A forward genetics approach to identify molecular drivers of liver cancer using Sleeping Beauty mouse models

Jesse Daniel Riordan
University of Iowa

Copyright 2013 Jesse Daniel Riordan

This dissertation is available at Iowa Research Online: https://ir.uiowa.edu/etd/5049

Recommended Citation
https://ir.uiowa.edu/etd/5049. https://doi.org/10.17077/etd.s8b7lqi

Follow this and additional works at: https://ir.uiowa.edu/etd
Part of the Cell Anatomy Commons, and the Cell Biology Commons
A FORWARD GENETICS APPROACH TO IDENTIFY MOLECULAR DRIVERS OF LIVER CANCER USING SLEEPING BEAUTY MOUSE MODELS

by

Jesse Daniel Riordan

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Anatomy and Cell Biology in the Graduate College of The University of Iowa

December 2013

Thesis Supervisor:  Associate Professor Adam J. Dupuy
This is to certify that the Ph.D. thesis of

Jesse Daniel Riordan

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Anatomy and Cell Biology at the December 2013 graduation.

Thesis Committee:

Adam J. Dupuy, Thesis Supervisor

John F. Engelhardt

Michael D. Henry

Paul B. McCray, Jr.

D. Thomas Rutkowski
To my family – past, present, and future
As a fraction of the lifespan of the universe as measured from its beginning to the 
evaporation of the last black hole, life as we know it is only possible for one thousandth 
of a billion billion billionth, billion billion billionth, billion billion billionth of a percent. 
And that’s why, for me, the most astonishing wonder of the universe isn’t a star or a 
planet or a galaxy. It isn’t a thing at all. It’s an instant in time. And that time is now. 
Humans have walked the earth for just the shortest fraction of that briefest of moments in deep time. But in our 200,000 years on this planet we’ve made remarkable progress. It was only two and a half thousand years ago that we believed that the sun was a god and measured its orbit with stone towers built on the top of a hill. Today the language of curiosity is not sun gods, but science.

Brian Cox, Wonders of the Universe
ACKNOWLEDGEMENTS

I would like to thank Adam Dupuy, my mentor, for helping me to focus my broad scientific interests and to determine the type of research that I enjoy most. I feel that my training in his lab has provided me with exceptional preparation for a successful scientific career. I would also like to thank my thesis committee members (John Engelhardt, Michael Henry, Paul McCray, and Tom Rutkowski) for providing valuable insights and guidance during my time as a graduate student. David Largaespada, Vincent Keng, and Bobbi Tschida have been extremely helpful both in assistance with experiments and discussion of research projects.

My wife Shylo has been tremendously supportive over the last six years, putting up with late nights and weekends in the lab, as well as science-associated frustrations. I am deeply grateful for her love and support. My parents have also been instrumental in my success; my fascination with science developed at an early age as a result of their encouragement. Finally, I would like to thank my son Max, whose smiling face greeting me at the top of the stairs when I get home can always turn a rough day around.
ABSTRACT

Each year liver cancer kills more than half a million people, making it the third leading cause of cancer-related death worldwide. Annual incidence continues to rise steadily, both domestically and globally, increasing the burden of this disease. Advancements in the ability to obtain detailed molecular profiles of tumors have led to the successful development of targeted therapies for a number of different cancers. Unfortunately, however, the molecular pathogenesis of liver cancer is poorly understood relative to many other types of malignancies. Thus, the identification of factors contributing to the development and progression of liver tumors is a major goal of current research.

In pursuit of this goal, I have utilized the Sleeping Beauty (SB) transposon system as a tool for forward genetic mutagenesis screening in mice. The SB system recapitulates the kinetics of spontaneous tumor development in humans by providing a stepwise accumulation of mutations. Micro-evolutionary processes within a developing tumor lead to the selective expansion of cells harboring mutations that confer some kind of selective advantage. By identifying the most prevalent mutation events within a specific tumor type across a large number of independent samples, a list of genes implicated as being involved in tumorigenesis can be generated. Using this approach, the Dlk1-Dio3 imprinted domain was identified as a site of frequent mutation in SB-induced hepatocellular carcinomas (HCCs). I discovered that the mechanistic basis for recurrent selection of transposon insertion within this domain in liver tumors involved activated expression of Retrotransposon-like 1 (Rtl1). I also found that RTL1 activation is a common event in human HCC, suggesting that it could potentially be beneficial as a therapeutic target in a subset of patients.

Etiological factors related to liver cancer development are varied, but are linked by the fact that each provides a chronic liver injury stimulus that promotes the
development of hepatic fibrosis. In fact, ~90% of human HCC occurs in this context, and yet the majority of mouse liver cancer models fail to account for this important environmental component of the disease. I have conducted a screen for genetic drivers of liver cancer in the presence or absence of hepatic fibrosis. Comparison of mutation profiles between fibrotic and non-fibrotic tumors revealed largely non-overlapping sets of candidate genes, indicative of a differential selective pressure for mutations depending on the fibrotic context of the liver. Driver mutations identified preferentially in the presence of liver fibrosis have a high likelihood of relevance to human disease, given the similarities in environmental context and kinetics of mutation acquisition. Consistent with this idea, multiple genes with well-established roles in human HCC were found to be preferentially mutated in SB-induced tumors developed in a fibrotic liver.

Before a candidate cancer gene identified in an animal model system can have an impact on human disease, its proposed role in tumorigenesis must be validated. Existing techniques for validation of putative liver cancer genes suffer from significant limitations including high cost, low throughput, and a level of complexity that prohibits widespread utilization. I have contributed to the generation of a novel tool for in vivo validation of candidate genes that is not subject to these limitations. By combining elements of recombinant adenoviral vectors and the piggyBac transposition system, we have generated a highly flexible gene delivery system with significant advantages over existing techniques. The Ad-PB system has broad accessibility and applicability, making it a valuable tool for advancing efforts to improve cancer therapies.
TABLE OF CONTENTS

LIST OF TABLES x

LIST OF FIGURES xi

CHAPTER

1. INTRODUCTION 1

Liver Cancer Background and Impact 1
Molecular Targeted Therapies 2
Sleeping Beauty Forward Genetic Mutagenesis 4
Hepatic Fibrosis and HCC 7
Validation of Candidate Genes 9
Research Goals 10

2. IDENTIFICATION OF RTL1, A RETROTRANSPOSON-DERIVED IMPRINTED GENE, AS A NOVEL DRIVER OF HEPATOCARCINOGENESIS 15

Abstract 15
Introduction 16
Results and Discussion 18

Determining the effect of transposon integration on Dlk1-Dio3 domain members 18
Integrated transposons directly drive Rtl1 expression 19
Rtl1 expression in cultured hepatocytes promotes growth in ECM 21
In vivo hepatic Rtl1 expression drives tumorigenesis 22
RTL1 activation in human HCC 23
Rtl1-expressing mouse HCCs resemble human S1 subclass 25
Potential of RTL1 as a therapeutic target and/or biomarker 26
Conclusion 27

Materials and Methods 27
Mice 27
Human tissue samples 27
RNA sequencing and data analysis 28
RT-PCR 28
Illumina sequencing of transposon insertions 30
Matrigel growth assay 30
Hydrodynamic gene delivery 31
Western blotting 31
Gene Set Enrichment Analysis (GSEA) 31
Immunohistochemistry 32
3. A SLEEPING BEAUTY MUTAGENESIS SCREEN IDENTIFIES MUTATIONS DRIVING HEPATOCellular CARCINOMA IN FIBROTIC LIVER

Abstract 44
Introduction 45
Results and Discussion 47
CCl4-induced liver fibrosis increases severity of a Sleeping Beauty HCC model 47
Identification of common insertion sites in SB-induced HCC 49
Fibrotic liver environment provides differential selective pressure for mutations 50
Analysis of candidate genes in human HCC 52
Conclusions 54
Materials and Methods 54
Mice 54
CCl4 injections 54
Histology 55
Illumina sequencing of transposon insertions 55

4. PIGGYBAC TRANSPOSITION FROM RECOMBINANT ADENOVIRUS AS A METHOD FOR GENE DELIVERY WITH STABLE EXPRESSION

Abstract 62
Introduction 63
Results and Discussion 65
Design of a hybrid piggyBac-adenoviral vector system 65
Efficient genomic integration in cultured cells 66
Application as a validation tool for putative oncogenes and tumor suppressors in vivo 68
Applications for gene therapy 69
Conclusions 70
Materials and Methods 70
Vector construction 70
Cell culture 71
Excision PCR 71
Flow cytometry 71

5. DISCUSSION

Involvement of the Dlk1-Dio3 Imprinted Domain in Cancer 75
Translation of Rtl1 to the Clinic 79
Toward an Improved Mechanistic Understanding of Fibrotic Liver Cancer 84
Implications for Gene Therapy 92
LIST OF TABLES

Table

1. Liver tumors developed in hydrodynamically injected mice 43
2. Scoring of liver fibrosis severity in CCl₄-treated mice 59
3. Liver tumors generated by SB mutagenesis screen 60
4. Common insertion sites identified in SB-induced liver tumors 61
5. Flow cytometry analysis of mCherry expression in transduced cells 74
## LIST OF FIGURES

Figure

1. Mechanisms of *Sleeping Beauty* transposon-induced mutation 14
2. *Dlk1-Dio3* domain transposon integration sites in SB-induced HCC and effects on domain expression 33
3. Heat maps depicting global differential transcript expression in SB-induced HCCs and normal liver 34
4. Integrated transposons drive overexpression of *Rtl1* 35
5. Transposon integrations are preferentially detected upstream of *Rtl1* in SB-induced HCCs 36
6. Rtl1 promotes growth of cultured hepatocytes in extracellular matrix 37
7. *In vivo* hepatic overexpression of *Rtl1* promotes tumorigenesis 38
8. Transgene expression in tumors induced by hydrodynamic injection 39
9. Expression of *RTL1* in human HCC 40
10. *Rtl1*-expressing mouse HCCs resemble human S1 subclass 41
11. Validation of S1 subclass protein expression in SB-induced HCC 42
12. Induction of hepatic fibrosis and steatosis in CCl₄-treated mice 56
13. Patterns of transposon insertion within frequently mutated genes 57
14. Proposed model for functional interaction between Nfe2, Nfe2l2, Fign, Keap1, and Cul3 58
15. Structure of Ad-PB vectors 72
16. PB transposition from the adenoviral genome 73
Liver Cancer Background and Impact

Liver cancer is the sixth most commonly diagnosed cancer worldwide, and it ranks third in annual fatalities at ~600,000 per year\(^1\). Hepatocellular carcinoma (HCC) comprises >80% of liver cancer cases. Etiological factors for HCC are varied and include infection with hepatitis B (HBV) or hepatitis C virus (HCV), aflatoxin exposure, alcohol-induced damage, and non-alcoholic fatty liver disease (NAFLD)\(^2\). A shared characteristic of these risk factors is the induction of liver damage, inflammation, and fibrosis, suggesting that these processes directly drive hepatocarcinogenesis. In contrast to most types of cancer, the incidence of HCC in the United States is on the rise, with age-adjusted incidence tripling over the last three decades\(^3\). Increased incidence in the US has been attributed largely to increased prevalence of HCV infection and obesity, which often leads to NAFLD. A gender bias is observed in HCC incidence, wherein men are three to five times more likely to develop the disease than women, a trend that is also observed in mouse models\(^4\).

As evidenced by the high ratio of HCC-related deaths to diagnosed cases, patients with this disease have a very poor prognosis. Less than half of patients diagnosed with HCC will survive the first year, and the five-year survival rate is among the lowest of any cancer type at ~10%\(^3\). The primary explanation for such poor survival rates is a lack of effective treatment options. Traditional cancer therapies such as radiation and chemotherapy show little efficacy. While surgical resection of the tumor mass can be curative, HCC is rarely diagnosed at an early enough stage for this to be possible; generally tumors have either grown too large or have disseminated to distant sites.
Molecular Targeted Therapies

Over the past decade, the focus of cancer therapeutics research has largely shifted to the identification and targeting of molecular mechanisms underlying tumor formation and progression. Such an approach has numerous advantages over traditional non-specific therapies, including the potential for increased efficacy and fewer negative side effects. The biggest factor contributing to an enhanced molecular characterization of cancer has been an exponential increase in the throughput of DNA sequencing technologies. This increased throughput has allowed researchers to obtain tumor molecular profiles of unprecedented depth, as exemplified by The Cancer Genome Atlas (TCGA) project. Begun in 2006 with a stated mission “to accelerate the understanding of the molecular basis of cancer through the application of genome analysis technologies, including large-scale genome sequencing”, this extensive collaborative project has generated immensely detailed profiles of large sample sets for more than twenty tumor types. These profiles include analyses of somatic mutations, copy-number alterations, RNA expression, and DNA methylation. The enormous amount of resources dedicated to TCGA demonstrates the intense emphasis that has been placed on identifying molecular drivers of cancer. Investments such as this are expected to lead to the development of significantly improved therapeutic approaches for cancer patients.

Already, molecular targeted therapies have been employed as a major treatment approach for several cancer types, most notably for chronic myelogenous leukemia (CML) and metastatic melanoma. Following its approval for clinical use in 2001, imatinib revolutionized the treatment of CML. The drug works by inhibiting the activity of Abelson tyrosine-protein kinase (ABL), which undergoes constitutive activation in CML as a result of a chromosomal translocation event that fuses the breakpoint cluster region (BCR) to ABL. This small molecule has transformed CML therapy; the rate of CML-related death for patients in the five years after starting treatment has decreased from 71% with standard chemotherapy therapy(5) to less than 5% with imatinib(6).
Activating mutations in the *BRAF* locus are the most common recurrent genetic aberration detected in malignant melanomas, present in ~50% of cases (data from the Catalogue of Somatic Mutations in Cancer (COSMIC) database(7)). These mutations lead to constitutive activation of serine/threonine-protein kinase B-raf, which promotes unchecked cell proliferation via the MAPK signaling pathway. The FDA has recently approved vemurafenib, a small molecule B-raf inhibitor, for treatment of patients with metastatic melanoma harboring a *BRAF* activating mutation. In a Phase III clinical trial, patients receiving vemurafenib showed relative reductions of 63% and 74% in the risk of death or tumor progression, respectively, as compared to patients receiving dacarbazine, a non-specific chemotherapeutic agent(8). The successes of imatinib and vemurafenib demonstrate the power of molecular targeted therapies to reduce cancer burden.

Currently, one molecular targeted therapy is approved for the treatment of HCC. Sorafenib, a small-molecule inhibitor of multiple protein kinases, is the only systemic therapy to provide any significant survival benefit to late-stage HCC patients to date. Targets of sorafenib include vascular endothelial growth factor receptor (VEGFR), platelet-derived growth factor receptor (PDGFR), stem cell factor receptor (c-Kit), and B-raf, though the exact mechanisms underlying the drug’s anti-tumor activity are not completely understood. In a Phase III clinical trial, median survival and time to tumor progression were extended by three months in HCC patients treated with sorafenib(9). While this result is a significant improvement over prior therapeutic approaches and shows promise that molecular targeted therapies are likely to be more effective for HCC than non-specific chemotherapeutic agents, the survival benefit is quite modest. In fact, approval of sorafenib for HCC treatment within the United Kingdom’s National Health Service was declined based on the viewpoint that the survival benefits of sorafenib did not justify its high cost (£3000 per patient per month)(10). It is becoming increasingly clear that a detailed understanding of the molecular events driving tumor formation and progression are key to the development of effective targeted cancer therapies.
Unfortunately, however, much remains to be learned about the molecular pathogenesis of HCC. While we do have a preliminary understanding of the genetic aberrations involved in HCC, most of what is currently known involves alterations in genes with well-established roles in cancer, and it is clear that many genes which contribute to the development and progression of this disease have yet to be identified (11). There is thus a critical need to identify novel drivers of HCC in order to develop more effective treatments for patients.

**Sleeping Beauty Forward Genetic Mutagenesis**

Molecular profiling of human tumor samples, such as that being done by TCGA, is of obvious benefit, as it represents the most direct method to identify genetic aberrations present in human cancer. There are, however, several drawbacks and challenges to this strategy. For one, a comprehensive and integrated approach is required to ensure that contributions of a given gene or pathway are not underestimated. There are many avenues by which a cancer cell can achieve the same functional output; for example, signaling from a protein kinase could be increased via acquisition of an activating mutation, copy number amplification, chromosomal translocation, alteration of promoter methylation, or activation of mRNA expression, among other mechanisms. Without an integrated analysis of all these types of alterations, the contribution of that protein kinase to human cancer would be underestimated. Additionally, utilization of a large sample set is critical to the successful molecular characterization of each tumor type. These requirements limit the extent to which comprehensive analysis of human tumors can be achieved, as the amount of resources needed and associated costs are tremendous. As the throughput of DNA sequencing continues to increase while the cost continues to drop, these limitations will be lessened.

A bigger challenge to the identification of molecular therapeutic targets is the immense level of complexity present in tumors. Most tumors contain hundreds to
thousands of genetic disruptions, yet the percentage of these events that actually contribute to tumor development and/or progression is estimated to be less than 1%(12). Identifying these “driver” events that confer selective advantage to tumor cells among the much larger number of “passenger” events that do not contribute to neoplastic transformation, but are propagated during tumor development due to their co-occurrence with drivers, has proven to be quite challenging. Distinguishing drivers from passengers is a critical step in the development of effective molecular targeted therapies. One way to achieve this goal is through the use of animal models of cancer as a complementary approach to direct sequencing of human tumors, which affords a level of experimental control not possible with human studies. Complex datasets generated by deep profiling of human tumors can more readily be deciphered through comparison to more precisely defined animal models of cancer.

The *Sleeping Beauty* (SB) transposon system has proven useful for identifying drivers of tumorigenesis in a wide variety of tissue types(13). By using SB transposons as insertional mutagens to conduct forward genetic cancer screens in mice, mutation events that are recurrently selected for during the process of tumorigenesis can easily be identified. The SB system is comprised of two major elements – a cut-and-paste DNA transposon and the transposase enzyme that binds to and mobilizes the transposon. Only when both components of the system are present in the same cell nucleus can mobilization, and therefore transposon-induced mutagenesis, occur. For cancer screens, mouse strains have been generated that harbor concatemers of transposons at specific sites in the genome. SB transposase (SBase) activity can be induced within distinct cell types of these strains’ offspring, either through breeding to an SBase-expressing strain(14) or to a tissue-specific Cre recombinase strain to activate a resident inducible SBase allele(15).

The SB mutagenesis system offers several advantages as a forward genetic screening tool. Mutagenic transposons have been engineered to be capable of inducing
both gain-of-function and loss-of-function mutations, allowing efficient identification of oncogenes and tumor suppressors, respectively. This occurs primarily through disruption of splicing events by elements within the transposon (Figure 1). Following excision of a transposon from the donor concatemer, re-insertion occurs at random throughout the genome due to the ubiquity of the recognition site for integration (a TA dinucleotide). This creates an unbiased mutation profile, which maximizes the potential to identify tumor-driving events throughout the entire genome. In contrast, some alternative systems utilized for insertional mutagenesis screens, including piggyBac transposons and retroviral vectors, display a bias for integration within coding elements or regions of active transcription(16, 17). Because the sequence of the transposon is known, insertion sites can easily be amplified following tumor development, allowing rapid generation of detailed mutation profiles. Another advantage of the SB system is its ability to closely recapitulate the process of tumorigenesis as it occurs in humans. Somatic mutations accumulate in a stepwise manner, driving a micro-evolutionary process within the developing tumor wherein those mutations that confer a selective advantage to cells are preferentially maintained. Positive selection for these mutation events leads to selective clonal expansion of the cells harboring them. Importantly, the resultant tumors are considerably less complex at the molecular level than human tumors, facilitating the identification of driver mutation events. This is accomplished by comparing insertion profiles from multiple independent tumors of a particular type in order to identify recurrent mutation events, termed common insertion sites (CISs). CISs represent promising candidate driver events, as their recurrence in multiple independent tumors suggests positive selection during cancer development, which is indicative of functional significance.
Hepatic Fibrosis and HCC

Around 90% of HCC develops in the context of chronic hepatic fibrosis(18), a condition characterized by the accumulation of excess extracellular matrix (ECM) and decreased cellularity within the liver. Induction of fibrosis occurs as a result of a chronic injury stimulus and the associated inflammatory environment. All of the major risk factors for HCC development, which include HBV or HCV infection, alcohol-induced damage, and NAFLD, promote sustained liver injury that leads to fibrosis. The mechanism of fibrotic development involves disruption of liver ECM dynamics. Normal turnover of ECM is impaired, and hepatic stellate cells are activated to deposit excess matrix proteins, particularly collagen. Together, these processes lead to the formation of scar tissue and increased liver stiffness.

In the continued presence of an injury stimulus, hepatic fibrosis is a progressive condition that increases in severity over time. A feed-forward loop emerges in which damaged hepatocytes release cytokines that recruit inflammatory cells, the inflammatory cells activate hepatic stellate cells to secrete collagen, and the activated stellate cells additionally secrete cytokines that recruit more inflammatory cells and modulate their activity(19). Patients eventually develop cirrhosis, the most advanced stage of fibrosis, generally 15-20 years after the condition’s initial onset. Cirrhosis is characterized by increased deposition of scar tissue, loss of structural architecture, portal hypertension resulting from intrahepatic resistance to blood flow, and the formation of regenerative hepatocyte nodules. Once established, cirrhosis is generally irreversible. With an annual incidence of ~4%, one third of cirrhotic patients will go on to develop HCC at some point in their life(20).

While a strong association between liver fibrosis and HCC clearly exists, the extent to which fibrosis directly contributes to hepatocarcinogenesis remains unclear. It is possible that their frequent co-occurrence results from induction by the same factors, rather than through a direct promotion of HCC by fibrosis. Several potential mechanisms
by which liver fibrosis may drive tumorigenesis have been proposed(21), though they require definitive experimental validation. Obtaining this validation has proven to be challenging, in part due to difficulties in experimentally separating the processes of inflammation and fibrosis. Proposed contributions of a fibrotic environment to HCC development include enhanced mitogenic and anti-apoptotic signaling through integrins, growth factor secretion by activated hepatic stellate cells, increased stromal stiffness leading to proliferation, and growth factor sequestration by the ECM resulting in autocrine and paracrine signaling.

Although the relative importance of each of the proposed mechanisms by which fibrosis may promote cancer is uncertain, the fact remains that the vast majority of human HCC occurs in the context of a fibrotic liver environment. The utility of mouse models for gaining insight into the molecular pathogenesis of tumor development, including HCC, is well established; however, the ability to obtain clinically relevant insights from these models is critically dependent on the degree to which they accurately recapitulate human disease. The majority of existing mouse models of liver cancer fail to account for the fibrotic context in which most human tumors develop(22). In Chapter 3, I summarize the results of a study comparing the molecular profiles of HCC developed through SB forward genetic mutagenesis screening in the presence or absence of experimentally induced liver fibrosis. Analysis of common transposon insertion sites from tumors developed in both environments revealed several candidate cancer genes preferentially mutated in the presence or absence of hepatic fibrosis. I propose that mutation events identified as drivers of HCC in a fibrotic environment are likely to be highly relevant to human disease and may shed light on fibrosis-dependent mechanisms of hepatocarcinogenesis.
Validation of Candidate Genes

A fundamental goal of cancer research at any level is to decrease the burden of this devastating disease on the human population. From experiments designed to understand a gene’s basic functions to large-scale clinical trials of a novel therapeutic agent, the ultimate shared aim of researchers is to improve our ability to effectively prevent and treat cancer. For cancer geneticists to maximize the likelihood of success in this endeavor, it is critically important to validate the proposed role of candidate cancer genes through direct manipulation in a setting that recapitulates human disease as closely as possible. Equally important is the identification of a link to human cancer, be it through shared effects on a gene’s activity between animal models and human tumors or shared functional consequences brought on by distinct genetic events in these two settings (e.g. activation of the same signaling pathway). For a candidate cancer gene identified in an animal model system to have a therapeutic impact for humans, these two criteria (functional validation and relevance to human cancer) must be met to warrant further pursuit of translational applications.

With an ever-increasing ability to deeply profile the molecular characteristics of human tumors and the availability of huge, publicly accessible databases such as those generated by TCGA, analyzing associations of candidate genes with human cancer is becoming a task that often requires little more than a computer. Validation of a candidate gene’s function in a context relevant to human cancer, however, remains a major experimental challenge. Large amounts of resources and technical expertise are generally required to perform experiments that adequately test a gene’s oncogenic or tumor suppressive potential. Existing techniques for validating putative liver cancer genes in vivo often rely on complex engineered mouse strains(23), transplantation experiments(24), strong predisposing alleles(25), induction of severe liver damage(26), or a combination of these factors. An alternative approach is to generate transgenic, knock-in, or knockout mouse strains, although the associated high cost and low throughput limit the extent to
which this can be done. There is currently a need for a liver cancer gene validation technique that is inexpensive, high-throughput, efficient, minimally invasive, and amenable to combination with additional genetic and environmental inputs. In Chapter 4, I describe the initial development and characterization of such a technique. Capitalizing on advantages provided by recombinant adenoviral vectors and the *piggyBac* transposon system, we have designed an approach for *in vivo* validation of candidate liver cancer genes with several advantages over existing techniques. Additionally, this approach could easily be adapted to selectively modulate expression of a candidate gene in any cell type that can be effectively transduced by adenovirus, which has implications beyond validation of candidate cancer genes.

**Research Goals**

The primary goal of my thesis research has been to identify novel molecular drivers of liver cancer. This is a complex task that requires a multifaceted approach. The major foci of my work have been: 1) mechanistic characterization and subsequent validation of a commonly selected mutation event in SB-induced HCCs, 2) generation and molecular profiling of an SB-induced liver cancer model that recapitulates the fibrotic environment present in the majority of human HCC, and 3) development of a novel gene delivery system for efficient *in vivo* validation of candidate liver cancer genes.

Previous work in the Dupuy lab identified transposon insertion within the *Dlk1-Dio3* imprinted domain as a highly recurrent mutation event in liver tumors developed through SB mutagenesis(15). One of my research goals was to clarify the mechanism of transposon-induced mutation in these tumors. I hypothesized that some molecular event was commonly being induced as a result of transposon insertion within the domain and that this event was a driving force for hepatocarcinogenesis, explaining the observed positive selection for *Dlk1-Dio3* insertions in HCC. In-depth molecular characterization of tumors with these insertions revealed that each of them displayed significant activation
of Retrotransposon-like 1 (Rtl1) expression driven by integrated transposons. This result suggested that the primary tumor-promoting event under positive selection in SB-induced HCCs was activation of Rtl1. Further analysis validated Rtl1 as a driver of hepatocarcinogenesis in vivo, identified a subset of human HCC with RTL1 overexpression, and suggested that the mechanism of tumorigenesis may involve degradation of the extracellular matrix by Rtl1 protein. Though further work is required for translation of these findings to a clinical application, my results implicate RTL1 as a promising molecular target to pursue therapeutically in HCC.

Another of my research goals was to generate a novel mouse model of liver cancer with high relevance to human disease. To do this, I conducted a liver-specific SB mutagenesis screen in mice with experimentally induced hepatic fibrosis. This model develops tumors in a fibrotic liver environment, which is the setting for most human HCC development. It also recapitulates the stepwise accumulation of somatic mutations that drives human tumorigenesis. Encouragingly, some genes known to play a role in human HCC were commonly mutated in tumors resulting from our screen. We additionally identified several novel candidate genes that may be relevant to the pathogenesis of human HCC. By comparing SB-induced mutation profiles between tumors developed in the presence or absence of hepatic fibrosis, we discovered several mutation events that are subject to differential selective pressure depending on the fibrotic context in which tumors develop. These findings may shed light on mechanisms of fibrosis-dependent HCC development, and the identified mutations represent logical candidates for follow-up validation experiments and comparisons to human HCC.

Experimental approaches such as forward genetic screens are valuable tools to identify candidate genes involved in a biological process of interest. The information gained from this type of approach serves as a starting point to generate testable hypotheses, and it requires further experimental follow-up to be validated. Such validation is especially important for putative cancer genes identified through genetic
screens. Without functional demonstration of a role in tumorigenesis, there is little justification for the pursuit of diagnostics or therapeutics based on a candidate. Existing techniques to functionally validate putative liver cancer genes in vivo suffer from limitations including high cost, low throughput, dependence on strong predisposing mutations, or induction of supraphysiological liver damage. An additional research goal of mine was to generate a new method for in vivo liver cancer gene validation that was not subject to these limitations. I have contributed to the development of a system that will allow efficient and stable gene delivery to mouse liver via simple tail vein injection. This system capitalizes on the ability to efficiently transduce hepatocytes with recombinant adenoviral vectors following tail vein injection, while also permitting maintained gene expression through stable genomic integration of a piggyBac transposon containing the gene of interest. One major advantage of this method over existing techniques is increased throughput. Its relative simplicity and low cost, combined with the possibility of testing pools of candidates in a single experiment, allows validation experiments to be conducted for several genes. Another advantage is the system’s flexibility. Because all of the components necessary for stable gene delivery are supplied in trans via tail vein injection, the procedure can be performed using any mouse strain of interest or can be combined with specific treatments to induce a phenotype of interest. These advantages of our method for liver cancer gene validation, as compared to existing techniques, make it feasible to functionally test a larger number of putative oncogenes and tumor suppressors over a broader range of genetic and environmental contexts. This will facilitate the identification of tumor-driving events that warrant further study and represent logical molecular targets for which to develop therapeutic agents, depending on relevance to human cancer.

Through my thesis research, I have made a significant contribution to the field of liver cancer genetics. My work has led to the identification of Rtl1 as a novel oncogene involved in HCC development, the generation of a mouse model of HCC with high
relevance to human disease that may provide insight into the role of fibrosis in driving hepatocarcinogenesis, and the development of a new method for candidate liver cancer gene validation \textit{in vivo} with significant advantages over existing techniques. Each of these accomplishments has contributed to the achievement of my primary goal of identifying novel molecular drivers of liver cancer, both directly and through the development of useful resources for other researchers with the same goal.
Figure 1. Mechanisms of *Sleeping Beauty* transposon-induced mutation

(A) Structure of the T2/Onc3 mutagenic transposon. Functional elements include splice acceptors in both transcriptional orientations, a cytomegalovirus early enhancer/chicken β-actin (CAG) promoter followed by a splice donor, and a bi-directional polyadenylation signal. (B-E) Mechanisms of transposon-mediated gene disruption. Transposon integration directly upstream of (B) or within (C) a gene in the same transcriptional orientation can cause CAG promoter/splice donor-driven overexpression of nearly full-length or truncated transcripts. This can produce gain-of-function mutations and may result in the activation of oncogenes. The splice acceptors and bi-directional polyadenylation signal of a transposon inserted within a gene in either transcriptional orientation can trap the promoter (D-E), which can result in loss-of function mutations, including inactivation of tumor suppressor genes. **SA**, splice acceptor. **CAG**, cytomegalovirus early enhancer/chicken β-actin promoter. **SD**, splice donor. **pA**, polyadenylation signal.
CHAPTER 2
IDENTIFICATION OF RTL1, A RETROTRANSPOSON-DERIVED IMPRINTED GENE, AS A NOVEL DRIVER OF HEPATOCARCINOGENSES

Abstract

We previously utilized a Sleeping Beauty (SB) transposon mutagenesis screen to discover novel drivers of HCC. This approach identified recurrent mutations within the Dlk1-Dio3 imprinted domain, indicating that alteration of one or more elements within the domain provides a selective advantage to cells during the process of hepatocarcinogenesis. For the current study, we performed transcriptome and small RNA sequencing to profile gene expression in SB-induced HCCs in an attempt to clarify the genetic element(s) contributing to tumorigenesis. We identified strong induction of Retrotransposon-like 1 (Rtl1) expression as the only consistent alteration detected in all SB-induced tumors with Dlk1-Dio3 integrations, suggesting that Rtl1 activation serves as a driver of HCC. While previous studies have identified correlations between disrupted expression of multiple Dlk1-Dio3 domain members and HCC, we show here that direct modulation of a single domain member, Rtl1, can promote hepatocarcinogenesis in vivo. Overexpression of Rtl1 in the livers of adult mice using a hydrodynamic gene delivery technique resulted in highly penetrant (86%) tumor formation. Additionally, we detected overexpression of RTL1 in 30% of analyzed human HCC samples, indicating the potential relevance of this locus as a therapeutic target for patients. The Rtl1 locus is evolutionarily derived from the domestication of a retrotransposon. In addition to identifying Rtl1 as a novel driver of HCC, our study represents one of the first direct in vivo demonstrations of a role for such a co-opted genetic element in promoting carcinogenesis.
Introduction

Hepatocellular carcinoma (HCC) is the third leading cause of cancer-related deaths worldwide(1). In contrast to the downward trends in incidence observed for most cancer types, that of HCC continues to rise, particularly in the United States(3). This is due in part to increases in obesity and hepatitis C viral infection, both of which have been implicated in HCC pathogenesis. Treatment options for patients are limited, particularly for those with advanced disease, and the five-year survival rate remains low at ~10%.

A major goal of HCC research is to develop therapies targeted at the molecular mechanisms underlying tumor development and progression. This type of approach is expected to be much more efficacious, increasing survival rates for HCC patients. Consistent with this idea, treatment with sorafenib, a multi-kinase inhibitor, has shown survival benefits for late-stage patients(9) – a rare achievement in HCC treatment. Nevertheless, sorafenib treatment is only able to extend median survival by three months, underlying the need for improved targeted therapies. Unfortunately, the molecular drivers of HCC remain poorly characterized, precluding the development of such therapeutics. Large-scale sequencing efforts currently being undertaken by The Cancer Genome Atlas (TCGA) project will likely characterize the recurrent genetic alterations present in human liver tumors and may identify novel therapeutic targets. However, it is becoming increasingly clear that human tumors are incredibly complex, and identifying molecular drivers of carcinogenesis among the larger number of background events has proven difficult. Comparative analysis of the information gained from human tumor profiling with data from animal models provides an improved ability to distinguish driver events contributing to human disease.

The Sleeping Beauty (SB) transposon mutagenesis system has proven useful for identifying drivers of tumorigenesis in a wide variety of tissue types(13). We have previously used SB mutagenesis to generate mice that developed HCC(15). Subsequent genetic analysis of SB-induced liver tumors identified the Dlk1-Dio3 imprinted domain
as a common target of transposon-induced mutations. This highly complex domain contains genes encoding protein-coding transcripts, long non-coding RNAs (lncRNAs), microRNAs (miRNAs), and small nucleolar RNAs (snoRNAs). Expression of domain members is regulated in an allele-specific manner and depends on epigenetic modifications established in the germline(27). Regulation of this expression pattern is maintained, at least in part, by multiple differentially methylated regions (DMRs) throughout the domain that are methylated on the paternally inherited allele. Maintenance of imprinting is critical for normal function, as evidenced by the fact that uniparental disomy (UPD) for either parental allele leads to severe and widespread developmental defects in both mouse models(28) and human patients(29).

A link between the Dlk1-Dio3 domain and HCC has previously been identified. Interestingly, it has been reported that adeno-associated viral (AAV) vector integration within the same region of the domain as the SB transposon integrations in our model is associated with HCC development in mice(30, 31). AAV integrations were found to alter expression of several domain members, preventing elucidation of a clear molecular mechanism of tumorigenesis. Other studies have also identified correlation between disrupted expression from the Dlk1-Dio3 domain and HCC(32-36), often with several domain members showing aberrant expression. The majority of these studies are correlative in nature, and no attempt is made to validate tumorigenic function of domain members through direct modulation of gene expression.

Here we describe a series of experiments that initially utilized deep-sequencing analyses to obtain detailed gene expression profiles of the SB-induced HCCs. This approach revealed that transposon integration within the Dlk1-Dio3 domain has variable effects on expression of several elements throughout the imprinted domain, but uniformly drives dramatic overexpression of Retrotransposon-like 1 (Rtl1). Validation experiments demonstrate that hepatic overexpression of Rtl1 promotes tumorigenesis in vivo.
Additionally, we find that *RTL1* is aberrantly expressed in ~30% of human HCC samples, suggesting that it may be a relevant therapeutic target.

*Rtl1* is a poorly characterized gene that encodes a predicted transmembrane protein with aspartic protease activity. Interestingly, this locus is derived from domestication of a sushi-ichi-related retrotransposon(37) and is unique to placental mammals(38). This study identifies *Rtl1* as a novel oncogene involved in hepatocarcinogenesis and suggests that its expression may be used as a prognostic indicator and/or targeted therapeutically to improve outcome for patients with HCC. It also represents one of the first direct *in vivo* demonstrations of a role for a co-opted genetic element in driving carcinogenesis.

**Results and Discussion**

*Determining the effect of transposon integration on Dlk1-Dio3 domain members*

We previously reported the identification of a 33 kilobase region of the imprinted *Dlk1-Dio3* domain as a common target of transposon insertion in an SB-induced model of HCC(15) (Figure 2A). Given the domain’s complexity and previous studies demonstrating altered expression of multiple domain members in response to insertion of exogenous DNA(30, 31, 39), we used both transcriptome and miRNA sequencing approaches to obtain expression profiles of eight SB-induced HCCs with *Dlk1-Dio3* integrations and six normal livers for comparison (Figures 2B-C and 3; full data sets can be viewed as supplemental material in Riordan et al., *PLoS Genetics* (40)). Expression of *Dlk1-Dio3* domain miRNAs was low to undetectable in normal liver. Similar results were detected for three of eight tumors, while the remaining five tumors displayed activated expression of several imprinted miRNAs. Thus, transposon insertion in the *Dlk1-Dio3* domain does not consistently alter miRNA expression. Interestingly, tumor samples with elevated expression of imprinted miRNAs also showed enhanced
expression of Meg3 and Rian, suggesting a possible transposon-mediated loss of imprinting effect. Dramatic activation of expression from the locus encoding Rtl1 and Rtl1 antisense (Rtl1as) was observed in all eight SB-induced HCCs, while no expression was detected in normal liver. Notably, elevated expression from this locus is the only event that was consistently observed in all SB-induced HCCs with Dlk1-Dio3 integrations (Figure 2B-C). Because transcription can occur on either strand at this locus(41), strand-specific RT-PCR was performed to determine whether the observed increase resulted from expression of Rtl1, Rtl1as, or a combination of both transcripts. As shown in Figure 4A, reads from the locus encoding Rtl1 and Rtl1as detected in HCCs were derived primarily from transcription of the protein-coding sense strand (i.e. Rtl1). The lack of detectable Rtl1 in normal liver suggests that transposon integration results in activation of a normally transcriptionally silent allele.

**Integrated transposons directly drive Rtl1 expression**

As we previously reported, SB transposon integration sites in HCC samples clustered near the 5’ end of Rian within the Dlk1-Dio3 domain(15). Our initial characterization of transposon integrations was performed using ligation-mediated (LM)-PCR followed by pyrosequencing. It has been shown that this approach yields suboptimal sequencing depth for confident identification of clonal insertion sites(42). To ensure adequate sequence coverage, the SB-induced HCCs were re-sequenced for the current study using the Illumina platform. Surprisingly, while integrations near the 5’ end of Rian were still found to be the most common event, a transposon orientation bias was revealed that had not previously been evident. For many of the tumors, multiple transposon integrations were identified in this region, and for each of the tumors at least one of these integrations was in the same orientation as Rtl1 (Figure 2A). To validate the significance of transposon integrations upstream of Rtl1 in SB-induced HCCs, insertion sites from a larger set of tumors, as well as some normal livers(43), were sequenced using
the Illumina platform. A quantitative analysis of all transposon integrations in the Dlk1-Dio3 domain for these samples is provided in Figure 5. Consistent with recent studies demonstrating minimal insertion bias for SB transposon integration(17, 44), background insertion sites identified in normal liver and HCC samples did not show any evidence for preferential integration within the Dlk1-Dio3 domain. In contrast, clonal sites identified in tumors were highly enriched upstream of Rtl1, suggesting positive selection for insertions in this region during the process of tumorigenesis. This analysis further confirmed that transposon integrations in the same transcriptional orientation as Rtl1 are preferentially detected specifically in HCCs. Based on these results, we hypothesized that the high levels of Rtl1 observed in tumors were driven directly by transposons integrated upstream. Amplification of transposon/Rtl1 fusion products from cDNA confirmed transposon-driven Rtl1 overexpression for each of the tumors harboring integrations in this region (Figure 4B). Two different sizes of fusion products were detected, representing direct splicing of the T2/Onc3 transposon into Rtl1 (smaller product) or inclusion of a cryptic upstream exon (larger product). Importantly, both fusion products encode the full Rtl1 open reading frame and are thus predicted to drive overexpression of functional Rtl1 protein.

Two additional Sleeping Beauty screens have been reported in which liver tumors were generated and characterized(24, 26). Neither of these studies identified the Dlk1-Dio3 domain as a common site of integration. Both screens utilized T2/Onc mice as the source of mutagenic transposons. This transposon is similar in structure to that of the T2/Onc3 strain used in our study, but a distinct promoter is included within the transposon. T2/Onc transposons contain the murine stem cell virus (MSCV) 5’ long-terminal repeat (LTR) promoter, while T2/Onc3 transposons contain the cytomegalovirus (CMV) enhancer/chicken β-actin (CAG) promoter. Differences in promoter activities likely affect the profile of mutations that are selected for in tumors resulting from SB mutagenesis. We suspect that the MSCV promoter may be too weak to overcome the
influence of imprinting within the Dlk1-Dio3 domain to drive sufficient hepatic Rtl1 expression to provide cells with a selective advantage and promote tumorigenesis. The CAG promoter, which has a much higher activity in epithelial cells like hepatocytes, may be better able to drive Rtl1 overexpression when integrated upstream, resulting in frequent selection of cells with such mutations in tumors. Consistent with this idea, insertional mutations upstream of Rtl1 have been linked to liver tumor development in three independent studies that utilized viral vectors containing promoters with high activity in hepatocytes(30, 36, 45).

*Rtl1 expression in cultured hepatocytes promotes growth in ECM*

Our RNA profiling analyses and fusion transcript detection led us to conclude that the primary tumor-driving event under positive selection in SB-induced HCCs is activation of Rtl1. While we cannot exclude the possibility that other domain members play a role independently and/or cooperatively with Rtl1, in our model it seems to be the dominant driver of hepatocarcinogenesis. It should be noted that other models of HCC have been described in which altered expression of maternal Dlk1-Dio3 domain members is observed in the absence of Rtl1 activation(46), suggesting that distinct roles may exist for both paternal and maternal components of the domain in different subtypes of HCC.

To study the effects of Rtl1 overexpression on hepatocyte growth and morphology in vitro, we stably overexpressed it in the murine hepatocyte cell line TIB-73. Importantly, this cell line is non-tumorigenic and lacks endogenous expression of Rtl1. Based on the predicted protein structure of Rtl1, which contains an extracellular protease domain, we hypothesized that its effects may be mediated via cleavage of a substrate within the extracellular matrix (ECM). To test this hypothesis, TIB-73 cells expressing either Rtl1 or an empty vector were embedded in a matrix of Matrigel, plated in 24-well plates, and cultured in serum-free medium. Two weeks after plating, cells expressing
Rtl1 had grown to form dozens of cyst-like colonies composed of several cells (Figure 6D). In contrast, cells lacking Rtl1 expression formed less than one colony per well on average, and colonies that did form were much denser and smaller (Figure 6C). These results demonstrate that Rtl1 expression promotes growth of hepatocytes in the presence of ECM in the context of physiologically relevant levels of growth factors, and they are consistent with our hypothesis that Rtl1 acts by cleaving an ECM component. ECM is an important aspect of the tumor microenvironment, particularly in the liver. The process of liver fibrosis, which involves ECM remodeling and expansion, is strongly linked to HCC, with nearly 90% of cases developing in this context(18). One mechanism by which fibrosis may contribute to the development of HCC is through sequestration of growth factors in the newly remodeled ECM(21). According to this model, subsequent release of growth factors through protease-mediated cleavage of ECM components promotes proliferation of adjacent hepatocytes. Our results suggest that Rtl1 may contribute to hepatocarcinogenesis via this mechanism.

In vivo hepatic Rtl1 expression drives tumorigenesis

We next sought to determine if Rtl1 overexpression is sufficient to promote hepatocarcinogenesis in vivo. Mice with stable hepatic expression of Rtl1 were generated by hydrodynamic tail vein injection of transposon-based expression constructs(47) into Fah-deficient male mice expressing SB transposase(26). Selective repopulation of the liver was achieved through inclusion of a separate Fah expression vector that allowed stably transfected cells to survive withdrawal of NTBC(48), an event that triggers the death of Fah-null hepatocytes. Mice were euthanized nine months post-injection to assess liver tumorigenesis. Of fourteen mice injected with Rtl1 overexpression constructs, twelve (86%) developed liver tumors, with an average of 2.9 tumors per mouse (Table 1 and Figure 7). In another experimental condition, a third construct encoding a short hairpin directed against Trp53 was additionally included. Loss of p53 function is one of
the most commonly observed molecular abnormalities in human HCC, occurring in
~30% of cases and making this a relevant context in which to validate putative oncogenes.

Of twelve mice injected with all three transposon constructs, ten (83%) developed liver
tumors, with an average of 4.3 tumors per mouse. Six of the mice from this cohort were
sacrificed at time points earlier than nine months. When considering only those mice that
were aged for nine months to allow direct comparison between the two experimental
groups, five of six (83%) mice with p53 knockdown in addition to Rtl1 overexpression
developed liver tumors, with an average of 6.7 tumors per mouse. This is significantly
higher (p=0.027) than the number of tumors per mouse developed with Rtl1
overexpression alone. Knockdown of p53 in tumors was assessed by western blot
(Figure 8A). Although efficiency was somewhat variable, the majority of tumors showed
significant knockdown. It has been shown that following liver repopulation, the Fah
mouse model is predisposed to tumor formation in the absence of any additional
transgene(23, 49). The tumors that develop in this context uniformly lack expression of
Fah. We assessed expression of both Rtl1 and Fah by RT-PCR in fourteen tumors
developed following hydrodynamic injection (Figure 8B). Of these fourteen tumors,
eleven were found to express both genes. This result suggests that while a small subset
of our tumors are likely background events developed independently of Rtl1 expression
due to the model’s predisposition, the majority of tumors were induced directly by
overexpression of Rtl1. Further evidence for the tumorigenic activity of Rtl1 in vivo
comes from a recently published study showing that liver tumors develop in mice
following hepatic lentiviral delivery(36).

**RTL1 activation in human HCC**

In order to determine the prevalence of RTL1 activation in human disease, RT-
PCR was performed on a collection of thirty-three human HCC RNA samples, along with
matched benign adjacent liver tissue (Figures 9A-B). A lack of significant expression
was observed for all but one of the benign liver samples. In contrast, significant activation of \textit{RTL1} was detected in 30\% (10/33) of analyzed tumors. To assess \textit{RTL1} expression in another set of human HCCs, we utilized RNASeq data available through The Cancer Genome Atlas (TCGA) consortium. Consistent with our initial analysis, \textit{RTL1} expression was found to be significantly activated in 30\% (10/33) of analyzed tumors (Figure 9C). Low-level expression was detected in two of the adjacent benign tissue samples for which sequence data was available. It should be noted that four of the tumor samples included in the TCGA dataset overlap with the initial set of 33 samples analyzed by RT-PCR. No expression of \textit{RTL1} was detected in these four samples by either analysis. A notable gender disparity is observed in human HCC, wherein men are around three times more likely to develop the disease than women(1). We analyzed our human expression data to determine if \textit{RTL1} overexpression was associated with tumors from one gender or the other, but failed to detect evidence of any bias. Based on the combined set of human samples that we analyzed, \textit{RTL1} was found to be overexpressed in samples from 12/38 males (32\%) and 8/24 females (33\%).

Unfortunately, there is very little existing data on the expression of \textit{RTL1} in disease states, including cancer. Most expression analyses utilize commercially available microarray platforms, the vast majority of which lack probes for \textit{RTL1}. While multiple studies have identified correlative links between disrupted expression of other \textit{DLK1-DIO3} domain members and HCC(30-35), expression of \textit{RTL1} has not typically been assessed. This may be due in part to the fact that \textit{RTL1} is a single exon gene, preventing straightforward design of primers that specifically amplify from cDNA and not genomic DNA. Notably, we have utilized a method for \textit{RTL1} expression analysis that adds a unique sequence tag during reverse-transcription(50), thus allowing specific amplification from cDNA and eliminating background amplification from genomic DNA.

In the setting of spontaneous hepatocarcinogenesis in humans, \textit{RTL1} activation may occur as a result of loss of imprinting (LOI) within the \textit{DLK1-DIO3} domain.
Epigenetic abnormalities are known to play a large role in driving tumor development and progression, in part through induction of LOI(51). A direct causal role for LOI in cancer was demonstrated by Holm et al., who showed that chimeric mice created using embryonic stem cells lacking imprinting-specific DNA methylation develop multiple tumor types with nearly complete penetrance(52). The most common tumor type observed was HCC, suggesting that LOI in the liver confers a strong predisposition to cancer. While expression from the Dlk1-Dio3 domain was not examined in the study, the results we present here suggest that hepatic activation of Rtl1 may be a driving factor in the HCCs that were developed. Interestingly, Wang et al. reported loss of methylation within the Rtl1 locus in mouse HCCs resulting from AAV integration(31), although effects on Rtl1 expression were not determined. To assess whether or not Rtl1 overexpression is associated specifically with altered expression of other imprinted genes in our SB-induced HCCs, analysis of variance (ANOVA) was conducted on the whole transcriptome to identify genes with differential expression between Rtl1-overexpressing tumors and normal liver. Following Bonferroni correction, 3 of 125 imprinted genes and 474 of 20,707 non-imprinted genes were identified as having significantly different expression between the two sample sets. By Fisher’s exact test, these proportions are not significantly different (p=0.760). This analysis shows that activation of Rtl1 does not correlate specifically with altered expression of other imprinted genes in our tumors.

*Rtl1*-expressing mouse HCCs resemble human S1 subclass

Next we sought to determine if Rtl1-induced HCCs in mice resemble a specific subtype of human HCC. An integrative meta-analysis of human HCC gene expression profiles has identified three major expression subtypes called S1, S2, and S3(53). Transcriptome sequencing data from the mouse HCCs overexpressing Rtl1 was used to determine the extent to which these SB-induced tumors resemble human HCC. Expression levels of genes defining the S1, S2, and S3 subclasses of human HCC were
assessed for each of the SB-induced tumors and normal liver samples. Unsupervised clustering of samples based on expression of constituent genes was performed individually for each subclass. The results show that the SB-induced tumors resemble human HCCs within the S1 subclass (Figure 10). This was further supported by Gene Set Enrichment Analysis (GSEA)(54, 55) that showed a statistically significant association (p=0.039) between Rtl1-induced HCCs and the S1 expression class. Immunohistochemistry was performed to validate protein expression of two S1 subclass genes in SB-induced HCC (Figure 11). This subclass of human HCC is associated with poor to moderate cellular differentiation, activation of the WNT signaling pathway, and early tumor recurrence.

**Potential of RTL1 as a therapeutic target and/or biomarker**

*Rtl1* is a poorly characterized gene that encodes a predicted transmembrane protein with aspartic protease activity. Knockout studies in mice have demonstrated a role in the placental feto-maternal interface(56), but functional studies in other tissues are lacking. Experiments to determine the necessity of Rtl1’s protease domain for its ability to promote tumorigenesis and to identify targets of its activity will help to clarify the oncogenic mechanism. If required, RTL1’s protease activity represents a promising target for therapeutic intervention in HCC patients. Pepstatin is a naturally occurring bacterial peptide that demonstrates broad potential to inhibit aspartic proteases(57). Additionally, more specific inhibitors have successfully been developed that target the activity of other aspartic proteases, including renin(58) and HIV-1 protease(59). It is also possible that *RTL1* expression could be a useful biomarker for HCC. Based on the human samples that we analyzed, its expression appears to be highly tumor-specific. Although low-level expression was detected in three non-tumor liver samples, all of the benign samples came from HCC patients and are therefore unlikely to be representative of truly normal liver.
Conclusion

In this study we identify Rtl1, a co-opted imprinted gene, as a novel driver of hepatocarcinogenesis. Mutations resulting in its overexpression were highly selected for in liver tumors developed using a forward genetic screen. While several correlative results linking the Dlk1-Dio3 domain to HCC development have been reported, our study provides direct evidence that modulation of a domain member in vitro and in vivo promotes a tumorigenic phenotype. We show here that overexpression of Rtl1 in cultured hepatocytes results in an increased growth ability in extracellular matrix. We also show that overexpression via hydrodynamic gene delivery results in highly penetrant liver tumor formation in mice. Additionally, a subset of human HCCs displays overexpression of RTL1, suggesting it may be a relevant therapeutic target for patients.

Materials and Methods

Mice

SB-induced mouse HCCs used in this study were generated as previously described(15). All tumors used in this study came from male mice and were collected using procedures approved and monitored by the Institutional Animal Care and Use Committees at the National Cancer Institute-Frederick and the University of Minnesota.

Human tissue samples

Paired tumor and benign liver tissues were obtained from 33 patients undergoing resections for HCC at Mayo Clinic between 1987 and 2003, snap-frozen in liquid nitrogen, and stored at −80 °C. The Mayo Clinic Institutional Review Board approved the study.
RNA sequencing and data analysis

Transcriptome sequencing. Total RNA was collected from SB-induced HCC and normal liver samples using the miRNeasy kit (Qiagen). Library preparation and sequencing were performed using Illumina’s mRNA-Seq workflow. For data normalization, the raw number of reads for each transcript was converted to reads per kilobase per million mapped reads (RPKM)(60). This was followed by \( \log_2 \) transformation of the RPKM value + 1. Unsupervised clustering was performed on samples based on normalized expression of genes with variation in Euclidean distance among samples of at least 2.5 standard deviations using Cluster 3 software(61). Heat maps were generated using Java TreeView software(62).

miRNA sequencing. Total RNA was collected from SB-induced HCC and normal liver samples using the miRNeasy kit (Qiagen). The flashPAGE Fractionator system (Life Technologies) was used to isolate RNAs shorter than 40nt. Library preparation and sequencing were performed using the SOLiD small RNA expression workflow (Life Technologies). For data normalization, the raw number of reads for each miRNA was converted to reads per 100,000 mapped reads. This was followed by \( \log_2 \) transformation of the normalized value + 1. Unsupervised clustering was performed on samples based on normalized expression of genes with variation in Euclidean distance among samples of at least 1.5 standard deviations using Cluster 3 software(61). Heat maps were generated using Java TreeView software(62).

RT-PCR

Strand-specific RT-PCR to detect expression of Rtl1 and Rtl1as. One nanogram total RNA was used as template for cDNA synthesis with AMV reverse transcriptase (New England Biolabs). The cDNA synthesis reaction was primed with oligonucleotides complementary to Gapdh (Gapdh_R: 5’–TGTAGGCCATGAGGTCCACCAC–3’) and
either Rtl1 (Rtl1_R: 5’–GGAGCCACTTCATGCCTAAGACGA–3’) or Rtl1as (Rtl1as_R: 5’–GTGGAGAACCTCGCTGTCATCGC–3’). PCR was performed with primers to detect transcripts for Gapdh (Gapdh_F: 5’–TTGTCTCCTCGACTTCAA–3’ and Gapdh_R (amplicon 150bp)), Rtl1 (Rtl1_F: 5’–TACTGCTCTTGAGAGGTGGACCC–3’ and Rtl1_R (amplicon 297bp)), or Rtl1as (Rtl1as_F: 5’–TCTCCACTCGAGGGGTACTCCACCT–3’ and Rtl1as_R (amplicon 298bp)).

Transposon/Rtl1 fusion transcript detection. One microgram total RNA was used as template for oligoD-primed cDNA synthesis with Superscript III reverse transcriptase (Life Technologies). Control reactions lacking the RT enzyme were also performed. PCR was performed with a forward primer within the transposon splice donor (SD_F: 5’–AAGCTTGTACTAGCACCAGACG–3’) and reverse primer within Rtl1 (Rtl1_R2: 5’–TTCCCTGGGCTGGCCACTATC–3’) (amplicon 394bp or 459bp, depending on splicing pattern).

Detection of RTL1 in human HCCs. Five hundred nanograms total RNA was used as template for cDNA synthesis with AMV reverse transcriptase (New England Biolabs). The cDNA synthesis reaction was primed with oligonucleotides complementary to TBP (TBP_R: 5’–GCCATAAGGCATCCCATTTGAC–3’) and RTL1 (RTL1_RT_tag: 5’–GTAATACGACTCCTAGGGCCTCGATAGGGGAGATGTTGC–3’). The RTL1_RT_tag primer adds a unique 22 base sequence (underlined) to the 5’ end of the newly synthesized cDNA. Control reactions lacking the RT enzyme were also performed. PCR was performed with primers to detect transcripts for TBP (TBP_F: 5’–GCTGAGAAGAGTGTGCTGGA–3’ and TBP_R (amplicon 204bp)) and RTL1 (RTL1_F: 5’–TTCTACTGGGGAGGAGTCCAGGA–3’ and RT_tag_R: 5’–
GTAATACGACTCACTATAGGGC–3’ (amplicon 238bp)). RT_tag_R binds to the unique sequence tag added during cDNA synthesis, minimizing the potential for amplification from contaminating genomic DNA. Formamide was included in the PCR mix at a final concentration of 3% or 5% for amplification of TBP or RTL1, respectively. Gel images were processed using VisionWorks®LS image analysis software (UVP) to obtain intensity values for each lane. For each sample, values were normalized by subtracting the - RT lane from the + RT lane, then dividing each corrected value by the average corrected intensity value in benign samples lacking detectable expression.

**Detection of Rtl1 and Fah in tumors induced by hydrodynamic injection.** One microgram total RNA was used as template for oligo(dT)-primed cDNA synthesis with Superscript III reverse transcriptase (Life Technologies). Control reactions lacking the RT enzyme were also performed. PCR was performed with primers to detect Rtl1 (Rtl1_F2: 5’–GTGGAGAACTTCGCTGTCATCGC–3’ and Rtl1_R2: 5’–TCTCCACTCGAGGGTACTCCACCT–3’ (amplicon 298bp)), Fah (Fah_F: 5’–CTTCTGCGACAATGCACCT–3’ and Fah_R: 5’–ACCACAATGGAGGAAGCTCG–3’ (amplicon 172bp)), or Tbp (Tbp_F: 5’–CTATCACTCCTGCCACACCA–3’ and Tbp_R: 5’–CAGTTGTCCGTGGCTCTTT–3’ (amplicon 189bp)).

**Illumina sequencing of transposon insertions**

DNA from SB-induced tumors was prepared for sequencing of transposon integration sites as previously described(42).

**Matrigel growth assay**

Stable cell lines were generated by delivery of piggyBac transposon constructs encoding either Rtl1 or an empty vector into TIB-73 (ATCC: BNL CL.2) cultured mouse hepatocytes. 24-well plates were coated with a thin layer of Matrigel basement
membrane mix (BD Biosciences) and allowed to set up for 30 minutes at 37°C. For each stable cell line, cells were trypsinized and washed with PBS before resuspension of 5,000 cells in additional Matrigel. The resuspended cells were plated on top of the thin layer of basement membrane mix and allowed to set up, followed by addition of serum-free, low-glucose DMEM (Life Technologies). Images were taken two weeks after plating.

**Hydrodynamic gene delivery**

Hydrodynamic tail vein injection into *Fah*-deficient male mice expressing SB11 transposase was performed as previously described (26). A plasmid expressing *Rtl1* from the human *PGK* promoter and flanked by SB transposon inverted repeat/direct repeats (IR/DRs) was generated by amplifying the open reading frame of *Rtl1* from C57Bl/6J mouse genomic DNA and subcloning it into pT2/PGK-pA. This plasmid was co-injected with PT2/PGK-FAHIL, a plasmid containing an SB IR/DR-flanked expression cassette for *Fah* and firefly luciferase. Some mice were additionally injected with pT2/shp53, a plasmid containing an SB IR/DR-flanked expression cassette for a short-hairpin RNA directed against *Trp53* (48, 63).

**Western blotting**

Total protein was collected from liver tumor samples by homogenization in RIPA lysis buffer. Samples were boiled for five minutes in a reducing buffer and SDS-PAGE was performed. Proteins were transferred to nitrocellulose membranes for blotting. Primary antibodies used were anti-p53 (Cell Signaling Technology #2524), anti-GFP (Clontech #632380), and anti-β-tubulin (Sigma-Aldrich #T4026).
**Gene Set Enrichment Analysis (GSEA)**

GSEA(54, 55) was performed using default parameters. Analyzed gene sets were comprised of all the genes defining human HCC subclasses S1, S2, and S3(53) for which mouse orthologs have been annotated.

**Immunohistochemistry**

Formalin-fixed, paraffin-embedded liver samples were sectioned to a thickness of 4µm and baked onto glass slides. Samples were de-paraffinized, rehydrated, and treated with citrate antigen unmasking solution (Vector Laboratories). Endogenous peroxidase activity was blocked by treatment with a 3% solution of hydrogen peroxide for fifteen minutes. The anti-rabbit ImmPRESS reagent kit (Vector Laboratories) was used for immunolabeling with primary antibodies anti-Fyb (Abgent #AJ1306a) and anti-Ier3 (Abgent #AP11790a). Both primary antibodies were diluted 1:100 and incubated with samples for one hour at room temperature. The ImmPACT DAB kit (Vector Laboratories) was used for detection. Sections were counterstained with hematoxylin QS (Vector Laboratories) and mounted in Permount (Fisher Scientific) for light microscopy.
Figure 2. Dlk1-Dio3 domain transposon integration sites in SB-induced HCC and effects on domain expression

(A) The Dlk1-Dio3 imprinted domain spans ~800 kilobases at the distal end of mouse chromosome 12 (human chr14q32). Three protein-coding genes are expressed from the paternal allele (Dlk1, Rtl1, and Dio3). The maternal allele encodes four IncRNAs (Meg3, Rtl1as, Rian, and Mirg), as well as several miRNAs and snoRNAs. SB transposon and AAV integration sites found to be associated with HCC development in mice are depicted. (B) Intensity plot showing normalized expression levels of long transcripts within and surrounding the Dlk1-Dio3 domain in SB-induced HCCs and normal livers. (C) Intensity plot showing normalized expression of Dlk1-Dio3 domain miRNAs. miRNAs contained within IncRNAs are indicated. miRNAs with no detected expression across all samples were omitted.
Figure 3. Heat maps depicting global differential transcript expression in SB-induced HCCs and normal liver

(A) RNASeq results. Unsupervised clustering was performed based on genes with normalized expression values varying among samples by at least 2.5 standard deviations. For genes with more than one associated transcript, the NCBI RefSeq accession number is indicated. *Rtl1* is indicated with a red arrow. (B) Small RNA sequencing results. Unsupervised clustering was performed based on miRNAs with normalized expression values varying among samples by at least 1.5 standard deviations. *Dlk1-Dio3* domain miRNAs are listed in red.
Figure 4. Integrated transposons drive overexpression of \textit{Rtl1}

(A) Strand-specific RT-PCR detected activation of \textit{Rtl1} expression in SB-induced HCCs, with minimal activation of \textit{Rtl1as} observed. No expression of either transcript was detected in normal liver. (B) Transposons integrated upstream of \textit{Rtl1} drive its expression by generating fusion transcripts. Transcription initiated from the CAG promoter within the transposon splices into \textit{Rtl1}, either directly or via inclusion of an upstream cryptic exon. PCR to detect fusion transcripts was performed on cDNA from SB-induced HCCs using the indicated primers. Transposon-driven expression of \textit{Rtl1} was detected for all of the tumor samples harboring \textit{Dlk1-Dio3} domain integrations. SD, splice donor.
Figure 5. Transposon integrations are preferentially detected upstream of *Rtl1* in SB-induced HCCs

(A) For this analysis, the *Dlk1-Dio3* domain was divided into eleven distinct regions defined by constituent genes and their promoter regions. (B) Quantification of *Dlk1-Dio3* domain transposon integrations in the livers of SB mice. A comprehensive analysis of all transposon insertions within chromosome 12 detected in six normal livers and thirty-four HCCs from SB mice. Insertions are grouped into three distinct categories based on whether they were detected in normal liver or tumor tissue and whether they were identified as clonal or background (i.e. subclonal) (none of the sites identified in normal tissue were identified as clonal). The bar graph shows percentages of all chromosome 12 insertions that fall within the intervals defined in panel A. The actual values used to generate the graph are shown in the table. (C) Analysis of this larger set of tumors confirms the selection and orientation bias for transposon integrations upstream of *Rtl1*. Filled arrowheads represent transposons with the same transcriptional as *Rtl1* and unfilled arrowheads represent transposons with the opposite orientation. Each arrowhead represents a clonal insertion detected in a separate tumor sample. IGR, intergenic region. bkg, background.
Figure 6. Rtl1 promotes growth of cultured hepatocytes in extracellular matrix

Two weeks after plating cultured hepatocytes in a matrix of Matrigel, cells transfected with an empty vector construct (A) failed to grow significantly. Cells transfected with an Rtl1 expression construct (B) grew to form several large colonies. (C-D) Rtl1 promotes growth of large cyst-like structures. Increased magnification reveals that cells lacking Rtl1 (C) form small, dense colonies, while those expressing Rtl1 (D) form large cyst-like colonies composed of several cells. (E) Quantification of colonies per well formed by each cell line in a 24-well plate. The results depicted are based on three experimental replicates per condition and are representative of experiments conducted on three separate days. Scale bars = 0.5cm (A-B) and 100µm (C-D).
Figure 7. *In vivo* hepatic overexpression of *Rtl1* promotes tumorigenesis

(A-C) Macroscopic images of whole livers containing solid tumors from mice injected with *Rtl1* overexpression constructs via hydrodynamic tail vein injection. Mice were euthanized and livers collected nine months post-injection. (D) A normal liver from a hydrodynamically injected mouse is shown for comparison. Injections were performed in the lab of David Largaespada at the University of Minnesota. Scale bars = 1cm.
Figure 8. Transgene expression in tumors induced by hydrodynamic injection

(A) Confirmation of Trp53 knockdown in tumors from mice injected with p53 short hairpin construct. Western blotting was used to detect the presence of the pT2/shp53 construct and its knockdown efficiency in tumors. Detection of GFP indicates presence of the construct, which also contains a GFP expression cassette. To assess the degree of knockdown, Trp53 signal for each sample was normalized to beta-tubulin signal from the same sample. For each tumor, this ratio was normalized to the ratio obtained for a tumor developed following hydrodynamic injection without the p53 hairpin construct (Rtl1 only tumor). These normalized values are plotted in the graph below. (B) RT-PCR on cDNA (RT +) confirmed expression of Rtl1 and Fah in eleven of fourteen tumors. Control reactions performed without reverse transcriptase (RT -) are also shown. Amplification of Tbp was included as a control for cDNA quality.
Figure 9. Expression of *RTL1* in human HCC

(A) *RTL1* expression in human HCC and matched benign liver samples was analyzed by RT-PCR. Plotted values represent normalized band intensities from imaged gels (B). In addition to RT-PCR (RT +), control reactions performed without reverse transcriptase (RT -) are shown. The threshold above which a sample was scored as positive for significant *RTL1* expression (dashed line) was set at three standard deviations above the average intensity value in benign samples lacking detectable expression. For the one patient with significant *RTL1* expression detected in benign tissue, the matched HCC sample also displayed expression (indicated with arrows). (C) Plot of *RTL1* expression in human HCC and normal liver samples based on RNASeq data available through TCGA. The threshold above which a sample was scored as positive for significant *RTL1* expression (dashed line) was set at one standard deviation above the average expression level in tumor-free liver. RSEM, RNASeq by Expectation Maximization. *TBP*, *TATA*-binding protein. B, benign. T, tumor.
Expression levels for the gene sets defining human HCC subclasses S1, S2, and S3 were analyzed in SB-induced HCCs and normal livers. Gene Set Enrichment Analysis (GSEA) was conducted for each subclass independently to assess the significance of differential expression between tumor and normal samples. Heat maps generated by GSEA are shown. This analysis revealed a significant ($p=0.039$) overexpression of the genes defining human subclass S1 in SB-induced HCCs, as compared to normal liver.
Figure 11. Validation of S1 subclass protein expression in SB-induced HCC

Immunohistochemistry was used to confirm altered expression of two proteins from the human HCC subclass S1 gene set. (A-B) Staining for FYN binding protein (Fyb) was performed on normal liver (A) and HCC tissue (B) from a mouse with SB-induced HCC. Though detected in both tissues, the staining pattern in normal liver is more diffuse. Regions of higher staining density are detected specifically in the tumor. (C-D) Staining for Immediate early response 3 (Ier3) was performed on sections from the same tissue samples shown in panels A-B. No significant expression was detected in normal liver, while several regions of high density staining were detected in the tumor. (E) Section of the same tumor shown in panels B and D for which the primary antibody was omitted. Scale bars = 100µm.
### Table 1. Liver tumors developed in hydrodynamically injected mice

<table>
<thead>
<tr>
<th>Mouse ID</th>
<th>Experimental Group</th>
<th>Days post-injection</th>
<th>Liver Nodules</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1381</td>
<td>Rtl1 only</td>
<td>240</td>
<td>4</td>
</tr>
<tr>
<td>M1382</td>
<td>Rtl1 only</td>
<td>240</td>
<td>1</td>
</tr>
<tr>
<td>M1383</td>
<td>Rtl1 only</td>
<td>240</td>
<td>0</td>
</tr>
<tr>
<td>M1384</td>
<td>Rtl1 only</td>
<td>240</td>
<td>6</td>
</tr>
<tr>
<td>M1375</td>
<td>Rtl1 only</td>
<td>251</td>
<td>1</td>
</tr>
<tr>
<td>M1376</td>
<td>Rtl1 only</td>
<td>251</td>
<td>3</td>
</tr>
<tr>
<td>M1482</td>
<td>Rtl1 only</td>
<td>251</td>
<td>5</td>
</tr>
<tr>
<td>M1483</td>
<td>Rtl1 only</td>
<td>251</td>
<td>1</td>
</tr>
<tr>
<td>M1484</td>
<td>Rtl1 only</td>
<td>251</td>
<td>4</td>
</tr>
<tr>
<td>M1485</td>
<td>Rtl1 only</td>
<td>251</td>
<td>3</td>
</tr>
<tr>
<td>M1491</td>
<td>Rtl1 only</td>
<td>251</td>
<td>5</td>
</tr>
<tr>
<td>M1492</td>
<td>Rtl1 only</td>
<td>251</td>
<td>4</td>
</tr>
<tr>
<td>M1501</td>
<td>Rtl1 only</td>
<td>251</td>
<td>0</td>
</tr>
<tr>
<td>M1502</td>
<td>Rtl1 only</td>
<td>251</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Avg</td>
<td></td>
<td>2.86</td>
</tr>
<tr>
<td>M1371</td>
<td>Rtl1 + shp53</td>
<td>120</td>
<td>0</td>
</tr>
<tr>
<td>M1372</td>
<td>Rtl1 + shp53</td>
<td>120</td>
<td>2</td>
</tr>
<tr>
<td>M1373</td>
<td>Rtl1 + shp53</td>
<td>120</td>
<td>2</td>
</tr>
<tr>
<td>M1374</td>
<td>Rtl1 + shp53</td>
<td>120</td>
<td>1</td>
</tr>
<tr>
<td>M1441</td>
<td>Rtl1 + shp53</td>
<td>184</td>
<td>2</td>
</tr>
<tr>
<td>M1442</td>
<td>Rtl1 + shp53</td>
<td>184</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Avg</td>
<td></td>
<td>2.00</td>
</tr>
<tr>
<td>M1391</td>
<td>Rtl1 + shp53</td>
<td>251</td>
<td>5</td>
</tr>
<tr>
<td>M1392</td>
<td>Rtl1 + shp53</td>
<td>251</td>
<td>0</td>
</tr>
<tr>
<td>M1394</td>
<td>Rtl1 + shp53</td>
<td>251</td>
<td>6</td>
</tr>
<tr>
<td>M1451</td>
<td>Rtl1 + shp53</td>
<td>254</td>
<td>5</td>
</tr>
<tr>
<td>M1454</td>
<td>Rtl1 + shp53</td>
<td>254</td>
<td>16</td>
</tr>
<tr>
<td>M1455</td>
<td>Rtl1 + shp53</td>
<td>254</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Avg</td>
<td></td>
<td>6.67</td>
</tr>
</tbody>
</table>
CHAPTER 3

A SLEEPING BEAUTY MUTAGENESIS SCREEN IDENTIFIES MUTATIONS DRIVING HEPATOCELLULAR CARCINOMA IN FIBROTIC LIVER

Abstract

Hepatocellular carcinoma (HCC) development is strongly associated with an underlying fibrotic liver environment. All of the major risk factors for HCC are associated with chronic liver injury leading to hepatic fibrosis; as such, ~90% of cases occur in this context. Despite its status as the third leading cause of cancer-related deaths worldwide, our understanding of the molecular pathogenesis of HCC is limited, particularly as compared to that of other prevalent cancers. Animal models of cancer have proven highly valuable for identifying molecular drivers of tumorigenesis. A key to the success of such a model is its ability to accurately recapitulate human disease. Unfortunately, the majority of existing animal models of liver cancer fail to replicate the fibrotic environment in which most human HCC develops. In order to discover molecular drivers of HCC relevant to human cancer, we conducted a liver-specific Sleeping Beauty (SB) transposon mutagenesis screen in mice with variable induction of hepatic inflammation and fibrosis caused by injection of carbon tetrachloride (CCl₄). By comparing mutation profiles between tumors developed in the presence or absence of fibrosis, we identified several candidate cancer genes preferentially mutated in a fibrotic liver environment. In addition to genes with an established role in human HCC, such as MET and β-catenin, we identified many novel candidate genes whose disruption was commonly selected for in fibrotic SB-induced tumors. These candidates, which include Fign, Gli2, and Nfe2, represent promising loci to study further and potentially target therapeutically for HCC.
Introduction

Each year ~600,000 people throughout the world die from liver cancer, making it the third leading cause of cancer-related death(1). Hepatocellular carcinoma (HCC), which accounts for >80% of liver cancer cases, most commonly occurs in patients afflicted with hepatic fibrosis(18). This condition is characterized by the accumulation of excess extracellular matrix (ECM) in the liver and a resultant increase in tissue stiffness. It also includes an angiogenic response that involves the formation of new blood vessels and remodeling of existing vasculature(64). During fibrosis development, normal turnover of ECM is impaired. At the same time, the liver’s population of stellate cells is expanded and activated to synthesize and deposit excess matrix proteins, resulting in substantial accumulation of ECM. Induction of fibrosis, which is essentially scarring of the liver, represents a wound-healing response to a liver injury stimulus, and it generally includes activation of inflammatory signaling. While acute injury can activate mechanisms of fibrogenesis, a sustained stimulus is required for significant accumulation to occur. In some cases, the inflammatory response to hepatic injury subsides as progressive fibrosis develops; in others it remains as a significant component of the damaged liver environment. All of the major risk factors for HCC, which include HBV or HCV infection, alcohol-induced liver damage, and NAFLD, provide a chronic injury stimulus that leads to significant hepatic fibrosis. Without resolution of the underlying injury stimulus, liver fibrosis progressively increases in severity. This can eventually lead to the development of cirrhosis, the most advanced stage of fibrosis. While milder cases of fibrosis can be reversed if the injury stimulus is removed, cirrhosis is generally considered to be irreversible. This condition is characterized by increased deposition of scar tissue, loss of structural architecture, portal hypertension resulting from intrahepatic resistance to blood flow, and the formation of regenerative hepatocyte nodules. HCC is a primary cause of death in patients with liver cirrhosis. One third of these patients will develop HCC in their lifetime, with an annual risk of ~4%(20).
Clearly there is a strong association between liver fibrosis and HCC, as ~90% of cases develop in this context(18). While this association has been recognized for decades, the extent to which fibrosis directly contributes to HCC pathogenesis remains unclear. Whether the two conditions co-occur as a result of shared etiologies or because of a direct role for fibrosis in promoting HCC is a question without a conclusive answer. Several potential fibrosis-dependent mechanisms of hepatocarcinogenesis have been proposed(21), but they require definitive experimental validation. Proposed contributions include enhanced mitogenic and anti-apoptotic signaling through integrins, growth factor secretion by activated hepatic stellate cells, increased stromal stiffness leading to proliferation, and growth factor sequestration by the ECM resulting in autocrine and paracrine signaling. Assessing the relative importance of these and other fibrotic contributions during HCC development is an active area of research with great potential to provide insight into molecular pathways that may be targeted prophylactically in patients with hepatic fibrosis to delay or prevent progression to HCC.

Mouse models serve as an invaluable tool for gaining an increased understanding of the molecular drivers of human cancer. When engineering such models, it is essential to recapitulate the pathogenesis of human disease as closely as possible in order to maximize the relevance and utility of experimental findings(65). Insertional mutagenesis screens using the *Sleeping Beauty* (SB) transposon system mimic the random, stepwise accumulation of somatic mutations that occurs in human cancer and are thus highly useful for cancer gene discovery(66). We present here the results of a liver-specific SB mutagenesis screen conducted in the presence or absence of chronic hepatic fibrosis induced by carbon tetrachloride (CCl₄) injection. Unlike most mouse models of HCC, our SB-induced model takes into account the fibrotic context in which the majority of human HCC cases are developed. By comparing transposon insertion profiles between tumors developed in treated and untreated mice, we have identified several mutation events that are subject to differential selective pressure depending on the fibrotic context.
in which tumors develop. Encouragingly, mutations in some genes known to be involved in human HCC, such as \textit{MET} and \textit{\(\beta\)-catenin}, were commonly detected in tumors developed in a fibrotic liver environment. This result highlights the relevance of our model to human cancer, and it provides support for the hypothesis that other commonly mutated genes identified by our screen, including \textit{Fign}, \textit{Gli2}, and \textit{Nfe2}, may be similarly relevant cancer genes.

In addition to generating a highly relevant mouse model of human HCC with which to identify novel drivers of cancer, our study provides opportunities to clarify the mechanistic role of hepatic fibrosis in HCC development. Mutation events subject to preferential selection in the absence of fibrosis may represent mechanistic targets or mediators of the fibrotic process; such mutations would be under less selective pressure in a fibrotic liver, as their functional consequences could be achieved by the induction of fibrosis alone without requiring additional mutation. Those events preferentially selected in the presence of fibrosis likely cooperate with or require a fibrotic environment to promote tumorigenesis. The set of commonly mutated genes identified by our screen provides a list of putative HCC drivers likely to be highly relevant to human disease. Additionally, these genes represent logical candidates to study for their role in fibrosis-dependent hepatocarcinogenesis.

\textbf{Results and Discussion}

\textit{CCl4-induced liver fibrosis increases severity of a Sleeping Beauty HCC model}

To generate a model of spontaneous HCC development, mice with Cre-inducible expression of SB transposase and harboring T2/Onc3 transposon concatemers (RosaSB\textsuperscript{\text{LsL/LsL}},T2/Onc3\textsuperscript{Tg/Tg})(15) were bred to mice expressing Cre recombinase from a liver-specific \textit{Albumin} promoter (\textit{Alb-Cre}\textsuperscript{Tg/Tg})(67). Resultant offspring frequently developed HCC as a result of restricted mutagenic SB transposon mobilization in the
liver. For this study, only male offspring were aged to assess tumor formation, as a previous SB HCC screen utilizing the Alb-Cre strain reported resistance to tumor development and long latency in female mice(68). Starting at eight weeks of age, liver inflammation and chronic hepatic fibrosis were induced in half (n=75) of the mutagenized cohort by intraperitoneal injection of carbon tetrachloride (CCl₄) twice weekly for a total of twelve weeks. This treatment regimen was sufficient to induce a significant fibrogenic response in all injected mice that was sustained until the experimental endpoint of 65 weeks, though the inflammation subsided. As shown in Figure 12, CCl₄ treatment resulted in variable induction of hepatic collagen deposition and steatosis. This variability likely results primarily from differences in recovery following CCl₄-induced liver damage, as has been observed in other studies of CCl₄-induced hepatotoxicity(69, 70), rather than differences in responsiveness to treatment. Histological tumor sections were analyzed by Masson’s trichrome staining and were assigned a score based on the severity of end-point fibrosis using the METAVIR method(71) (Table 2). Analysis by Fisher’s exact test failed to reveal any statistically significant associations between the fibrotic index and mutation profile of tumors.

All mice in the study were aged until moribund or until formation of palpable tumors. Compared to uninjected mice (n=77), CCl₄-treated mice displayed a tumor phenotype with increased severity (Table 3). Induction of chronic fibrosis led to increased penetrance and multiplicity of tumors, as well as decreased latency. All of the CCl₄-treated mice developed liver tumors, with an average latency of 48.5 weeks. Only 44% of untreated mice developed tumors within the 65-week aging period, with an average latency of 64.5 weeks. Of the mice that developed HCC, the average multiplicities were 5.0 and 1.7 tumors per mouse for the CCl₄-treated and uninjected cohorts, respectively. A control cohort (n=15) of RosaSBase^{Lsl/LsL;T2/Onc3Tg/Tg} mice lacking SB-induced mutagenesis was subjected to CCl₄-induced liver fibrosis to determine the background level of tumorigenesis in this context. When euthanized at 65
weeks of age, a total of seven liver tumors were collected from two of the mice (13%). Of note, the average size of these background tumors induced by CCl4 treatment alone was much smaller than the average tumor size in SB mutagenized mice of both cohorts.

**Identification of common insertion sites in SB-induced HCC**

To identify recurrent mutation events implicated as drivers of hepatocarcinogenesis, transposon integration sites were amplified by ligation-mediated polymerase chain reaction (LM-PCR) and sequenced. Genomic DNA was extracted from 47 and 179 liver tumors induced in the absence or presence of CCl4-induced hepatic fibrosis, respectively. DNA fragments containing transposon/genome junctions were selectively amplified by LM-PCR and sequenced using the Illumina HiSeq platform. Clonally expanded insertion sites were defined based on criteria including their detection in amplicons from both ends of the transposon at an abundance of at least 0.1% of the total reads for a given sample (42). Only these clonally expanded sites were included in further analyses, as they most confidently represent the mutation events that have been subject to positive selection during tumor development and progression. This is an often-overlooked factor in forward genetic screens that is important to minimize the false discovery rate for driver mutation events. To increase our ability to make statistically relevant comparisons between mutation profiles of the two cohorts, insertion sites from an additional 36 SB-induced HCCs developed in the absence of liver fibrosis (43) were included in downstream analyses, bringing the total number to 83. Similarity in the characteristics of the insertion profiles from the two groups of untreated SB-mutagenized liver tumors were verified.

Clonally expanded transposon insertion sites detected in liver tumors from CCl4-treated and untreated mice were analyzed separately to identify common insertion sites (CISs) in each cohort. CISs represent sites in the genome that contain SB transposon insertions at a significantly higher frequency than predicted by random chance, as
calculated using Monte Carlo simulation(15). We additionally utilized a gene-centric CIS (gCIS) classification method(72) to identify putative HCC drivers. Through these analyses, 25 candidate genes were identified in fibrotic liver tumors and 8 candidates were identified in tumors developed in the absence of a fibrotic microenvironment (Table 4). It should be noted that CCl₄ is a non-mutagenic agent when used in mouse studies(73), so chemically induced mutation events are not a complicating factor in our analyses.

Based on the injury stimulus and induction of hepatocyte turnover provided by CCl₄ treatment, it may be expected that fewer mutations would be required to transform hepatocytes in fibrotic liver as compared to non-fibrotic liver. This does not seem to be the case in our model, as the average number of clonally expanded mutations did not significantly differ between CCl₄-treated (10.9 clonal mutations) and untreated (11.4 clonal mutations) mice (p=0.698). This result is consistent with our assumption that CCl₄ provides a minimal mutagenic contribution to tumor development in our model.

*Fibrotic liver environment provides differential selective pressure for mutations*

While four genetic loci (Rian/Rtl1, Hras1, Kras, and Snd1) were identified as putative liver cancer genes in both cohorts of mice, the candidate lists were otherwise non-overlapping. This result is indicative of differential selective pressure for transposon-induced mutations depending on the fibrotic state of the liver. We hypothesize that recurrent mutations identified in the context of hepatic fibrosis are likely to be most relevant to human HCC due to the similarities in microenvironments. Consistent with this hypothesis, *Met* and *β-catenin* were identified as CISs preferentially selected in the presence of fibrosis. Both of these genes have well-established roles in human HCC(74, 75).
Our detection of Met as a CIS preferentially mutated in HCC developed in a fibrotic liver environment makes intuitive sense. All but one of the SB transposons at this locus are integrated upstream of and in the same transcriptional orientation as the Met open reading frame (ORF) (Figure 13A). This pattern of insertion is indicative of a gain-of-function mutational mechanism wherein transcription initiated within integrated transposons splices into downstream exons and drives overexpression of the wild-type ORF, as has previously been observed for other oncogenes (40, 76). Importantly, the overexpressed Met protein, which serves as a cell-surface receptor for hepatocyte growth factor (HGF), is still dependent on ligand binding for activation of its tyrosine kinase activity and induction of proliferative signaling pathways. In a basal state, there is little functional hepatic HGF/Met signaling, demonstrated by the fact that liver-specific knockout of Met produces no overt phenotype (77). Under these conditions, a mutation resulting in overexpression of ligand-dependent Met would not be subject to strong positive selection. Upon induction of liver injury and hepatic fibrosis, the liver’s population of stellate cells is dramatically expanded and activated to secrete HGF, among other cytokines (78). In this environment, overexpressed Met would be functionally active, allowing it to serve as a driver of tumorigenesis. This context provides a strong positive selective pressure for mutations causing Met overexpression, which explains its preferential detection as a CIS in tumors from the CCl4-treated cohort of mice in our study. A single exception to the pattern of Met mutations in our screen is an insertion downstream of the first coding exon, which is predicted to drive overexpression of an in-frame Met ORF lacking the amino-terminal 429 amino acids. It is unclear whether or not truncated Met overexpression is a driving mechanism for this tumor, which also contains a clonally expanded insertion upstream of the Rtl1 locus.

Mutations predicted to drive overexpression of Gli2 were commonly detected in SB-induced liver tumors developed in a fibrotic context, but never detected in tumors developed in the absence of fibrosis (Figure 13B). Gli2 is a transcription factor activated
downstream of Sonic hedgehog signaling. The Hedgehog (Hh) pathway is activated in response to many types of liver injury, where it plays a role in tissue regeneration(79). Under normal conditions, Hh signaling subsides once the liver has been repaired. Hh pathway activity has also been shown to be critical for both activation and survival of hepatic stellate cells, the primary mediators of liver fibrosis(80). Inhibition of Hh signaling in a mouse model of fibrotic HCC development was shown to reduce the severity of fibrosis and tumor development(81). These results suggest that failure to down-regulate Hh signaling following repair of liver injury may promote the progressive development of severe fibrosis leading to hepatocarcinogenesis. In this case, a mutation event that leads to overexpression of Gli2 would serve to uncouple the activation of Hh pathway transcriptional targets from upstream receptor signaling. Such a mutation would confer a particular selective advantage in a liver environment with an initial injury stimulus to activate Hh signaling, preventing subsequent cessation of signaling and promoting increased fibrotic severity and eventual tumor development.

Analysis of candidate genes in human HCC

The most frequently mutated locus identified in our screen was Rian/Rtl1. We have previously shown that transposons integrated within this locus commonly drive overexpression of Rtl1, which acts as a driver of hepatocarcinogenesis(40). In the same study, we found that ~30% of analyzed human HCC samples (n=62) display activation of RTL1 expression, implicating it as a gene relevant to human disease.

As mentioned above, β-catenin and MET are each known to play a role in the pathogenesis of human HCC(74, 75). According to the Catalogue of Somatic Mutations in Cancer (COSMIC) database(7), mutations in these genes have been detected in 19% and 3% of analyzed human HCCs, respectively. In addition to the Wnt/β-catenin and HGF/Met pathways, activation of Ras/MAPK signaling has been identified as a frequent occurrence in human HCC(82). Mutations rates of 2-3% are reported for NRAS and
KRAS by the COSMIC database. Our SB mutagenesis screen identified Kras and Hras1 as CISs in tumors from both cohorts of mice.

Among the most frequently detected mutated genes in SB-induced fibrotic liver tumors was Nuclear factor, erythroid derived 2 (Nfe2) (Figure 13C). Transposon insertions occurred in a pattern predicted to result in overexpression of the wild type ORF. A recent exome sequencing study detected activating mutations in NFE2-like 2 (NFE2L2; also known as NRF2) in 6% of analyzed human HCC samples(83). NFE2 and NFE2L2 have highly homologous basic leucine zipper domains with shared DNA binding sites(84) and transcriptional targets(85). Both are involved in cellular responses to oxidative stress, a condition common to hepatic fibrosis developed in humans or induced by CCl4 injection in mice. We hypothesize that activation of Nfe2 in the livers of mice may promote hepatocarcinogenesis via a mechanism that functionally recapitulates the effect of NFE2L2 activating mutations in human HCC. Further analysis of our fibrosis-specific CIS list identified two additional mutation events that may be involved in this mechanism – activation of Fidgetin (Fign) (Figure 13D) and inactivation of Cullin 3 (Cul3) (Figure 13E). Cul3 is part of an E3 ubiquitin ligase complex that cooperates with adaptor protein Keap1 to target substrate proteins for proteasomal degradation. Of note, recurrent inactivating mutations in KEAP1 have recently been detected in human HCC(86). Nfe2l2 is a well-known substrate of Cul3 in mammalian systems(87). Interestingly, the AAA-ATPase Fidgetin-like 1 has been demonstrated to be a Cul3 target in Caenorhabditis elegans(88), suggesting that its ortholog Fign may be a Cul3 target in mammals. We propose a hypothetical model that links Nfe2, Nfe2l2, Keap1, Cul3, and Fign via a conserved functional mechanism during liver tumorigenesis (Figure 14). In this model, an altered expression pattern of Nfe2/Nfe2l2 target genes is the shared driving force for transformation that can be achieved through activation of Nfe2, Nfe2l2, or Fign. Alternatively, the same outcome can be achieved through inhibition of
Keap1 or Cul3. Further experimentation is required to validate the interconnectedness of this proposed gene network and its putative role in promoting liver cancer.

Conclusions

We present here the results of a liver-specific *Sleeping Beauty* transposon forward genetic mutagenesis screen conducted in the presence or absence of hepatic fibrosis. This approach identified several candidate genes predicted to be involved in the development of HCC, many of which were differentially detected in tumors depending on the fibrotic context. The list of putative liver cancer genes includes known effectors of human HCC, as well as novel putative drivers that represent good candidates for further study. Our results provide some insight into the mechanistic role of fibrogenesis in HCC, suggesting that oxidative stress pathways and growth factor signaling from hepatic stellate cells may be particularly important mediators of fibrosis-dependent hepatocarcinogenesis.

Materials and Methods

Mice

*Alb-Cre*<sup>Tg</sup>/<sup>Tg</sup> mice (Jackson Labs) were crossed to *RosaSBase<sup>Lsl/Lsl</sup>;T2/Onc<sup>Tg</sup>/<sup>Tg</sup>* mice (15) to generate an SB-induced model of HCC. Male offspring were aged and monitored for liver tumor development. All animal studies were approved and monitored by the Institutional Animal Care and Use Committee at the University of Iowa.

CCL<sub>4</sub> injections

Starting at eight weeks of age, mice were given intraperitoneal injections of a solution containing 10% CCL<sub>4</sub> in mineral oil to induce hepatic inflammation and fibrosis. Mice were injected with 2.5µl per gram body weight twice a week for a total of twelve weeks.
Histology

Formalin-fixed, paraffin-embedded liver samples were subjected to Masson’s trichrome staining to assess the degree of fibrosis. A score was assigned to each sample based on the severity of fibrosis, according to the METAVIR method(71).

Illumina sequencing of transposon insertions

DNA from SB-induced tumors was isolated and sequenced to identify transposon integration sites as previously described(43). A list of clonally expanded insertion sites was defined for each tumor and statistical analysis to identify common insertion sites was performed as previously described(42, 72).
Figure 12. Induction of hepatic fibrosis and steatosis in CCl₄-treated mice

As compared to normal liver from an uninjected mouse (A), Masson’s trichrome staining revealed variable induction of hepatic fibrosis and steatosis in CCl₄-treated mice (B-D). Collagen deposition (blue staining) ranged from moderate (B-C) to severe (D). No evidence of steatosis was observed for some injected mice (B), while others showed mild (C) to severe (D) lipid droplet accumulation. Staining was performed by the University of Iowa Histology Research Lab. Scale bars = 20μm.
Figure 13. Patterns of transposon insertion within frequently mutated genes

Transposon insertion sites identified in SB-induced liver tumors within the Met (A), Gli2 (B), Nfe2 (C), Fign (D), and Cul3 (E) loci are shown as arrowheads. The direction of the arrowhead indicates the integrated transposon’s transcriptional orientation. Closed arrowheads represent insertions identified in tumors from CCl₄-treated mice, while open arrowheads represent those from uninjected mice.
Figure 14. Proposed model for functional interaction between Nfe2, Nfe2l2, Fign, Keap1, and Cul3

(A) Under basal conditions, Nfe2 and Nfe2l2 are recruited by Keap1 for ubiquitination by Cul3, marking them for proteasomal degradation. (B) Fign serves as a competitive Cul3 target, freeing Nfe2 and Nfe2l2 to enter the nucleus and affect transcription. (C) Oxidative stress causes dissociation of the ubiquitination complex, freeing Nfe2 and Nfe2l2 to enter the nucleus. Ub, ubiquitin moiety.
Table 2. Scoring of liver fibrosis severity in CCl4-treated mice

<table>
<thead>
<tr>
<th>META VIR Score</th>
<th>Associated Phenotype</th>
<th>Number of Mice</th>
<th>Average Number of Tumors/Mouse</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0</td>
<td>No fibrosis</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Portal fibrosis without septa</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F1</td>
<td>Portal fibrosis with rare septa</td>
<td>22</td>
<td>2.2</td>
</tr>
<tr>
<td>F2</td>
<td>Numerous septa without cirrhosis</td>
<td>31</td>
<td>3.0</td>
</tr>
<tr>
<td>F3</td>
<td>Cirrhosis</td>
<td>10</td>
<td>3.5</td>
</tr>
</tbody>
</table>

Note: Scoring was performed based on Masson’s trichrome staining performed by the University of Iowa Histology Research Lab
Table 3. Liver tumors generated by SB mutagenesis screen

<table>
<thead>
<tr>
<th></th>
<th>Untreated SB Mice</th>
<th>CCl₄-treated SB Mice</th>
<th>CCl₄-treated Control Mice</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of Mice with Tumors</td>
<td>44%</td>
<td>100%</td>
<td>13%</td>
</tr>
<tr>
<td>Average Number of Tumors/Mouse</td>
<td>1.7</td>
<td>5.0</td>
<td>3.5</td>
</tr>
<tr>
<td>Total Number of Tumors</td>
<td>59</td>
<td>378</td>
<td>7</td>
</tr>
<tr>
<td>Average Age at Endpoint (Weeks)</td>
<td>64.5</td>
<td>48.5</td>
<td>65</td>
</tr>
</tbody>
</table>

Note: Some tumors have not yet undergone insertion site profiling
Table 4. Common insertion sites identified in SB-induced liver tumors

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Insertions in Fibrotic Tumors</th>
<th>Insertions in Non-Fibrotic Tumors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rian/Rtl1\textsuperscript{a,b}</td>
<td>81</td>
<td>58</td>
</tr>
<tr>
<td>Met\textsuperscript{a}</td>
<td>25</td>
<td>2</td>
</tr>
<tr>
<td>Fign\textsuperscript{a}</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>Gli2\textsuperscript{a}</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Nfe2\textsuperscript{a}</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Arithl\textsuperscript{a}</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Dpyd\textsuperscript{a}</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Arhgap42\textsuperscript{a}</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Setd2\textsuperscript{a}</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Hras1\textsuperscript{a,b}</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Eml4\textsuperscript{a}</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Ptpkr\textsuperscript{a}</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Rfwd2\textsuperscript{a}</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Ctnnb1\textsuperscript{a}</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Zfp217\textsuperscript{a}</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Kras\textsuperscript{a,b}</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Snd1\textsuperscript{a,b}</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Chchd3\textsuperscript{a}</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Cul3\textsuperscript{a}</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Zbtb20\textsuperscript{a}</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Arhgef12\textsuperscript{a}</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Cdhl2\textsuperscript{a}</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Mitf\textsuperscript{a}</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Myo1b\textsuperscript{a}</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Rbm6\textsuperscript{a}</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Immp2l\textsuperscript{b}</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Usp32\textsuperscript{b}</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Atxn2\textsuperscript{b}</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Ptprc\textsuperscript{b}</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Identified as a CIS in CCl4-treated cohort (n=179)

\textsuperscript{b}Identified as a CIS in untreated cohort (n=83)
CHAPTER 4

PIGGYBAC TRANSPOSITION FROM RECOMBINANT ADENOVIRUS AS A METHOD FOR GENE DELIVERY WITH STABLE EXPRESSION

Abstract

Much research has gone into the development of a hybrid system that combines the broad tropism and efficient transduction of adenoviral vectors with the ability to achieve somatic integration and stable expression of cargo genes. Such a system would be highly useful for studying gene function in vivo, allowing higher-throughput experiments than can be achieved through the generation of transgenic or targeted animal strains. It would also have the potential to be used in gene therapy applications requiring efficient delivery and maintained gene expression. Despite substantial effort, current approaches often rely on multi-step processes to accomplish stable gene integration, with reported efficiencies generally below 1%. Those that have achieved higher efficiencies are limited by low cargo capacity or a dependence on target cell division for integration. There is a need for a hybrid gene delivery system that provides all of the following: 1) efficient delivery to target cells, 2) stable expression of cargo genes in a high percentage of cells (whether mitotic or not), 3) large cargo capacity, 4) flexibility to use with a wide range of strains and/or additional treatments, and 5) simple experimental technique. We describe here the generation and initial characterization of a hybrid system with all of these characteristics that combines recombinant adenoviral vector delivery with piggyBac transposition for stable integration. The considerable potential utility of this Ad-PB system for in vivo cancer gene validation, therapeutic gene delivery, and other functional gene studies warrants further development and characterization to maximize its impact and benefit.
Introduction

Recombinant adenoviral (Ad) vectors are widely used in biological research to deliver transgenes to a cell type of interest *in vitro* or *in vivo* (89). As compared to other viral systems, Ad vectors provide many advantages. These include a broad tissue tropism, efficient transduction of target cells regardless of mitotic state, large cargo capacity, low immunogenicity (depending on the specific vector used), and the ability to generate high titer viral preparations. A major limitation of Ad vectors, however, is that aside from very rare instances of chromosomal integration of the entire vector (90, 91), expression of cargo genes is transient and dependent on the maintenance of episomal viral vectors. This limitation precludes the use of standard Ad vectors for studies of gene function in dividing cells, as episomal vectors are progressively diluted with each cell cycle. It also constrains their utility for therapeutic gene delivery to situations where stable, long-term expression is not required.

To overcome the limitations imposed by transient gene expression, several hybrid Ad vector systems have been developed that contain cargo genes embedded within an element capable of genomic integration. This approach retains the advantages that Ad vectors provide in terms of delivery to target cells, while also allowing transgene expression to be maintained long-term. Hybrid systems described to date have utilized adeno-associated virus (AAV), retrovirus, lentivirus, *Sleeping Beauty* (SB) transposons, L1 retrotransposons, or PhiC31 integrase to achieve stable integration (92, 93). While each of these systems has been demonstrated to successfully mediate genomic integration of transgenes, reported efficiencies are typically below 1% of transduced cells (94-97). Hybrid Ad vector systems incorporating lentivirus or foamy virus have reported efficiencies of 20-30% in cultured cells when using a multiplicity of infection (MOI) that would be achievable *in vivo* (92, 98); however, these systems are hampered by low cargo capacity and a reliance on actively-dividing cells, respectively. Another strategy used to achieve long-term expression from Ad vectors utilizes elements derived
from Epstein-Barr virus (EBV) to generate self-replicating episomes containing the transgene. While relatively efficient EBV-based episome production and long-term expression can be achieved with this system(99), episomes are depleted over time in the absence of selection, resulting in loss of transgene expression(100). There is currently a need for a system that allows high efficiency delivery and stable expression of large transgenes into a wide range of cell types *in vivo*. Such a system would be of broad utility for functional studies and therapeutic delivery of genes of interest.

Research into the molecular mechanisms underlying cancer development is one area of study that would benefit greatly from a gene delivery system with the qualities described above. In-depth analyses are providing increasingly detailed characterization of the molecular aberrations present in human tumors, often revealing a surprisingly complex genetic landscape. The percentage of detected aberrations that have a functional role in promoting tumor development and/or progression is estimated to be less than 1%(12), making the identification of these “driver” mutations among the much larger number of “passenger” mutations critically important. As a complementary approach to human tumor analysis, animal model systems are effective tools to identify candidate cancer genes(65). To maximize the usefulness of information gained from animal models of cancer, these candidates must be validated as drivers of tumorigenesis *in vivo*. The time, money, and resources required for proper validation significantly limit the number of candidates and environmental contexts that can be tested. An efficient and flexible system for stable gene delivery *in vivo* would significantly increase the throughput of validation experiments, leading to more rapid identification of promising therapeutic targets for human cancer.

We have developed a novel gene delivery system with significant advantages over existing approaches by incorporating *piggyBac* (PB) transposons into Ad vectors. This design allows transgenes to be integrated into the genome of target cells expressing PB transposase (PBase), which can be incorporated into the model system or co-delivered.
Compared to other transposition systems, PB has particularly high activity(101) and large cargo capacity(102), making it a flexible tool for transgene delivery. We present here the strategy used in designing this Ad-PB system, as well as initial experiments characterizing its efficiency.

**Results and Discussion**

*Design of a hybrid piggyBac-adenoviral vector system*

Two distinct types of Ad-PB vectors are described in this study. The first, schematized in Figure 15A, is a fluorescent reporter construct designed to allow the efficiency of stable transgene integration to be determined. The construct (Ad-PB-flox-mCherry) contains PB inverted terminal repeats (ITRs) inside of the adenoviral ITRs. The PB ITRs serve as recognition sites for PBase, which mediates excision and reinsertion of any DNA sequence between the ITRs (*i.e.* within the transposon). An expression cassette encoding mCherry fluorescent protein driven by the human elongation factor-1 alpha (EF-1α) promoter is contained within the transposon. An additional expression cassette encoding cytomegalovirus (CMV) promoter-driven green fluorescent protein (GFP) is included outside of the PB transposon. LoxP sites were placed just outside of the ITRs at both ends of the PB transposon, as well as in between the EF-1α promoter and mCherry coding sequence. These sites were included to allow deletion of transposons contained within the Ad genome upon activation of Cre recombinase. Importantly, any transposons excised by PBase and inserted into the cellular genome prior to activation of Cre would be protected from deletion, as they would only retain one LoxP site. This feature was included so that expression of mCherry from episomal Ad genomes could be eliminated following cellular transduction, facilitating accurate quantification of signal from integrated PB transposons.

An example of the second type of Ad-PB vector is schematized in Figure 15B. This construct (Ad-PB-EF1α-LsL) was designed to allow stable delivery of short hairpin
RNAs (shRNAs) and/or Cre-inducible transgene overexpression cassettes. In this way, both gain-of-function and loss-of-function studies investigating the effects of modulating gene expression can be performed with the same basic vector. The EF-1α promoter contained within the PB transposon includes an intron that gets spliced out following transcription. Embedding an shRNA or microRNA (miRNA) within the intron leads to post-splicing recognition by the cell’s endogenous miRNA processing pathway, allowing functional RNA interference (RNAi) to occur(103). Downstream of the EF-1α promoter is a Lox-stop-Lox (LsL) cassette, which consists of LoxP-flanked (floxed) GFP followed by three polyadenylation signals. This design prevents expression of a transgene downstream of the LsL cassette until activation of Cre and excision of the cassette. Inclusion of GFP within the LsL cassette allows the efficiency of Ad transduction to be quantified. Upon activation of Cre, the cassette is excised, leading to EF-1α-driven transgene expression and elimination of GFP expression. We have designed a series of constructs utilizing this Ad-PB vector that encode shRNAs and expression cassettes designed to promote liver tumor formation following in vivo delivery.

**Efficient genomic integration in cultured cells**

The non-tumorigenic murine hepatocyte cell line TIB-73 was used to test the stable integration efficiency of our hybrid Ad-PB system in vitro. TIB-73 cell lines with stable expression of Cre-ER<sup>T2</sup>, hyperactive PBase (hyPBase), or both were generated by lentiviral transduction. Cre-ER<sup>T2</sup> is a ligand-inducible version of Cre that can only enter the nucleus and mediate DNA recombination upon interaction with tamoxifen(104). In this way, recombination activity can be temporally regulated. HyPBase is a version of PBase generated by introducing activating mutations into the mammalian codon-optimized PBase (mPB)(105), resulting in a >10-fold increase in transposition activity(106). An additional TIB-73 line with expression of mPB was generated to assess the effect of PBase activity level on integration efficiency. Each of the four cell lines was
transduced with the Ad-PB-flox-mCherry virus at an MOI of 2. Cells were passaged at 72 hours post-transduction and mCherry fluorescence was detected by flow cytometry analysis (Table 5). The efficiency of adenoviral transduction at this MOI ranged from 68-81%. To confirm transposition activity, excision PCR was conducted on total DNA isolated from cells with or without hyPBase expression. As shown in Figure 16, a product derived from a portion of the adenoviral genome outside of the transposon was detected in both samples, while an excision-specific product was detected only in the presence of hyPBase. At this point the cell lines expressing Cre-ER\textsuperscript{T2} were each split into two separate plates. For one plate, 4-hydroxytamoxifen (4-OHT) was added to the media at a final concentration of 1 µM to activate Cre-ER\textsuperscript{T2}, while the other plate was left untreated. At six days post-transduction, effectiveness of 4-OHT treatment in reducing mCherry expression was confirmed by epifluorescence microscopy, but was not quantified. All cell lines were passaged out to twenty days post-transduction (eight total passages), at which point fluorescence was once again detected by flow cytometry analysis.

Analysis of fluorescence expression data revealed a stable integration rate of 9.5-10.1% for TIB-73 cells expressing hyPBase (Table 5). The rate for cells expressing mPB was found to be 3.2%, demonstrating that increased transposition efficiency can lead to a corresponding increase in stable integration. This suggests that efficiency could be further improved by increasing levels of hyPBase, which could be achieved by co-delivery of a hyPBase expression vector to cells transduced with Ad-PB vectors. We have developed an AAV-hyPBase vector for this purpose. Despite a relatively low MOI, the Ad-PB system achieved a stable integration rate 10-fold higher than a similar system utilizing SB transposition(95, 107). This likely reflects the increased transposition activity of PBase relative to SBase(101), as well as the fact that the Ad-SB system depends on Cre-mediated circularization of transposon substrates for transposition to
occur(95). Quantitative experiments with higher MOIs and AAV-mediated delivery of hyPBase are underway and are expected to yield higher stable integration efficiencies.

Application as a validation tool for putative oncogenes and tumor suppressors in vivo

The Ad-PB system could prove particularly useful for in vivo candidate cancer gene validation experiments. Due to the requisite high rate of cell division associated with neoplastic development, validation studies require stable maintenance of transgene expression. A population of cells transduced with a standard Ad vector harbors a fixed number of episomes, and therefore the number of Ad genomes per cell decreases by half for each cell doubling that occurs. This would rapidly lead to loss of transgene expression in the majority of cells as a tumor mass began to develop, likely precluding further progression. Genomic integration of the transgene cassette would prevent this proliferation-associated loss of expression.

As an example of the Ad-PB system’s utility for this type of study, we will consider validation of liver cancer genes. Existing techniques for validating putative liver cancer genes in vivo often rely on complex engineered mouse strains(23), transplantation experiments(24), strong predisposing alleles(25), induction of severe liver damage(26), or a combination of these factors. An alternative approach is to generate transgenic, knock-in, or knockout mouse strains, although the associated high cost and low throughput limit the extent to which this can be done. There is currently a need for a liver cancer gene validation technique that is inexpensive, high-throughput, efficient, minimally invasive, and amenable to combination with additional genetic and environmental inputs. The Ad-PB system fulfills all of these requirements. The costs associated with development and maintenance of a new mouse strain are more than ten times higher than that of producing a recombinant adenovirus. The ability to generate pooled populations of Ad-PB vectors with distinct transgenes further decreases the cost
per gene for validation and increases the throughput with which candidates can be evaluated. Of the tissue types susceptible to transduction with Ad vectors, liver is among the most efficiently targeted, and this can be achieved through standard tail vein injection. Because all of the necessary components for stable gene expression can be delivered in a single injection, Ad-PB vectors can be combined with any existing mouse model system. This facilitates assessment of gene function in a wide variety of contexts, such as in the presence of predisposing mutations or environmental factors.

For a proof-of-principle experiment, we have engineered Ad-PB-EF1α-LsL vectors encoding Cre-inducible constitutively active NRAS (NRAS<sup>G12V</sup>)(108) or β-catenin (CTNNB1<sup>ΔN90</sup>)(109) along with an shRNA directed against Trp53 (shp53)(63). These transgenes have been demonstrated to promote liver tumor development when delivered via hydrodynamic tail vein injection(48, 110). The tumorigenic vectors will be administered to mice via standard tail vein injection to determine whether or not HCC can be induced via transgene delivery with the Ad-PB system. A vector encoding Cre-inducible mCherry without shp53 will be used as a negative control for tumorigenesis.

Applications for gene therapy

An obvious application of a stable gene delivery system is the administration of therapeutic genes to correct inherited disorders. Unlike cancer gene validation experiments, in which there is a strong positive selection pressure for cells expressing the transgene, highly efficient stable gene expression is typically a requirement for gene therapy. We have observed a rate of 10% stable transduction with the Ad-PB system in cultured cells, which is unlikely to be adequate for most gene therapy applications. There are, however, several simple modifications that could be made to the system that would be expected to increase its efficiency, perhaps to a level suitable for therapeutic gene delivery. These modifications will be discussed in Chapter 5.
Conclusions

We describe here the initial development and characterization of a hybrid Ad-PB system with broad utility for stable transgene delivery that provides several advantages over existing approaches combining Ad vectors with stable integration machinery. The Ad-PB system can be used to achieve efficient, stable delivery of large transgenes to a wide variety of cell types in vivo, including non-dividing cells. It is simple and flexible enough to be combined with existing genetic and environmental modifiers, permitting functional gene studies to be conducted in multiple distinct contexts. By decreasing the cost and increasing the throughput of cancer gene validation studies, particularly in the liver, the Ad-PB system will facilitate more rapid and efficient identification of promising candidates for targeting therapeutically in human cancer.

Materials and Methods

Vector construction

Expression vectors encoding Cre-ER\textsuperscript{T2}, mPB, and hyPBase were generated using the pHAGE2 self-inactivating, non-replicative lentiviral vector backbone(111). Shuttle vectors were generated containing the functional elements of Ad-PB-flox-mCherry and Ad-PB-EF1\textalpha-LsL flanked by homology arms for recombination into the pAd-nGFP adenoviral genome backbone(112). Homology arms were PCR-amplified using 10ng pAd-nGFP as template with the following primers: shared 3’ homology arm (3’HArm\textsubscript{F}: 5’–ACATGTAAGCGACGGATGTG–3’ and 3’HArm\textsubscript{R}: 5’–GATCCAAGCTTCTGACTGAC–3’ (amplicon 326bp)), Ad-PB-flox-mCherry 5’ homology arm (flox5’HArm\textsubscript{F}: 5’–GATCTCCCGATCCCCCTATG–3’ and flox5’HArm\textsubscript{R}: 5’–GCGTATATCTGGCCCGTACA–3’ (amplicon 219bp), and Ad-PB-EF1\textalpha-LsL 5’ homology arm (LsL5’HArm\textsubscript{F}: 5’–TCCGATTCGACAGATCACTG–3’ and LsL5’HArm\textsubscript{R}: 5’–GGCTCAGGAAAGCAAGATCA–3’ (amplicon 362bp)).
Homologous recombination to generate the final Ad-PB vector plasmids was performed using a previously described recombineering system(113).

Cell culture

TIB-73 cells (ATCC: BNL CL.2) were maintained in high-glucose DMEM (Life Technologies) supplemented with 10% fetal bovine serum. Lentiviral vectors were transfected into 293FT cells (ATCC: PTA-5077) along with plasmids encoding viral packaging proteins using Effectene transfection reagent (Qiagen). TIB-73 cell lines with stable expression of Cre-ER\textsuperscript{T2}, mPB, and hyPBase were generated by lentiviral transduction followed by antibiotic selection with 1 mg/ml G418 or 1 µg/ml puromycin. 4-hydroxytamoxifen (Sigma-Aldrich) dissolved in ethanol was added to culture media at a final concentration of 1 µM to activate Cre-ER\textsuperscript{T2}.

Excision PCR

Genomic DNA was extracted from transduced TIB-73 cells using the GenElute Mammalian Genomic DNA Miniprep kit (Sigma-Aldrich). PCR was performed using 50ng DNA as template to amplify fragments from the Ad-PB vector. Primers were designed to amplify either an excision-specific product (Ad-PB_Ex_F: 5’–ACCGTTTACGTGGAGACTCG–3’ and Ad-PB_Ex_R: 5’–GCCGATATACTATGCCGATGA–3’ (amplicon 248bp)) or a portion of the Ad genome present regardless of excision (Ad-PB_F: 5’–CACCAAAATCAACGGGACTT–3’ and Ad-PB_R: 5’–CATGGTGGAGCCTGCTTT–3’ (amplicon 227bp)).

Flow cytometry

Fluorescent protein expression analysis was conducted at the University of Iowa Flow Cytometry Facility using a Becton Dickinson LSR II cytometer.
Figure 15. Structure of Ad-PB vectors

(A) Ad-PB-flox-mCherry vector. Within the adenoviral ITRs, this construct includes a CMV-GFP cassette and a PB transposon containing EF-1α promoter-driven mCherry. LoxP sites (black triangles) are included flanking the transposon and in between the EF-1α promoter and mCherry cassette. (B) Ad-PB-EF1α-LsL vector. Inside of the PB transposon, this construct contains EF-1α promoter-driven mCherry preceded by a Lox-stop-Lox cassette containing GFP. An shRNA can be cloned within the intron of the EF-1α promoter.
Figure 16. PB transposition from the adenoviral genome

Excision PCR was performed on total DNA extracted from TIB-73 cells after transduction with Ad-PB-flox-mCherry. (A) Primers used to detect an excision product or the adenoviral genome are depicted in black and grey, respectively. The excision product primers will only amplify a product following excision of the PB transposon, while the adenoviral genome primers will amplify a product regardless of excision. (B) Excision was detected only in cells expressing hyPBase, while the Ad genome was detected in cells with or without PBase. Depicted results are representative of three independent biological replicates.
Table 5. Flow cytometry analysis of mCherry expression in transduced cells

<table>
<thead>
<tr>
<th>TIB-73 Cell Line</th>
<th>% Fluorescent (Day 3)</th>
<th>% Fluorescent (Day 20)</th>
<th>% Stable Transduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cre-ERT2</td>
<td>68.1</td>
<td>0.2</td>
<td>-</td>
</tr>
<tr>
<td>Cre-ERT2 + 4-OHT</td>
<td>-</td>
<td>1.4</td>
<td>-</td>
</tr>
<tr>
<td>hyPBase</td>
<td>78.2</td>
<td>7.9</td>
<td>10.1</td>
</tr>
<tr>
<td>hyPBase; Cre-ERT2</td>
<td>80.5</td>
<td>7.8</td>
<td>9.7</td>
</tr>
<tr>
<td>hyPBase; Cre-ERT2 + 4-OHT</td>
<td>-</td>
<td>9.3</td>
<td>11.6</td>
</tr>
<tr>
<td>mPB</td>
<td>75.4</td>
<td>2.4</td>
<td>3.2</td>
</tr>
</tbody>
</table>
CHAPTER 5
DISCUSSION

Involvement of the Dlk1-Dio3 Imprinted Domain in Cancer

The Dlk1-Dio3 imprinted domain has been identified as a region of interest in a variety of cancers. Located at chromosome 14q32 in humans and distal chromosome 12 in mice, the domain spans ~850 kilobases and encodes several distinct types of genes. Three protein-coding transcripts are expressed from the paternally inherited allele, while the maternal allele encodes four lncRNAs, a C/D snoRNA cluster, and >50 miRNAs. Epigenetic modifications established in the germline regulate allele-specific expression of domain members, a pattern that is maintained largely through differential methylation of control regions throughout the domain(27). Failure to establish imprinting within the Dlk1-Dio3 domain leads to severe and widespread developmental defects in both mouse models(28) and human patients(29), demonstrating the importance of the domain and its proper regulation in a variety of processes.

The majority of studies linking the Dlk1-Dio3 domain to cancer are strictly correlative, however, multiple groups have performed experiments that demonstrate a direct role for a domain member in promoting tumorigenesis. While they are primarily limited to liver cancer, these direct studies lend support to hypotheses of Dlk1-Dio3 domain involvement in other cancers based on correlative evidence. Due to its complexity in terms of constituent genetic elements and mechanisms to regulate their expression, the majority of reports linking the Dlk1-Dio3 domain to cancer development identify several loci within the domain that are concurrently dysregulated in tumors. This type of result indicates either: 1) a highly complicated process wherein cooperation of several distinct domain members is required to promote transformation or 2) direct involvement of a single (or a few) member(s) in driving tumor development with relatively inconsequential changes in expression of additional members resulting from
shared regulatory pathways. Given the intricate patterns of coordinated gene regulation common to imprinted domains, the latter seems more likely. Definitive results from direct studies individually modulating activity of domain members, including my work on Rtl1, is in agreement with this conclusion.

Among the most intriguing results linking Dlk1-Dio3 domain disruption to liver cancer was the discovery that HCCs developed in mice following hepatic delivery of AAV gene therapy vectors recurrently harbored genomic insertions of the vector within the imprinted domain(30). This result suggested a strong positive selective pressure for Dlk1-Dio3 mutations in hepatocytes, a hypothesis confirmed by a follow-up study showing that gene targeting to introduce an AAV vector into the mouse Dlk1-Dio3 domain in 0.01% of hepatocytes promoted liver tumor development with 100% penetrance(31). Detailed gene expression profiling revealed broad upregulation of Dlk1-Dio3 domain members including mRNAs, miRNAs, and snoRNAs; however, no direct assessment of the molecular mechanism of tumorigenesis was performed. Rtl1 expression was not determined due to its exclusion from the microarray platform utilized in the study. Interestingly, an independent study recently reported detection of Dlk1-Dio3 domain AAV vector integration in HCCs arising in mice following delivery of a distinct gene therapy vector(45). Together with our identification of the domain as a CIS in SB-induced HCC, these results demonstrate a strong association between the integration of exogenous DNA in this region and HCC development.

Other groups have reported correlations between liver cancer and dysregulated expression of Dlk1-Dio3 domain miRNAs(34, 114-116), lncRNAs(32, 117), and protein-coding mRNAs(33, 35, 36, 118). All three of the studies that analyzed Rtl1(33, 36, 114), which is normally not expressed in adult liver, detected its expression in liver tumors, although upregulation of Rtl1as was found to be more significant in one study(114). While oncogenic or tumor suppressive activities have been proposed for multiple Dlk1-Dio3 domain members based on cell culture and xenograft assays, Rtl1 is the only
member to date that has been validated as a driver of liver cancer through direct modulation and subsequent spontaneous tumorigenesis in vivo\cite{36, 40}. Based on the paucity of data on \textit{Rtl1} expression in HCC and the \textit{in vivo} validation experiments that have been completed, I would speculate that many of the links identified between the \textit{Dlk1-Dio3} domain and HCC can be explained by \textit{Rtl1}’s activity as a driver of hepatocarcinogenesis. Some groups that have studied domain members in HCC may conduct follow-up experiments based on my work to assess \textit{Rtl1} expression in their sample sets. If frequent upregulation in tumors is reported, particularly in human HCC samples, it will further support my conclusion that activation of this locus is an important driver of liver cancer. Additionally, it will provide impetus to pursue \textit{RTL1} as a therapeutic target for human HCC.

The \textit{Dlk1-Dio3} domain may play an even broader role in cancer development, potentially with distinct members driving transformation in different cell types. Beyond the liver, disruption of the \textit{Dlk1-Dio3} domain has been reported in cancers of several tissues including pituitary\cite{119-121}, brain\cite{122}, mammary\cite{123, 124}, lung\cite{123, 125}, skin\cite{126}, colon\cite{123}, pancreas\cite{123}, adrenal\cite{127}, blood\cite{128-131}, ovary\cite{132}, and kidney\cite{132-135}. As with the liver studies, the majority of these reports are strictly correlative. Further experiments are required to clarify the extent of the \textit{Dlk1-Dio3} domain’s contribution to transformation of these tissues. It is interesting to note that expression from the \textit{Dlk1-Dio3} domain has been found to distinguish induced pluripotent stem (iPS) cells that are most similar to embryonic stem (ES) cells\cite{136}. The majority of iPS cells in the study displayed epigenetic silencing of \textit{Dlk1-Dio3} domain expression, which was associated with low developmental potential and an inability to support the production of mice derived entirely from iPS cells. In contrast, non-silenced iPS cells or those with expression rescued through epigenetic modification were competent to produce all-iPS cell mice. These results demonstrate a functional role for the \textit{Dlk1-Dio3} domain in embryogenesis, a process that shares many characteristics with tumor
development. There are several instances of genes and signaling pathways with well-established roles in both normal development and tumor formation, exemplified by the Wnt, Hedgehog, and Notch pathways.

The most commonly reported Dlk1-Dio3 domain disruption in cancer is activation of Dlk1 expression. One factor complicating mechanistic interpretation of this finding is that Dlk1 has been widely used as a biomarker for cells in an undifferentiated state(137-139), a phenotype often associated with cancer. The frequent observation of Dlk1 activation in tumors could result from this association, rather than a role in driving cellular transformation. On the other hand, expression may be associated with both prevention of differentiation and development of cancer due to a shared functional mechanism in each process. Consistent with this idea, expression of Dlk1 has been found to be important for the regeneration of liver(140) and skeletal muscle tissues(141). This finding suggests a functional role for Dlk1 in the proliferation of cells during tissue regeneration that could potentially contribute to tumor growth when inappropriately activated.

Prior to the publication of the data in Chapter 2, several groups had reported links between the Dlk1-Dio3 domain and liver cancer, both in human tumors and mouse models. My work implicating Rtl1 as a driver of hepatocarcinogenesis provides insight into the molecular mechanism behind these identified links. Although our results do not rule out the involvement of other domain members in HCC development, they do demonstrate definitively that activation of Rtl1 can drive this process. These results represent an important advancement to the field, as they provide a mechanism by which disruption of the Dlk1-Dio3 domain contributes to cancer. As the vast majority of studies in this area have been correlative, the prevalence of Rtl1 activation in tumors other than HCC and functional contributions of domain members besides Rtl1 remain to be determined. Without direct validation experiments to confirm a gene’s role in tumorigenesis, simple observation of its aberrant expression in cancer is insufficient to
demonstrate functional significance. An increase in the number of researchers conducting such validation studies would be enormously beneficial for determining which genes are actually important for tumor development and progression. This in turn would lead to improved efficiency and efficacy of molecular targeted therapy development. A major goal of mine in contributing to the development and characterization of the Ad-PB system described in Chapter 4 is to generate a tool for in vivo cancer gene validation that is simple, flexible, and accessible enough to be utilized by researchers in all types of lab settings to facilitate the development of improved therapies for cancer patients.

**Translation of Rtl1 to the Clinic**

Our lab identified the Dlk1-Dio3 imprinted domain as a commonly disrupted site in HCCs induced by SB transposon insertional mutagenesis. My work demonstrated that the tumor-driving event under positive selection in this model was overexpression of Rtl1. Further analysis validated Rtl1 as a driver of hepatocarcinogenesis in vivo, identified a subset of human HCC with RTL1 overexpression, and suggested that the mechanism of tumorigenesis may involve degradation of the extracellular matrix by Rtl1. These results implicate RTL1 as a strong candidate oncogene in human HCC, and they warrant its further study as a potential therapeutic target.

Through analysis of two independent sets of human HCCs, I identified activation of RTL1 in 30% of cases. This result is highly significant, as the most commonly mutated genes in HCC – TP53, ARID1A, and CTNNB1 – are disrupted in 31%, 29%, and 19% of cases, respectively(7). It suggests that activation of RTL1 may be among the most common driving events in human HCC. With such high prevalence of aberrant activation, one would expect that disruption of RTL1 would have been detected in human HCC before now. The main reason that this has not occurred has to do with the format of high-throughput analytical tools that have been used. Because it is a relatively
uncharacterized gene with some sequence homology to other retrotransposon-derived elements in the genome, RTL1 has been excluded from nearly all array-based analyses. Exome sequencing, the most common method to detect DNA mutations, depends on array hybridization to enrich coding sequences, and the vast majority of gene expression data on human HCC comes from microarray analyses. For this reason, RTL1 has eluded detection as a commonly disrupted locus in human HCC. Increased throughput and decreased cost have led to growing prevalence of direct transcriptome and whole genome sequencing datasets. It will be interesting to see whether or not these analyses detect RTL1 as one of the most common disruptions in human HCC.

For a molecule to be useful as a therapeutic target, it needs to be druggable, meaning that its activity can be modulated through administration of a drug. It is quite likely that RTL1 will be a good candidate for drug development, and there may already be drugs that are capable of inhibiting its activity. My discovery that Rtl1 overexpression allows cultured hepatocytes to proliferate and become polarized in a 3D Matrigel growth assay led me to hypothesize that its mechanism of action involves liberation of matrix-embedded growth factors via proteolytic degradation of an ECM component. Consistent with this hypothesis, follow-up experiments conducted in depleted Matrigel demonstrated a dependence of the phenotype on growth factor levels. Based on my hypothesis, the extracellular aspartic protease domain of Rtl1 mediates substrate cleavage and subsequent growth factor release. Accordingly, inhibition of the protease domain would be predicted to abolish Rtl1’s growth-promoting activity. Aspartic protease domains are a quite common feature in nature. Both broadly acting and specifically targeted small molecule inhibitors have been developed to inhibit their activity(57-59). The ability of these inhibitors to impair Rtl1’s activity could be tested in vitro and in vivo using the models described in our study. A particularly interesting experiment would be to introduce aspartic protease inhibitors at various time points in a mouse model of Rtl1-induced HCC to determine the feasibility of preventing tumor formation or inducing tumor regression.
If a treatment regimen beginning after a tumor has already formed can successfully induce regression, it would suggest that aspartic protease inhibitors might be useful for treating patients that have HCCs with activated RTL1. If treatment initiated prior to macroscopic tumor formation can prevent tumorigenesis, it would suggest that patients at high risk of HCC development (e.g. with a cirrhotic liver) might benefit from prophylactic treatment. Assuming protease inhibitors are found to be effective, the same mouse model of Rtl1-induced HCC with drug treatment could be combined with the SB mutagenesis system to identify mutation events that confer resistance. By activating hepatic transposon mobilization (e.g. with 4-OHT injection into T2/Onc3 X Alb-Cre-ER\textsuperscript{T2} mice) prior to or concurrent with drug administration, mutation events that allow tumors to escape treatment could be identified.

Another possibility for therapeutic inhibition of RTL1 involves epigenetic modification. I suspect that activation of RTL1 in human HCC may result from loss of imprinting (LOI) within the DLK1-DIO3 domain. LOI is a common occurrence in cancer(51), and epigenetic disruption of the domain has been observed in human HCC(117). In fact, one HCC study identified specific loss of methylation within the RTL1 locus(31). If aberrant epigenetic modifications are responsible for RTL1 activation in tumors, they could potentially be reverted through drug treatment. Several compounds have been developed to regulate epigenetic marks including histone acetylation, histone methylation, and DNA methylation, with some demonstrating efficacy in cancer clinical trials(142). To date, only one clinical trial has investigated the use of epigenetic therapy for HCC(143). Treatment of late-stage HCC patients with Belinostat, a histone deacetylase (HDAC) inhibitor, did not result in significant improvement of patient outcome. Of note, treatment of iPS cells with valproic acid, another HDAC inhibitor, was found to reactivate expression of silenced Dlk1-Dio3 domain members(136). In a case where a patient’s tumor was driven in part by activation of RTL1 or another domain member, this type of treatment could exacerbate the patient’s condition rather than
improving it. Examples like this demonstrate the importance of understanding the molecular underpinnings of tumor development on a case-by-case basis so that appropriate targeted therapies can be utilized to maximize the likelihood of successful treatment. Modifiers of DNA methylation may be more effective than histone modifiers in manipulating expression of RTL1 given its regulation by DMRs and the pattern of methylation loss observed in human HCC.

Though not always required for therapeutic efficacy, a detailed mechanistic understanding of a molecule’s involvement in disease pathogenesis can be highly beneficial. More experiments must be done to clarify Rtl1’s mechanism of action in promoting cellular proliferation and liver tumorigenesis. I have begun to generate some tools for this purpose. First, I have created a construct encoding a version of Rtl1 containing a point mutation within the catalytic domain (Rtl1D684A) that would be predicted to lack protease activity. I am also generating a panel of Rtl1 mutants in which each predicted functional domain has been deleted. These expression constructs will be used to determine the necessity of each domain for Rtl1’s ability to promote cellular growth phenotypes in vitro and in vivo. My current hypothesis is that the protease domain is the primary mediator of Rtl1’s effects. If this is the case, it will be useful to identify target substrates. The 3D Matrigel growth assay that I have used is a good place to start, as it is a relatively simple system. The list of protein components present in Matrigel and growth media is defined and fairly short. Comparative proteomic analysis of Matrigel following cell culture in the presence or absence of Rtl1 overexpression could lead to identification of differential processing of one or more components of the ECM, which would be implicated as targets of Rtl1’s protease activity. Either 2D PAGE or mass spectrometry would be able to identify differentially processed proteins. To confirm cleavage by Rtl1, these candidate targets could be validated using gel zymography.
I have found that overexpression of Rtl1 in cultured hepatocytes is insufficient to promote colony formation in growth factor-depleted Matrigel, demonstrating a dependence on one or more of these factors. A series of experimental conditions in which each growth factor is individually added back to depleted Matrigel could be used to identify the factor(s) required for growth promotion by Rtl1. One of the primary components in complete Matrigel is epidermal growth factor (EGF). Interestingly, activating mutations in its receptor, \textit{Egfr}, were the most common event identified in a mouse model of HCC induced by T2/Onc SB transposons\(^{(26)}\). Activation of EGFR signaling has also been detected in a significant portion of human HCC cases\(^{(68)}\). As discussed in Chapter 2, I suspect that the T2/Onc screen failed to detect \textit{Rtl1} mutations because of a relatively weak promoter element. In contrast, our liver cancer screen with T2/Onc3 SB transposons frequently identified \textit{Rtl1} mutations, but not \textit{Egfr} mutations. This may also be due to differences in promoter activities between the two transposons. T2/Onc transposon insertions within the \textit{Egfr} locus were most frequently detected in intron 24 and in the opposite transcriptional orientation as \textit{Egfr}. Transposon insertions caused gain-of-function mutations via the generation of a C-terminally truncated Egfr protein with ligand-independent activity. Insertion of a T2/Onc3 transposon in the same position within the \textit{Egfr} locus would be more likely to cause a loss-of-function mutation due to competition between the endogenous promoter and the strong EF-1\(\alpha\) promoter within the transposon, along with transposon-initiated transcription of the antisense strand at the \textit{Egfr} locus. I hypothesize that this is the reason that \textit{Egfr} mutations have not detected in our SB-induced HCC screens utilizing T2/Onc3 transposons. If EGF is found to be the primary growth factor responsible for Rtl1’s phenotypic effects, it would provide a functional link between the most frequently mutated genes in two independent mouse models of SB-induced HCC.

Aside from being targeted therapeutically, \textit{RTL1} could serve as a biomarker or prognostic indicator for patients with liver disease. My analyses of mouse and human
liver samples demonstrate that expression is absent in normal liver. This is a desirable quality for a disease biomarker, as it increases detection sensitivity. Additionally, presence versus absence of expression is much more easily interpreted than variations in expression level between normal and diseased tissue. It remains to be determined if RTL1 expression is activated as part of the fibrogenic response to liver damage. If not, it could potentially be an indicator of progression to malignant disease in patients with liver fibrosis. In this case, analysis of RTL1 expression would be a beneficial addition to the set of diagnostic tests used to monitor fibrotic patients at high risk for HCC. Through transcriptome profiling of SB-induced HCCs, we identified an association between tumors driven by activation of Rtl1 and human HCC categorized within the molecular S1 subclass. General characteristics of this subclass include poorly differentiated tumors and early tumor recurrence following treatment(53). Given these associations, RTL1 could potentially be used as an indicator of poor prognosis.

My thesis work has identified Rtl1 as a novel driver of hepatocarcinogenesis and as a promising molecule to pursue therapeutically for human HCC. It has shed light on complex results from numerous other studies on the Dlk1-Dio3 imprinted domain, suggesting that Rtl1 may be the most functionally important element with regard to liver cancer. I have also generated preliminary data suggesting a potential mechanism of action for Rtl1 in promoting cell proliferation and tumorigenesis. This information could prove to be highly useful for developing molecular targeted therapies aimed at improving HCC patient outcome.

**Toward an Improved Mechanistic Understanding of Fibrotic Liver Cancer**

I conducted an SB mutagenesis screen to model liver cancer development in the presence or absence of hepatic fibrosis. Molecular analysis of tumors from the screen identified several interesting candidate cancer genes whose mutation was subject to
differential selective pressure depending on the fibrotic context. Though further work must be done to validate that these mutation events directly drive tumorigenesis and that they are relevant to human cancer, my work provides a list of genes that warrant further study for their role in hepatocarcinogenesis. Mutations identified preferentially in fibrotic liver tumors are of particular interest. Because the tumors developed in a context similar to that of most human HCCs, the mutations implicated as drivers are likely to be highly relevant to human disease. My results represent a significant advancement in the molecular characterization of HCC, especially considering that the majority of current animal models do not recapitulate the fibrotic environment common to most human tumors. They may also provide an opportunity to gain insight into the mechanistic contributions of hepatic fibrosis to liver tumor development, an area of study in need of clarification.

Validation experiments are currently in progress to confirm \textit{in vivo} tumor-promoting activity for a subset of CISs identified by our screen. Hydrodynamic gene delivery into the \textit{Fah\textsuperscript{-/-}} mouse model will be used for this purpose. Following liver repopulation by hepatocytes expressing a gene of interest, half of the mice will be injected with CCl\textsubscript{4} to induce liver fibrosis. I hypothesize that genes identified by our screen as commonly mutated preferentially in the presence of fibrosis were identified as a result of a specific selective pressure provided by the fibrotic environment. Based on this hypothesis, I predict that these mutations will be more potent drivers of liver tumorigenesis in the presence of fibrosis than in its absence. In this case, validation experiments should demonstrate a more severe tumor phenotype in CCl\textsubscript{4} treated mice, as compared to untreated mice. In contrast, mutations identified preferentially in mice lacking hepatic fibrosis are predicted to be stronger drivers of tumorigenesis in the absence of CCl\textsubscript{4} treatment, producing a more severe phenotype in untreated mice. Mutations identified by the screen in both cohorts of mice with no apparent bias will also
be tested *in vivo*, with the expectation that tumor severity will be similar between CCl$_4$-treated and untreated mice.

I have proposed a functional interaction between *Nfe2*, *Fign*, and *Cul3*, three genes with mutations identified by our screen in fibrotic liver tumors. In addition to being completely specific to tumors developed in the presence of fibrosis, these mutations were identified in an entirely non-overlapping set of tumors. This finding is consistent with a shared functional role, as acquisition of a mutation in one of the three genes would eliminate the selective pressure for a mutation in either of the other two. In order to test the hypothesized interactions, a phenotypic outcome based on the proposed shared functional mechanism must be identified. Due to homology between *Nfe2* and *Nfe2l2*, a gene with a well-characterized role in cellular response to oxidative stress, along with the fact that CCl$_4$ treatment induces oxidative stress(144), I hypothesize that the shared functional mechanism is likely to involve this pathway. To test this hypothesis, I have generated a reporter construct that contains a portion of the mouse *Rbx1* promoter upstream of firefly luciferase. This reporter has been shown to be highly responsive to experimental modulation of Nfe2l2 levels(145). I will use this construct to first confirm that *Nfe2* and *Nfe2l2* share *Rbx1* as a transcriptional target in the setting of cultured hepatocytes. Assuming this is the case, I will selectively modulate activity of *Fign* and *Cul3* to determine whether or not luciferase levels change in the predicted manner. My hypothesis is that overexpression of *Nfe2* or *Fign* will activate the *Rbx1* reporter, as will inhibition of *Cul3*. This outcome would suggest that the mutations identified in these three genes by our screen functionally converge on the activation of a transcriptional pattern associated with oxidative stress. Further support for this conclusion could be obtained by comparing gene expression profiles between normal liver and SB-induced liver tumors with mutations in *Nfe2*, *Fign*, or *Cul3*. The identification of a shared set of up-regulated genes associated with oxidative stress would be indicative of a common tumor-driving mechanism involving this pathway. An important control for this
experiment would be the inclusion of tumors developed in CCl₄-treated mice, but lacking mutations in any of the three genes of interest. Because CCl₄ treatment can induce oxidative stress on its own, the expression of some genes involved in response to this environment would be expected to change significantly. Gene expression changes shared by tumors with *Nfe2*, *Fign*, or *Cul3* mutations, but not by other tumors developed in CCl₄-treated mice, would be the most likely mechanistic mediators of a common tumorigenic pathway for these three events. If such a functional connection is identified between CISs detected through SB screening, it is likely to prompt investigation of similar interconnectedness within CIS lists generated by other screens.

A clear link between chronic liver fibrosis and HCC exists, and mechanisms of fibrotic contribution to hepatocarcinogenesis have been proposed. However, little direct experimentation has been conducted to confirm or refute these proposed mechanisms. By providing a genome-wide, unbiased set of mutations for selection during tumor formation in the presence or absence of hepatic fibrosis, our SB screen has the potential to shed light on processes driving fibrosis-dependent liver cancer. Mutation events preferentially identified in tumors developed in the context of fibrosis are likely to cooperate with or require such an environment to promote tumorigenesis. In contrast, genes preferentially mutated in the absence of fibrosis may function as mechanistic targets or mediators of the fibrotic process. The functional outcome of such mutations may be achieved during fibrogenesis, resulting in a decreased selective pressure for their maintenance and a failure of cells harboring them to undergo clonal expansion, explaining the observed bias for their detection in non-fibrotic liver tumors. Our screen identified four genes as CISs in non-fibrotic, but not fibrotic liver tumors, a result that demonstrates a selection bias for mutation of these genes in the absence of fibrosis. This result could indicate a positive selection pressure for mutation of these genes in normal liver that becomes a neutral selection pressure upon the induction of fibrosis due to functional redundancy. Alternatively, it could indicate a negative selection pressure for
these mutations in a fibrotic liver. Analysis of expression levels for each of these genes in fibrotic versus normal liver may provide some clarification. For example, a significant change in expression in response to the induction of fibrosis would indicate that the gene could be a mediator of the process, suggesting a probable lack of positive selection pressure for mutation in this context. Of the mutation events preferentially identified in the presence of fibrosis, some of the most common are predicted to be involved in growth factor signaling between hepatic stellate cells and hepatocytes (Met and Gli2) or cellular response to oxidative stress (Fign, Nfe2, and Cul3). Both of these processes occur during the development and maintenance of chronic liver fibrosis, and both have been proposed, among other processes, as potential mediators of hepatocarcinogenesis(18, 21). My results provide support for the direct involvement of stellate cell growth factor signaling and oxidative stress response in promoting HCC.

A major impediment to the definitive assignment of direct mechanistic roles for fibrosis in driving HCC stems from difficulties in experimentally separating the processes of inflammation and fibrosis(21). Specific and direct tumor-driving functions have been demonstrated for various inflammatory signaling molecules in the liver(146), but contributions of fibrosis are less well-characterized. In humans, hepatic fibrosis develops as a result of a sustained injury stimulus, which involves a strong inflammatory response. Inflammation and fibrosis are similarly linked in most animal models of liver disease (including CCl4-induced fibrosis), making the dissection of fibrosis-specific molecular components involved in tumor development in these contexts challenging. There are some existing mouse models that could make this experimental goal more accessible. Transgenic mice with liver-specific overexpression of either platelet-derived growth factor (PDGF)-B or –C develop hepatic fibrosis in the absence of liver damage or inflammation(147, 148). It would be informative to conduct an SB mutagenesis screen for drivers of hepatocarcinogenesis in these strains. The list of common mutations
identified in this context would likely contain genes involved in fibrosis-specific mechanisms of tumor development not dependent on inflammatory signaling.

Another interesting line of experimentation would involve generating SB-induced HCCs in a wide range of different liver fibrosis models. Several models are currently available, including those induced by chemical agents, surgical procedures, gene knockout or overexpression, and environmental factors such as diet(149). Each distinct model has a unique set of characteristics and accurately recapitulates certain elements of human chronic fibrosis. Comparison of HCC driver mutations across different models could provide valuable insight into fibrotic liver tumor development, and it would allow specific hypotheses to be generated and tested based on the particular model system(s) in which a mutation was identified. Perhaps the most beneficial information provided by this approach would be the list of common HCC drivers shared by the majority of mouse models of liver fibrosis. Their identification as functional promoters of tumor development in the context of fibrosis, regardless of etiology, would make such mutation events highly promising targets for further study and possible design of therapeutic strategies.

In our SB screen for drivers of tumorigenesis in a fibrotic liver, transposon mobilization was initiated through Alb-Cre-mediated activation of SBase expression. Alb-Cre is first expressed in the liver at embryonic day 15.5 (as determined by The Jackson Laboratory Cre Repository), so mice in our screen were subject to ongoing liver-specific transposition from this point on. Treatment with CCl₄ was started at eight weeks of age. As a result, SB-induced mutation events were already present at the time of fibrosis induction. It is possible that mutation profiles could be affected by the relative timing of mutagenesis and fibrosis. Experimental systems could be designed to more precisely control the induction of each process, thereby allowing an investigation into the importance of timing in the model. The time course of fibrosis development could easily be controlled by selecting different starting points for CCl₄ injection. Induction of
mutagenesis could be more tightly controlled by substituting \textit{Alb-Cre-ER}^{T2} for \textit{Alb-Cre}. In this way, SBase expression could be activated in a liver-specific manner only upon administration of tamoxifen. By making this modification to the system, SB-induced mutagenesis could be activated and followed by different fibrosis-free intervals prior to \textit{CCl4} injection. Alternatively, transposition could be initiated at various time points after the induction of fibrosis, allowing mutagenesis to begin at various stages of fibrotic development. The ability to precisely control the timing of fibrosis and transposition in this system could lead to the classification of driver mutations based on the relative timing of their acquisition during tumor pathogenesis, which could have important implications for therapeutic approaches targeting these drivers.

An unanswered question in the liver cancer field is whether or not maintenance of a fibrotic environment is required for tumor development to progress. One potential method to address this question would be to transplant tumors developed in a fibrotic context into mice with or without liver fibrosis. If maintenance is required, the transplanted tumor would only be expected to grow in the mouse with liver fibrosis. Another possibility is that the tumor could initiate fibrogenesis when transplanted into the liver of a mouse lacking fibrosis. There are also protocols to induce reversible hepatic fibrosis in mice, including a shorter-term \textit{CCl4} injection regimen than what was used in our SB screen. These protocols could be used to investigate the ability of a tumor arising in fibrotic liver to continue growing after reversion of fibrosis. Another method for fibrotic reversion involves intravenous injection of liposomes containing small interfering RNAs (siRNAs) directed against \textit{heat shock protein 47 (Hsp47)} to deplete the liver of stellate cells(150). Injection of mice following induction of tumorigenesis in a fibrotic liver environment could be utilized to investigate the necessity of sustained fibrosis for tumor maintenance and/or progression. Each of the methods that I have proposed to assess the importance of maintained fibrosis for liver tumor development would be most useful if tumor formation and development could be monitored in live animals. This
could be achieved through the incorporation of a luciferase reporter gene into the system to allow bioluminescence imaging (BLI). For the transplantation experiment, donor mice could be engineered to express luciferase throughout the liver, for example by including a Cre-inducible luciferase reporter allele and using the Alb-Cre strain to activate it. For tumor models induced by delivery of an oncogene or tumor-suppressor-directed shRNA, a constitutively expressed luciferase cassette could be included in the construct to allow BLI monitoring of any cells expressing the transgene of interest. A determination of whether or not liver tumors developed in a fibrotic context require maintenance of such an environment for continued development or whether there are specific molecular signatures associated with this trait would be highly beneficial for guiding therapeutic approaches in human patients. Recent evidence suggests that reversion of even severe hepatic fibrosis may be possible in patients with liver disease(151). If it is discovered that a specific molecular signature is associated with tumor regression following reversion of fibrosis, it would indicate that surgical resection may not be necessary for a patient harboring a tumor with that signature, as long as fibrosis can be reversed.

My research on molecular drivers of liver cancer in the context of hepatic fibrosis provides valuable information to others working in this field. The set of tumors that I profiled represents the largest SB-induced liver cancer screen conducted to date. The screen identified several genes likely to be involved in human HCC. Additionally, comparison of driver mutation profiles between tumors developed in the presence or absence of chronic hepatic fibrosis has the potential to help clarify the role that fibrosis plays in directly promoting hepatocarcinogenesis. The follow-up experiments that I have proposed may provide further insight into this process, and they could aid in the development of improved therapeutics for patients with fibrotic liver disease aimed at either the prevention of progression to cancer or the regression of existing tumors.
Implications for Gene Therapy

Experimental evidence from mouse studies strongly indicates that insertion of exogenous DNA within the Dlk1-Dio3 imprinted domain is associated with the development of HCC. Unbiased transposon mutagenesis screens conducted by our lab have identified integrations within the domain as the most common mutation event present in HCCs(40). Additionally, two independent studies delivering therapeutic gene products to mouse liver using recombinant AAV vectors reported the formation of HCCs with vectors commonly inserted within the same region of the domain(30, 45). These findings have implications for human gene therapy approaches, which commonly utilize AAV vectors due to their purported safety. The use of SB transposons for therapeutic gene delivery is also emerging as a popular strategy(152). A case of HCC was recently reported in a woman that had received liver-targeted adenoviral gene therapy fourteen years prior to tumor development(153). The authors conclude that insertional mutagenesis was unlikely to be a contributing factor, although no attempt was made to detect integrated vectors. Although chromosomal integration of adenoviral vectors is a very rare event(90, 91), a strong positive selection pressure could make such an event highly significant. Insertion within the Dlk1-Dio3 domain has been demonstrated to provide a strong enough positive selection pressure to drive hepatocarcinogenesis when occurring in only 0.01% of hepatocytes(31). In light of these results, it may be worthwhile to assess Dlk1-Dio3 domain status in this patient’s tumor.

Along with the viral vector studies, my results suggest that care must be taken when designing therapeutic vectors for hepatic delivery to avoid activation of expression from the Dlk1-Dio3 domain. Each of the DNA elements whose insertion within the domain has been associated with HCC development contains enhancer and promoter elements with strong activity in the liver, which likely explains the observed increase in expression of genes surrounding insertion sites in tumors. One way to avoid position effects resulting from insertion of therapeutic transgenes is to flank the expression
cassette with insulator elements. Unfortunately, due to their small cargo capacity, this is not possible with AAV vectors. It would be easily achievable, however, with the Ad-PB system that I describe in Chapter 4. Advancement of the current system and adaptation to allow efficient delivery of therapeutic genes could provide a novel technique for gene therapy with significant advantages over existing approaches.

For the initial characterization of stable integration efficiency with the Ad-PB system, we used an MOI of two. For in vivo delivery, especially in humans, it is advisable to use the lowest possible dose that provides therapeutic efficacy. An MOI of two is easily achievable in vivo and is considerably lower than what is commonly used. Our results demonstrate that even at this low dose, the efficiency of stable transgene insertion is much higher than can be achieved with most existing hybrid Ad vector systems, a finding that bodes well for adaptation to gene therapy approaches. Experiments are currently underway to assess the efficiency of stable integration from Ad-PB vectors when delivered at MOIs of ten or twenty, doses commonly used with other hybrid Ad vectors. Delivery of an increased number of viral genomes is predicted to result in higher stable transduction efficiency.

Besides delivering more viral particles, other methods to increase the efficiency of the Ad-PB system in vivo include decreasing immune-mediated clearance, increasing the copy number of transposons included within the viral genome, and improving the delivery and/or activity of PBase. For the experiments conducted so far, we have utilized a recombinant adenoviral vector of human serotype 5 (Ad5) with deletion of the E1 and E3 genes, commonly referred to as a “first-generation” Ad vector(89). This type of vector retains low-level expression of viral genes, generating an immune response in vivo that leads to its clearance. To avoid this issue, high-capacity (HC) Ad vectors could be utilized instead. HC-Ad vectors lack all viral genes, containing only the ITRs and packaging signal. Because they don’t express viral genes, immune-mediated clearance is much lower for these vectors, as compared to first-generation Ad vectors. Packaging of
HC-Ad vectors is more complex and requires Cre recombinase, so inducible HC-Ad-PB vectors would have to be engineered with recognition sites for a different recombinase, such as Flp or Dre.

The efficiency of stable transgene delivery could potentially be increased by including a second or third PB transposon within the Ad genome. An Ad-PB vector harboring two transposons with fluorescent reporter genes is currently being generated to determine whether or not this is the case. The cargo capacity of the current Ad-PB system is ~10kb, which is enough space to include most transgene cassettes, including flanking insulators. However, a capacity of 10kb is insufficient to incorporate multiple transposons with insulators for most transgenes. For this to be possible, an HC-Ad vector would need to be used, which has a cargo capacity of ~37kb.

In addition to modifications made to the Ad-PB transposon vector, alterations to the other component of the transposition system, PBase, could further improve efficiency of the system. It has been demonstrated that engineered modifications to PBase that increase its efficiency commonly promote higher levels of protein expression(105, 106). Unlike the SB system, PB transposition does not appear to be affected by overproduction inhibition, a phenomenon in which transposition frequency decreases as transposase expression is increased excessively(44, 105). These results suggest that increased expression of PBase in cells transduced with Ad-PB vectors could improve stable integration efficiency. Our initial characterization experiments were conducted in cells stably expressing hyPBase from an integrated lentiviral vector. An AAV vector expressing hyPBase is currently being generated for in vivo characterization experiments. Because they are substantially less immunogenic, AAV vectors can be delivered at higher MOIs than Ad vectors in vivo(154). Delivery of PBase with AAV vectors will ensure high protein levels in cells transduced with Ad-PB vectors, facilitating efficient transposition. Li and colleagues showed that the efficiencies of PBase-mediated excision and re-integration can be differentially modulated through mutagenesis by generating a
version of PBase capable of efficient excision but incapable of re-integration(155). Presumably, a similar approach could be taken to selectively increase the integration efficiency of PBase, which would be beneficial for gene therapy approaches.

A major concern with any therapeutic gene delivery system that involves integration of exogenous DNA into the genome is insertional mutagenesis. It would be highly advantageous to have the ability to target transposon insertions to specific sites in the genome, thereby avoiding disruption of coding elements. Multiple approaches have shown that fusion of PBase to specific DNA binding domains can bias insertion site preference to a region of interest(156, 157). While there is certainly room for improvement in specific targeting efficiency, the fact that it can be achieved at all provides a significant biosafety advantage for human gene therapy approaches.

The Ad-PB system represents a novel gene delivery tool that has the potential to be highly useful for the stable transfer of therapeutic genes. Though still in the early stages of development and characterization, the system has already shown significant advantages over other techniques currently utilized for gene therapy. Further refinement should lead to improved efficiency and biosafety, making the Ad-PB system a promising candidate for future gene therapy applications.

**Conclusions**

It is an exciting time to be a researcher in the field of cancer genetics. Both at the bench and in the clinic, the pursuit of molecular drivers of tumorigenesis has become a major focus. This approach has tremendous potential to improve patient outcome through the development of improved screening and therapeutic methodologies. Indeed, the successful development of molecular targeted therapies with greater efficacy and fewer side effects, as compared to the previous standard of care, has already been achieved for certain subsets of some major cancers. At the same time, the lack of effective targeted therapies for the majority of cancers and the frequent evolution of
resistance to treatment indicate that much remains to be learned about the molecular mechanisms of tumor development and progression if the impact of disease is to be minimized. This is particularly true for HCC, a malignancy for which the only therapy that has shown any benefit for late stage patients provides a modest increase in survival of ~10 weeks.

My primary goal in conducting my thesis research has been to contribute to an increased understanding of the molecular pathogenesis of liver cancer. I have achieved this goal through the identification of several promising candidate genes likely to play a role in HCC development and through the development of a novel tool for meaningful validation of such candidates. By conducting a forward mutagenesis screen in the presence of liver fibrosis, I have discovered several putative liver cancer genes with a high likelihood of relevance to human HCC. This work has additionally provided some insight into the role that fibrosis plays in directly promoting hepatocarcinogenesis, an area of study that needs more clarification. I have validated a tumor-promoting function for Rtl1, one of the candidate genes identified by SB screening for liver cancer genes, and demonstrated that its human ortholog is similarly disrupted in human HCC. My most significant contribution to liver cancer genetic research may turn out to be my involvement in the development of the Ad-PB system for validation of candidate cancer genes. The relative efficiency, simplicity, flexibility, and accessibility of this system as compared to other validation tools could lead to its utilization by a large number of researchers in a wide variety of subfields of cancer research. Increasing the ability of researchers to conduct meaningful validation experiments in vivo would make a significant positive impact on efforts to develop more effective cancer therapies.
REFERENCES


115. Xu, W. P., Yi, M., Li, Q. Q., Zhou, W. P., Cong, W. M., Yang, Y., Ning, B. F.,
Yin, C., Huang, Z. W., Wang, J., Qian, H., Jiang, C. F., Chen, Y. X., Xia, C. Y.,
LIN28A-NF-kappaB regulatory circuit contributes to the development of
hepatocellular carcinoma. *Hepatology*

factor-4alpha reverses malignancy of hepatocellular carcinoma through regulating
miR-134 in the DLK1-DIO3 region. *Hepatology*

117. Anwar, S. L., Krech, T., Hasemeier, B., Schipper, E., Schweitzer, N., Vogel, A.,
the DLK1-MEG3 locus in human hepatocellular carcinoma. *PLoS One 7*, e49462

S., Demetris, A. J., Nalesnik, M., Yu, Y. P., Ranganathan, S., and Michalopoulos,
carcinomas and hepatoblastomas. *Hepatology 44*, 1012-1024

of the imprinted DLK1-MEG3 locus in human clinically nonfunctioning pituitary
adenomas. *Am J Pathol 179*, 2120-2130

MEG3 expression and intergenic differentially methylated region
hypermethylation in the MEG3/DLK1 locus in human clinically nonfunctioning
pituitary adenomas. *J Clin Endocrinol Metab 93*, 4119-4125

121. Butz, H., Liko, I., Czirjak, S., Igaz, P., Korbonits, M., Racz, K., and Patocs, A.
(2011) MicroRNA profile indicates downregulation of the TGFbeta pathway in
sporadic non-functioning pituitary adenomas. *Pituitary 14*, 112-124

122. Yin, D., Xie, D., Sakajiri, S., Miller, C. W., Zhu, H., Popovicci, M. L., Said, J. W.,
and associated with oncogenic activities. *Oncogene 25*, 1852-1861

123. Yanai, H., Nakamura, K., Hijioka, S., Kamei, A., Ikari, T., Ishikawa, Y.,
surface antigen on foetal hepatic stem/progenitor cells, is expressed in
hepatocellular, colon, pancreas and breast carcinomas at a high frequency. *J
Biochem 148*, 85-92

region repress the epithelial-to-mesenchymal transition by targeting the TWIST1


