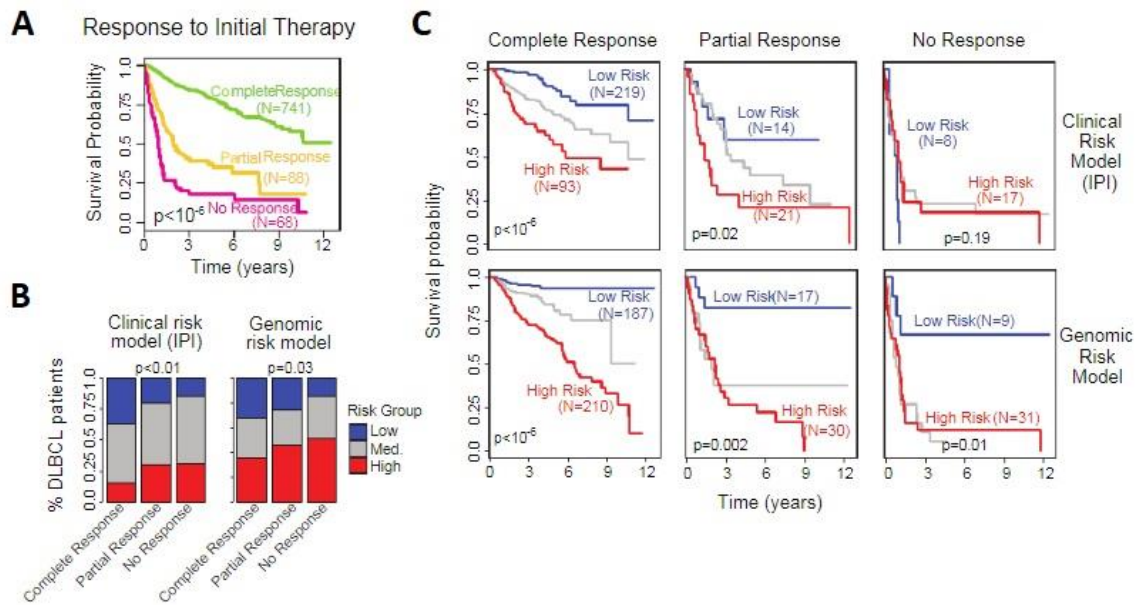


Figure 21: Genomic risk model identifies high risk groups in all clinical response groups. (A) Survival plot for response to initial therapy (logrank test, $p < 10^{-6}$). **(B)** Prediction of response to initial therapy using clinical risk model (left) and genomic risk

model (right) (chi-square test). (C) Survival plots showing the stratification of genomic and clinical risk models for each of the individual responses to therapy (logrank test).

2.4 Discussion

The genetic heterogeneity of DLBCL has led to diverse approaches to classification and prognostication and presented challenges in the development of effective new therapies. By enrolling 1,001 DLBCL patients, we designed a study that is adequately powered to identify genetic drivers and their clinical significance. CRISPR screening provided a powerful approach to probing the functional role of the genetic drivers, identifying the driver genes that directly affect cell survival.

Of the 150 driver genes that we identified, 27 genes including *SPEN*, *KLHL14*, and *MGA* have not been previously implicated in DLBCL to our knowledge. In addition, a number of genes such as *CDKN2A* and *RB1* that were thought to be altered in DLBCLs through copy number alterations were also revealed to be recurrent targets for function-altering mutations. While there is considerable interest in nominating genes for therapeutic targeting using mutational and ontology data, our data indicate a considerable need for caution. For instance, CRISPR-based knockout of several therapeutically targetable *NOTCH2*, *PIK3CD*, and *JAK2* did not have a significant impact in the growth of DLBCL cells. These findings suggest that such oncogenes may play a role in early pathogenesis or another cancer-related function that may not be easy to

target directly in DLBCL. On the other hand, functional screening established a role for a number of genetic drivers as potentially targetable genes in DLBCL.

Our study informs therapeutic development in several ways. First, our data provide a comprehensive approach to risk stratification, distinguishing DLBCL patients who benefit from standard therapy from those who are expected to experience poor outcomes. Second, our data indicate that nearly 36% of DLBCL patients have a genetic alteration that may inform sensitivity to currently available therapies. Our genomic risk model can be applied clinically using existing assays, through the measurement of cell-of-origin, *BCL2* and *MYC* expression, and targeted sequencing of a panel of DLBCL driver genes. Our data also provide the context for understanding outcomes in clinical trials of targeted therapy. For instance, while *MTOR* mutations are prevalent in DLBCL, those patients tend to have generally good outcomes, especially in the GCB subtype. Since patients with relapsed DLBCL comprise the vast majority of patients who are enrolled in clinical trials, many such trials may be missing the DLBCL population that harbors the target (e.g., *MTOR*) but respond well to standard therapy, potentially contributing to a negative trial that fails to show efficacy of these agents. Finally, we anticipate that our work will provide a starting point for hypothesis-driven biological modeling of different gene combinations that contribute to poor-risk DLBCLs to better understand their combined effects and aid in developing new therapeutic approaches.

Genetic heterogeneity is a defining feature of DLBCL. Our study demonstrates the benefits of applying genomic methodologies to large, well-characterized cohorts of cancers to discern patterns that are not possible with smaller studies. Our approach has identified clinical and functional drivers of DLBCL that are prerequisites to improving outcomes in the disease.

3. The Role of *RHOA* in Diffuse Large B Cell Lymphoma

3.1 Introduction

One of the striking findings of the CRISPR screen (see: chapter 2) was the potential role of *RHOA* as an essential gene, suggesting a possible oncogenic role in DLBCL. *RHOA* is known to be required for early B cell development, and loss of *RHOA* reduces BAFF-mediated Akt activation [139]. Recurrent *RHOA* mutations have also been identified in other B cell lymphomas, with enrichment in R5Q mutations [140,141]. Figure 22A shows the mutational distribution of *RHOA* in DLBCLs, with mutations occurring in equal proportions in both ABC and GCB subtypes. The R5Q hotspot was identified with approximately 30% of *RHOA* mutations occurring at this position. Interestingly, computational and biochemical modeling of this change suggest different outcomes of this change. Some studies have suggested a deleterious effect, as it is involved with GEF binding, while others posit that this mutation would cause *RHOA* to become more promiscuous with other GEFs [141,142]. The overrepresentation of R5Q mutations coupled with the CRISPR screen hit suggests that *RHOA* is acting as an oncogene in the context of DLBCL, and elucidating its role in the disease will give greater understanding of the underlying biology of the disease and provide possible therapeutic targets.

3.2 Methods

3.2.1 RHOA knockdown

The inducible TRIPZ *RHOA* shRNA set, RHS4740-EG387 glycerol set, was purchased from Dharmacon. To generate lentiviral particles, *RHOA* shRNA, pPAX2 and pVSV-G plasmids were used in a ratio of 2:2:1 to transfect 293T cells with lipofectamine 2000 following the vendor's instruction. 48 hours after transfection, supernatant was collected and filtered through 0.45 µm filter. Human Raji cells were infected by mixing the cells and filtered supernatant together. 1-2 µg/ml Puromycin was added 48 hours after infection and cells were selected for 5-7 days. To induce knock down, 0.5-1 µg/ml doxycycline was added to the media. 3-5 days later, proteins were collected for western blots to confirm the *RHOA* knock down. V3THS_379598 and V3THS_411340 produced the most effective knockdown.

3.2.2 Western blot

All western blots and analysis were performed using standard procedures. Briefly, cell lysates were made by pelleting 1×10^6 cells and adding 100µL of RIPA buffer. 90µL of 2x Laemmli buffer and 10µL of beta mercaptoethanol were added. 10µL of sample were loaded into a 4-15% tris-glycine polyacrylamide gel. Samples were separated by electrophoresis and transferred to a PVDF membrane. Blots were then blocked in a 5% non-fat dry milk, 0.1% tween-20, PBS solution. The following antibodies

were used: RHOA, β -Actin (Table 2). All primary antibodies were used at 1:1000 dilution.

3.2.3 Cell cycle analysis

Both knockdown and scramble control inducible shRNA Raji lines were used for cell cycle analysis. 1×10^6 cells per condition were resuspended in 50 μ L Hank's Balanced Salt Solution (Thermo Fisher), with 2% FBS. 1 mL of ice cold 70% ethanol was added dropwise while vortexing. Cells were then pelleted and washed twice with HBSS + 2% FBS. Cells were then resuspended in 10 μ g/mL 4',6-Diamidino-2-Phenylindole, Dihydrochloride (DAPI) for 30 minutes at room temperature. Cell cycle analysis was performed as previously described for DAPI staining. Flow cytometry was performed using a Sony LE-SH800 flow cytometer and cell sorter.

3.2.4 Mouse strains and breeding

Mice were housed in a Duke University Medical Center Division of Laboratory Animal Resources facility, and experiments and animal protocols were approved by the Duke University and Medical Center Animal Care and Use Committee. The Rhoa conditional knockout strain was a gift from Dr. Yi Zheng of Cincinnati Children's Hospital Medical Center. The AID-Cre transgenic strain for GCB specific Cre expression and the Mb1-Cre knock-in strain for pan B cell specific Cre expression were provided by Dr. Yuan Zhang (Duke University, Durham, NC, USA).

To evaluate the function of RHOA in the context of loss in germinal center B cells, *Rhoa* +/fl mice were crossed with *Rhoa* +/fl AID-Cre Tg/0 mice. Resulting Cre+ wildtype and Cre+ knockout mice were used for collection of splenic and Peyer's patch derived lymphoid cell isolations. Genotyping was performed using Qiagen HotStarTaq master mix. Primers used for the Mb1-Cre knockin allele genotyping were 5'- CCC TGT GGA TGC CAC CTC-3' and 5'-GTC CTG GCA TCT GTC AGA G-3' and the product amplified using touchdown PCR (annealing temperature declining from 66° to 58° over nine cycles and then at 57°C for 41 cycles). The AID-Cre transgene was identified using primers 5'-CCG TAA CCT GGA TAG TGA AAC AG-3' and 5'-CCA TGC GAG TCT TAA GAT GTT G-3', and an annealing temperature of 64°C. Primers for *Rhoa* genotyping were 5'-TCT CTG CAC TGA GGG AGT TAG G-3' and 5'-GTA CAT ACA GGG AAT GGA AAC AAG G-3', with an annealing temperature of 58° for 33 cycles.

3.2.5 Sheep red blood cell (SRBC) immunization

Purified sheep red blood cells were washed three times with PBS then resuspended in PBS at a concentration of 5x10⁸ cells/ml. At 8 weeks of age mice were injected intraperitoneally with 1x10⁸ sheep red blood cells to stimulate germinal center formation and were sacrificed 8 - 10 days later for lymphocyte isolation.

3.2.6 Flow cytometry of murine cells

Mouse spleen and Peyer's patches were harvested in RPMI-1640 medium containing 10% (vol/vol) FBS for experiments assessing lymphocyte populations, or in RPMI-1640

containing 0.5% fatty acid-free BSA for experiments assessing F-actin. Single-cell suspensions were made by dissociating the tissues between ground glass slides and filtering the suspension through 70 μm cell strainers. Red blood cells were lysed, and remaining cells were washed twice with PBS/2% FBS. Prior to staining, samples were blocked with anti-mouse CD16/CD32 for 5 minutes at room temperature.

For the analysis of B cell populations, spleen cells (1-2x10⁶ per staining) and Peyer's patch cells (1/5 of harvested population per staining) were blocked using CD16/32 then incubated with antibodies to surface antigens for 20 minutes at 4° (Table: 2). Cells were then washed twice, resuspended in PBS/2% FBS and analyzed by flow cytometry using a Sony LE-SH800 flow cytometer and cell sorter. For lineage analysis, B cells were identified as B220+, T cells were identified as CD3+, and myeloid were identified as CD11b+. Identification of B cells, germinal center (GC) B cells, and dark zone (DZ) and light zone (LZ) GC B cell subpopulations were based on the staining and gating strategy as previously described. Briefly, within the lymphocyte gate total B cells were B220+, and GC B cells were B220+FAS+CD38/CD4^{low} with DZ GC B cells additionally CD83^{low}CXCR4^{hi} and LZ GC B cells CD83^{hi}CXCR4^{low}. For GCB quantification, staining and gating were identified as B220+IgD^{lo}FAS+GL7+.

For detection of F-actin, cells were blocked, stained for surface GC B markers as above, fixed using BD Cytofix/Cytoperm™ and permeabilized using BD Perm/Wash™ (BD Pharmingen), stained with Acti-stain 488 phalloidin (Cytoskeleton, Inc.) for 30

minutes at room temperature, and washed in BD Perm/Wash™. Cells were then washed twice, resuspended in PBS/2% BSA and analyzed by flow cytometry using a Sony LE-SH800 flow cytometer and cell sorter.

Table 2: Antibodies used in RHOA study

List of antibodies used for RHOA study. WB = Western Blot, FC = Flow cytometry.

Antibody	Clone	Conjugate	Application	Source
β-Actin	13E5	N/A	WB	Cell Signaling Technology
RHOA	67B9	N/A	WB	Cell Signaling Technology
Acti-Stain	Phalloidin	AF488	FC	Cytoskeleton Inc.
B220/CD45R	RA3-6B2	BV421	FC	BioLegend
CD3ε	145-2C11	FITC, PE	FC	BioLegend
CD4	RM4-5	FITC	FC	BioLegend
CD16/32	93	N/A	FC	BioLegend
CD38	90	FITC	FC	BioLegend
CD83	Michel-19	BIOTIN	FC	BioLegend
CXCR4	L276F12	PE	FC	BioLegend
Fas	Jo2	PE-Cy7	FC	BD Pharmingen
GL7	GL7	AF647	FC	BioLegend
IgD	11-26c.2a	PE	FC	BioLegend
Streptavidin	N/A	APC	FC	BD Pharmingen

3.3 Results

3.3.1 *In vitro* models of RHOA gain and loss

We first interrogated the mutational profiles of *RHOA* in the compiled 1001 patient samples from chapter 2. Though we found some mutations across the gene body, there was a single hotspot mutation at R5Q that was overrepresented (Fig 22A). This, coupled with the CRISPR screen showing *RHOA* as an essential gene, suggests RHOA is acting as an oncogene in the context of DLBCL. We next overexpressed both the wildtype form of the protein and the R5Q mutant form of the protein in SUDHL4 cells. We observed significantly increased proliferation in overexpressed form and even greater proliferation in the R5Q mutant overexpressed form (Fig 22B)

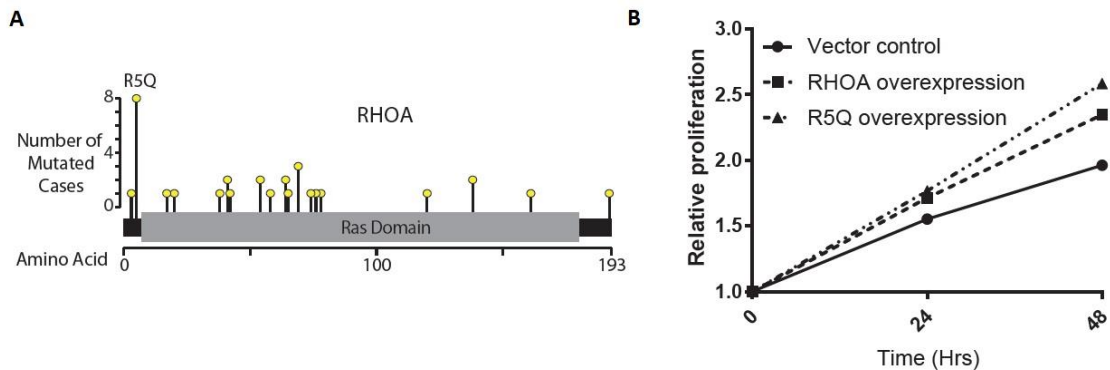


Figure 22: R5Q mutations in DLBCL. (A) Lollipop plot of mutations found in 1,001 DLBCL cases. Bar height indicates number of mutated cases. Bar position indicates affected amino acid. Ras domain indicated by grey box. (B) Relative proliferation of SUDHL4 cell with either vector control (circles), wildtype RHOA overexpression construct (squares), or R5Q mutated RHOA overexpression (triangles) grown over 48 hours. Both overexpression constructs display significant overexpression ($p < 0.05$). Proliferation was measured every 24 hours.

To interrogate the role of *RHOA* in B cell lymphomas we then knocked down *RHOA* expression in the SUDHL4 DLBCL cell line using 2 doxycycline inducible shRNAs. Knockdowns were then confirmed upon doxycycline induction (Figure 23A). We then examined the effects of *RHOA* knockdown on cell cycle progression using scramble control as well as induced and uninduced cells. Upon 72-hour induction of *RHOA* shRNA knockdown, cells began to accumulate in G2/M phase and consequently there was a reduction in G1 phase compared to uninduced (Figure 23B left) and scramble control (Figure 23B right). This data is summarized for all shRNAs, induction conditions, and controls in Figure 23C. This suggests a reduction in the ability of *RHOA* knockdown cell to efficiently pass through the G2/M checkpoint and slower overall cell cycle progression. Cell morphology was also altered in *RHOA* knockdown cells. Upon induction of *RHOA* shRNAs, cells began to lose their uniform shape and their diameter increased dramatically (Figure 23D bottom), while control cells maintained their

uniform size (Figure 23D top).

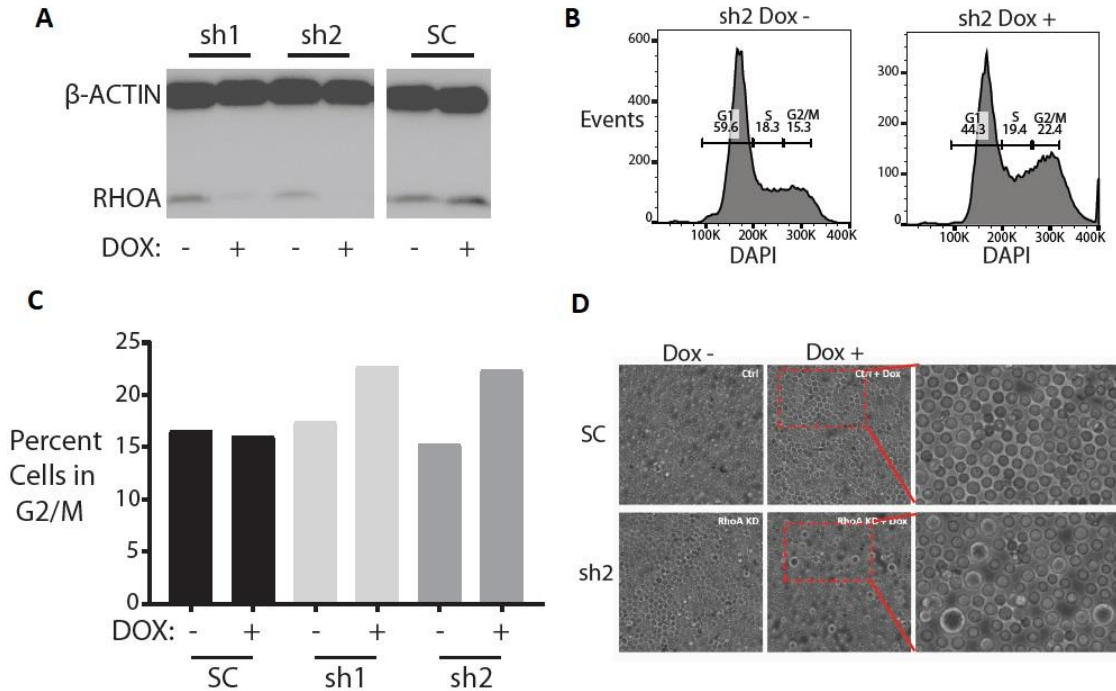


Figure 23: RHOA knockout disrupts cell cycle. (A) Western blot of inducible RHOA knockdown in RAJI cells with either shRNA1 (sh1), shRNA2 (sh2), or scramble control (SC). Doxycycline induction for 12 hours (+) or untreated (-). Blotted for RHOA (22kDa) and beta-actin (42kDa). (B) Cell cycle analysis of RAJI cells stained with DAPI. Gating was performed with wild type controls. Gates based on cell cycle position for G1, S, and G2/M. Both performed with sh2, with induction indicated by sh2+. (C) G2M percentages from all RAJI knockdowns of RHOA (sh1, sh2) and scramble control (SC). Analysis based on DAPI staining wild type controls. Induction by doxycycline (+) or untreated (-). (D) Microscopy of RHOA knockdown (sh2) or scramble control (SC) cells after 48 hours. Induction by doxycycline (+) or untreated (-).

3.3.2 *In Vivo* mouse models of *Rhoa* loss

We further investigated the *in vivo* role of *Rhoa* in normal B cells by breeding *Rhoa* conditional knockout mice with Mb1-Cre transgenic mice, in which Cre is restricted to the B cell lineage [143]. The breeding scheme led to mice with *Rhoa* specifically deleted in

B cells (Figure 24A shows final breeding scheme for mice used in Mb1-Cre experiments). We then compared B, T, and myeloid cell population in young adult mice who were wild type or deleted for *Rhoa*. We assessed these cell types in dissected Peyer's patches by flow cytometry. We found a dramatic reduction in B cell population and subsequent increase in T cell percentages in *Rhoa* deleted mice (Figure 24B)

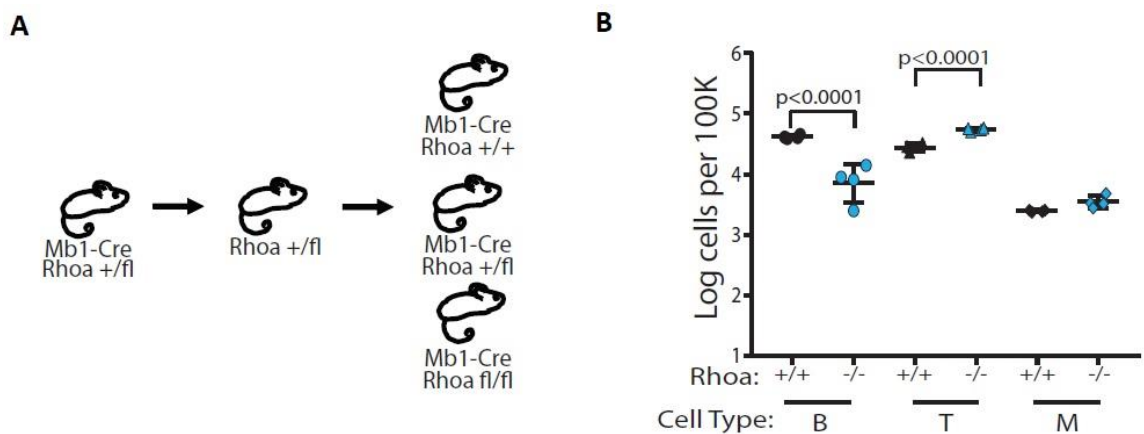


Figure 24: RHOA knockout in B lineage. (A) Schematic of breeding scheme for mice with knockout of RHOA in all B lineage tissues under the Mb1-Cre driver. “+” indicates wildtype allele. “fl” indicates floxed allele. (B) Relative proportion of B cells (B220+), T cells (CD3+), and myeloid cells (CD11b+) in wildtype (+/+) and RHOA fl/fl knockout (-/-) mice. All mice contained one copy of Mb1-Cre. Relative proportion is denoted as log cells per 100,00 single-cells analyzed.

To further refine the role of RHOA in the cell type of origin of DLBCL, we knocked out *Rhoa* in mouse GC B cells using floxed *Rhoa* and a GC B specific AID-Cre (Figure 25A). In these mice we saw a reduction in GC B cells as a proportion of total B cells (Figure 25B). The overall proportion of dark zone and light zone cells as a proportion of total GCB cells were altered in the knockout from their normal wild type

counterparts (Figure 25C), suggesting a dysregulated germinal center. Indeed, the normal ratio of approximately 2:1 of dark zone to light zone was significantly altered in the knockout (Figure 25D). *Rhoa* knockout GCB cells also had reduced filamentous actin (Figure 25E-F), a possible explanation of cell morphology changes in the human cell knockdown.

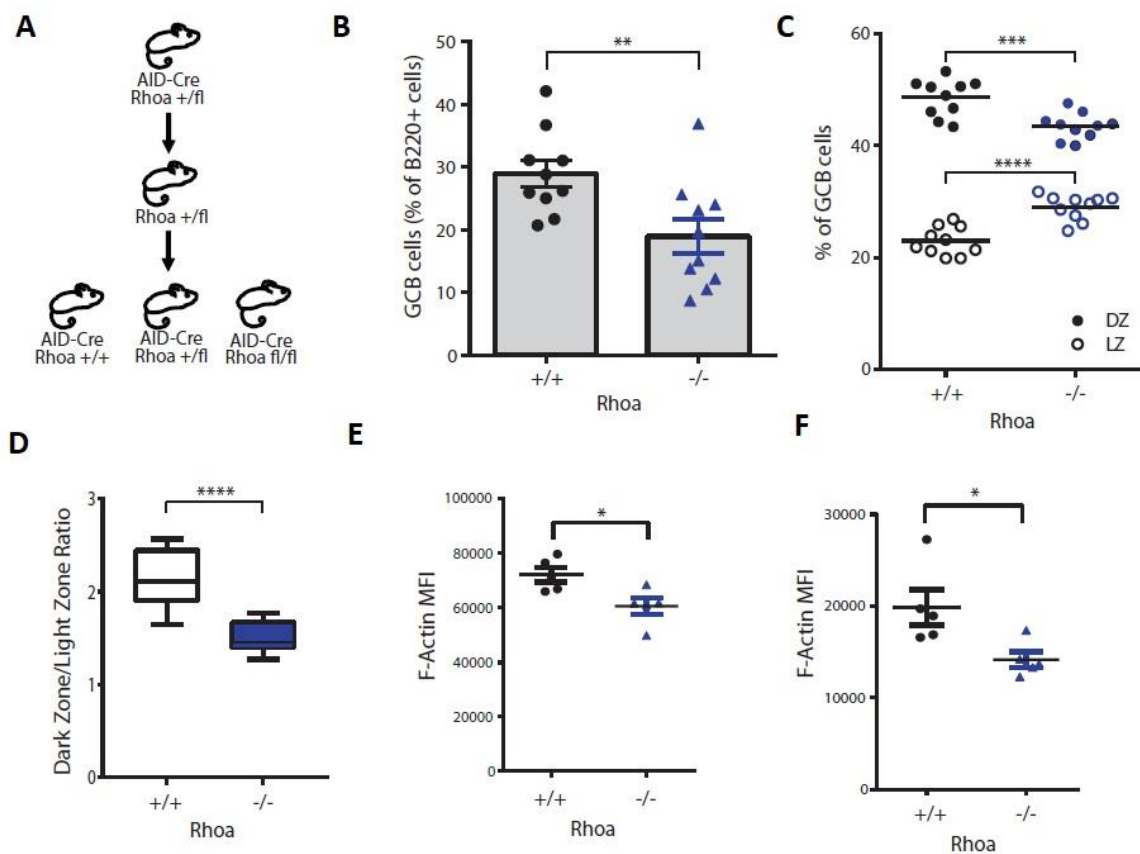


Figure 25: *In vivo* knockout of RHOA in GCB cells. (A) Schematic of mouse breeding scheme for floxed-*Rhoa* knockout in GCB cells by *Aid*-promoter-driven Cre recombinase.

“+” indicates wildtype allele. “fl” indicates floxed allele. (B) GCB cells (B220⁺ FAS⁺ CD38^{low} CD4^{low}) in *Rhoa* wildtype (+/+) mice and *Rhoa* knockout (-/-) mice as a percentage of all B220⁺ cells (N=10 per genotype, p<0.01). (C) Percent of dark zone (B220⁺ FAS⁺ CD38^{low} CD4^{low} CD83^{low} CXCR4^{hi}) and light zone (B220⁺ FAS⁺ CD38^{low} CD4^{low} CD83^{hi} CXCR4^{low}) as a percent of all GCB cells in *Rhoa* wildtype and knockout (N=10 per genotype, p<0.001). (D) Dark zone to light zone ratio for wildtype and *Rhoa* knockout

GCB cells (DZ/LZ, N=10 per genotype, $p < 0.0001$). **(E)** Mean fluorescent intensity of filamentous-actin in wild type and *Rhoa* knockout GCB cells (experiment 1, N=5 per genotype, $p < 0.05$). **(F)** Mean fluorescent intensity of filamentous-actin in wild type and RHOA knockout GCB cells (experiments=2, N=5 per genotype, $p < 0.05$).

3.4 Discussion

These data indicate that *RHOA* plays a role in B cell development, germinal center maintenance, and GC B cell function. In normal cells, RHOA seems to play a critical role in survival of the cell, with a great reduction in cells lacking RHOA signaling. The exact role of that RHOA plays downstream in cell survival and germinal center cycling is unknown, but obvious targets would be actin dynamics and migration or invasion. Other GEFs generally or B cells specific homing mechanisms, such as CXCR4, are also possible players in this mechanism.

The effects of *RHOA* on both ABC and GCB DLBCL indicates that it has a broad role in these tumors regardless of subtype. Specifically, the role of the R5Q mutation needs functional biochemical study to understand how this mutation interacts with GEFs and other pathway members. Hotspot mutations can lead to specifically tailored therapeutics, as functionality of the mutant form can be targeted. This data together shows that *RHOA* is critical to B cell biology and DLBCL pathogenesis, and it can be regarded as a legitimate therapeutic target in DLBCL.

Germinal center B cells are highly dependent on BCR signaling, but how dependent on focal adhesion related integrin signaling remains unclear. Pro-growth signaling in these cells are required for these cells to remain alive, both *in vitro* and *in vivo*. GC B cells are primed for apoptosis without stimulation of the B cell receptor, but GCB cells do not normally undergo anoikis, or cell death driven by the lack of ECM engagement. These cells often exist in an ECM free state and are often in a single-cell state. GC B cells do require cellular attachment to dendritic cells in the germinal center during affinity selection. A ring of associated integrins is engaged with the B cell receptor at the immune synapse during selection. The exact role of the pro-growth signaling at the immune synapse is unclear, but integrin signaling alone is insufficient to keep these cells alive. It is also unclear the degree to which B cell malignancies, which are almost all antigen independent, are dependent on integrin signaling, which would be mediated by FAK. In this study, I determine the role of FAK signaling in normal and malignant B cells and define its role in the B cell receptor signaling pathway.

4.2 Methods

4.2.1 FAK gene expression and TP53 Status

Gene expression information for normal plasma, memory, germinal center, and naïve B cells was taken from the IMMGEN database

(<http://www.immgen.org/databrowser/index.html>). This data was mouse gene

expression data from microarrays. The gene expression for *Ptk2* was then normalized to the lowest expression of the cellular subtypes, plasma cells.

Human gene expression data was compiled from our internal DLBCL data set (See: Chapter 2). Gene expression was calculated in FPKM to normalize across all samples. Associated mutations in *TP53* from the same samples was also gathered. Only high-quality calls for mutations were used.

4.2.2 Dose response curves

PF-578223 was acquired from Tocris Bioscience, and diluted to 10mM concentration in DMSO. Drug was plated in a 384-well plate using the LabCyte Echo 550 acoustic dispensing system. Drug was dispensed in a range of concentration from 10uM to 0.1nM in 10-fold dilutions, leading to 6 concentrations with a final volume of 25uL. Each well was then backfilled with DMSO to a final concentration of 0.1% carrier. A control well for live cells was used with the 0.1% DMSO. A control well for dead cells was plated at 10% DMSO by volume. Each concentration and control were plated in quadruplicate. Healthy OCI-LY3 and SU-DHL-4 cells were diluted to a concentration of 2.5×10^6 cells/mL. 25uL of these cells were then plated and mixed into each well of drug and control. Cells were incubated for 48 hours at 37°C in a cell culture incubator. 2 hours before the incubation was complete, 2.5uL of Alamar Blue reagent was added to all wells. Cells were then returned to the incubator for 2 hours. Cells were then measured for fluorescence on a Tecan Infinite 200 PRO microplate reader at 560 excitation and 590

emission. Dead control was subtracted from all wells as background fluorescence. All drug treated cells were divided by the live control to normalize across samples and replicate time points.

4.2.3 Western blot

All western blots and analysis were performed using standard procedures. Briefly, cell lysates were made by pelleting 1×10^6 cells and adding 100uL of RIPA buffer. 90uL of 2x Laemmli buffer and 10uL of beta mercaptoethanol were added. 10uL of sample were loaded into a 4-15% tris-glycine polyacrylamide gel. Samples were separated by electrophoresis and transferred to a PVDF membrane. Blots were then blocked in a 5% non-fat dry milk, 0.1% tween-20, PBS solution. Antibodies used include: SYK, pSYK, AKT, pAKT, ERK, pERK, FAK, pFAK, anti-rabbit IgG, Beta-tubulin, and H3 (See: Table 3). All antibodies were incubated at 1:1000.

Table 3: Antibodies used in FAK study

WB = Western Blot, FC = Flow cytometry.

Antibody	Clone	Conjugate	Application	Source
Histone H3	3H1	HRP	WB	Cell Signaling Technology
FAK	EP695Y	N/A	WB	ABCAM
Acti-Stain	Phalloidin	AF488	FC	Cytoskeleton Inc.
B220/CD45R	RA3-6B2	BV421	FC	BioLegend
CD3 ϵ	145-2C11	FITC, PE	FC	BioLegend
CD4	RM4-5	FITC	FC	BioLegend
CD16/32	93	N/A	FC	BioLegend

CD38	90	FITC	FC	BioLegend
CD83	Michel-19	BIOTIN	FC	BioLegend
CXCR4	L276F12	PE	FC	BioLegend
Fas	Jo2	PE-Cy7	FC	BD Pharmingen
GL7	GL7	AF647	FC	BioLegend
IgD	11-26c.2a	PE	FC	BioLegend
Streptavidin	N/A	APC	FC	BD Pharmingen
pTyrosine	PY20	N/A	WB	ABCAM
AKT (pan)	C67E7	N/A	WB	Cell Signaling
P-AKT S437	D9E	N/A	WB	Cell Signaling
p44/42	137F5	N/A	WB	Cell Signaling
P-p44/42 T-202 Y204		N/A	WB	Cell Signaling
SYK	D3ZIE	N/A	WB	Cell Signaling
P-SYK	C87C1	N/A	WB	Cell Signaling

4.2.4 Knockdown of FAK

FAK knockdown constructs were acquired from Dharmacon. These constructs bicistronically express a shRNA hairpin and a puromycin resistance cassette. Lentiviral particles were made of these constructs. Briefly, the knockdown constructs were grown in competent *E. coli*, then isolated from the bacteria. HEK-293T cells were then co-transfected with these constructs (individually and a scramble control), VSV-G, and PAX2 constructs. Viral supernatants were collected over 2 days. The supernatants were then filtered and concentrated. RAJI cells, viral particles, and polybrene were mixed together and centrifuged at 1000g for 1 hour. Cells were then grown for 24 hours, then exposed to puromycin to select for infected cells. Selected cell lines were then blotted for

FAK to determine knockdown. Of the 6 knockdown constructs, sh2 and sh4 had the most reduction in FAK. These and the scramble control were used for all later experiments. Activation of RAJI cells was performed by incubation of cells in normal media with goat-anti-human IgM antibodies for 15 minutes. Antibodies used in Table 3.

4.2.5 Immunofluorescence

Immunofluorescence was performed on RAJI cells activated with goat-anti-human IgM for 15 minutes. Briefly, cells are activated with IgM then spun down and placed on ice. Cold PFA is added for fixation for 20 minutes. The pellet is resuspended in permeabilization solution (4% formaldehyde, 0.03M sucrose, 1x PBS). Cells are then blocked for 15 mins in block buffer (5% horse serum, 1% BSA, 1x PBS). Cells are then stained with primary antibody overnight, and secondary is added for 2 hours. Cells are then resuspended in SlowFade Gold and placed on a slide with coverslip. Images were taken on a confocal microscope at the Duke University Light Microscopy Core Facility.

4.2.6 Mouse model of *Fak* loss in GC B cells

Mice were housed in a Duke University Medical Center Division of Laboratory Animal Resources facility, and experiments and animal protocols were approved by the Duke University and Medical Center Animal Care and Use Committee. The *Fak* conditional knockout strain was a gift from Dr. Jun-Lin Guan of the University of Cincinnati. The *Cy-Cre* transgenic strain for GCB specific Cre, restraining knockout to only GC B cells and their developmental lineage.

Mice were bred with a single copy of *C γ -Cre* per mouse and either wild-type, heterozygote, or knockout for FAK using the transgene as a knockout. Due to the restriction of *C γ -Cre* to gamma-immunoglobulin, all samples were only taken from splenic B cells. Mice were genotyped using the same protocol as the *Rhoa* mice (See: Chapter 2), with the following primers: P1, 5'-GCTGATGTCCCAAGCTATTCC-3'; P2, 5'-TGGCCTGCTATGGATTTCGC-3'; and P3, 5'-AGGGCTGGTCTGCGCTGACAGG-3'.

Mouse spleen and Peyer's patches were harvested in RPMI-1640 medium containing 10% (vol/vol) FBS for experiments assessing lymphocyte populations. Single cell suspensions were made by dissociating the tissues between ground glass slides and filtering the suspension through 70 μ m cell strainers. Red blood cells were lysed, and remaining cells were washed twice with PBS/2% FBS. Prior to staining, samples were blocked with anti-mouse CD16/CD32 for 5 minutes at room temperature.

For the analysis of B cell populations, spleen cells were used at 2×10^6 per staining. Cells were then washed twice, resuspended in PBS/2% FBS and analyzed by flow cytometry using a Sony LE-SH800 flow cytometer and cell sorter. Identification of B cells, germinal center B cells, and dark zone and light zone GC B cell subpopulations was based on the staining and gating strategy as previously described. Briefly, within the lymphocyte gate total B cells were B220+, and GC B cells were B220+FAS+CD38/CD4^{low} with DZ GC B cells additionally CD83^{low}CXCR4^{hi} and LZ GC B

cells CD83^{hi}CXCR4^{low}. For GCB quantification, staining and gating were identified as B220+IgD^{low}FAS+GL7+.

4.3 Results

4.3.1 Identification of FAK in B cell receptor signaling

Gene expression analysis of the B cell lineage has shown that *FAK* (*PTK2*) is upregulated specifically in the germinal center B cell, with a greater than 2-fold increase over other developmental stages both up- and downstream (Fig 27). Germinal center B cells are dependent on a number of cytokine and extra-cellular signals to escape programmed apoptosis within the germinal center. Cultured cell lines have escaped these programs as they became malignant and then cultured cells. They exist in an ECM and dendritic cell free environment in suspension, thus leading to minimal integrin growth stimuli.

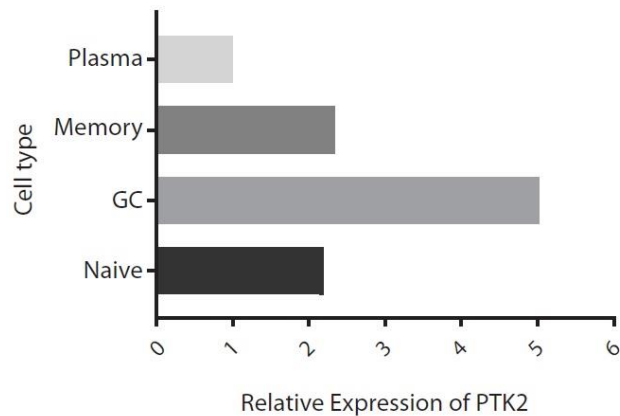


Figure 27: Relative expression of FAK (PTK2) in B cell lineage. All expression is relative to plasma cell expression. Sample include plasma cells (plasma), memory B cells (memory), germinal center B cells (GC), and naïve B cells (Naive).

To see if there was any effect on survival of FAK inhibition in this minimal-integrin condition, 2 cell lines were treated with an FAK inhibitor (PF-578223) for 48 hours across a range of doses from 10uM to 0.1nM. Results from this show that there is an effect in some B cell lines from as little as 0.1nM but most cell lines respond in the 1uM range (Figure 28). This suggest FAK is critical for survival and growth of GCB lymphomas, but it also suggests that it may perform this function in an integrin independent manner.

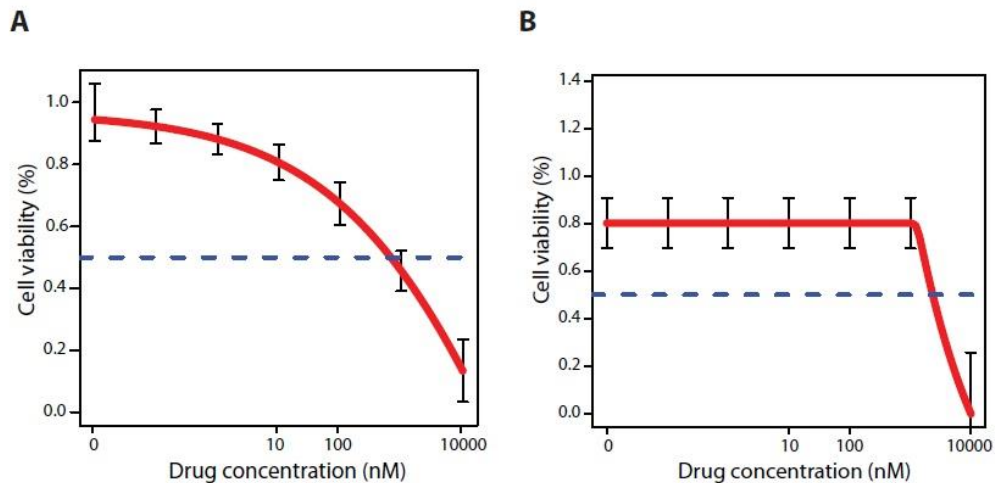


Figure 28: Dose response to FAK inhibition. Dose response curves of (A) OCI-LY3 and (B) SU-DHL-4 cells to inhibition of FAK by PF-578223. Cell viability was measured by incubation of Alamar Blue for 4 hours. Cell viability is relative to vehicle (DMSO) control. Cell viability is represented by red line. 50% viability is denoted by dashed line.

The most obvious target for understanding survival and signaling in B cells is the B cell receptor. In order to stimulate B cell receptor signaling, BCRs were activated using anti-IgM antibodies, aggregating and activating the BCR complex. Baseline levels of

phospho-FAK are low in unstimulated cells, and levels of total FAK remain constant upon FAK inhibition by PF-578223 (Figure 29, lane 1 and 2). Upon activation for 15 minutes, levels of phospho-FAK dramatically rise, but are abolished when pre-treated with PF-578223 (Figure 29, lane 3 and 4). This shows that FAK is a target of BCR activation, but its place in the signaling pathway is not known.

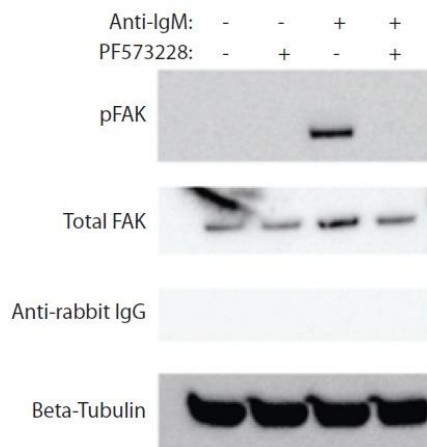


Figure 29: Effects of BCR activation on FAK. Western blot of immunoprecipitation of FAK by anti-FAK antibody (Total FAK) and reprobe with phospho-tyrosine antibody (pFAK) at ~250kDa from RAJI cells. Anti-Rabbit IgG control for non-specific immunoprecipitation. Beta-tubulin control (~50kDa). Cells with Anti-IgM+ were treated with Anti-human-IgM for 15 minutes. Cells with PF-573228+ were treated with drug for 2 hours pre-exposure to IgM. Cells with Anti-IgM- were given equal volume vehicle (PBS) and cells with PF-573228- were given vehicle control (DMSO) in equal volume.

In order to interrogate where in the pathway FAK lies, the canonical growth effectors were measured with FAK inhibition. Using DMSO as a carrier control, normal induction of ERK and AKT, here measured by phospho-ERK and phospho-AKT, is shown. Inhibition of FAK by PF-573228 shows not only reduction of ERK and AKT

induction by BCR activation, but inhibition of baseline pERK and pAKT (Figure 30B). This places FAK in the BCR pathway, with downstream growth and survival signaling being dependent on FAK activation. To further refine where in the signaling cascade FAK lies, levels of phospho-SYK were measured with FAK inhibition, as SYK lies at the intracellular surface of the BCR. Upon activation, pSYK levels rise, but with FAK inhibition, pSYK levels remain constant when activated (Figure 30A). This suggests that FAK lies at the intracellular surface in conjunction with SYK and LYN. FAK is known to associate with SRC, which is in the same family as SYK, suggesting there may be additional FAK functionality with other members of SRC-like kinases.

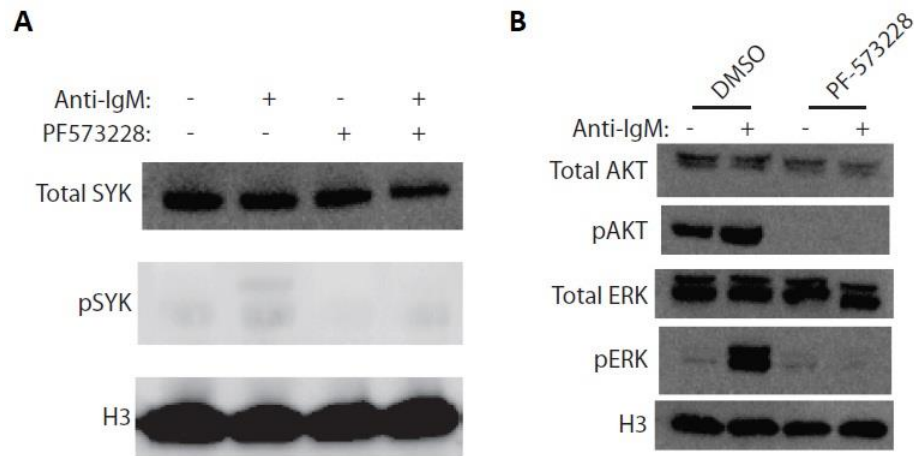


Figure 30: FAK inhibition reduces downstream BCR signaling. Western blot of RAJI cells treated with PF-573228 for 2 hours (+) or vehicle control (-, DMSO), and then treated with anti-human-IgM (+) or vehicle (-, PBS) for 15 minutes. **(A)** Samples probed for total SYK (72kDa), phospho-SYK (pSYK, 72kDa), and histone H3 (H3, 15kDa). **(B)** Samples probed for total AKT (~60kDa), phospho-AKT (pAKT, ~60kDa), total ERK 1/2 (Total ERK, ~42/44 kDa), phospho-ERK 1/2 (pERK, ~42/44 kDa), and histone H3.

In order to confirm that these effects were linked to FAK and not off target effects, a knockdown line of *FAK* in the same cell type using shRNAs against *FAK* were developed (Figure 31A). Knockdown of *FAK* showed similar results, with reduced pAKT and pERK upon activation with anti-IgM (Figure 31B). Immunofluorescence of activated cells also shows formation of BCR complexes, with FAK spread across the cytoplasm (Figure 31C).

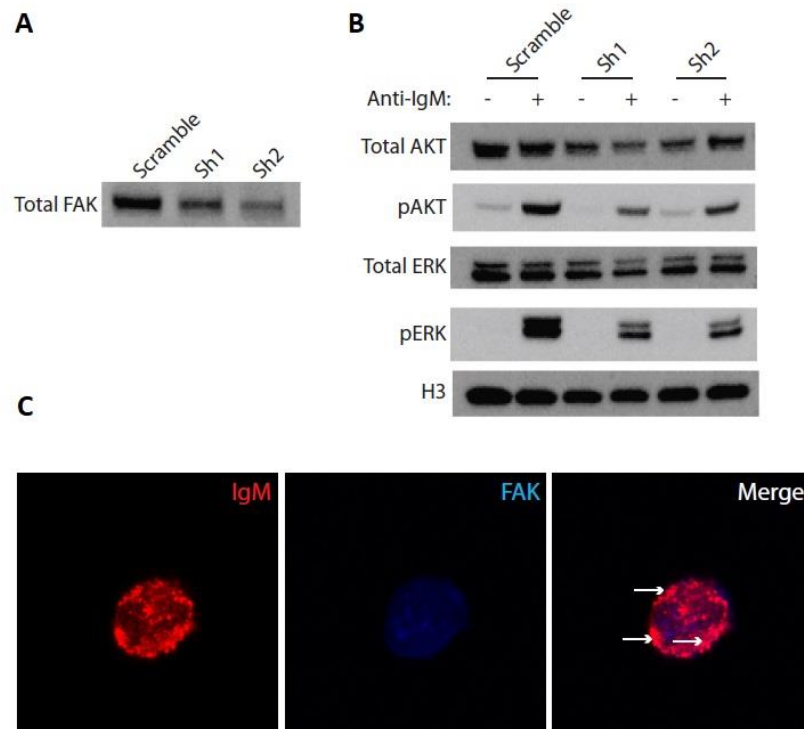


Figure 31: FAK knockdown reduces BCR signaling. FAK knockdown by 2 shRNA constructs (Sh1 and Sh2) and a scramble control (Scramble) in RAJI cells. **(A)** Western blot for total FAK (~125kDa). **(B)** Samples treated with anti-human-IgM for 15 minutes (+) or vehicle control (PBS, -). Blotted for Samples probed for total AKT (~60kDa), phospho-AKT (pAKT, ~60kDa), total ERK 1/2 (Total ERK, ~42/44 kDa), phospho-ERK 1/2 (pERK, ~42/44 kDa), and histone H3. **(C)** Confocal microscopy of immunofluorescent staining of FAK knockdown RAJI cells. Staining of IgM (red) and FAK (blue) images merged (right). BCR aggregated SMAC complexes shown with white arrows.

Together, these data show a clear role for *FAK* in germinal center derived B cells. *FAK* lies in the B cell receptor pathway, likely associated with SYK and LYN at the intracellular surface of the BCR complex. Inhibition, both chemical and genetic, reduces downstream growth and survival signaling. This novel role of *FAK* reveals a new way in which BCR activation may lead to other downstream signaling as well.

4.3.2 *In Vivo* mouse models of *Fak* loss

In order to understand the role of *Fak* in the context of normal B cell function, a mouse model was developed with knockout of *Fak* (*Ptk2*) specifically in germinal center B cells. A breeding schematic can be found in figure 32. Briefly, mice with *Lox* sites flanking *Fak* were bred with mice with *C γ -Cre*. *C γ -Cre* is a Cre recombinase under the promoter of *C γ* immunoglobulin. This drives a tissue specific recombination of *FAK* at the *LoxP* sites only in germinal center B cells. Further differentiated B cells, such as plasma cells and memory B cells, would also carry this mutation. Knockout mice are healthy and age appropriately and do not show an increase in infection or other disease.

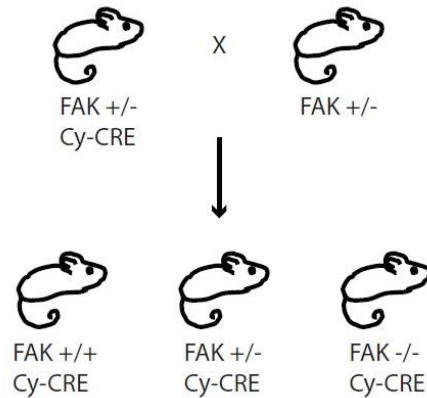


Figure 32: Mouse breeding schematic for *Fak* knockdown in GCB cells. “+” represents wildtype allele, “-” represents floxed allele. “Cy-Cre” represents Cre-recombinase under the Cy-immunoglobulin promotor.

Phenotypically, these mice show a similar result as the in vitro data in all post GCB cells and in cell types that differentiate from them. Normal mouse spleen when stimulated with antigen shows normal boundaries with clear germinal centers formed in the interior (Figure 33A). *Fak* knockout spleens show cell death near the outer edge of the spleen, with disorganized and disrupted germinal centers within the organ (Figure 33B). This reduction may be due to improper association with dendritic cells and signaling through the BCR once these cells undergo class switch, activating the *C γ -Cre*. Further characterization of the cells present in the spleen shows a reduction in the percentage of GC B cells in both *Fak* heterozygous and *Fak* knockout mice (Figure 33C). To elucidate whether *Fak* knockout in GC B cells has consequences for further differentiated cells, levels of serum immunoglobulin were measured. Knockout mice had reduced levels of both IgG and IgM, suggesting that plasma cells that differentiate from the GC B cells are impaired and do not produce the appropriate amount of serum

immunoglobulin (Figure 33D-E). This phenotype is independent of the number of plasma cells, which remains constant. Together, this data establishes a similar pattern of GC B cell death as with the *in vitro* data when FAK is inhibited, either chemically or genetically.

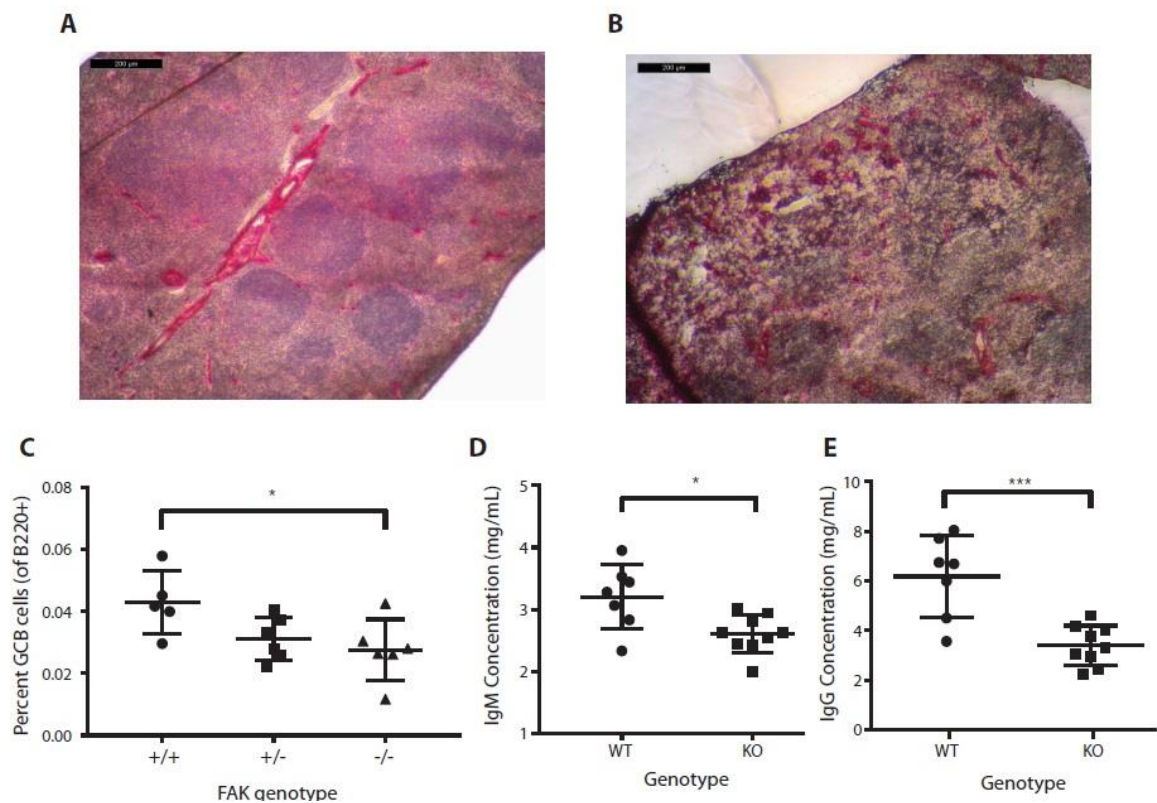


Figure 33: *In vivo* knockout of *Fak* in GCB cells. Hematoxylin and eosin staining of (A) wildtype mouse spleen and (B) *Fak* knockout mouse spleen. (C) Percent of GCB cells (B220⁺ FAS⁺ CD38^{low} CD4^{low}) as a percent of total B cells (B220⁺) from wildtype (+/+), heterozygous (+/-) and knockout (-/-) mouse splenic isolates. Serum levels of (D) IgM and (E) IgG in wildtype (WT) and *Fak* knockout (KO) mice measured by ELISA.

4.4 Discussion

FAK has long been known to be involved in integrin signaling, cell polarization, and related growth, but this is usually focused on solid tissues and tumors of endothelial origins. The *FAK*-Src axis of signaling is well understood in this context, with α/β integrins physically associated with *FAK*, which in turn associate with Src and other growth factors. Cells that do not attach and have these tonic signals undergo anoikis and begin to rapidly die. B cells, however, are not dependent on this signaling, yet *FAK* levels are high in GC B cells. Placing *FAK* in the context of GC B cells suggests the involvement of other pathways, and here we have placed it in the B cell receptor pathway. While this may seem like a large departure from normal function, there is evidence that *FAK* is physically associated with LYN and the associated kinase, SYK, is a close family member of Src kinase.

A sparse schematic of *FAK* in the B cell receptor pathway can be found in figure 34. The placement of *FAK* in the BCR signal transduction pathway itself is shown by the abolishment of downstream growth signals in AKT/ERK activation. It also was shown to be an early membrane/BCR associated effector shown by blocking SYK activation as well as its previous physical association with LYN [144]. Blocking these signals either chemically or genetically will block all downstream signals even at the intracellular surface of the BCR.

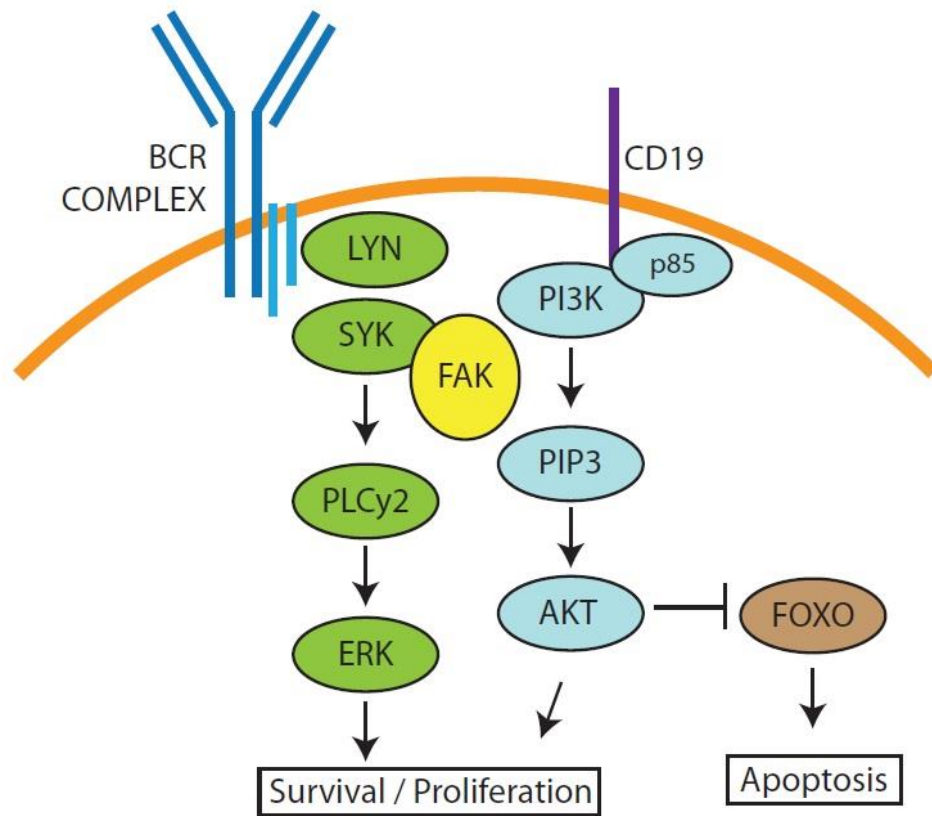


Figure 34: Proposed model of FAK involvement in BCR signaling. BCR complex (dark blue) attached in plasma membrane (orange). CD19 (purple) also attached in membrane. MAPK pathway (green) and AKT pathway (light blue) lead to growth and proliferation. Apoptotic genes (brown) blocked by activated signal transduction. FAK (yellow) associated with LYN-SYK complex near BCR intracellular interface.

Together, this data clearly places FAK in the B cell receptor pathway. FAK is necessary to basic functions of the signaling cascade and necessary for the activation of downstream processes, such as growth and proliferation. This makes FAK an attractive target for the development of therapies for malignancies that arise from germinal center B cells, such as GCB diffuse large B cell lymphoma and Burkitt lymphoma. The BCR and the underlying stimulatory pathway are attractive targets, as B cells are reliant on this

tonic signaling and are often mutated to keep this signaling active. CAR-T therapies leverage this, using CD19, a BCR associated surface protein, to target these CD19+ malignant cells. Other factors also point to a utility for a FAK inhibitor in DLBCL. Mutational profiles (see: chapter 3) show that FAK is rarely mutated in DLBCL, and unlike many other targets, would be more easily targetable as it is in the normal conformation. Also, a large number of targeted inhibitors have been developed for FAK for use in solid tumor metastatic disease. These compounds exist across the development cycle with some being through the safety testing for clinical use, and thus these drugs may be quickly repurposed for GCB derived malignancies.

5. Gene Expression Perturbations upon Drug Treatment Enable Identification of Resistance Mechanisms and Novel Combination Therapies in Lymphoma

5.1 Introduction

Identification of novel combination therapies has long been a goal of small molecule screening in cancer. Combination therapies allow for lower dose single agent treatment [145], while preventing single event resistance formation [146]. Much of the work in this area has followed phenotypic and mechanistic finding of resistance in specific cancer contexts, such as PARP inhibition in CHK1 inhibition-resistant cancer [147]. Such methods are bound by known interaction networks and drug activity, and they are often not generalizable outside of the context in which they are tested.

Traditional combinatorial approaches based on cell viability assays contain so many variables (dosage, number of cell lines, subtypes modeled, number of drugs, etc.) that the matrix of possibilities expands to untenable levels of assay numbers. Most studies of this type remain small, while limiting key variable that explore breadth of activity or explore compounds with less specified activity [148]. The revolution in genomics technology has added a host of powerful tools that can assay cell state in parallel, yet the fundamental relationship between cell state, small molecule perturbation, and disease remains obscured.

Many efforts have been put forward to try to understand this relationship between disease, therapeutic agents, and combination therapies. The Connectivity Map

used gene expression microarray data to create signatures for both disease and drug treatment [149], finding correlations between drug activity and expression were discovered that placed drugs of varying structure and disease target into functional categories based on gene expression. Other studies using computational methods to identify compound synergy in high-throughput have identified previously known interactions, but lack rigorous validation of new targets [150,151]. The largest effort to date to identify combination therapeutics using gene expression paired drug exposure data was the NCI-DREAM initiative [152]. 31 teams submitted methods that would predict synergy of a single dataset of 14 drugs on a single cell line. The top performing model used combination of similarity between drugs and dose response to predict combinations. The top performing algorithm had only 19% precision, so there is room for improvement of these tools.

The overall study design for defining single agent effectiveness, gene by perturbant gene expression, and predicting combination synergy is shown in figure 35. In order to more broadly capture disease heterogeneity of both mutational spectrum and disease subclass 6 cell lines were used from both ABC and GCB subtypes of DLBCL. We then defined single agent effectiveness across cell lines and performed gene expression analysis on each cell line versus compound combination. Here we found novel mechanisms of both resistance and sensitivity to single agents, and we found stratification of single agents based on their gene expression profiles. We then developed

a method based on reversal of disease gene expression (RDGE) that combines single agent data to predict possible drug combinations. We validated some these combinations *in vitro*, and moved one combination (panobinostat with ruxolitinib) to a successful *in vivo* model, showing both synergy of the combination and low toxicity. The findings of this work suggest cRDGE is a valuable model for using gene expression data to predict dual drug synergy with high precision.

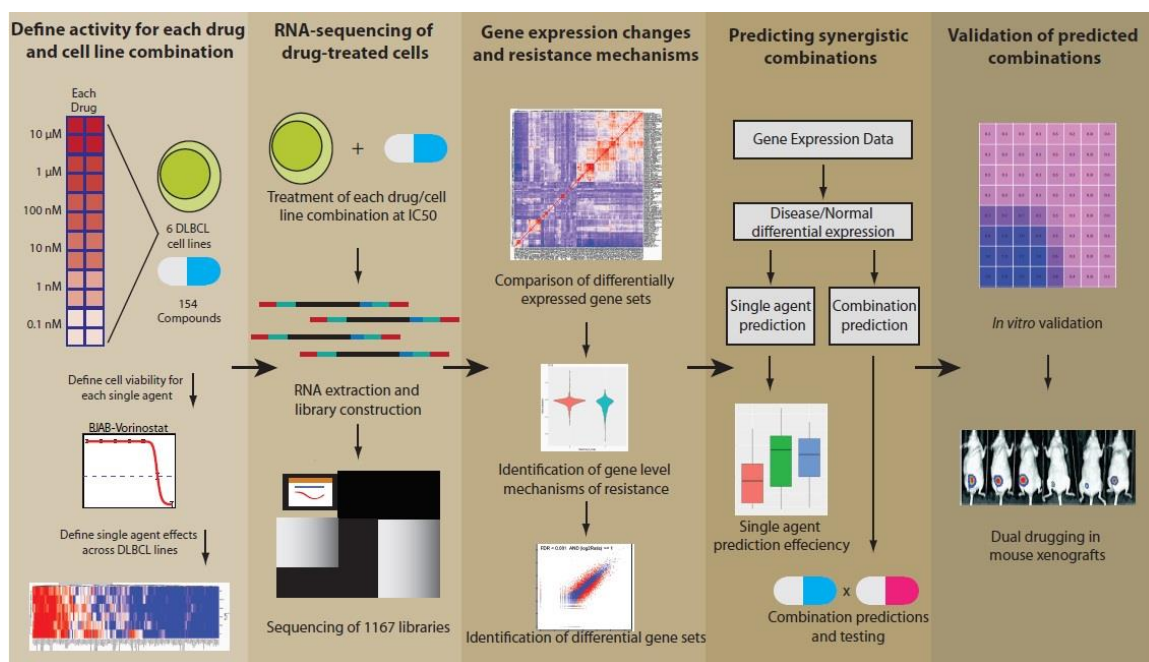


Figure 35: Study design expression-based predictions of synergy. (Left) Defining activity for each drug by cell line combination. Exposure of all 6 cell lines to 154 FDA-approved compounds and defining single-agent effects. (Second from left) RNA sequencing of each drug by cell line combination. Drugs are treated at IC₅₀, RNA is extracted, sequencing libraries are made from this RNA and libraries are sequenced. (Middle) Gene expression changes are measured. Comparison of gene expression changes by drug yields mechanism of resistance and synergy for single agents. (Second from right) Predicting synergistic combinations. Gene expression changes are developed in to a model based on reversal of disease gene expression to predict synergy between compounds. (Right) Validation of predicted combinations. Predicted combinations are validated *in vitro* and in *in vivo*.

5.2 Methods

5.2.1 Cell lines and culture

OCI-LY3 were grown in RPMI 1640 with 20% HI-Fetal bovine serum and 1% Penicillin-Streptomycin (5,000 U/mL. Pfeiffer, SU-DHL-4, BJAB, TMD8, and HBL-1 were grown in RPMI 1640 with 10% HI-Fetal bovine serum and 1% Penicillin-Streptomycin (5,000 U/mL). All cell lines were grown at 37C with 5% CO₂. OCI-LY3, TMD8, SU-DHL-4, Pfeiffer, and HBL-1 are derived from males, and BJAB is derived from a female. Cell lines were authenticated through STR sequencing from the Duke University DNA Analysis Core Facility.

5.2.2 Compounds

Small molecule and chemotherapeutic compounds (drugs) were dissolved in DMSO and stored at -20 C in small batches to avoid water contamination. Compounds were stored at 10mM or the highest concentration to 10mM as solubility allows. Cytokines were dissolved in water to the highest concentration solubility will allow. Cytokines were prepared fresh before samples were treated. A full list of all perturbants and can be found in table 4.

Table 4: Compounds

Drug #	Compound	Drug #	Compound	Drug #	Compound
1	Hydroxyurea	52	Floxuridine	103	Paclitaxel
2	Dacarbazine	53	Cytarabine hydrochloride	104	Vinorelbine tartrate
3	Azacitidine	54	Dexrazoxane	105	Melphalan hydrochloride
4	Streptozocin	55	Nelarabine	106	Etoposide
5	Vorinostat	56	Megestrol acetate	107	Trametinib
6	Mitomycin	57	Axitinib	108	Vinblastine sulfate
7	Fludarabine phosphate	58	Raloxifene	109	Temsirolimus
8	Erlotinib hydrochloride	59	Afatinib	110	Abiraterone
9	Topotecan hydrochloride	60	Ixabepilone	111	Tamoxifen citrate
10	Regorafenib	61	Thiotepa	112	Dabrafenib mesylate
11	Allopurinol	62	Methoxsalen	113	Vincristine sulfate
12	Arsenic trioxide	63	Thalidomide	114	Imiquimod
13	Decitabine	64	Pentostatin	115	Crizotinib
14	Cladribine	65	Pomalidomide	116	Epirubicin hydrochloride
15	Exemestane	66	Bendamustine hydrochloride	117	Teniposide
16	Mitotane	67	Mitoxantrone	118	Sirolimus
17	Capecitabine	68	Ibrutinib	119	Plerixafor
18	Olaparib	69	Pralatrexate	120	no drug
19	Dasatinib	70	Romidepsin	121	Nintedanib
20	Vemurafenib	71	Aminolevulinic acid hydrochloride	122	Ruxolitinib
21	Fluorouracil	72	Lomustine	123	PKC412
22	Temozolomide	73	Procarbazine hydrochloride	124	Panobinostat
23	Carmustine	74	Gemcitabine hydrochloride	125	Sonidegib
24	Ifosfamide	75	Chlorambucil	126	Masitinib
25	Anastrozole	76	Carboplatin	127	Palbociclib
26	Clofarabine	77	Gefitinib	128	Thapsigargin
27	Celecoxib	78	Methotrexate	129	Entospletinib
28	Pemetrexed disodium salt	79	Enzalutamide	130	BGJ398
29	Pazopanib hydrochloride	80	Omacetaxine mepesuccinate	131	AMG337
30	Vandetanib	81	Cabozantinib	132	Volitinib
31	Thioguanine	82	Estramustine phosphate sodium	133	E7820
32	Busulfan	83	Valrubicin	134	Guadecitibine
33	Cyclophosphamide	84	Everolimus	135	ENMD-2076
34	Cisplatin	85	Triethylenemelamine	136	ARN-509
35	Letrozole	86	Ponatinib	137	LY2228820
36	Belinostat	87	Lapatinib	138	BVD-523
37	Sunitinib	88	Carfilzomib	139	AZD-1208
38	Vismodegib	89	Dactinomycin	140	Selinexor
39	Imatinib	90	Amifostine	141	PF-03084014
40	Idarubicin hydrochloride	91	Bosutinib	142	Volasertib
41	Mercaptopurine	92	Ceritinib	143	Defactinib
42	Altretamine	93	Docetaxel	144	Venetoclax
43	Uracil mustard	94	Plicamycin	145	EPZ-6438
44	Tretinoin	95	Oxaliplatin	146	Galunisertib
45	Lenalidomide	96	Daunorubicin hydrochloride	147	LCL161
46	Pipobroman	97	Irinotecan hydrochloride	148	Veliparib
47	Bortezomib	98	Cabazitaxel	149	Elesclomol
48	Idelalisib	99	Bleomycin sulfate	150	LY2606368
49	Sorafenib	100	Zoledronic acid	151	VTX-2337
50	Nilotinib	101	Doxorubicin hydrochloride	152	Filanesib
51	Mechlorethamine hydrochloride	102	Fulvestrant		

5.2.3 Dose response curves

Drugs were plated in a 384-well plate using the LabCyte Echo 550 acoustic dispensing system. Drug was dispensed in a range of final concentration from 10 μ M to 0.1nM in 10-fold dilutions, leading to 6 concentrations (10 μ M, 1 μ M, 100nM, 10nM, 1nM, and 0.1nM) with a volume of <1% of assay volume. Each well was then backfilled with DMSO to a final concentration of 1% carrier. A control well for live cells was used with the 1% DMSO. A control well for dead cells was plated at 10% DMSO by volume. Each concentration and control were plated in quadruplicate. Healthy OCI-LY3 and SU-DHL-4 cells were diluted to a concentration of 2.5×10^6 cells/mL. 25 μ L of these cells were then plated and mixed into each well of drug and control. Cells were incubated for 48 hours at 37°C in a cell culture incubator. 2 hours before the incubation was complete, 2.5 μ L of Alamar Blue reagent was added to all wells. Cells were then returned to the incubator for 2 hours. Cells were then measured for fluorescence on a Tecan Infinite 200 PRO microplate reader at 560 excitation and 590 emission. Dead control was subtracted from all wells as background fluorescence

5.2.4 Drug treatment and RNA-seq library construction

All DMSO-dissolved compounds were plated in plated at their respective IC₅₀ concentration for each cell line and compound pair. Compounds that did not have an effect on cell viability were dosed at 10 μ M. Plating was performed with the LabCyte Echo 550 acoustic dispensing system. Compounds were plated into round bottom 96-

well cell culture plates. 100uL of cells were plated in appropriate growth media at 1×10^6 cells/mL. For drugs, cells were treated for 6 hours at 37C. For cytokines, cells were serum starved for 2 hours, then cytokine was added and incubated for 2 hours. DMSO only and serum starved controls were added to each plate for each cell line. All samples were centrifuged at 300g for 6 minutes, and placed on ice. Remaining media was aspirated, and cells were lysed in Lysis Buffer (Zymo).

RNA was extracted using the Zymo Quick-RNA ZR-96 RNA extraction kit. Representative samples were taken and measured for RNA content and integrity with the Bioanalyzer 2100 RNA Nano Kit. RNA was normalized across all sample plates for input into library construction.

RNA sequencing libraries were constructed for all cell line and perturbant combinations. Samples were processed using the KAPA stranded mRNA seq kit. Briefly, RNA was mRNA selected using poly-T beads. RNA was then fragmented at 95C for 6 minutes. Then cDNA was synthesized, and A overhangs were added. Custom adapters were ligated to the ends of each molecule (See: Chapter 2). Each adapter contains sequencing primers, and each molecule was tagged with an 8-mer unique barcode sequence for multiplexed sequencing. Samples were pooled in equimolar amounts and sent for sequencing at the HudsonAlpha institute for Biotechnology. Samples were sequenced on the HiSeq 2000.

5.2.5 Combination predictions

Data was compiled for several publicly available datasets from previous studies as well as all data produced in this study. Expression and single agent sensitivity data was used from the DREAM dataset [152]. Normal tissue gene expression data from Zheng et al. (Blood 123, no. 19. May 2014) was also used for comparison.

Reversal of disease gene expression analysis was performed as previously described in Chen et al. and related publications [153,154]. Briefly, a signature for drug treatment was created using gene expression comparisons between perturbant treated versus vehicle control treated. Tumor versus normal gene expression signatures were computed using tumor expression for each cell line versus that of the normal B cell tissue.

Calculations for reversal of disease gene expression was calculated similarly to Chen et al [153]. Briefly, gene were ranked by expression by drug signature. Enrichment scores for each up and down gene was computed by rank list. Absolute enrichment was then calculated for each up and down-regulated for each profile. Then RGEN is estimated as: $RGEN = (ExpressionSig)UP - (ExpressionSig)DOWN$.

A linear model based on pairs of drugs was developed to find combination synergy. Given a matrix of gene-expression changes (fold-change), Delta, we computed how well a pair of drugs T1 and T2 predict normal (N) vs disease (D) expression using the following linear model:

$$\Delta_{N/D} \sim \Delta_{T1/ctrl} + \Delta_{T2/ctrl}$$

Goodness of fit was computed as $R^2_{T1,T2}$. Separate linear models were also computed for T1 and T2. Predicted synergy for the combinations was computed as the difference between combination R^2 and the maximum of the single agent R^2 s, using the following model:

$$\text{Predicted synergy}_{T1,T2} = R^2_{T1,T2} - \max(R^2_{T1} - R^2_{T2})$$

5.2.6 Combination screening

Compounds were plated into 384-well plates with the LabCyte Echo 550 acoustic dispensing system. One of the compounds is across the plate, with concentration in log2 dilution down the plate (highest concentration in row A). The other compound is plated across the plate with log2 dilution from left to right (highest concentration in column 1). All pairwise combinations are plated in quadruplicate. A carrier control (1% DMSO) and dead control (10% DMSO) were also plated on each plate. Cells were then added and mixed to each well at 2.5×10^5 cells per mL. Cells were incubated for 48 hours at 37°C in a cell culture incubator. 2 hours before the incubation was complete, 2.5uL of Alamar Blue reagent was added to all wells. Cells were then returned to the incubator for 2 hours. Cells were then measured for fluorescence on a Tecan Infinite 200 PRO microplate reader at 560 excitation and 590 emission. Average fluorescence for the dead control was subtracted from all wells as background fluorescence.

5.3 Results

5.3.1 Single agent screening reveals sensitive and resistant drug classes in DLBCL

We first sought to determine the effectiveness of a panel of 152 FDA-approved compounds (Table 4) on 6 human DLBCL cell lines (OCI-LY3, TMD8, SU-DHL-4, Pfeiffer, BJAB, and HBL-1). These compounds included broad chemotherapeutics as well as targeted inhibitors. Cell viability testing of these drug/cell line pairs shows a wide range of single agent activity across the cell panel (Figure 36A). The compounds fall into 3 main classes: pan-effective, selective, and resistant drugs. Pan-effective drugs included DNA damage agents (e.g. doxorubicin) as well as histone deacetylase inhibitors (e.g. belinostat). Selective drugs were found to be effective in some cell lines and resistant in others. Targeted agents against the same pathway, such as the mTOR inhibitors everolimus, temsirolimus, and sirolimus, showed similar effectivity within the same cell line, suggesting a resistance mechanism to pathway inhibition in this cell line. Certain drugs were also only effective in one subtype of DLBCL, such as BTK inhibition by ibrutinib in the GCB subtype (Fig 36B). Additionally, about one-third of the drugs screened showed no viability response in any DLBCL line.

Figure 36: Single agent effects reveal classes of effective drugs in DLBCL. (A) Heatmap of all drugs (left) by cell lines (bottom). Color of interaction indicates cell viability. Cell lines used include OCI-LY3 (L), TMD-8 (T), BJAB (B), Pfeiffer (P), SU-DHL-4 (S), and HBL-1 (H). ABC cell lines are in orange. GCB cell lines are in green. Drug categories (right) include Pan-effective (red), selective (yellow) and resistant (blue). **(B)** Dose response curves of pan-effective vorinostat (red) and selective ibrutinib (yellow), against all cell lines (right).

5.3.2 Gene expression changes reveal molecular mechanisms of sensitivity and resistance in DLBCL

Previous studies examining effects of expression across a drug screen have used either a single or a small number of fixed concentrations that are not unique to each compound. Here, we examined expression profiles of compound/cell line pairs that were exposed to compounds at their determined IC50 concentration. Pairwise correlation of gene expression changes across cell lines for each drug-drug pair show distinct patterns of correlated and anti-correlated expression based on the class of compound used (Figure 37). DNA damage agents clustered together with agents that had no cell viability effect, suggesting these compounds do not have a large effect on overall expression in the cell. Subgroups of correlated expression emerged for targeted inhibitors to both similar targets and pathways. All four HDAC inhibitors (panobinostat, vorinostat, belinostat and romidepsin) had very tightly correlated gene expression after treatment, even though the effects of HDAC inhibition are genome wide. Similarly, drugs targeting different nodes in the same pathway, like mTOR and PIK3CD inhibitors cluster together, in spite of having very different IC50 values.

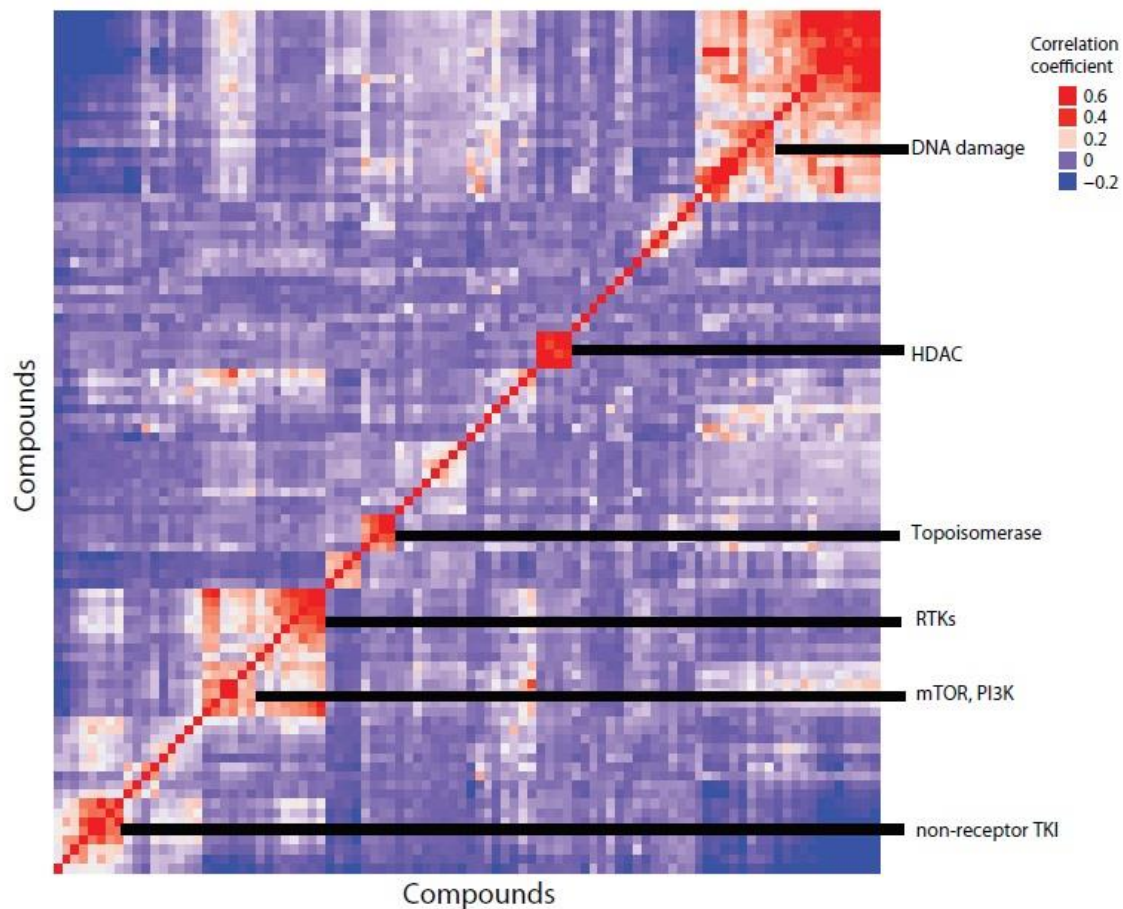


Figure 37: Compound gene expression signatures reveal classes of drugs. Heat map of all compound by compound gene expression correlations. Color determined by correlation coefficient from most correlated (red) to least (blue). Compounds were then clustered based on relatedness to each other.

We also sought to investigate the relationship between gene expression and resistance mechanisms. While expression of no genes were singly predictive of either resistance or sensitivity, sensitive drugs significantly reduced *MYC* expression compared to resistant cell line expression (Fig 38A). We also found that gene expression changes in the target of the drug may also be predictive of efficacy. In sensitive cell lines,

treatment with the inhibitor ibrutinib reduced expression of its target, *BTK* (Fig 38B). In resistant cells, expression of *BTK* increased upon treatment. We also observed a relationship between targeted inhibitors and expression of their target. Expression of *BRAF* was directly related to the efficacy of the inhibitor of BRAF, vemurafenib (Fig 39C). Higher expression of *BRAF* correlates with a higher A_{max} for the inhibitor, and vice versa.

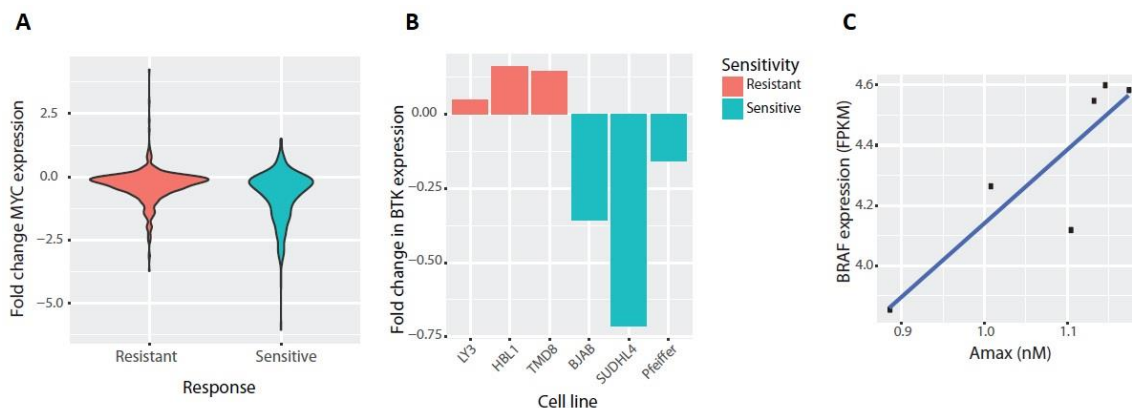


Figure 38: Gene expression reveals mechanisms of resistance and sensitivity. (A) Violin plot of fold-change in MYC expression in all resistant (pink) versus sensitive (teal) interactions across all drug by cell line combinations ($p < 0.05$). **(B)** Fold change in *BTK* expression across cell lines from upon treatment with BTK inhibitor (ibrutinib). **(C)** BRAF expression in untreated cell lines versus A_{max} of vemurafenib.

5.3.3 Predicting Synergistic Combinations Using Expression and Activity Data

We next sought to determine if expression-based signals from single agent expression and activity data could be used to predict synergy in combination drugging. We created a method for using disease-based gene expression compared to normal and using that to predict single agent and combination effectiveness. The outline of the model can be seen

in figure 39. Briefly, this model uses gene expression data of drug treated cell versus their untreated control and compares that to normal cell expression to find a reversal of the disease gene expression. This reversal of disease gene expression (RDGE) was then be used to make predictions about resistant and sensitive single agents which align with the dose response data collected in the DREAM study (Fig 40A $p < 0.01$).

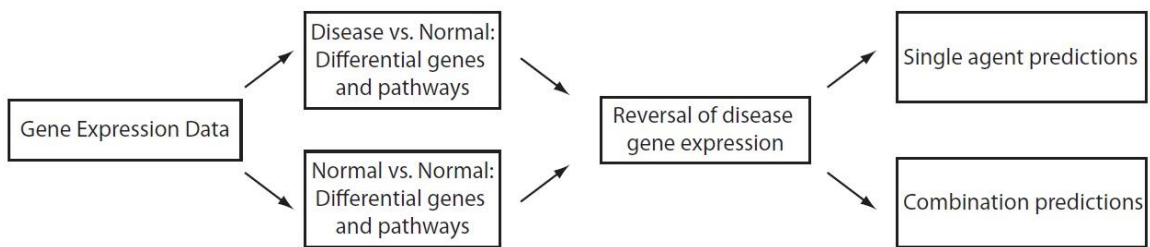


Figure 39: Schematic for cRDGE model

We then extended this method to develop a method we are calling combination reversal of disease gene expression (cRDGE). This method creates a linear model that will give a numerical prediction score that can be used to rank possible synergistic combination of drugs. We then applied this method to the DREAM data again, testing for possible synergies. This method outperformed previous efforts from the DREAM challenge teams, with a precision of 75% in our top scoring predictions (Fig 40B).

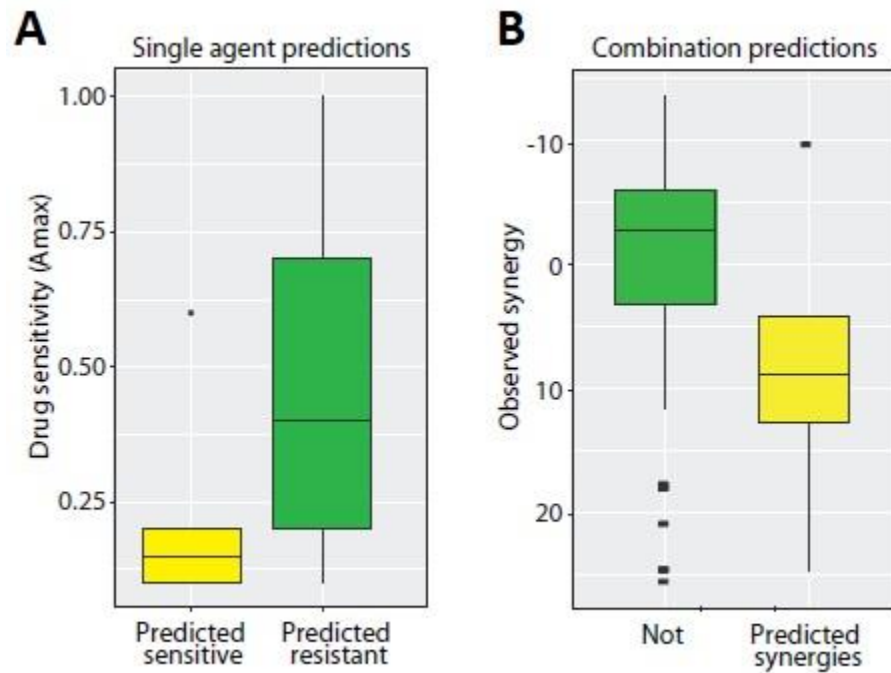


Figure 40: Validation of cRDGE model in DREAM data. (A) Drug sensitivity (A_{max}) of predicted sensitive and predicted resistant drug interactions based on expression in cRDGE model ($p < 0.01$) **(B)** Observed synergy scores from DREAM dataset with synergy predicted by cRDGE model.

The method proved to be predictive in other datasets, but we next sought to empirically validate these results within our own dataset. We first used top scoring therapeutic combinations in the cRDGE method and tested them using dual dose response combination plating. We observed synergy between vincristine sulfate and filanesib (Figure 41A), as well as between gemcitabine hydrochloride and dactinomycin (Figure 41B). We also observed synergy between the kinesin spindle protein (KSP) inhibitor filanesib and epidermal growth factor receptor (EGFR) inhibitor LY2606368

(Fig 41C). Further analysis shows that synergy is strong by both Bliss modeling (Figure 41D) and highest single agent (HSA) synergy (Figure 41E) and, with synergy scores of 4.084 and 6.902 respectively.

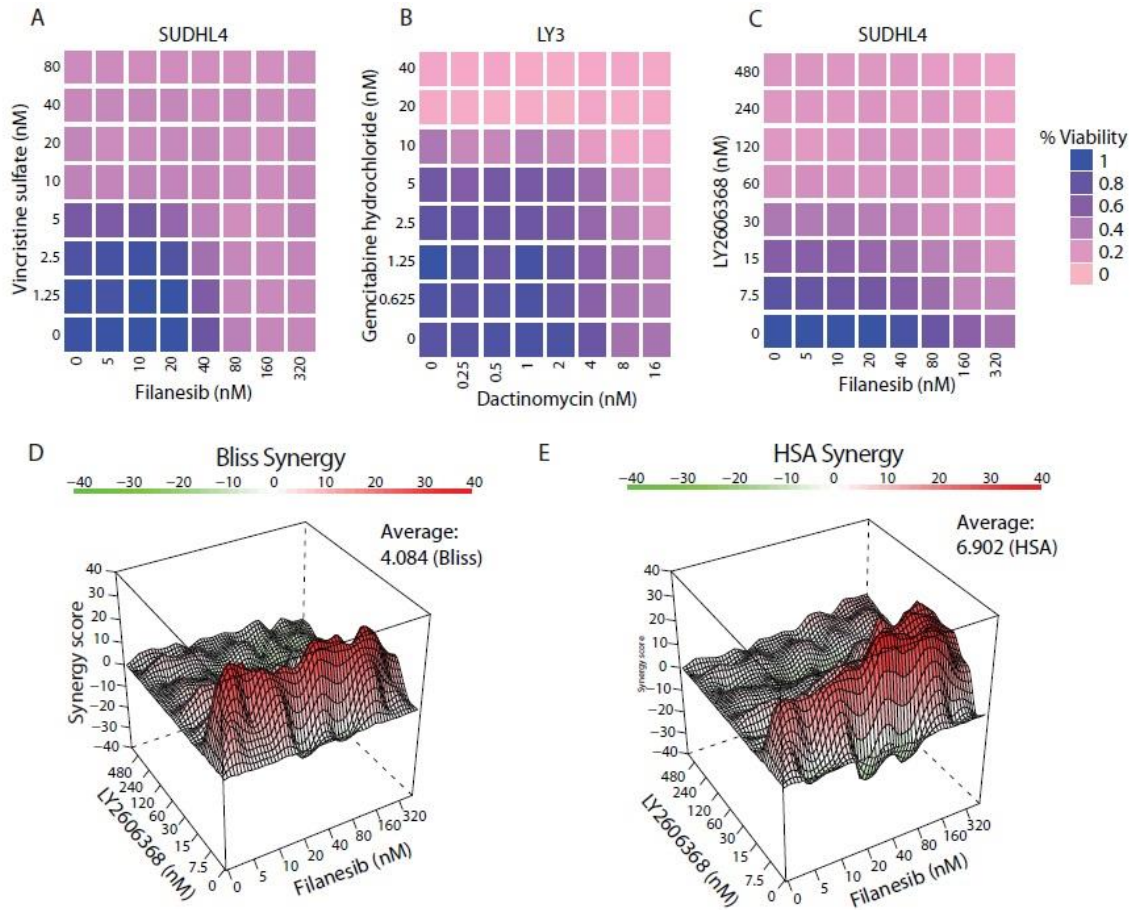


Figure 41: Figure 42: Validation of synergy of cRDGE predicted combinations. (A-C) cell viability plots of 2-way dose response to indicated cell line and drug combination. Color represents cell viability. **(D)** Bliss synergy and **(E)** HSA synergy of filanesib combined with LY2606368 (Bliss = 4.084, HSA = 6.902).

5.3.4 HDAC and JAK/STAT combination therapy

One of the highest cRDGE synergy scores was for the combination of HDAC inhibition (panobinostat, vorinostat, belinostat, or romidepsin) and JAK/STAT inhibition by ruxolitinib. The most effective combination was panobinostat and ruxolitinib, which will be the combination used going forward. We observed combination synergy in all 6 cell lines with panobinostat and ruxolitinib (Figure 43), with synergy scores 2.1 to 6.4 by Bliss modeling, 3.4 to 5.3 for HSA models, and 4.9 to 7.7 for ZIP models.

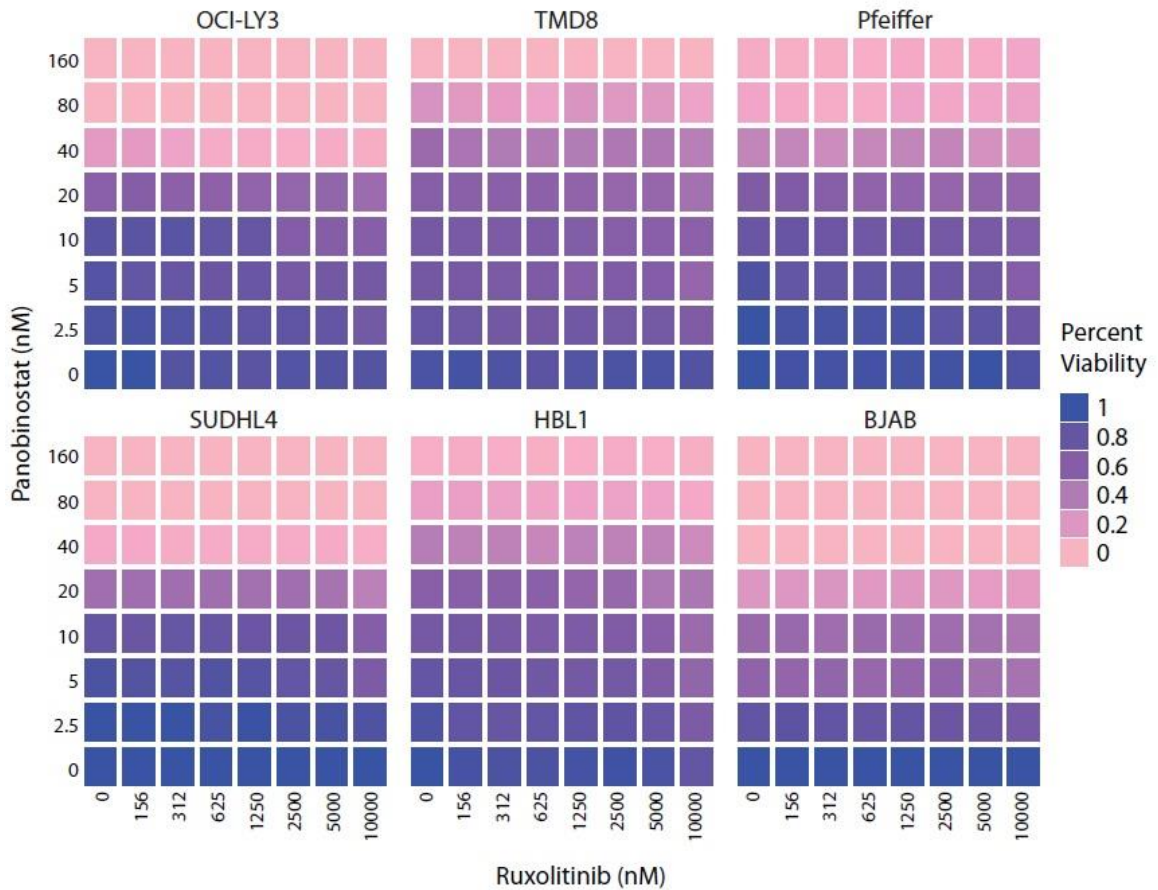


Figure 43: Validation of panobinostat and ruxolitinib synergy *in vitro*: cell viability plots of 2-way dose response to indicated cell line and drug combination indicated. Color represents cell viability.

Next, we examined the effects of combination drugging on STAT signaling.

Testing of one in-group cell line (OCI-LY3) and one cell line outside of the original dataset (Farage), we found that at high dosage, panobinostat alone will reduce STAT signaling near the IC50 of the cell (Figure 44). In combination with low dose of ruxalitinib, cells are sensitized, and lower doses of panobinostat are required to reduce

STAT signaling. Noticeably, ruxolitinib alone does not ablate STAT signaling alone at 100nM, and it must be augmented with HDAC inhibition to begin to lower pSTAT3.

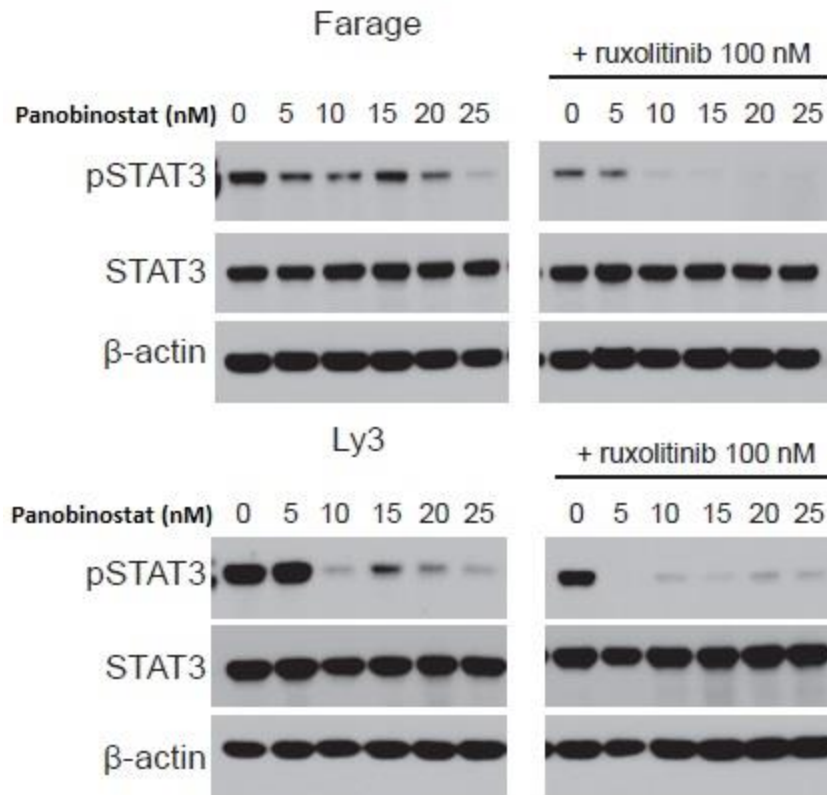


Figure 44: Sensitizing STAT inhibition with panobinostat and ruxolitinib synergy. Western blots for Farage (top) and OCI-LY3 (bottom). Blots have dose increase for panobinostat indicated for all blots. Ruxolitinib was added at a single dose (100 nM) to right blots. Blots measured phospho-STAT3 (pSTAT3), total STAT3 (STAT3), and beta actin.

While single agent sensitivity for ruxolitinib is low, panobinostat is effective in all 6 cell lines tested. To extend this examine broad sensitivity across B cell lymphomas, we extended the scope of the single agent activity testing of panobinostat on a panel of 56 cell lines that covered a wide range of B cell lymphomas (Figure 45). All cell lines

tested had IC50 values below 110nM, with 96% of cell lines have an IC50 value below 30nM.

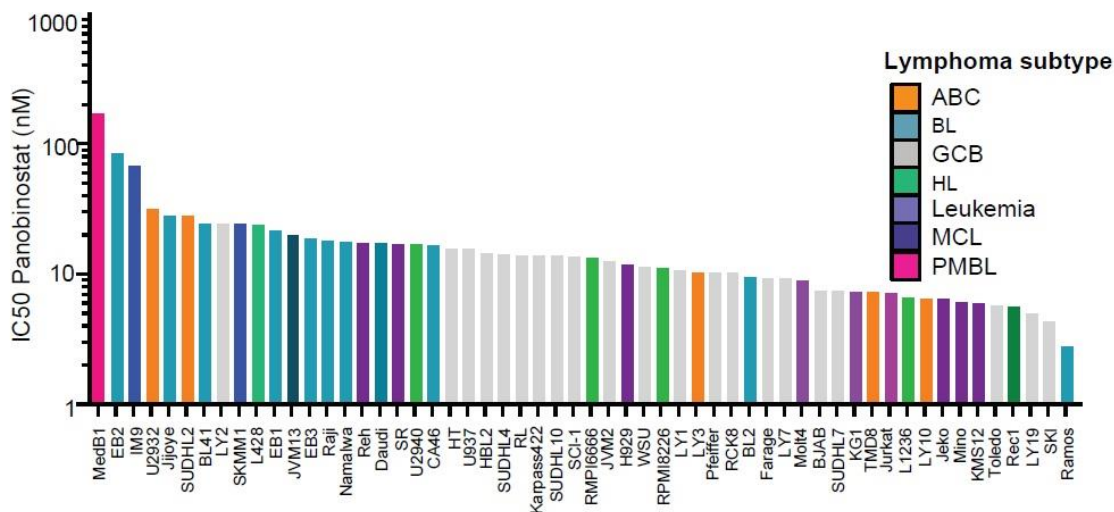


Figure 45: Panobinostat sensitivity across B cell malignancies. IC50s of cell lines (x-axis) treated with panobinostat. Color indicates disease type. Diseases include ABC DLBCL (ABC), Burkitt Lymphoma (BL), GCB DLBCL (GCB), Hodgkin’s Lymphoma (HL), Leukemia, Mantle Cell Lymphoma (MCL), and Primary Mediastinal B cell Lymphoma (PMBL).

We then extended the synergy testing to cell lines outside the original group to test for synergy. Here we found synergy with both the in-group cell lines, OCI-LY3 and TMD-8, as well as those outside the original group, SU-DHL-2 and Farage (Figure 46). Synergy measured by the Chou-Talalay method determined combination index for all lines between 0.15 and 0.3.

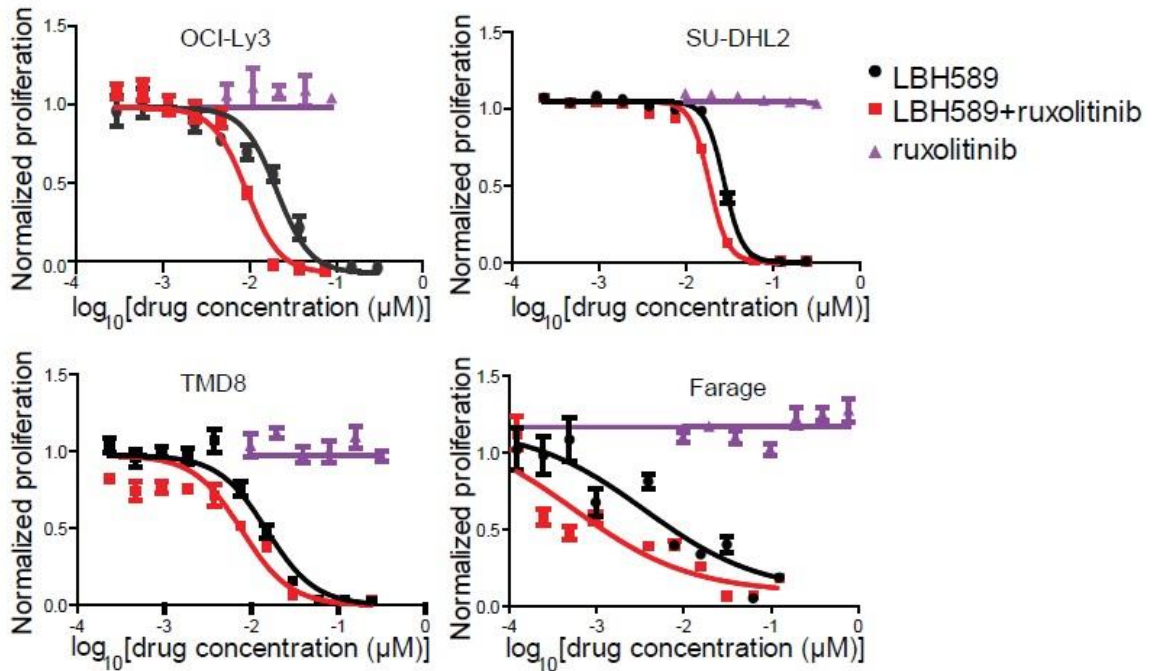


Figure 46: Panobinostat with ruxolitinib synergy. 4 cell lines (labeled above) were treated with either increasing dose panobinostat (LBH589) alone (black), ruxolitinib alone at 100nM (purple), or ruxolitinib at 100nM with increasing dose of panobinostat (red). Panobinostat concentrations are noted below each curve.

The effectiveness of in vitro data provided an initial starting point for testing combination therapy of DLBCL with panobinostat and ruxolitinib in xenograft models. Xenografts were established using stably expressing luciferase in Farage cells, implanted in SCID mice. 4 cohorts of mice were used that were treated with carrier alone, ruxolitinib alone, panobinostat alone, or combination therapy (Figure 47A). Both single agents had modest effects on tumor size, but we found significant synergy with the combination (Figure 47B-C, $p < 0.01$). Nearly 30% of mice treated with the combination therapy were found to be tumor free at the end of 16 days, compared to none of the

single agent or carrier alone treated mice. In vivo modeling also allowed for testing for hematological toxicity. The mice generally performed well when treated with single agent and combination therapy. Hemoglobin levels and ANC remained constant with all therapies, though platelet levels dropped significantly (Figure 47D). This level of thrombocytopenia is clinically tolerable with monitoring.

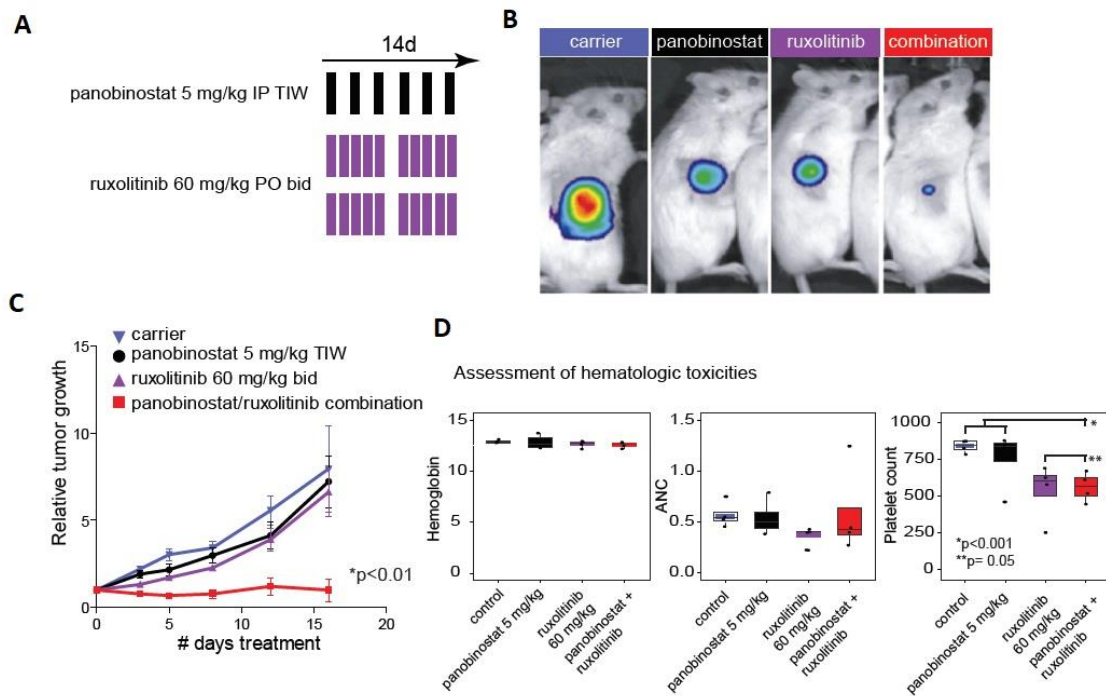


Figure 47: Synergy in xenograft models. (A) Drug and dose scheduling for mice in xenograft models. (B) Tumors size images from luciferase expressing implanted Farage cells measured by luciferase imaging for carrier alone, panobinostat treated, ruxolitinib treated or combination treated at 16 days. (C) All mouse tumor size over 16 days of treatment. (D) Assessment of hematological toxicities.

5.4 Discussion

Combination therapies are a major target in cancer therapeutic research as they often provide a more tolerable therapeutics that can overcome primary resistance in cancers

that do not respond to frontline therapy. Significant effort has been put forth, both publicly and privately, to identify combination therapies in cancer. Traditional screening is inefficient at the scale necessary to identify combinations with all variables tested. Here we assemble the largest expression-based drug screen assayed in DLBCL. We then define the cRDGE method for effective identification of synergy between small molecules using only drug response data paired with expression data at a single dose. Analysis of previous data using drug perturbed gene expression data yields a higher hit rate using the cRDGE method. Using the method, we identify the synergistic combination of panobinostat and ruxolitinib, and show preliminary efficacy both *in vitro* and *in vivo*.

Our data also shows that expression-based screening is also powerful in understanding how mechanisms of resistance and sensitivity can be assayed. We find baseline gene expression of targets is correlated with sensitivity in targeted inhibitors such as BRAF. We also find that downregulation of gene target upon drug treatment is correlated to sensitivity to the drug, while upregulation portends sensitivity in BTK. Non-target gene expression of drivers, such as MYC in DLBCL, can be used as a broad proxy for sensitivity for all drug types, which can be used for enriching studies for more actionable drugs. These methods are extendable to drug screening across cancer type. While exact targets, especially driver genes, may change, the underlying biology can be used to define resistant and sensitive effects of drugs.

Unlike most expression-based studies, our study used 6 cell lines of the same disease, modeling the well-known genetic heterogeneity of the disease. This allowed us to identify subtype specific differences in single agent activity, resistance mechanisms, and expression differences in response to drugging. While some drugs recapitulated this subtype specific division of DLBCL, other single agent and combination synergy was seen in subgroups that cross subtypes. Much like ER+ breast cancer or castration resistant prostate cancer, this therapy associated subtyping may be useful in defining clinical subgroups that are responsive to specific therapies.

This study serves as a pilot for deeper investigation into combination therapies into DLBCL and other cancer types. Advances in microfluidics and reductions in sequencing costs allow for massive parallelization of these types of studies to much larger numbers of drugs, increasing the investigative space, as well as number of cell lines, increasing power and accuracy in heterogeneous disease. Advances in single-cell technology also make it possible to use primary tumor sample from biopsy or patient derived xenografts to be used as starting material into these screens, allowing for intratumor heterogeneity to be assayed alongside drug response. Resistance mechanisms across subpopulations of the tumor as it exists can be assayed in parallel. The methodology of this study is can be used to both look backward at previous studies to leverage these drug-expression datasets to mine for possible combination therapies,

and going forward to design better, larger studies that can find possible combination therapies within the striking genetic and phenotypic heterogeneity of cancer.

6. Conclusions

As new sequencing technologies emerge and larger, deeper efforts into cancer genomics are pursued, the future holds much promise in new experimental and informatic approaches. Here, I discuss several of these exciting new avenues of research and how they may be used for future therapeutic development.

Cancer genomics: wider and deeper

As sequencing cost continue to drop, sample numbers, sample types, and depth of sequencing can increase without costs rising in turn. We leveraged this to study the largest cohorts of samples in DLBCL genomics and therapeutic screening. New technology, such as the NovaSeq from Illumina, allow for even more data output for similar costs. Sample sizes can become even larger, enabling more high-powered studies to be performed and subsets of disease to be analyzed at high depth. Often, the bottleneck in genomics comes at the computational processing of this huge amount of data. New technologies in this area, including the Firehose analysis pipeline from the Broad and DeepVariant from Google AI, allow for automated processing of this data with fewer computational recourses and validated pipelines.

The research presented in this dissertation contains the largest sequencing study in DLBCL genomics, yet much information remains to be gleaned from even these samples. Advances in whole genome sequencing and non-coding functionality assays allow us to map and analyze variation outside of the 1% often assayed through exome

sequencing. Exome sequencing allows us to enrich for more interpretable and actionable sequence variation, but the continuous decrease in sequencing costs allows for similar numbers of samples to be run at the whole-genome scale. Whole-genome sequencing has produced valuable data in certain cancer types [155], but most cancer entities have not received this level of attention. Whole genome data adds many layers of functional data including transcription factor binding regions and enhancer regions. These regions have proven to have effects on cancer and new technology has allowed for parallel evaluation of these types of functional regions [156].

With these advancements in both sequencing technology and computational analysis methods, these types of studies can be performed at scale. Large studies have been possible only in large multi-institutional groups, such as TCGA. These tools allow for smaller groups to perform these types of deep genomic studies with systematic rigor and data analytics.

Advances in single-cell sequencing technology have opened up a new frontier in understanding cancer with broad applicability across cancer research. Ready-to-use systems, such as the 10x Genomics Chromium platform, allow reliable, streamlined processes for profiling single-cell gene expression profiles. Single-cell sequencing allows for tumors sequencing to be dissociated into individual cells, removing the requisite aggregate nature of traditional RNA-seq approaches. This allows for malignant cells to be totally isolated from other stromal cells, eliminating the need for tumor purity

estimates and deconvolution of contaminating cells to inform expression sequencing.

This pure population can then be analyzed for subpopulations of the tumor alone.

Subpopulations can be analyzed for cancer stem cell identification, resistant clone isolation, and overall rates of intratumor heterogeneity. These single-cell expression profiles can then be reassembled into whole tumor profiles, but without the stromal cell contamination that traditional RNA-seq methods carry.

Deep sequencing: more than mutations

The deep catalogue of sequenced tumors paired with the relative ease of performing genomic profiling of new samples allows for studies powered to answer questions that have been elusive in previous studies. One of these questions is how the evolution of cancer occurs naturally within the body. Deep sequencing for allele frequencies and newer single-cell sequencing allows for identification of outgrowing subpopulations, giving insight into mutation order and comparative fitness. Single-cell technology is especially useful here, as it can give mutations at a cell specific manner. These mutations can be placed in the context of a hierarchy, and advantaged/disadvantaged populations can be identified. These disadvantaged populations can contain cells with alterations in pathways targeted by therapeutics. Identification of these subpopulations would be a boon in clinical settings where this data can inform therapeutic options.

RNA-sequencing also provides a wealth of data yet to be fully utilized in all cancer types. Allele-specific expression of mutant or otherwise altered alleles can be

identified by phased analysis of RNA-seq reads. This allows us to understand how these mutant alleles are expressed, either high, making it a higher priority oncogenic target, or lower, suggesting tumor suppressor function. Bulk and single-cell RNA-seq can also provide information on the invading stromal cells. These can be immune cells, vasculature, or other invading support cells. These signatures have already been associated with survival and specific changes in the tumor biology. Single-cell sequencing allows for individualized counts and deconvolution of cell types, which may prove important to outcomes in cases of immune invasion.

RNA-seq also allows for a more refined analysis of cell-of-origin. Cell-of-origin has been shown to influence aggressiveness of disease, but rare cells-of-origin may yet be obscured. Single-cell sequencing of normal tissues normal tissues may yet reveal cell types that have not been identified, which may explain some of the variation in cancer phenotype from cells of what are currently understood to be the same type. These rare cells may have markers that have not been explored in these tissues, but gene expression analysis could reveal these subtypes of cells. Associating this data with known cancer expression profiles could identify these relationships. Subsetting this group would allow for more power studies to be performed on them, allowing for identification of novel targets for therapeutics.

Focal adhesion: a model solution

Focal adhesion related genes exist in a unique space within cancer biology. They can regulate nearly all processes related to cancer, such as growth, migration, immune evasion, and apoptosis, yet they are extremely difficult to study due to their interactive nature with normal stroma. Focal adhesion kinase alone has been shown to regulate most of the hallmarks of cancer, manipulate the tumor microenvironment, and overexpress in nearly 80% of cancers. GNA13 has been shown by our lab to be involved in lymphomagenesis by altering migration and proliferation signaling. The problem with studying these types of pathways is that they are inextricably tied to tumor-stromal and tumor-microenvironmental interactions.

The necessities of other non-malignant cells in the tumor model makes understanding the full scope of these genetic alterations difficult. Manipulation of cell lines will give limited insight into these interactions, as malignant cells alone cannot model this interaction. One promising avenue in hematological malignancies is the growing success of humanized mouse models. These models can reconstitute the entire hematopoietic lineage of human cells into a mouse model. Introducing genetic knockouts, overexpression, or mutated forms of focal adhesion related genes gives a more accurate representation of the disease with intact immune-immune contacts. While these models are not perfect as they contain slightly different stromal cells, niche requirements, and signaling molecules, a vertebrate model with a human immune system can much more accurately model hematopoietic malignancies. Indeed, this can be extended into solid

tumor xenograft models as well. Immune escape and modulation is well studied cancer phenotype, and cell line or patient derived xenografts can be placed in a mouse model with a human immune system, giving a more accurate model of tumor-immune interactions. As these many of these interactions center around T cell involvement and its related focal adhesion, these models would be informative to these interactions in way that we currently cannot model effectively or at scale.

Precision medicine: genomics and therapeutics

As we move into an era where clinical sequencing of patient tumors becomes the routine, many questions arise around what data we should collect and how to then use this data. Currently, only a handful of well characterized mutations are sequenced for prognostic and therapeutic decisions. As large-scale genomics with clinical implications comes to the fore, we can expand on known variant targets as well as introduce more data types and techniques to the clinic.

Much like our Genomic Classification Model from chapter 2, integrative genomics can build on existing, well-characterized clinical tests and integrate many genomic features to increase the predictive powers of these models. Single mutations, copy number change, and expression differences have all proven to be diagnostic in some form, and they have been implemented in clinical tests. These tests are often individually predictive for survival or specific drugs, but combining these data points into predictive and therapeutic models represents a major goal of this type of research.

This requires very large datasets to capture the full heterogeneity of the disease of interest. While this is feasible for common cancer types, disease subtypes and rare cancer entities will be challenging to get cohorts large enough for this type of analysis.

Genomic sequencing may also be useful for patients with relapsed disease or tumors resistant to conventional therapies. Patients with these types of tumors have extremely poor outcomes, and any increase in predictive and therapeutic power is welcome. Large genomic datasets with paired clinical data can be used to create models for identification of the patients before treatment, allowing for their entry into clinical trials or other treatment options. Identification of actionable targets through initial sequencing at diagnosis allows for a more effective, tailored treatment to be used, increasing the likelihood of success by leveraging these genomic models. Patient tumors can also be used as patient derived xenografts (PDX) or placed into personal drug screens to develop individually specified therapies. PDX models for tumors can be used to grow and study the patient tumor, and they can be used as treatment models to find effective therapies. These tumors, or even the primary sample, can be placed into a miniaturized drug screen that looks for possible therapeutic agents that work on a particular cancer type. Using single-cell or low-input expression sequencing, methods such as cRDGE can be used to predict effective therapeutics in these cancers that are tailored to the specific patient, their tumor, and even resistant clones within the tumor. The challenges of finding functional genomic variation that is predictive of outcomes

and targeted therapeutics for these genetic variants are high, but the outcomes may change the way we think about and treat cancer.

References

1. SEER. *SEER Cancer Statistics Review, 1975-2014. SEER Cancer Statistics Review, 1975-2014, National Cancer Institute. Bethesda, MD (2017).*
2. McArthur, G. A. *et al.* Safety and efficacy of vemurafenib in BRAFV600E and BRAFV600K mutation-positive melanoma (BRIM-3): extended follow-up of a phase 3, randomised, open-label study. *Lancet Oncol.* **15**, 323–332 (2014).
3. Tokheim, C. J., Papadopoulos, N., Kinzler, K. W., Vogelstein, B. & Karchin, R. Evaluating the evaluation of cancer driver genes. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 14330–14335 (2016).
4. Jiang, Y., Soong, T. D., Wang, L., Melnick, A. M. & Elemento, O. Genome-Wide Detection of Genes Targeted by Non-Ig Somatic Hypermutation in Lymphoma. *PLoS One* **7**, e40332 (2012).
5. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396**, 643–649 (1998).
6. Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
7. Roulland, S. *et al.* t(14;18) translocation: A predictive blood biomarker for follicular lymphoma. *J. Clin. Oncol.* **32**, 1347–1355 (2014).
8. Rowley, J. D. A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243**, 290–293 (1973).
9. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).
10. Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).
11. Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).
12. Rosenwald, A. *et al.* The Use of Molecular Profiling to Predict Survival after Chemotherapy for Diffuse Large-B-Cell Lymphoma. *N. Engl. J. Med.* **346**, 1937–1947 (2002).

13. Wright, G. *et al.* A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9991–6 (2003).
14. Lenz, G. *et al.* Oncogenic CARD11 Mutations in Human Diffuse Large B Cell Lymphoma. *Science* **319**, 1676–1679 (2008).
15. Lohr, J. G. *et al.* Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 3879–3884 (2012).
16. Pasqualucci, L. *et al.* Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat. Genet.* **43**, 830–837 (2011).
17. Zhang, J. *et al.* Genetic heterogeneity of diffuse large B-cell lymphoma. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 1398–403 (2013).
18. Healy, J. A. *et al.* GNA13 loss in germinal center B cells leads to impaired apoptosis and promotes lymphoma in vivo. *Blood* **127**, 2723–2731 (2016).
19. Maverakis, E. *et al.* Metastatic melanoma – A review of current and future treatment options. *Acta Dermato-Venereologica* **95**, 516–524 (2015).
20. Sala, E. *et al.* BRAF Silencing by Short Hairpin RNA or Chemical Blockade by PLX4032 Leads to Different Responses in Melanoma and Thyroid Carcinoma Cells. *Mol. Cancer Res.* **6**, 751–759 (2008).
21. Hatzivassiliou, G. *et al.* RAF inhibitors prime wild-type RAF to activate the MAPK pathway and enhance growth. *Nature* **464**, 431–435 (2010).
22. Nazarian, R. *et al.* Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature* **468**, 973–977 (2010).
23. Straussman, R. *et al.* Tumour micro-environment elicits innate resistance to RAF inhibitors through HGF secretion. *Nature* **487**, 500–504 (2012).
24. Wilson, T. R. *et al.* Widespread potential for growth-factor-driven resistance to anticancer kinase inhibitors. *Nature* **487**, 505–509 (2012).
25. Weisberg, E., Manley, P. W., Cowan-Jacob, S. W., Hochhaus, A. & Griffin, J. D. Second generation inhibitors of BCR-ABL for the treatment of imatinib-resistant chronic myeloid leukaemia. *Nature Reviews Cancer* **7**, 345–356 (2007).

26. Hantschel, O., Rix, U. & Superti-Furga, G. Target spectrum of the BCR-ABL inhibitors imatinib, nilotinib and dasatinib. in *Leukemia and Lymphoma* **49**, 615–619 (2008).
27. Le, X. F., Pruefer, F. & Bast, R. C. HER2-targeting antibodies modulate the cyclin-dependent kinase inhibitor p27Kip1 via multiple signaling pathways. *Cell Cycle* **4**, 87–95 (2005).
28. Williams, S. I. & Rogers, C. E. *HER2 and cancer: Mechanism, testing and targeted therapy. HER2 and Cancer: Mechanism, Testing and Targeted Therapy* (2011).
29. Molina, M. A. *et al.* Trastuzumab (Herceptin), a humanized anti-HER2 receptor monoclonal antibody, inhibits basal and activated HER2 ectodomain cleavage in breast cancer cells. *Cancer Res.* **61**, 4744–4749 (2001).
30. Sharma, M. K. *et al.* Distinct genetic signatures among pilocytic astrocytomas relate to their brain region origin. *Cancer Res.* **67**, 890–900 (2007).
31. ENCODE. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
32. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
33. Polak, P. *et al.* Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature* **518**, 360–364 (2015).
34. Osborne, C. K., Yochmowitz, M. G., Knight, W. A. & McGuire, W. L. The value of estrogen and progesterone receptors in the treatment of breast cancer. *Cancer* **46**, 2884–2888
35. Early Breast Cancer Trialists Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet* **351**, 1451–1467 (1998).
36. Gregory, C. W. *et al.* A mechanism for androgen receptor-mediated prostate cancer recurrence after androgen deprivation therapy. *Cancer Res.* **61**, 4315–4319 (2001).
37. Zhang, J. *et al.* Genetic heterogeneity of diffuse large B-cell lymphoma. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 1398–403 (2013).
38. Author, G., Scheinman, R. I., Cogswell, P. C., Lofquist, A. K. & Baldwin, A. S. Role

- of Transcriptional Activation of I κ B α in Mediation of Immunosuppression by. *Source Sci. New Ser. J. Pet. Meteorit. D. E. Grady, J. Geophys. Res* **270**, 283–286 (1995).
39. Weller, M. Glucocorticoid Treatment of Primary CNS Lymphoma. *J. Neurooncol.* **43**, 237–239 (1999).
 40. Shaw, T., Quan, J. & Totoritis, M. C. B cell therapy for rheumatoid arthritis: the rituximab (anti-CD20) experience. *Ann. Rheum. Dis.* **62 Suppl 2**, ii55-9 (2003).
 41. Weiner, G. J. Rituximab: mechanism of action. *Semin. Hematol.* **47**, 115–23 (2010).
 42. Kochenderfer, J. N. *et al.* Chemotherapy-refractory diffuse large B-cell lymphoma and indolent B-cell malignancies can be effectively treated with autologous T cells expressing an anti-CD19 chimeric antigen receptor. *J. Clin. Oncol.* **33**, 540–9 (2015).
 43. Lee, D. W. *et al.* T cells expressing CD19 chimeric antigen receptors for acute lymphoblastic leukaemia in children and young adults: a phase 1 dose-escalation trial. *Lancet* **385**, 517–528 (2015).
 44. Wilson, A. & Trumpp, A. Bone-marrow haematopoietic-stem-cell niches. *Nat. Rev. Immunol.* **6**, 93–106 (2006).
 45. Lu, T. T. & Cyster, J. G. Integrin-mediated long-term B cell retention in the splenic marginal zone. *Science* **297**, 409–12 (2002).
 46. Wang, X., Rodda, L. B., Bannard, O. & Cyster, J. G. Integrin-mediated interactions between B cells and follicular dendritic cells influence germinal center B cell fitness. *J. Immunol.* **192**, 4601–9 (2014).
 47. Arana, E. *et al.* Activation of the Small GTPase Rac2 via the B Cell Receptor Regulates B Cell Adhesion and Immunological-Synapse Formation. *Immunity* **28**, 88–99 (2008).
 48. Spaargaren, M. *et al.* The B cell antigen receptor controls integrin activity through Btk and PLCgamma2. *J. Exp. Med.* **198**, 1539–50 (2003).
 49. Jaffe, A. B. & Hall, A. RHO GTPASES: Biochemistry and Biology. *Annu. Rev. Cell Dev. Biol.* **21**, 247–269 (2005).
 50. Koopman, G. *et al.* Adhesion of human B cells to follicular dendritic cells involves both the lymphocyte function-associated antigen 1/intercellular adhesion molecule 1 and very late antigen 4/vascular cell adhesion molecule 1 pathways. *J.*

Exp. Med. **173**, 1297–304 (1991).

51. Koopman, G. *et al.* Adhesion through the LFA-1 (CD11a/CD18)-ICAM-1 (CD54) and the VLA-4 (CD49d)-VCAM-1 (CD106) pathways prevents apoptosis of germinal center B cells. *J. Immunol.* **152**, 3760–7 (1994).
52. Batista, F. D. & Neuberger, M. S. B cells extract and present immobilized antigen: implications for affinity discrimination. *EMBO J.* **19**, 513–20 (2000).
53. Karnoub, A. E., Symons, M., Campbell, S. L. & Der, C. J. Molecular basis for Rho GTPase signaling specificity. *Breast Cancer Research and Treatment* **84**, 61–71 (2004).
54. Valencia, A., Chardin, P., Wittinghofer, A. & Sander, C. Perspectives in Biochemistry The ras Protein Family: Evolutionary Tree and Role of Conserved Amino Acids. *Biochem. ©* (1991).
55. Adamson, P., Paterson, H. F. & Hall, A. Intracellular localization of the P21(rho) proteins. *J. Cell Biol.* **119**, 617–627 (1992).
56. Sah, V. P., Seasholtz, T. M., Sagi, S. A. & Brown, J. H. The role of Rho in G protein-coupled receptor signal transduction. *Annu. Rev. Pharmacol. Toxicol.* **40**, 459–89 (2000).
57. Kjølner, L. & Hall, A. Signaling to Rho GTPases. *Experimental Cell Research* **253**, 166–179 (1999).
58. Fessler, M. B. *et al.* Dual role for RhoA in suppression and induction of cytokines in the human neutrophil. *Blood* **109**, 1248–56 (2007).
59. Teo, C. R., Casey, P. J. & Rasheed, S. A. K. The GNA13-RhoA signaling axis suppresses expression of tumor protective Kallikreins. *Cell. Signal.* **28**, 1479–1488 (2016).
60. Eva, A., Vecchio, G., Rao, C. D., Tronick, S. R. & Aaronson, S. A. The predicted DBL oncogene product defines a distinct class of transforming proteins. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 2061–5 (1988).
61. Hart, M. J., Eva, A., Evans, T., Aaronson, S. A. & Cerione, R. A. Catalysis of guanine nucleotide exchange on the CDC42Hs protein by the dbl oncogene product. *Nature* **354**, 311–4 (1991).
62. O'Connor, K. & Chen, M. Dynamic functions of RhoA in tumor cell migration and

- invasion. *Small GTPases* **4**, 141–7 (2013).
63. Wang, H.-R. *et al.* Regulation of cell polarity and protrusion formation by targeting RhoA for degradation. *Science* **302**, 1775–9 (2003).
 64. Strutt, D. I., Weber, U. & Mlodzik, M. The role of RhoA in tissue polarity and Frizzled signalling. *Nature* **387**, 292–295 (1997).
 65. Arthur, W. T. & Burridge, K. RhoA inactivation by p190RhoGAP regulates cell spreading and migration by promoting membrane protrusion and polarity. *Mol. Biol. Cell* **12**, 2711–20 (2001).
 66. McBeath, R., Pirone, D. M., Nelson, C. M., Bhadriraju, K. & Chen, C. S. Cell Shape, Cytoskeletal Tension, and RhoA Regulate Stem Cell Lineage Commitment. *Dev. Cell* **6**, 483–495 (2004).
 67. Straight, A. F. & Field, C. M. Microtubules, membranes and cytokinesis. *Current Biology* **10**, (2000).
 68. Eggert, U. S., Mitchison, T. J. & Field, C. M. Animal Cytokinesis: From Parts List to Mechanisms. *Annu. Rev. Biochem.* **75**, 543–566 (2006).
 69. Drechsel, D. N., Hyman, A. A., Hall, A. & Glotzer, M. A requirement for Rho and Cdc42 during cytokinesis in *Xenopus* embryos. *Curr. Biol.* **7**, 12–23 (1997).
 70. Jantsch-Plunger, V. *et al.* CYK-4: A Rho family gtpase activating protein (GAP) required for central spindle formation and cytokinesis. *J. Cell Biol.* **149**, 1391–404 (2000).
 71. O’Keefe, L., Somers, W. G., Harley, A. & Saint, R. The pebble GTP exchange factor and the control of cytokinesis. *Cell Struct Funct* **26**, 619–26. (2001).
 72. Dopeso, H. *et al.* Mechanisms of inactivation of the tumour suppressor gene RHOA in colorectal cancer. *British Journal of Cancer* **118**, 106–116 (2018).
 73. Rodrigues, P. *et al.* RHOA inactivation enhances Wnt signalling and promotes colorectal cancer. *Nat. Commun.* **5**, (2014).
 74. Wang, K. *et al.* Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat. Genet.* **46**, 573–582 (2014).
 75. Kakiuchi, M. *et al.* Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. *Nat. Genet.* **46**, 583–587 (2014).

76. O'Hayre, M. *et al.* Inactivating mutations in GNA13 and RHOA in Burkitt's lymphoma and diffuse large B-cell lymphoma: a tumor suppressor function for the Gα13/RhoA axis in B cells. *Oncogene* **35**, 3771–3780 (2016).
77. Yoo, H. Y. *et al.* A recurrent inactivating mutation in RHOA GTPase in angioimmunoblastic T cell lymphoma. *Nat. Genet.* **46**, 371–375 (2014).
78. Palomero, T. *et al.* Recurrent mutations in epigenetic regulators, RHOA and FYN kinase in peripheral T cell lymphomas. *Nat. Genet.* **46**, 166–170 (2014).
79. Sakata-Yanagimoto, M. *et al.* Somatic RHOA mutation in angioimmunoblastic T cell lymphoma. *Nat. Genet.* **46**, 171–175 (2014).
80. Guan, J.-L. & Shalloway, D. Regulation of focal adhesion-associated protein tyrosine kinase by both cellular adhesion and oncogenic transformation. *Nature* **358**, 690–692 (1992).
81. Parsons, J. T. Focal adhesion kinase: the first ten years. *J. Cell Sci.* **116**, 1409–1416 (2003).
82. Webb, D. J. *et al.* FAK–Src signalling through paxillin, ERK and MLCK regulates adhesion disassembly. *Nat. Cell Biol.* **6**, 154–161 (2004).
83. Mitra, S. K., Hanson, D. A. & Schlaepfer, D. D. Focal adhesion kinase: in command and control of cell motility. *Nat. Rev. Mol. Cell Biol.* **6**, 56–68 (2005).
84. Tomar, A. & Schlaepfer, D. D. Focal adhesion kinase: switching between GAPs and GEFs in the regulation of cell motility. *Current Opinion in Cell Biology* **21**, 676–683 (2009).
85. Peng, L. *et al.* Secreted LOXL2 is a novel therapeutic target that promotes gastric cancer metastasis via the Src/FAK pathway. *Carcinogenesis* **30**, 1660–1669 (2009).
86. Dia, V. P. & Gonzalez de Mejia, E. Lunasin potentiates the effect of oxaliplatin preventing outgrowth of colon cancer metastasis, binds to α5β1 integrin and suppresses FAK/ERK/NF-κB signaling. *Cancer Lett.* **313**, 167–80 (2011).
87. Naora, H. & Montell, D. J. Ovarian Cancer Metastasis: Integrating insights from disparate model organisms. *Nat. Rev. Cancer* **5**, 355–366 (2005).
88. Lin, T.-H. *et al.* D-pinitol Inhibits Prostate Cancer Metastasis through Inhibition of αVβ3 Integrin by Modulating FAK, c-Src and NF-κB Pathways. *Int. J. Mol. Sci.* **14**,

9790–9802 (2013).

89. Hochwald, S. N. *et al.* A novel small molecule inhibitor of FAK decreases growth of human pancreatic cancer. *Cell Cycle* **8**, 2435–2443 (2009).
90. Hsia, D. A. *et al.* Differential regulation of cell motility and invasion by FAK. *J. Cell Biol.* **160**, 753–767 (2003).
91. Parsons, J. T., Martin, K. H., Slack, J. K., Taylor, J. M. & Weed, S. a. Focal adhesion kinase: a regulator of focal adhesion dynamics and cell movement. *Oncogene* **19**, 5606–5613 (2000).
92. Sieg, D. J. *et al.* FAK integrates growth-factor and integrin signals to promote cell migration. *Nat. Cell Biol.* **2**, 249–256 (2000).
93. Carragher, N. O. & Frame, M. C. Focal adhesion and actin dynamics: A place where kinases and proteases meet to promote invasion. *Trends in Cell Biology* **14**, 241–249 (2004).
94. Westhoff, M. A., Serrels, B., Fincham, V. J., Frame, M. C. & Carragher, N. O. SRC-mediated phosphorylation of focal adhesion kinase couples actin and adhesion dynamics to survival signaling. *Mol. Cell. Biol.* **24**, 8113–8133 (2004).
95. Sebolt-Leopold, J. S. Development of anticancer drugs targeting the MAP kinase pathway. *Oncogene* **19**, 6594–6599 (2000).
96. Braun, C. *et al.* Expression of calpain I messenger RNA in human renal cell carcinoma: correlation with lymph node metastasis and histological type. *Int J Cancer* **84**, 6–9 (1999).
97. Frame, M. C. Src in cancer: Deregulation and consequences for cell behaviour. *Biochimica et Biophysica Acta - Reviews on Cancer* **1602**, 114–130 (2002).
98. Kurenova, E. *et al.* Focal adhesion kinase suppresses apoptosis by binding to the death domain of receptor-interacting protein. *Mol. Cell. Biol.* **24**, 4361–71 (2004).
99. Frisch, S. M., Vuori, K., Ruoslahti, E. & Chan-Hui, P. Y. Control of adhesion-dependent cell survival by focal adhesion kinase. *J. Cell Biol.* **134**, 793–799 (1996).
100. Agochiya, M. *et al.* Increased dosage and amplification of the focal adhesion kinase gene in human cancer cells. *Oncogene* **18**, 5646–5653 (1999).
101. McLean, G. W. *et al.* Decreased focal adhesion kinase suppresses papilloma

- formation during experimental mouse skin carcinogenesis. *Cancer Res.* **61**, 8385–8389 (2001).
102. Bergh, J., Norberg, T., Sjögren, S., Lindgren, A. & Holmberg, L. Complete sequencing of the p53 gene provides prognostic information in breast cancer patients, particularly in relation to adjuvant systemic therapy and radiotherapy. *Nat. Med.* **1**, 1029–1034 (1995).
 103. Rodrigues, N. R. *et al.* p53 mutations in colorectal cancer. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 7555–9 (1990).
 104. Ng, S. B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272–276 (2009).
 105. Bainbridge, M. N. *et al.* Whole exome capture in solution with 3 Gbp of data. *Genome Biol.* **11**, (2010).
 106. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–303 (2010).
 107. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
 108. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
 109. Cibulskis, K. *et al.* MuTect - brief summary. *Nat. Biotechnol.* **31**, 213–219 (2013).
 110. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164–e164 (2010).
 111. Winer, J., Jung, C. K. S., Shackel, I. & Williams, P. M. Development and Validation of Real-Time Quantitative Reverse Transcriptase–Polymerase Chain Reaction for Monitoring Gene Expression in Cardiac Myocytes *In Vitro*. *Anal. Biochem.* **270**, 41–49 (1999).
 112. Gardner, T. S., Di Bernardo, D., Lorenz, D. & Collins, J. J. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science (80-.)*. **301**, 102–105 (2003).
 113. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. Quantitative Monitoring of

- Gene Expression Patterns with a Complementary DNA Microarray. *Science* (80-.). **270**, 467–470 (1995).
114. Golub, T. R. *et al.* Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* (80-.). **286**, 531–527 (1999).
 115. van 't Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).
 116. Shipp, M. A. *et al.* Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat. Med.* **8**, 68–74 (2002).
 117. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).
 118. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
 119. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
 120. Zhang, J. *et al.* The genomic landscape of mantle cell lymphoma is related to the epigenetically determined chromatin state of normal B cells. *Blood* **123**, 2988–96 (2014).
 121. Morin, R. D. *et al.* Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* **476**, 298–303 (2011).
 122. Reddy, A. *et al.* Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell* **171**, 481–494.e15 (2017).
 123. Magi, A. *et al.* EXCAVATOR: Detecting copy number variants from whole-exome sequencing data. *Genome Biol.* **14**, (2013).
 124. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
 125. Papaemmanuil, E. *et al.* Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N. Engl. J. Med.* **374**, 2209–2221 (2016).
 126. Project, T. I. N.-H. L. P. F. A Predictive Model for Aggressive Non-Hodgkin's Lymphoma. *N. Engl. J. Med.* **329**, 987–994 (1993).

127. Scott, D. W. *et al.* Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. *Blood* **123**, 1214–7 (2014).
128. Hans, C. P. *et al.* Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood* **103**, 275–82 (2004).
129. Leiserson, M. D. M., Reyna, M. A. & Raphael, B. J. A weighted exact test for mutually exclusive mutations in cancer. *Bioinformatics* **32**, i736–i745 (2016).
130. Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, M. D. C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**, 267–273 (2014).
131. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science (80-.)*. **343**, 84–87 (2014).
132. Aguirre, A. J. *et al.* Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. *Cancer Discov.* **6**, 914–29 (2016).
133. Monti, S. *et al.* Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. *Blood* **105**, 1851–61 (2005).
134. Lenz, G. *et al.* Stromal Gene Signatures in Large-B-Cell Lymphomas. *N. Engl. J. Med.* **359**, 2313–2323 (2008).
135. Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science (80-.)*. **315**, 972–976 (2007).
136. Green, T. M. *et al.* Immunohistochemical double-hit score is a strong predictor of outcome in patients with diffuse large B-cell lymphoma treated with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone. *J. Clin. Oncol.* **30**, 3460–7 (2012).
137. Johnson, N. A. *et al.* Concurrent expression of MYC and BCL2 in diffuse large B-cell lymphoma treated with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone. *J. Clin. Oncol.* **30**, 3452–9 (2012).
138. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B (Statistical Methodol.* **67**, 301–320 (2005).

139. Zhang, S., Zhou, X., Lang, R. A. & Guo, F. RhoA of the Rho family small GTPases is essential for B lymphocyte development. *PLoS One* **7**, (2012).
140. Richter, J. *et al.* Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. *Nat. Genet.* **44**, 1316–1320 (2012).
141. Rohde, M. *et al.* Recurrent RHOA mutations in pediatric Burkitt lymphoma treated according to the NHL-BFM protocols. *Genes Chromosom. Cancer* **53**, 911–916 (2014).
142. Snyder, J. T. *et al.* Structural basis for the selective activation of Rho GTPases by Dbl exchange factors. *Nat. Struct. Biol.* **9**, 468–475 (2002).
143. Hobeika, E. *et al.* Testing gene function early in the B cell lineage in mb1-cre mice. *Proc. Natl. Acad. Sci.* **103**, 13789–13794 (2006).
144. Mlinaric-Rascan, I. & Yamamoto, T. B cell receptor signaling involves physical and functional association of FAK with Lyn and IgM. *FEBS Lett.* **498**, 26–31 (2001).
145. Nelson, H. S. Combination therapy of bronchial asthma. *Allergy Asthma Proc.* **22**, (2001).
146. Yonesaka, K. *et al.* Activation of ERBB2 signaling causes resistance to the EGFR-directed therapeutic antibody cetuximab. *Sci. Transl. Med.* **3**, (2011).
147. Rawlinson, R. & Massey, A. J. γ H2AX and Chk1 phosphorylation as predictive pharmacodynamic biomarkers of Chk1 inhibitor-chemotherapy combination treatments. *BMC Cancer* **14**, (2014).
148. Puri, N. & Salgia, R. Synergism of EGFR and c-Met pathways, cross-talk and inhibition, in non-small cell lung cancer. *J. Carcinog.* **7**, 9 (2008).
149. Lamb, J. *et al.* The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science (80-.).* **313**, 1929–1935 (2006).
150. Fitzgerald, J. B., Schoeberl, B., Nielsen, U. B. & Sorger, P. K. Systems biology and combination therapy in the quest for clinical efficacy. *Nature Chemical Biology* **2**, 458–466 (2006).
151. Zhao, X. M. *et al.* Prediction of drug combinations by integrating molecular and pharmacological data. *PLoS Comput. Biol.* **7**, (2011).

152. Bansal, M. *et al.* A community computational challenge to predict the activity of pairs of compounds. *Nat. Biotechnol.* **32**, 1213–1222 (2014).
153. Chen, B. *et al.* Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. *Nat. Commun.* **8**, (2017).
154. Chen, B. *et al.* Relating Chemical Structure to Cellular Response: An Integrative Analysis of Gene Expression, Bioactivity, and Structural Data Across 11,000 Compounds. *CPT Pharmacometrics Syst. Pharmacol.* **4**, 576–584 (2015).
155. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
156. Vockley, C. M. *et al.* Massively parallel quantification of the regulatory effects of noncoding genetic variation in a human cohort. *Genome Res.* **25**, 1206–1214 (2015).

Biography

Nicholas Samuel Davis was born in Decatur, Alabama on March 25, 1988 to Samuel Edgar Davis and Cecelia Dianne Johnson Davis. He attended Austin High School in Decatur, Alabama. He then attended Auburn University in Auburn, Alabama, where he received his Bachelor of Science in Cellular and Molecular Biology. Nick began his Ph.D. in the University Program in Genetics and Genomics in 2012. He rotated with Paul Magwene and Kris Wood before joining the lab of Sandeep Davé for his dissertation research.

Publications:

(* indicates authors contributed equally)

1. Reddy, A.*, Zhang, J.*, **Davis, N.***, *et al.* Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell* 171, 481-494 e15 (2017).
2. Ramaker, R., Bowling, K., Lasseigne, B., Hagenauer, M., Hardigan, A., **Davis N.**, *et al.* Post-mortem molecular profiling of three psychiatric disorders. *Genome Med* 9, 72 (2017).
3. Moffitt, A. *et al.* Enteropathy-associated T cell lymphoma subtypes are characterized by loss of function of SETD2. *J Exp Med* 214, 1371-1386 (2017).
4. Kirby, M. *et al.* Genome-wide DNA methylation measurements in prostate tissues uncovers novel prostate cancer diagnostic biomarkers and transcription factor binding patterns. *BMC Cancer* 17, 273 (2017).
5. McKinney, M., *et al.* The Genetic Basis of Hepatosplenic T-cell Lymphoma. *Cancer Discov* 7, 369-379 (2017).

6. Kirby, M., Ramaker, R., Gertz, J., **Davis, N.**, *et al.* RNA sequencing of pancreatic adenocarcinoma tumors yields novel expression patterns associated with long-term survival and reveals a role for ANGPTL4. *Mol Oncol* 10, 1169-82 (2016).
7. Healy, J., Nugent, A., Rempel, R., Moffitt, A, **Davis, N.**, *et al.* GNA13 loss in germinal center B cells leads to impaired apoptosis and promotes lymphoma in vivo. *Blood* 127, 2723-31 (2016).
8. Varley, K., Gertz, J., Roberts, B., **Davis, N.**, *et al.* Recurrent read-through fusion transcripts in breast cancer. *Breast Cancer Res Treat* 146, 287-97 (2014).
9. Gertz, J., Varley, K., **Davis, N.**, *et al.* Transposase mediated construction of RNA-seq libraries. *Genome Res* 22, 134-41 (2012).
10. Encode Consortium (**Davis, N.** HAIB) An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).