Assessment of DNA copy number alterations in mouse and human hepatocellular carcinoma

Grace Chappell

A technical report submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in Public Health in the Department of Environmental Sciences and Engineering

April 14, 2014

Approved by,

Ivan Rusyn, M.D., Ph.D.

Igor Pogribny, M.D., Ph.D.

Rebecca Fry, Ph.D.
Abstract

Grace Chappell

Assessment of DNA copy number alterations in mouse and human hepatocellular carcinoma

(Under the direction of Ivan Rusyn, M.D., Ph.D.)

Hepatocellular carcinoma (HCC) is a prevalent human cancer with rising incidence worldwide. Human HCC is typically associated with chronic liver inflammation and cirrhosis, which is typically the consequence of toxic insults, disturbances in metabolism, or viral infection. To better understand the pathogenesis of liver cancer, we studied molecular mechanisms in the development of HCC using a human-relevant mouse model of fibrosis-associated hepatocarcinogenesis. Because the increased incidence of liver tumors in this model was associated with epigenetic events indicative of genomic instability in a previous study, DNA copy number alterations (CNAs), a prominent feature of genomic instability and a common characteristic of cancer in both humans and model organisms, were evaluated using array comparative genomic hybridization and computational tools. A considerably higher frequency of CNAs was found in the tumors than in the matched fibrotic liver tissues, indicative of a profound increase in genomic instability with HCC progression. The same computational analysis was then conducted on publicly available data from human HCC patients, comparing human tumor and non-tumor cirrhotic tissue, as well as the profiles of human and mouse tumors. The results indicate that structural aberrations are profoundly increased in neoplastic tissues as compared to pre-neoplastic, damaged tissue in both species, confirming the utility of this particular mouse model and generating a list of inter-species relevant genes with aberrant copy number.
Acknowledgements

First and foremost, I am extremely grateful to my advisor, Dr. Ivan Rusyn, for giving me the opportunity to return to graduate school after several years away from academia, and participate in research in his lab. The guidance, assistance, support and (persistent) encouragement that Ivan has provided me over the past two years has proven to be invaluable. I look forward to a few more years working for Ivan and feel confident that the future holds many more exciting and educational experiences. I also have immense appreciation for Dr. Igor Pogribny, who allowed me the opportunity to spend time in his lab for my practicum study. Dr. Pogribny also has made himself extremely available, albeit remotely, for my questions and concerns during the design and execution of experiments. I am grateful for my third committee member, Dr. Rebecca Fry, for taking time to meet with me on several occasions to chat about science. I also was fortunate enough to have Dr. Fry as a lecturer in several of my classes during my time at UNC; during these classes I learned about genetic and epigenetic alterations as a result of exposure to toxic substances in the environment, a major interest of mine and focus of my research.

A huge thanks goes to Grace Silva, who I have referred to as “the other Grace” countless times when presenting and discussing my project, for being an amazing collaborator. I am extremely grateful for her patience in explaining bioinformatics programming techniques, which is nearly a foreign language to me, and her enthusiasm for the project. Without Grace’s assistance this project would not have been possible and I am forever indebted to her for her hard work.

I would also like to thank all of the members of the Rusyn lab, past and present, who have given me support, feedback, companionship, and assistance over the past two years.
Table of Contents

List of Tables v
List of Figures vi
List of Abbreviations vii
Chapter 1: Literature Review 1
Chapter 2: Introduction 9
Chapter 3: Materials and Methods 11
Chapter 4: Results 20
Chapter 5: Discussion, Limitations, Conclusions, and Future Directions 31
Chapter 6: Practicum Report 40
References 44
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Selected genes and their Applied Biosystems qRTPCR primer assay numbers</td>
<td>18</td>
</tr>
<tr>
<td>2. Expression of selected set of genes</td>
<td>30</td>
</tr>
</tbody>
</table>
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Chromosome ideograms demonstrating CNAs due to large-scale structural variation</td>
<td>3</td>
</tr>
<tr>
<td>2.</td>
<td>Mouse model of fibrosis-associated HCC</td>
<td>11</td>
</tr>
<tr>
<td>3.</td>
<td>Representative lowess-normalized intensity output</td>
<td>13</td>
</tr>
<tr>
<td>4.</td>
<td>Example of pathology report data</td>
<td>15</td>
</tr>
<tr>
<td>5.</td>
<td>SWITCHdna plots of mouse tumor and matched non-tumor fibrotic liver samples</td>
<td>21</td>
</tr>
<tr>
<td>6.</td>
<td>SWITCHdnaplus plots of mouse tumor compared to matched non-tumor fibrotic liver samples</td>
<td>22</td>
</tr>
<tr>
<td>7.</td>
<td>SWITCHdna plots of human tumor samples from patients with HCC and cirrhosis or fibrosis</td>
<td>23</td>
</tr>
<tr>
<td>8.</td>
<td>SWITCHdna plot of non-tumor cirrhotic liver tissue in patients with HCC</td>
<td>24</td>
</tr>
<tr>
<td>9.</td>
<td>SWITCHdnaplus plots of comparing humans HCC in a background of cirrhosis or fibrosis</td>
<td>25</td>
</tr>
<tr>
<td>10.</td>
<td>SWITCHdnaplus plots comparing segments with homologous human and mouse genes</td>
<td>26</td>
</tr>
<tr>
<td>11.</td>
<td>Expression of a set of up-regulated genes located in gained segments</td>
<td>27</td>
</tr>
<tr>
<td>12.</td>
<td>Expression of a set of down-regulated genes located in gained segments</td>
<td>28</td>
</tr>
<tr>
<td>13.</td>
<td>Comparison of the expression level of a set of genes located in gained segments in tumor and non-tumor mouse tissue</td>
<td>29</td>
</tr>
</tbody>
</table>
List of Abbreviations

BAC: bacterial artificial chromosome
CBS: circular binary segmentation
CIN: chromosomal instability
CNA: copy number alteration
CCl₄: carbon tetrachloride
DAVID: Database for annotation, visualization and integrated discovery
DEN: N-nitrosodiethylamine
FDR: False Discovery Rate
HCC: hepatocellular carcinoma
HBV: hepatitis B virus
Lowess: locally weighted scatterplot smoothing
mRNA: messenger RNA
NCTR: National Center for Toxicological Research
NTP: National Toxicology Program
qRT-PCR: quantitative reverse transcription polymerase chain reaction
SNP: single nucleotide polymorphism
SWITCHdna: SupWald Identification of Changes in DNA
TCGA: The Cancer Genome Atlas Project
UMD: UNC microarray database
Chapter 1

Literature Review

The Importance of Modeling Human HCC

Hepatocellular carcinoma (HCC) is one of the most prevalent, life-threatening human cancers (Center and Jemal 2011). While the overall cancer incidence and death rates are steadily declining, the incidence of HCC continues to increase (Center and Jemal 2011; AACR Cancer Progress Report Writing Committee et al. 2013). The development and progression of HCC is a complex and multistep process characterized by the progressive, sequential evolution of morphologically distinct pre-neoplastic lesions (formed as a result of chronic liver injury, inflammation, hepatocellular degeneration and necrosis, hepatocellular regeneration and small cell dysplasia, followed by the appearance of low- and high-grade dysplastic nodules), which eventually culminates in the formation of HCC (Farazi and DePinho 2006; Aravalli et al. 2013).

In humans, the majority (70-90%) of HCC cases are associated with advanced liver fibrosis or cirrhosis (Farazi and DePinho 2006), with HCC frequently arising in a liver with chronic inflammation and fibrosis/cirrhosis. In fact, the incidence of HCC in non-cirrhotic liver is estimated at about 15-20% of all HCC, and only 10-12% in perfectly healthy livers. Liver fibrosis and cirrhosis associated with HCC are typically the consequence of toxic insults, disturbances in metabolism or viral infection (Fattovich et al. 2004). Examples of these include exposure to or ingestion of alcohol, fatty liver disease, and chronic hepatitis B or C viral infection, respectively. While the histopathologic features of HCC is well established, the molecular mechanisms of the cancer-promoting effects of the main etiological factors, including cirrhosis, are not well understood (Aravalli et al. 2013; Zucman-Rossi 2010). An understanding of the molecular mechanisms that regulate the pathogenesis and progression of HCC is critical for prevention of this disease and development of effective therapies (Forner et al. 2012). While investigation of
these mechanisms using human HCC samples is desirable, few epidemiological studies have established both the causality and molecular underpinnings of the disease. Thus, animal models that resemble human HCC development may provide important molecular mechanisms involved in the various etiological factors linked to HCC (Heindryckx et al. 2009). Because most human cases of HCC develop in highly fibrotic or cirrhotic liver, characterizing the cellular and tissue features in which tumors develop is as important as identifying specific genes and pathways that are involved in tumorigenesis, and mouse models that emulate human liver disease that leads to HCC are, therefore, essential in the study of hepatocellular carcinogenesis (Fausto and Campbell 2010).

Many mouse models of HCC have been previously constructed, with various levels of similarity to the human manifestation of the disease (Fausto and Campbell 2010). One commonly used mouse liver cancer model is a single low dose injection of the genotoxic carcinogen N-nitrosodiethylamine (DEN) into 14-day-old male mice (Vesselinovitch and Mihailovich 1983). Additionally, repeat dosing of the pro-fibrogenic agent carbon tetrachloride (CCl4) also results in development of HCC (Fujii et al. 2010). To model key pathophysiological events of human cirrhosis-associated hepatocarcinogenesis, a combination of genotoxic, e.g., DEN, and non-genotoxic, e.g., CCl4, insults has been used to study the mechanisms involved in the development of fibrosis-promoted HCC in mice (Uehara et al. 2013). This particular model emulates the co-morbidity features often observed in human HCC patients and offers a relevant method for the study the mechanisms involved in hepatocellular carcinogenesis.
What is a Copy Number Alteration?

Copy number alterations (CNAs) occur as a consequence of errors during cell, and thus DNA, replication. CNAs refer to DNA segments (ranging from 1 kilobase to several megabases in size) that are present at a variable copy number when compared to a reference genome (Redon et al. 2006). After normal cell division, the expected DNA copy number value is two. Changes in DNA copy number results in either higher (copy number gain) or lower (copy number loss) values than the expected value. Many CNVs have no apparent influence on phenotype, while others have been definitively linked with disease, including cancer (Clancy 2008). In fact, structural and numerical chromosomal changes occur in the majority of cancer cells (Beroukhim et al. 2010)

Figure 1. Chromosome ideograms demonstrating CNAs due to large-scale structural variation.
Copy Number Alterations and Cancer

Somatic DNA alterations are frequent in cancer. CNAs represent one type of genetic aberration commonly involved in tumorigenesis, including hepatocellular carcinogenesis (Wang et al. 2013). An accumulation of these errors leads to a phenotype referred to as chromosomal instability (CIN), which is one manifestation of genomic instability (Gordon et al. 2012). Genomic instability is considered one of the hallmarks of cancer and is characteristic of all types of cancer (Hanahan and Weinberg 2011). An association between changes in DNA copy number values and the development and progression of cancer has been observed in various types of cancer, and in mouse models as well as in human clinical samples (Zucman-Rossi 2010; Wang et al. 2013; Weigman et al. 2012; Aleksic et al. 2011). These copy number alterations are thought to represent a type of oncogenic driver in the progression of cancer. For example, sites of DNA copy number gains are known to harbor oncogenes, while other sites of DNA copy number losses are found in tumor suppressor genes (Wang et al. 2013; Weigman et al. 2012).

An important consideration of the multistep process of carcinogenesis is that the standard mutation rate cannot explain the extensive number of mutations present in many types of cancer cells (Tomlinson and Bodmer 1999; Vogelstein et al. 2013; Nowak et al. 2002). The acquisition of some form of chromosomal instability is likely a necessary event in tumor development, probably occurs very early, and presents an explanation of the numerous karyotypic aberrations that are often observed in malignant tumors (Aleksic et al. 2011). The subsequent process of clonal expansion by a stepwise accumulation of numerical chromosome changes would, at least in some cases, be expected to give rise to subpopulations of neoplastic cells related by the same karyotypic abnormalities (Heim 1996). It has also been postulated that chromosome instability caused by mitotic errors drives cancer evolution, with pressure to avoid cytogenetic chaos by natural selection. The non-random and tumor-specific distribution of chromosomal gains and losses is the consequence of chromosome instability, as well as the
selection for specific phenotypes among major transcriptomic changes (Hu et al. 2013; Ried et al. 2012). Interestingly, although mouse models of epithelial carcinomas show various results in the similarity of CNA profiles to human solid tumors (in frequency and in genes involved), an in vitro study using mouse cells demonstrated conservation of chromosomal location and distribution of chromosomal gains and losses relative to those commonly seen in human cancer (Padilla-Nash et al. 2012). Additional studies of animal models of specific cancer types and sub-types are warranted, as addressed in the present study.

While genomic imbalances in cancer cells have been definitively shown to be non-random and involved in tumorigenesis, the cause for particular regions to be frequently, and seemingly preferentially, lost or gained remains poorly understood. It is likely that many of these regions are recurrent, whether within one specific carcinoma or across types, as a consequence of tumor evolution, which induces a selective pressure on CNAs with a growth advantage, whether due to the amplification of a region that harbors growth factors (e.g. EGFR) or the loss of genes that regulate apoptosis (e.g. PTEN) (Huang et al. 2011; Villanueva et al. 2007). Furthermore, CNAs that are specifically associated with relapse, survival, or specific cancer subtypes have been identified and demonstrated to be preferential for genes and/or pathways that have a causative role in the cancer sub-classed or clinical outcomes (Mullighan et al. 2008; Villanueva et al. 2008). Intuitively, CNAs that confer transcriptional changes that lead to the loss of cell viability would be selected against. Further, a study of CNAs in non-small cell lung cancer found that CNAs were more frequent in regions that did not contain genes, with this tendency decreasing with progression (Huang et al. 2011).

Genomic imbalances are common in HCC, with most cases harboring multiple chromosomal abnormalities (Villanueva et al. 2008). Multiple studies have been conducted using sets of clinical human samples to evaluate the molecular mechanisms involved in tumor development at both the gene expression and DNA copy number levels. Frequent and
recurrent CNAs have been found in multiple cancer types, with those found in colorectal (Bond et al. 2014; Diep et al. 2006), breast (Weigman et al. 2012; Perou et al. 2000), renal, prostate (Wyatt et al. 2013), and hematopoietic cancer being the best characterized (Kim et al. 2013). Several such CNAs have been identified as having diagnostic and prognostic value and utility in clinical oncological diagnostic practices (Teixeira and Heim 2005). These CNAs provide key information regarding the pathogenetic mechanisms of tumorigenesis and are used for pathogenetic inferences in cancer cytogenetics. HCC, however, does not have as well an established molecular classification system as many other cancers, nor are there generally accepted CNA profiles for use in diagnostic procedures. Nonetheless, CNAs have been studied in HCC and, as clinical data accumulates, a list of important, frequent CNAs continues to strengthen. For example, HCC recurrence has been associated with loss at 13q (Kusano et al. 2002), and a gain of 3q is indicative of relatively poor recurrence-free and overall survival (Poon et al. 2006). The most commonly observed chromosomal gains and losses observed in human HCC are: gains of chromosome 1q, 6p, 8q, 17q, 20q, and losses of chromosome 1p, 4q, 6q 8p, 13q, 16q, and 17p (Wang et al. 2013; Jia et al. 2011; Midorikawa et al. 2007; Homayounfar et al. 2013).

**CNAs and Genes: What Drives Cancer?**

The genomic regions that are aberrantly copied in cancer cells contain thousands of genes, some of which are known to be associated with cancer and others that are unrelated. Within chromosomal regions of CNA, genes exist that are unrelated to carcinogenic processes but are subject to up-regulation due to proliferative advantage (Villanueva et al. 2007). While examples exist of genes for which dose and expression are directly correlated and are also known to be involved in carcinogenesis (e.g. *EGFR* and lung carcinoma, *ERBB2* and breast...
carcinoma), there are many genes found in variable copy number in cancerous tissue whose role in carcinogenesis remains poorly understood or completely unknown. This is especially true for HCC, a heterogeneous cancer for which there does not exist any particular sub-type associated with an aberrant gene dose. Essentially, although several studies have found genes or clusters of genes associated with survival or prognosis (Roessler et al. 2012), there remains a lack in the understanding of which of the genes found in gained or lost segments in liver tumors are involved in tumorigenesis, and which genes are “passengers” located in gained or lost segments of genomic imbalances. Ongoing research on human clinical samples is necessary to identify candidate genes of interest in tumor development, and ongoing molecular studies are required to characterize the molecular basis of gene involvement in hepatocellular carcinogenesis. When data are available, gene expression can be associated with gene dose, enabling the identification of cancer-and CNA-associated genes. In previous work conducted using aCGH in breast cancer cells, CNA copy number was shown to influence gene expression across a large range of CNAs, with 62% of highly amplified genes also being expressed at an elevated level (Pollack et al. 2002).

A general consensus for CNAs in mouse models of HCC has not yet been established, although some studies have shown similarities in mouse HCC and those observed in humans: Aleksic, et al demonstrated chromosomal instability as an early event in HCC in mice treated with the genotoxic agent DEN and the tumor promoter phenobarbital (Aleksic et al. 2011). Zimonjic, et al found several recurrent, non-random CNAs in Myc-transgenic mice (Zimonjic et al. 2009).

Based on these results and others, CNAs are an important molecular feature of HCC. In addition to genetic aberrations (mutations), epigenetic alterations (DNA and histone methylation), and expression changes, CNAs should be considered in analysis of the underlying molecular mechanisms of HCC.
**CNAs in HCC: Application to Environmental Disease and Public Health**

HCC occurs as the result of several etiological factors, including viral infection, toxic insult, alcohol consumption, and metabolic disease (Fattovich et al. 2004). Because the liver is an organ that is often affected by ingestion of toxic substances, HCC is of particular interest in the field of toxicology. To this aim, effective animal models of human disease are especially important in the hazard assessment of particular chemical carcinogens. Furthermore, the rising incidence of HCC worldwide presents a major public health concern (Forner et al. 2012), stressing the importance of effective models of the disease and efforts to better understand the molecular mechanisms involved in the pathogenesis of HCC. Specifically, therapies targeted at particular genes are some of the most effective anti-cancer therapeutic strategies, and identifying genomic signatures for particular subtypes of any cancer aids in the development of such strategies. In addition, efforts to develop novel anticancer compounds that directly target the mitotic errors that cause CIN have been previously made (Jallepalli and Lengauer 2001). The prevalence of CNAs in HCC supports the merit of such pharmacological efforts.
Chapter 2

Introduction

HCC is a complex and heterogeneous disease, and ongoing research to better understand the underlying molecular mechanisms is imperative. Recurrent CNAs represent a feature of HCC that is frequently reported in human clinical cases. CNAs are a type of structural variant that occur as one of the consequences of genomic instability, which is one of the “hallmarks of cancer” (Hanahan and Weinberg 2011). The involvement of CNAs in hepatocellular carcinogenesis has been demonstrated, although the exact mechanistic role of such alterations is not clear. The use of animal models and bioinformatics tools can aid in identification of common CNAs, as well as characterization of specific CNAs as they are related to specific subtypes and etiologies of HCC.

In a previous study, we found that epigenetic alterations indicative of genomic instability were prevalent in the tumor tissues of mice treated with DEN+CCl₄, as compared to mice treated with DEN alone, CCl₄ alone, or the vehicle-controls (Chappell et al. 2014). Genomic instability manifests in various forms, one of which is CIN, which leads to copy number alterations (Janssen and Medema 2013). Therefore, the main hypothesis of the present study was that CNAs would be frequent in the liver tumors of mice that also had liver fibrosis, relative to a reference genome (liver tissue of vehicle-control mice from the same study).

The relevance of the fibrosis-associated mouse model of human liver cancer can be further confirmed by additional comparative molecular studies. Human copy number data can be used to evaluate and identify common CNAs in specific groups, such as those patients who have a diagnosis of cirrhosis, or those diagnosed with viral hepatitis. Comparison between human data and that of model organisms can confirm or scrutinize the robustness of the animal model of disease.
One of the challenges in interpreting the involvement of CNAs in oncogenic processes is delineating between oncogenic “drivers” and “passenger” genes, because both will likely be present in large segments of aberrantly copied genomic DNA (Fausto and Campbell 2010). A goal of the present study was to apply molecular and statistical techniques to obtain a list of genes present in aberrant copy numbers in both human and mouse liver tumors in a background of cirrhosis (human) and fibrosis (both mouse and human). Genes found in segments of genomic gain or loss in both species, and in the pathologic state most commonly seen in HCC patients, represent candidates for further study in mechanisms of HCC.

The overarching goal of the present study was to evaluate CNAs in a mouse model of human-relevant HCC, to evaluate CNAs in a set of human HCCs in patients diagnosed with liver disease, and to compare the findings. The first objective was to characterize CNAs in both tumor and non-tumor, surrounding fibrotic liver tissue in mice using array comparative genomic hybridization (aCGH) and bioinformatics tools for analysis. The next objective was to compare the findings in the mouse to CNAs found in human clinical samples. Finally, analysis of the genomic regions of overlap between mouse fibrosis-associated HCC and human cirrhosis-associated HCC was conducted.

In the present study, both human data from clinical HCC patients and mouse data from a controlled experiment were utilized in the analysis of CNAs. To this end, the inter-species approach enabled the characterization of CIN involvement in HCC and the generation of a list of aberrantly copied genes observed in both clinical and experimental conditions.
Chapter 3

Materials and Methods

Animals

The in-life portion of this study, mouse treatments, tissue collection protocols, and the incidence of neoplastic liver lesions are detailed in Uehara et al (Uehara et al. 2013). Briefly, male B6C3F1/J mice were allocated randomly to one control and three experimental groups. At two weeks of age, mice from two of the experimental groups were injected i.p. with DEN (1 mg/kg) in sterile phosphate buffered saline (PBS; 15 ml/kg). Mice from the control group and the remaining experimental group were injected with sterile PBS only. At eight weeks of age, mice from the control and the DEN-treated groups were injected i.p. two times per week with sterile olive oil (15 ml/kg). Mice from the remaining two experimental groups were injected i.p. two times per week with CCl₄ (0.2 ml/kg) diluted in sterile olive oil for an additional 14 weeks. In summary, the groups were treated with either PBS+olive oil, DEN+olive oil, PBS+CCl₄, or DEN+CCl₄. All mice were sacrificed at 22 weeks of age.

Figure 2. Mouse model of fibrosis-associated HCC. Dosing schematic and gross images taken from Uehara, et al (Uehara et al. 2013).
**DNA extraction and Array Comparative Genomic Hybridization**

Genomic DNA was extracted from frozen liver samples from vehicle-control and DEN+CCl₄-treated mice using a DNEasy kit (Qiagen, Valencia, CA). A total of 18 tumor samples was included in the study, as well as 18 matched non-tumor samples taken from fibrotic, non-tumorous, surrounding liver tissue. The 36 total samples were randomized into 3 batches of 12 to be run together for aCGH analysis. DNA from healthy livers of 6 vehicle control mice was pooled together and used as the reference genome in the aCGH experiments. The selection of these 6 controls, rather than all 7 from the original study, was based on availability of frozen tissue at the time of the initiation of the present study. Similarly, a total of 18 tumor and matched non-tumor samples were used (not the total 27 mice included in the DEN+CCl₄ treatment group) based on the availability of separated tumor and non-tumor liver tissues in these mice. Genomic DNA was enzymatically cut, subjected to a low level of amplification, quantified, fluorescently labeled with Cy3 (green, reference DNA) and Cy5 (red, sample DNA) dyes and then hybridized to a SurePrint G3 4x180K Mouse Genome CGH Microarray (Agilent, USA). This array includes content sourced from the UCSC mm9 (NCBI Build 37), July 2007. This particular array was chosen for use in this experiment based on its coverage of mouse genome, uniform spacing, most current NCBI build incorporation (second only to the 1M array), and cost efficiency.

**Normalization and segmentation of array data**

After the arrays were scanned, files were uploaded to the UNC Microarray Database (UMD). Files uploaded to UMD are lowess (locally weighted scatterplot smoothing) normalized, a method to adjust for intensity-dependent variation in dye bias for experiments in which two fluorescent dyes are used, was performed. Normalized values are presented as LogRRatio, calculated as: Log₂(observed intensity/reference intensity). After normalization, analysis of signal intensities across the genome were used to identify regions with CNAs.
Segmentation

Circular binary segmentation is a typical method proposed by Olshen et al. (Olshen et al. 2004) used to search for change points in an ordered sequence of values, which define segments with a different distribution of values (often measured by having different means). This method recursively tests if each new segment or breakpoint should be introduced inside an existing segment based on the differences in the distribution of values between the newly introduced segment and the remainder of the existing segment, or between the two resulting segments separated by the proposed breakpoint.

Figure 3. Example of LOWESS-normalized array data. aCGH data for chromosome 19 from two individual mice showing normalized values presented as LogRRatio, calculated as: $\log_2(\text{observed intensity/referece intensity})$. These values were then used for segmentation.

Recurrent copy number alterations were identified by defining altered segments of DNA and calculating their frequency among the sample set. SWITCHdna (SupWald Identification of dna-copy Changes) (Weigman et al. 2012) and CBS (circular binary segmentation) (Olshen et al. 2004) are methods used to identify regions of copy number variation based on intensity data.
generated from copy number platforms (such as aCGH or single nucleotide polymorphism (SNP) arrays). SWITCHdna identifies transition points by calculating an F-score at each array data point (oligonucleotide, SNP, or BAC, depending on the platform utilized), and then compares these scores to surrounding points in the genome. CBS is a permutation-based algorithm that identifies segments by detecting change-points using a maximal-\( t \)-test. Both methods define segments of change, which each have an average intensity value, size, and a corresponding statistical significance score.

**Human HCC Data Used in CNA analysis**

Human copy number data was downloaded from The Cancer Genome Atlas (TCGA), a multi-institutional effort to understand the molecular basis of cancer using genome sequencing and bioinformatics techniques, supervised by the National Cancer Institute (http://cancergenome.nih.gov/). Genome-wide SNP array data is collected and processed by circular binary segmentation by The Broad Institute. Additionally, pathology reports and other clinical information are publicly available for download through TGCA. To select patient data for use in the analysis of cirrhosis-associated HCC, the following inclusion criteria were applied: 1.) a diagnosis of either cirrhosis or fibrosis as detailed in the pathology report of the patient and 2.) circular binary segmented copy number data available for download.
Figure 4. Example of pathology report data inclusion. Pathology reports for two human HCC patients included in the TCGA database were reviewed. Samples taken from liver tumors arising in a background of fibrosis or cirrhosis were selected for analyses, representing the most common conditions involved human HCC.

The human HCC data available from TCGA was obtained from various institutions. The copy number data was collected and processed by the Broad Institute, using the Affymetrix genome-wide human SNP array 6.0.

**Graphing Tools**

To evaluate the segments gained and lost among all of the samples in each group (mouse tumor, mouse matched non-tumor, human fibrosis-associated tumor, human cirrhosis-associated tumor, human non-tumor), normalized data were analyzed using an R-coded statistical program called SWITCHdna. SWITCHdna compiles segments and plots the frequency of samples within a subset population that have significant copy number alterations at each identified segment. More information and a download of the R code can be found at: [https://genome.unc.edu/pubsup/SWITCHdna/](https://genome.unc.edu/pubsup/SWITCHdna/).
To investigate CNAs associated specifically with tumors and non-tumor tissues, additional statistical analyses were performed on the copy number data. SWITCHdnaplus, an additional code based on SWITCHdna, performs a two-tailed t-test on the copy number values for each array data point for all of the samples from tumors against the copy number values for that same data point in non-tumor surrounding tissue. This analysis was conducted in tumor and matched non-tumor samples in both humans and mice, and then between tumors from patients with HCC and cirrhosis versus tumors from patients with HCC and fibrosis. The resulting $P$ values were adjusted by the Benjamini-Hochberg method (Benjamini and Hochberg 1995) to correct for FDR in multiple-hypothesis testing, and genes with $P$ values <0.05 were then considered significantly different between the two groups.

To compare human and mouse CNAs, SWITCHdnaplus has the capability of identifying mouse segments that contain genes that are homologous to the human genome, and then comparing the copy number values between those segments in the mouse and human samples. A graphical output for this analysis “re-maps” the mouse homologous genes in human order and then highlights segments with copy number gains or losses in both sample sets.

**Expression Analysis and Quantitative Reverse Transcription**

The genes used for expression analysis were selected from a list of all genes that were present in gained or lost segments in tumors from both human cirrhosis patients and DEN+CCl₄-treated mice. The SWITCHdnaplus output includes a list of the genes contained in each segment. Comparing and matching the list of genes contained in the segments for which both mouse and human tumors had a significant gain or loss (human gene list), with those gained or lost specifically in mouse tumors (mouse gene list), yielded a list of shared genes with aberrant copy number. This gene list (2,347 genes) was then uploaded to the DAVID
Database for Annotation, Visualization and Integrated Discovery v6.7 online functional annotation tool that can be used to aid in understanding the biological meaning behind large lists of genes (Huang da et al. 2009). From this list, a small set of genes was selected for a representative study of the association between gene dose and gene expression. Genes were selected based on a known association with carcinogenesis (Akt1, Erbb2, Jrk, Map3k6, Rara, Tnf, Vegfa, Wnt1, Wnt10b, DNA damage and repair response mechanisms (Gadd45b and Xrcc1), or involvement in the pathway of a gene found to have significant change in expression (Tgfb1, Tgfbr1, Tgfbr2).

Total RNA was extracted from frozen tumor and non-tumor fibrotic liver from mice treated with DEN+CCl4, and normal liver of vehicle-control mice using a Qiagen RNeasy kit according to the manufacturer’s instructions (Qiagen, Valencia, CA). Samples from the same Complementary DNA (cDNA) was synthesized from 10 μg total RNA using a High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Carlsbad, CA). Gene expression was then determined by quantitative reverse-transcription PCR (qRT-PCR) using gene expression assays (Applied Biosystems). All genes and primers are listed in Table 1. Each sample was run in duplicate. Reactions were performed in a 96-well assay format using a 7900HT Fast Real-Time PCR System (Applied Biosystems). The mRNA level of the housekeeping gene Gusb1 was evaluated in tandem with the experimental runs. The relative amount of each mRNA transcript was determined using the $2^{-\Delta\Delta Ct}$ method (Schmittgen and Livak 2008).
Table 1. Selected genes and their Applied Biosystems qRT-PCR primer assay numbers.

<table>
<thead>
<tr>
<th>AppliedBiosystems Assay ID</th>
<th>Gene Symbol</th>
<th>Gene Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mm01331626_m1</td>
<td>Akt1</td>
<td>thymoma viral proto-oncogene 1</td>
</tr>
<tr>
<td>Mm00658541_m1</td>
<td>Erbb2</td>
<td>v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog</td>
</tr>
<tr>
<td>Mm00435123_m1</td>
<td>Gadd54b</td>
<td>growth arrest and DNA-damage-inducible 45 beta</td>
</tr>
<tr>
<td>Mm01197698_m1</td>
<td>Gusb</td>
<td>glucuronidase, beta</td>
</tr>
<tr>
<td>Mm00492767_s1</td>
<td>Jrk</td>
<td>Jerky</td>
</tr>
<tr>
<td>Mm00436264_m1</td>
<td>Rara</td>
<td>Retinoic acid receptor, alpha</td>
</tr>
<tr>
<td>Mm00451387_m1</td>
<td>Spats1</td>
<td>spermatogenesis associated, serine-rich 1</td>
</tr>
<tr>
<td>Mm00441724_m1</td>
<td>Tgfb1</td>
<td>Transforming growth factor beta 1</td>
</tr>
<tr>
<td>Mm00436971_m1</td>
<td>Tgfbr1</td>
<td>Transforming growth factor beta receptor 1</td>
</tr>
<tr>
<td>Mm00436978_m1</td>
<td>Tgfbr2</td>
<td>Transforming growth factor beta receptor 2</td>
</tr>
<tr>
<td>Mm00443258_m1</td>
<td>Tnf</td>
<td>Tumor necrosis factor</td>
</tr>
<tr>
<td>Mm01281449_m1</td>
<td>Vegfa</td>
<td>vascular endothelial growth factor A</td>
</tr>
<tr>
<td>Mm01300555_g1</td>
<td>Wnt1</td>
<td>wingless-related MMTV integration site 1</td>
</tr>
<tr>
<td>Mm00442104_m1</td>
<td>Wnt10b</td>
<td>wingless-related MMTV integration site 10b</td>
</tr>
<tr>
<td>Mm00494229_m1</td>
<td>Xrcc1</td>
<td>X-ray repair complementing defective repair in Chinese hamster cells 1</td>
</tr>
</tbody>
</table>

After computing the relative expression of each gene compared to the housekeeping gene, statistical analysis was conducted using GraphPad Prism 5 (San Diego, California). Differences between the control liver tissue and tumors from the DEN+CCl₄ were evaluated by an unpaired Student’s t-test, with Welch’s correction employed because equal variance was not
assumed between the two groups. Analysis to compare the mRNA level of the tumors, matched non-tumor, and tumor tissues was conducted using one-way ANOVA. Results are presented as mean ± SD. P-values < 0.05 were considered significant.
Chapter 4

Results

Array Comparative Genomic Hybridization of Mouse Liver Tumors and Non-tumor Tissue

A total of 18 tumors and matched non-tumor surrounding fibrotic liver tissues were subjected to aCGH. All 36 samples were successfully hybridized, were found to have high quality data after the scanning of each array, and were, thus, included in the CNA analysis. Each individual sample array file was uploaded to the UMD, where lowess normalization was conducted. Additionally, gene list annotation using the appropriate Agilent design file for the array was conducted after uploading to the UMD. After these procedures, normalized and annotated array data was available for download and use in additional analyses. A heatmap was created using the intensity data to visually confirm that there were no outliers or samples with unsuccessful fluorescence among the set.

Identification of Segments with CNAs

SWITCHdna analysis demonstrated that CNAs were more frequent in tumors than in the matched fibrotic liver tissue. Interestingly, copy number gains were predominant in the tumors, while both gains and losses were observed (albeit at very low frequencies) in the non-tumor tissues. A similar pattern has been observed in other tumor types, with the supposition that copy number losses are more deleterious to cells (Huang et al. 2011). Furthermore, several segments of gain were in the same location for both the non-tumor fibrotic samples and the tumor samples, indicating that these particular CNAs may occur at early stages of tumorigenesis.
Figure 5. SWITCHdna plots of mouse tumor and matched non-tumor fibrotic liver samples. The above plots show frequency of regions with CNAs among the sample set of mice 18 mice, plotted in genomic order. CNAs were more frequent in tumors (top panel) than in samples from matched fibrotic liver tissue.

**Identification of CNAs Specific to Tumors and Fibrotic Tissues**

SWITCHdnaplus was employed to identify regions that are specific to the tumors or the non-tumor fibrotic liver tissue (by two-tailed t-test). Several segments of copy number gain that were present at a frequency of 15% or greater, and were also significantly different between the two types of tissue (tumor vs. non-tumor fibrotic liver) in the tumor DNA, were identified. In contrast, only one segment that was significantly different in copy number value was also present at a frequency of about 15% in the non-tumor tissue. This further confirms the results of the SWITCHdna analysis, which was demonstrative a higher frequency, as well as diversity, of
CNAs in tumor tissue relative to the surrounding non-tumorous, fibrotic liver. This indicates that an increased frequency, as well as diversity, of CNAs is present in tumor tissues relative to the fibrotic, non-neoplastic tissue.

Figure 6. SWITCHdnaplus plots of mouse tumor compared to matched non-tumor fibrotic liver samples. The above plots show the regions that are specific to the type of DNA tested; tumor vs. non-tumor. SWITCHdnaplus adds the feature of highlighting regions that are significantly different between the two types of tissue, AND appear at a frequency of at least 15%.

These results indicate that several regions of gain are significant to the tumor tissues, while only one small region of gain was significant to the non-tumor tissue, and no losses were specific to either group. This demonstrates that an increased frequency, as well as diversity, of CNAs are present in tumor tissues relative to the fibrotic, non-neoplastic tissue.

**CNAs in a Set of Human Patients with HCC and Cirrhosis or Fibrosis**
As total of 30 samples with segmented copy number data from HCC patients with cirrhosis, and 19 with fibrosis, were included in the study. As shown in Fig. 7, the SWITCHdna plots show characteristic gains and losses of HCC (loss of 1p, 8p, 17p; gain of 1q, 6p, 8q, 20q). Furthermore, a roughly similar CNA profile was observed in tumors that arose in either a fibrotic or cirrhotic liver, indicating that chromosomal instability is likely an early event in hepatocellular carcinogenesis.

Figure 7. SWITCHdna plots of human liver tumors in cirrhotic and fibrotic livers. The above plots show frequency of regions with CNAs among human HCC samples from using TCGA data. The top plot shows CNAs in HCC patients with cirrhosis. The bottom plot shows CNAs in HCC patients with fibrosis.

Matched, non-tumor liver samples were available for a small set of patient samples through TCGA. Similar to the approach taken with mouse tissues, segmented data was analyzed for these matched, diseased liver tissues (outside the tumor margin). Interestingly,
very few small segments were identified with CNAs, and these segments were found at a frequency of less than 15% in the sample set.

![SWITCHdna plot of non-tumor cirrhotic liver tissue in patients with HCC.](image)

Figure 8. SWITCHdna plot of non-tumor cirrhotic liver tissue in patients with HCC. A lack of CNAs (very few existed, and those that were present were very small segments at a frequency lower than 15%) was observed in the matched, non-tumor tissue of HCC patients with a diagnosis of cirrhosis.

Next, to identify potential segments with CNAs specific either to fibrosis- or cirrhosis-associated HCC in human patients, copy number data was analyzed using SWITCHdnaplus in the same manner as in the mouse comparative analysis between tumor and non-tumor DNA. The two CNA profiles were largely similar, although several segments specific to each subset of samples did appear (Fig. 9). However, nearly all of the segments that were identified as specific to either fibrosis or cirrhosis-associated HCC were in proximal genomic regions. Because there are only two groups being compared by a t-test, any regions segmented by CBS and present at a frequency above 15% that significantly differed in copy number value between the two groups is highlighted. The fact that of the regions that are not highlighted in both patient groups are spatially similar indicates that altered segments differed slightly in size but that similar regions were particularly susceptible to copy number alteration.
Figure 9. SWITCHdnaplus plots of human liver tumor samples from cirrhotic and fibrotic livers. The above plots show the “subtype specific” frequency of segments with CNAs, after comparing tumors from patients with fibrosis to tumors from patients with cirrhosis. Most of the regions appear to be in relative proximity, suggesting that these regions may be particularly susceptible to CNAs throughout the progression of disease.

Finally, the results of the analyses on CNAs in both mouse and human samples were compared. The segments containing CNAs in mouse tumors and human tumors in patients with cirrhosis were compared. Many of the regions of gain or loss is human HCC were also observed in our mouse model of HCC in a background of severe liver disease. Of the segments that were gained or lost at a frequency of 15% or greater in the human samples, 51% of the segments in human tumor samples from patients with HCC and cirrhosis were also observed in the tumors from mice with liver fibrosis. A little over a third (33.2%) of all genes in segments in the mouse fibrosis-associated tumors were also present in the human cirrhotic tumors.
Figure 10. Comparative analysis of human and mouse CNAs with homologous genes: mouse segments compared to all gains/losses in tumors from human cirrhosis patients. Segments with CNAs at the same locations between groups (human cirrhosis against mouse tumor/non-tumor) are highlighted.

**Gene Expression Analysis**

To evaluate the relationship between gene dose and gene expression, qRTPCR was conducted on a set of genes that were found to be altered in copy in both human and mouse liver tumors in the experimental (mouse) and downloaded (human) data used in the present study. Consistent with a previous study (Pollack et al. 1999), which found that 62% of genes with altered dose had a positive correlation with gene expression, the expression of some genes was altered while that of others was unchanged or was altered in the opposite “direction” of the CNA. Of the set of genes evaluated in the present study, all located within regions of copy number gain in the mouse DEN+CCl4 tumor samples as well as in regions of copy number gain in the cirrhosis + HCC human data, 9/13 genes (69%) had an increased level of mRNA as compared to liver tissue from vehicle-control mice (Fig. 11). Four genes that were present in gained segments were down-regulated in the tumor tissues relative to the controls, evidenced by a lower mRNA level relative to that of vehicle-control animals (Fig. 12). An n of 4 vehicle control samples were used in the qRTPCR experiments. DNA from 6 animals was used for the
reference sample pool in the aCGH experiment; one sample was not included in the qRTPCR experiments because of a lack in material, and one was excluded from the analysis after it consistently showed Cp values below (i.e., expression level above) one standard deviation from the mean of the vehicle control sample group for multiple genes.

Figure 11. Expression of genes located in segments of copy number gain in fibrosis-associated HCC in mice. mRNA level is shown as fold change relative to vehicle-control mice, as evaluated by qRTPCR. Results are presented as mean ± SD, n=4 for control group, n=18 for tumor samples. * q-value < 0.10.
Figure 12. Expression of genes that were located in segments of copy number gain do not show a direct positive relationship with expression. mRNA level is shown as fold change relative to vehicle-control mice, as evaluated by qRT-PCR. Results are presented as mean ± SD, n=4 for control group, n=18 for tumor samples. * q-value < 0.10.

It is worth noting that the expression level of Wnt1 and Wnt10b was also increased in the tumors. However, because the expression level of these genes was too low to detect in the vehicle-control samples, relative analysis was not possible. This conclusion that the expression of these genes is increased is based on the fact that the mRNA was detectable in the tumor tissues versus undetectably low in the controls (data not shown).

To further investigate if the gains in expression in the genes found in segments of copy number gain was indeed due to an increase in gene “dose,” 3 of these genes were also evaluated in non-tumor fibrotic tissue. Tnf, and Map3k6 are both found in segments that were
increased in both non-tumor and tumor tissues in DEN+CCl₄-treated mice, although the frequency of the gained segments among the samples was higher in tumors. *Tgfbr2* is a receptor of *Tgfb1* and has been shown to have a highly correlated expression profile to that of *Tgfb1*, a gene that with copy number gain in fibrosis-associated tumors. The expression of these genes was not significantly up-regulated in non-tumor tissues, while the increase in level of all 3 genes was significantly higher in the tumors in comparison to the non-tumor tissues, as well as liver from the vehicle-control mice.

Figure 13. Expression of genes located in regions gained at a low frequency in non-tumor fibrotic liver tissue and at a high frequency in tumors. mRNA level is shown as fold change relative to vehicle-control mice, as evaluated by qRT-PCR. Results are presented as mean ± SD, n=4 for control group, n=18 for tumor samples. * P-value < 0.05.

A list of the genes evaluated for expression change, based on location in segments of CNA, is presented below detailing fold change and significance.
Table 2.

Expression of genes in mouse liver tumors as evaluated by qRT-PCR.*

<table>
<thead>
<tr>
<th>Gene</th>
<th>Fold change relative to control</th>
<th>Significant (s)/ Non-significant (ns) change</th>
<th>q-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akt1</td>
<td>1.15±0.73</td>
<td>ns</td>
<td>0.5385</td>
</tr>
<tr>
<td>Erbb2</td>
<td>1.21±0.54</td>
<td>ns</td>
<td>0.3977</td>
</tr>
<tr>
<td>Gadd54b</td>
<td>1.37±0.68</td>
<td>ns</td>
<td>0.3977</td>
</tr>
<tr>
<td>Jrk</td>
<td>0.44±0.29</td>
<td>ns</td>
<td>0.1334</td>
</tr>
<tr>
<td>Map3k6</td>
<td>5.60±2.79</td>
<td>s</td>
<td>0.0004</td>
</tr>
<tr>
<td>Rara</td>
<td>0.60±0.28</td>
<td>s</td>
<td>0.0226</td>
</tr>
<tr>
<td>Slc27a1</td>
<td>1.36±0.88</td>
<td>ns</td>
<td>0.2054</td>
</tr>
<tr>
<td>Tgfb1</td>
<td>3.16±1.56</td>
<td>s</td>
<td>0.0004</td>
</tr>
<tr>
<td>Tgfbr1</td>
<td>1.91±1.27</td>
<td>s</td>
<td>0.0208</td>
</tr>
<tr>
<td>Tgfbr2</td>
<td>6.52±3.82</td>
<td>s</td>
<td>0.0043</td>
</tr>
<tr>
<td>Tnf</td>
<td>3.02±1.64</td>
<td>s</td>
<td>0.001</td>
</tr>
<tr>
<td>Vegfa</td>
<td>0.67±0.25</td>
<td>s</td>
<td>0.0787</td>
</tr>
<tr>
<td>Xrcc1</td>
<td>0.50±0.25</td>
<td>ns</td>
<td>0.1263</td>
</tr>
</tbody>
</table>

* represented as mean fold change ± standard deviation. These genes were located in segments of copy number gain in both fibrosis- and cirrhosis-associated HCC in mouse and human samples, respectively. A threshold of 10% was set as statistically significant after FDR correction.
Chapter 5

Discussion, Limitations, Conclusions, and Future Directions

Hepatocellular carcinogenesis is a complex process that is the consequence of multiple molecular events that lead to the initiation, promotion, and progression of tumor cells (Aravalli et al. 2013; Zucman-Rossi 2010; Forner et al. 2012). Accumulated evidence has emphasized the importance of a distinct set of events that are required for carcinogenesis (Hanahan and Weinberg 2011; Laurent-Puig et al. 2001). Specifically, distinct cellular capabilities that enable tumorigenesis, or “hallmarks of cancer,” are increasingly recognized as essential processes in carcinogenesis (Hanahan and Weinberg 2011). To gain a comprehensive perspective of the mechanisms involved in specific cancer types, as well as various etiologies within a type of cancer, investigation of molecular underlying factors of cancer should consider genetic aberrations, epigenetic alterations, gene expression, and copy number alterations.

While the use of human clinical samples and epidemiological data is desirable, such resources are not always available. Therefore, animal models of disease are necessary. This is especially true for the analysis of chemical carcinogens and other toxic substances, as reliable human data on exposure, dose, and response is extremely limited. Mouse liver has only been shown to exhibit cirrhosis and HCC after administration of chemicals at high necrogenic doses (Becker 1983). Thus, the model used in the analyses presented here, in which advanced fibrosis was induced by chronic injection of CCl₄, offers an experimentally robust and relatively accurate representation of common human HCC etiology.

An interesting finding of this study in the context of previous work conducted with these mouse samples was the observed association between CIN and epigenetic alterations that are indicative of genomic instability. Analysis of CNAs confirmed the existence of genomic instability in the liver tumors of DEN+CCl₄-treated mice, as suggested by previous findings in
the same tissues (Chappell et al. 2014). Indeed, we found a higher frequency of CNAs in the tumor DNA as compared to the vehicle control by analysis with SWITCHdna (Fig. 5). The results from that study, together with that of the present work, suggest a link between epigenetic alterations in the form global DNA hypomethylation and loss of methylation at histones that are involved in chromatin structure and stability. This is in concordance with a recent study that evaluated CIN and global DNA hypomethylation in human samples of HCC (Nishida et al. 2013). That study found a significant correlation between global DNA hypomethylation and loss of 6p, 8p, 13q, and 17p. Although the mechanistic link between these two phenotypes is not clear, one potential explanation that has been proposed suggests that activation of hypomethylated repetitive DNA elements are within genes may cause chromosomal rearrangements (Gaudet et al. 2003). This indicates the potential association between DNA hypomethylation and CIN during clonal expansion and tumorigenesis. This theory is further supported by the fact that both CIN and epigenetic alterations are generally considered early events in carcinogenesis.

CNAs were evaluated in both tumor tissue and surrounding, non-tumor liver tissue in mice with fibrosis, as well as in HCC patients with fibrosis or cirrhosis. Interestingly, CNAs were very infrequent in the non-tumor surrounding liver and were present in genomically small segments. This observation was made in both mouse and human samples (Figs. 6 and 8, respectively), suggesting that CNAs are not a prominent feature of damaged, non-cancerous liver tissue. Because cirrhosis (and fibrosis in mice) is considered to be involved in the promotion of carcinogenesis (Zhang and Friedman 2012), the profound lack of CNAs in this potentially pre-neoplastic tissue was somewhat surprising. Although genomic instability has been indicated as an essential and early event in tumorigenesis, our findings indicate that this instability is not occurring in pre-neoplastic tissue.
Previous work using these mouse samples demonstrated a lack of mutations commonly observed in HCC in both humans and mice (Chappell et al. 2014). This finding, together with those of the present study, suggest that CNAs occur relatively early in the carcinogenic process as compared to common activating or inactivating mutations. This is consistent with other reports of CIN as an early event in tumorigenesis (Nowak et al. 2002).

An important component of the present study is the relevance of the mouse model to human data. Many common segments with altered copy number, as well as genes within those segments, were observed between the mouse model and the set of human cirrhosis-associated HCC samples (Fig. 10). Animal models, most often rat or mouse, are commonly used to understand the molecular pathogenesis of HCC (Heindryckx et al. 2009), as well as to test chemicals and drugs for potential cancer hazard (Wells and Williams 2009). Although the liver is the most common tissue for tumor development in experimental rodent studies of chronic exposure to xenobiotics (Hoenerhoff et al. 2011), most chronic rodent cancer studies fail to induce liver fibrosis or cirrhosis. The lack of fibrosis and cirrhosis in most positive National Toxicology Program (NTP) 2-year cancer bioassay studies in rodents is in stark contrast to human HCC, in which liver cirrhosis is the most common histopathological feature observed in subjects with HCC, and is an important mechanism of hepatocarcinogenesis (Farazi and DePinho 2006). The present analysis of CNA profiles shows a particular type of molecular similarity between human HCC samples and this particular mouse model, which used B6C3F1 mice, the most commonly used mouse strain in NTP 2-year cancer bioassays. Additional studies are necessary to further molecularly characterize this mouse model in relation to specific features of human HCC.

Gene expression analysis revealed that a majority of the expression of a representative set of genes correlated with CNAs; genes in gained segments had an increased level of mRNA relative to that in liver of vehicle-control mice (Fig. 11). A positive correlation between copy
Number and transcription has been reported in various cancer types, analyzed by various experimental and statistical methods, with the most detailed association studies having been conducted in colorectal cancer (Ried et al. 2012). Altered gene expression has been shown to occur in approximately half of the genes with copy number alterations in human cancer studies, both in clinical samples (Pollack et al. 2002; Qu et al. 2010), as well as in in vitro studies with immortalized human cell lines (Phillips et al. 2001). These results, together with those of the present study, indicate that CNAs can and do alter gene expression, but that this particular mechanism of altered gene expression is tumor- and tissue-type specific. Statistical or experimental methods can be applied to better understand the causal relationship between CNAs and gene expression changes. Analyses of other potential mechanisms of gene expression changes, including mutation and epigenetic silencing, of both genes of interest and their upstream effectors will enable a more comprehensive evaluation of any particular gene that is present in an altered copy number in cancerous tissue. Pollack, et al (Pollack et al. 2002) found that 62% of genes with altered copy number had a positive correlation with gene expression, and then conducted an estimation of the fraction of gene expression variation that was specifically attributable to CNAs. This secondary analysis showed that only ~12% of variation in mRNA levels was directly attributed to CNAs, although they state that this estimation “represents a significant underestimate” due to the conservative nature of the analysis, and global variation in tumor cells as well as the presence of non-tumor cells in the samples.

It is essential to consider pathways and gene-gene interactions when evaluating the role of genes in cancer, and when making assumptions based on copy number. For example, Vegfa is a gene commonly up-regulated in cancer and was present in a frequently gained segment in both mouse and human HCC in our sample sets. However, Vegfa gene expression was actually significantly reduced in the mouse tumor samples, compared to vehicle-controls (Fig. 12). Interesting, the gene Tgfb1, which has been shown to inhibit Vegf, was also located in a
segment of copy number gain in the mouse tumors. Therefore, the expression of Tgfb1, as well as two of its receptors, Tgfbr1 and Tgfbr2, was evaluated. Tgb1, Tgfbr1 and Tgfbr2 were all significantly up-regulated, which likely explains the down-regulation of Vegfa in the mouse tumor samples. Additionally, conflicting results have been reported in regard to the role of Tgfb1 in HCC; it has been shown to play a tumor-suppressing role in early stages of carcinogenesis and to enhance tumor growth and malignancy at later stages (Reichl et al. 2012). There are likely other examples of such association between genes in various pathways, and, thus, the genes harbored in gained and lost segments should be carefully evaluated in the formulation of conclusions about their individual involvement in hepatocellular carcinogenesis.

**Limitations**

While large sample sizes are typically ideal in any study that aims to uncover recurrent molecular characteristic of disease, this is not always attainable. A mouse sample size of 18 and a human sample size of 49 (30 patients with cirrhosis, 19 with fibrosis) was available for this study. A larger sample size may have given us more power in our analysis.

The sampling method of tumors from the mouse livers used in the analysis presents another potential limitation. Genomic DNA from one tumor from each mouse was used in our study. HCC is a heterogeneous cancer that presents with diverse molecular profiles across patients (or animals) and even within an individual (or animal). Tumor heterogeneity is thought to occur from the same selective processes that lead to common CNAs (Hu et al. 2013); as cells with genomic imbalances proliferate and accumulate, they acquire new mutations and CNAs (Vogelstein et al. 2013), and clonal expansion may occur at slightly different points in “time” in the evolution of that clonal cell line or selection for different combinations of transcriptional elements of cellular processes may occur. There are 4 main types of genetic
heterogeneity in cancer, as detailed in a recent review by Vogelstein et al (Vogelstein et al. 2013): intratumoral, existing among the cells within one tumor; intermetastatic, heterogeneity in one patient among different metastatic lesions; inrametastatic, variation among cells within a metastasis; and interpatient, heterogeneity among tumors in different patients. Most applicable to our study would be intratumoral, because it is possible that we did not obtain a comprehensive view of the CNA profile within one liver (one mouse) by only sampling DNA from one tumor. The liver tumors in our study were primary, and interpatient (or in our case, “interanimal”) heterogeneity was of less concern because the goal was to find the most common CNAs among a set of tumors that were assumed to have some level of heterogeneity.

The same concerns and arguments are applicable to the human samples used in this study; single samples were taken from primary HCCs from each patient. While interpatient heterogeneity is extremely important and has contributed to many of the efforts towards the individualized medicine movement, without a clear understanding of the role of CNAs in HCC, studies should first aim to uncover common CNAs among groups of patients and within types and subtypes of HCC and second apply these findings to individual cases.

The mouse tumors were isolated and separated into separate vials at sacrifice, which likely confer some bias for the largest tumors. However, because the analysis was aimed at identifying the CNAs of highest frequency across tumors in animals with the same tumor etiology, this per-animal sampling method was considered acceptable. Furthermore, as reported in the original study that constructed this mouse model, the within-animal (within-liver) variation in the tumors was low, as evaluated by histology by a trained veterinary pathologist (T. Uehara) (Uehara et al. 2013).

Ongoing analyses of CNAs in various cancer types are now employing single-cell analysis techniques to better understand the spectrum or CNAs within a tumor and how this
heterogeneity is relevant to disease phenotype and clinical outcome. A lower level of heterogeneity is expected for our study as compared to a human population, due to the use of inbred mice that shared an identical HCC etiology. This assumption can be addressed by comparison of different mouse strains and chemical challenges, which is one of the future aims mentioned below.

**Conclusions**

Our results demonstrate an accumulation of CNAs liver tumors in a mouse model of fibrosis-associated HCC, designed to emulate the most commonly etiology of human HCC. Previous work conducted using this mouse model demonstrated increased epigenetic alterations indicative of genomic instability in tumor samples relative to control tissues or in liver tissues of mice treated with either DEN or CCl₄ alone. This finding, together with a lack of mutations, suggest that CIN is also a feature of the tumor cells, preceding common mutations even in cases of treatment with a known genotoxic agent (DEN). Indeed, CNAs were more prevalent in tumors as compared to surrounding, fibrotic liver tissue, indicating that structural aberrations are profoundly increased in neoplastic tissues as compared to pre-neoplastic tissue. Similarly, tissue samples from non-tumor, cirrhotic, surrounding liver in human HCC patients had very small and infrequent CNAs, while the tumors had profound alterations in copy number of many segments, harboring thousands of genes.

Approximately half of the significantly gained or lost segments observed in human cirrhosis-associated liver tumor, and approximately one quarter of the genes, were also observed in the samples of fibrosis-associated liver tumors in mice. This demonstrates the similarity of this mouse model in genomic features to that of tumors in human patients with common etiology of HCC, and, thus, the utility of the model. Additionally, a set of genes prone
to CNAs involved in the pathogenesis HCC across species was generated. Several genes found in gained segments were found to have increased expression. However, careful attention to gene networks and pathways is essential, as discrepant results regarding copy number and expression are almost as common as correlated results.

**Future Work**

Additional work related to this project will potentially strengthen the conclusions and offer additional understanding of the role of CNAs in HCC. Conducting a similar analysis using aCGH and SWITCHdna techniques and liver samples from various mouse models of HCC will allow us to better characterize the CNAs observed in HCC. Using additional mouse models of fibrosis-associated HCC and comparative analysis to human data, it is possible to establish mouse models of specific types of human HCC, which is important for future research on mechanisms of carcinogenesis, prognostic factors, and development of targeted therapy strategies. We hope to collaborate with the NTP to obtain mouse tissues from previous studies of exposures to various types of carcinogenic agents. The NTP banks tissues from animal exposures for future use as additional molecular techniques and pertinent endpoints are developed. Further, comparative analysis of CNAs in mouse tumors that arise in healthy livers to those that arose in fibrotic liver may identify CNAs specific to the different types of pathogenesis. This comparative study could be expanded to TCGA data from human HCC patients without fibrosis or cirrhosis, potentially identifying molecular events or markers specific to cirrhosis-associated HCC.

Gene expression analysis is also very important in understanding how CNAs are related to disease. While we presented important evidence that CIN exists in tumors compared to healthy liver tissue, as well as to unhealthy surrounding liver tissue in HCC patients, specific
genetic alterations in addition to these genomic changes are important in understanding the role of CNAs in tumorigenesis. Although we have shown some gene expression data for our mouse model, we did not evaluate human expression data in this study. Expression data for the human samples we used is available through TCGA and can be downloaded and analyzed.

Finally, similar to the above-mentioned future direction of expanding our analysis to additional mouse models and/or exposures, a comparison between tumor samples from human HCC patients with various etiologies could be conducted. For example, HCC patients with or without viral hepatitis, alcoholic steatosis, exposure to particular hepatocarcinogen, and tumors that arose in a healthy liver. A recent study with a similar aim found that HCC in patients with alcoholic liver cirrhosis had a lower frequency of CNAs as compared to tumors in patients with several other etiological factors (Homayounfar et al. 2013). Another study characterized recurrent CNAs and associated them with HBV and Aflatoxin B1 exposure status in a small set of human patients (Qi et al. 2013). Both intra- (among various human etiologies) and inter- (comparative to mouse models) species analyses could be conducted using a data from a broader variety of human pathological specimens.
Chapter 6

Practicum Study

I completed my public health practicum at the National Center for Toxicological Research (NCTR), which is a division of the FDA, in Jefferson, Arkansas. During this time, I learned several molecular techniques for the analysis of both genetic and epigenetic alterations in tissues under the direction of Dr. Igor Pogribny and the assistance of Dr. Volodymyr Tryndyak. I conducted a mutation assay, gene expression analysis, and a gene methylation-specific assay for the gene h-ras, using tissues from the same mouse model as presented in the present study. h-ras is one of the most commonly mutated genes in mouse HCC, and the ras family of genes is involved in HCC across species. Identifying the involvement of this gene (or lack thereof) further characterized this mouse model of HCC. We found that h-ras was not mutated, nor was the expression level of the gene altered in the tumor or non-tumor tissues in the livers of mice with fibrosis-associated HCC. This was an exciting finding because it indicated that mechanisms other than common gene mutations are essential in the development of HCC, indicating that other causes of genomic instability occur earlier than, and possibly independent of, common mutations. This finding played a role in the work presented in the chapters above in that additional molecular characterizations were warranted to better understand the molecular underpinnings of HCC in fibrotic liver.

I learned valuable molecular techniques, procedural planning, and trouble-shooting methods. These techniques and lessons can be broadly applied to public health-focused initiatives because the experiments conducted analyzed samples from a mouse model of human disease with direct toxicological application; liver tumors are the most common cancer found in 2-year cancer bioassays conducted by the NTP, and the liver is a tissue that is highly susceptible to carcinogenesis in humans as the result of toxic insult. Furthermore, while visiting
NCTR I had the opportunity to meet several staff members and scientists, as well as the head of the head of the Division of Biochemical Toxicology (Dr. Frederick Beland). From these meetings, conversations, and observations, I was able to get an idea of the types of research projects are conducted at NCTR, and how they pertain to toxicology and public health.

Before beginning my practicum study, I selected the following competencies for specific emphasis during the experience, as quoted in italicized text:

- **Communication and informatics**
  - “*Engage in collective information sharing, discussion and problem solving.*” This goal was achieved by sharing of information and plans with the other members of the laboratory. The opportunity to visit a lab group and observe their techniques, procedures, and dynamic is an invaluable experience for a scientist in training. When unusual results or unacceptable quality control metrics were observed, myself and one or two other lab members was assessed the situation and considered the options for improvement. While in the lab, some of my experiments required repeating with different conditions to optimize the analysis. For example, time and quantity of a restriction enzyme in a PCR experiment to select for mutant DNA sequences had to be altered to obtain quality results. Additional clean-up steps for the PCR products were also required to obtain more reliable results. I gained experience in the process of trouble-shooting and problem solving in my own laboratory experiments while at NCTR.

- **Leadership**
  - “*Exercise productive organizational, time-management and administrative skills.*” While at NCTR, I was working under supervision and guidance. However, I was responsible for managing my own time and ensuring that the goals of my project were finished in the time while I was there. Sometimes this mean an earlier
arrival or later departure than my supervisor. Organization was very important while at the practicum site, as well as in the weeks prior to and following the practicum study. Before departing for NCTR, it was required to plan all experiments and prepare samples for shipment to the lab site. Additionally, reagents and supplied needed to be purchased prior to my arrival at the practicum. Taking care of these organizational and administrative tasks was a lesson for me in project management. A large amount of data was generated while at NCTR, as well as additional experimental products. The data needed to be carefully labeled, stored, and cataloged for future use. Additional samples were also produced during the study (clean-up products, PCR products, fragmented DNA, cDNA), which required careful labelling and storage so that they could be returned to UNC and be utilized in future experiments.

- Professionalism and Ethics
  - “Apply evidence-based concepts in public health decision-making.” My project at NCTR addressed mechanistic information of a model of human cancer. Mouse models are used to assess risk of cancer-causing agents, and an understanding of the relevance of the model to human disease, the causative effects of exposure on the model, and the molecular underpinnings of the pathological findings are all imperative for a better understanding of chemical carcinogenesis. The analysis of tissues taken from a mouse model of fibrosis-associated HCC enabled the construction of a sequence of events in liver tumorigenesis, the identification of putative events in the pathway to cancer, as well as additional hypotheses for future studies.

- Environmental Sciences Competencies
  - “Describe the direct and indirect human, ecological and safety effects of major environmental and occupational agents.” The main goal of NCTR is to assess
hazard and risk of human exposure to harmful agents. By working in the lab at NCTR, I was able to take part in and observe efforts to evaluate the effects of exposure to environmental agents on laboratory animals, which is a method to indirectly assess the effects of such chemicals on humans.

• “Describe federal and state regulatory programs, guidelines and authorities that control environmental health issues.” NCTR is a division of the U.S. Food and Drug Administration, which is the federal regulatory agency that oversees testing of the effects of potentially and known harmful substances. The efforts of NCTR are aimed at assessing the effects of exposure to toxic substances environmentally and nutritionally. These efforts are laboratory-based, using in vivo, in vitro, and in silico methods to evaluate the effects of exposure and their relation to disease. Many regulatory decisions have been and continue to be made based on experimental data generated at NCTR.

• “Explain the general mechanisms of toxicity in eliciting a toxic response to various environmental exposures.” My study specifically focused on a multi-hit model of cancer, specifically HCC. From the study it was evident that two “hits” profoundly increased tumorigenesis in the mice. However, mutations commonly found in HCC were not observed in the study. Epigenetic alterations were significantly altered in the tumors, indicating that epigenetic effects of a chemical may be just as, if not more, important that mutations.

• “Develop a testable model of environmental insult.” Using in vivo studies with inbred mice allows for testable, controlled, and repeatable experiments. Using animal models enables controlled conditions and known exposure (duration, dose) to particular agents. Data from exposure assessments in animals can aid in better understanding the molecular mechanisms involved in toxicity, which can be applied to humans.
References


