TRANSCRIPTIONAL ALTERATIONS DURING MAMMARY TUMOR PROGRESSION IN MICE AND HUMANS

By

Karen Fancher

B.S. Hartwick College, 1995

A THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

(Interdisciplinary in Functional Genomics *and* in Biomedical Sciences)

The Graduate School

The University of Maine

December, 2008

Advisory Committee:

Barbara B. Knowles Ph.D., Professor, The Jackson Laboratory, Co-Advisor

Gary A. Churchill Ph.D., Professor, The Jackson Laboratory, Co-Advisor

Keith W. Hutchison Ph.D., Professor of Biochemistry and Molecular Biology, The University of Maine, Thesis Committee Chair

M. Kate Beard-Tisdale Ph.D., Professor of Spatial Information Science and Engineering, The University of Maine

Shaoguang Li M.D., Ph.D., Associate Professor, The Jackson Laboratory

LIBRARY RIGHTS STATEMENT

In presenting this thesis in partial fulfillment of the requirements for an advanced degree at The University of Maine, I agree that the Library shall make it freely available for inspection. I further agree that permission for "fair use" copying of this thesis for scholarly purposes may be granted by the Librarian. It is understood that any copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Signature:

Date:

TRANSCRIPTIONAL ALTERATIONS DURING MAMMARY TUMOR PROGRESSION IN MICE AND HUMANS

By Karen Fancher

Thesis Co-Advisors: Dr. Barbara B. Knowles and Dr. Gary A. Churchill

An Abstract of the Thesis Presented
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
(Interdisciplinary in Functional Genomics *and* in Biomedical Sciences)
December, 2008

Family history, reproductive factors, hormonal exposures, and subjective immunohistochemical evaluations of *in situ* lesions, and to a lesser extent age, remain the best clinical predictors of an individual's risk of developing breast cancer. Identification of early markers predictive of impending invasive breast cancer from *in situ* carcinoma is a long-term goal. The latent mammary cancer transgenic mouse model of human breast cancer, C57BL/6J-Tg(WapTAg)1Knw (Waptag1), develops characteristic stages of tumorigenesis in a highly predictable manner: atypical hyperplasia advances to ductal carcinoma *in situ* (DCIS), which progresses to papillary adenocarcinomas and/or solid, invasive tumors. Microarray analyses of whole mammary glands and tumors across these stages, to detect transcriptional changes throughout tumorigenesis, revealed marked, phased stage-specific changes. In contrast, results from the laser capture microdissected tumor cells depict a moderately constant characteristic tumorigenic profile, irrespective of stage. Evaluation of differences in whole glands with those of microdissected samples suggests that paracrine signaling between tumor and stromal cells substantially alters the tumor microenvironment, early in progression. Strikingly, comparison of statistically significant microarray results between Waptag1 DCIS and human DCIS revealed 2,097 overlapping early transcriptional changes. When compared with species-specific controls, common abundant early gene alterations were associated with cell cycle, cell division, and DNA replication Gene Ontology categories, with a notable decrease in genes involved in aerobic energy metabolism, and significant increased transcription of retrotransposons and chromosome

modification genes. Based on these initial experimental results, retrotransposons were identified as a potential marker for testing in several mouse models and in biopsies derived from breast cancer patients. Analysis of data from five independent mouse models of mammary cancer and five human breast cancer datasets revealed over expression of retrotransposons, mainly Class I and Class II LTR elements, as well as LINEs and SINEs, when compared with normal samples. Cross-species comparison of gene expression profiles suggests epigenetic alterations and chromatin remodeling changes coincide with retrotransposon over expression. Through validation of such mutual human and mouse changes in gene expression, these novel putative markers may allow earlier detection and therapeutic intervention, possibly reducing the incidence of invasion and metastases in patients with breast cancer.

ACKNOWLEDGEMENTS

The research presented in this thesis was funded by CA34196, CA88327, and CA89713 from the National Institutes of Health; National Science Foundation IGERT Program Fellowship 0221625 (KSF, BBK-PI); Maine Cancer Foundation (BBK-PI); The University of Maine Graduate School of Biomedical Sciences and The Judith A. Lese Breast Cancer Foundation, Inc.

Research presented here may be part of ongoing collaborations. Karen Fancher is indebted to Drs. D. Craig Allred and Sangjun Lee for generously sharing the human DCIS raw microarray data. She would like to acknowledge the exceptional opportunities she was offered by her mentors, Drs. Barbara B. Knowles and Gary A. Churchill, during her graduate studies. She is appreciative of the efforts of her graduate committee members as well. She is abundantly grateful to Drs. Igor Mikaelian, Carol Bult, Barbara Tennent, Alexei Evsikov, and Carrie Evsikova as well as summer student, Hannah Tetrault; their support helped her to accomplish this goal.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1. Tumor Progression

Within a conducive environment, malignant cells arise as a consequence of multiple germline and somatic mutations over time. As cells accumulate mutations, they clonally expand[1] until a significant population of neoplastic cells results in a detectable lesion[7-10]. Progression from the primary lesion to invasive and/or metastatic cancer requires additional molecular changes, including loss of cell-cell adhesion, access to the blood stream, ability to embolize, and a microenvironment favoring growth at the distant site. Several classical 'Hallmarks of Cancer' are considered the minimum requirements for progression: growth signaling independence; resistance to anti-growth signals; escape from apoptosis; continuous angiogenesis; uncontrolled replication/division; tissue invasion/metastasis[11]. A seventh hallmark of cancer, deregulation of metabolic pathways, has also been proposed as a necessity for tumorigenesis[12-20,21]. Dysfunction of over a dozen tumor suppressor genes and activation of numerous oncogenes and their related pathways are associated with cancer. Although the initiation, mechanisms and pathways may differ among cells, tissues, and species it is the acquisition of specific combinations of capabilities within tumor cells (or in the surrounding stroma) which characterizes cancer as a life-threatening disease.

1.2. Human Breast Carcinogenesis

Environmental factors, genetic, genomic, and epigenetic alterations are all possible causes of cancer. Identification of early indicators of impending invasive breast cancer from carcinoma *in situ* biopsy material is one of the current challenges facing the research community. Despite technological breakthroughs and numerous research discoveries made over the last 30 years, the best predictors of a woman's risk of developing breast cancer remain family history, reproductive factors, and hormonal exposure, as well as age, to a lesser extent. Poor prognosis for human breast cancer has been based on lymph node assessment and subjective immunohistochemical evaluation of molecular changes in biopsied samples: for example, lack of estrogen receptor expression (ER-); abundant epidermal growth factor receptor, v-erb-b2 erythroblastic leukemia

viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (ERBB2/HER2/NEU) expression (HER2+); lack of tumor suppressor p53 (TP53); and expression of the myelocytomatosis (MYC) oncogene. At the transcript level, the following genes are frequently down-regulated in human cancers, to initiate, promote, and/or enhance tumorigenesis: tumor suppressors, such as TP53 or retinoblastoma (RB1); cyclin dependent kinase, CDKN2A (also known as p19/INK4A); adenomatosis polyposis coli (APC); and phosphatase and tensin homolog (PTEN). However, a 'one gene at a time' approach, based on results from developed tumors, does not sufficiently predict whether an early lesion will become invasive. To identify and validate potentially valuable markers for *early* pre-cancer progression, detection methods have to be affordable and easy to perform. Earlier detection and diagnosis of pre-malignant lesions and carcinoma *in situ*, may prevent the high incidence of progression to invasive cancer and subsequent metastases.

Large scale gene expression profiles have provided valuable genetic markers of carcinoma *in situ*, invasive and non-invasive human breast cancers, and predictors of relapse/recurrence and metastases. However, such studies are challenging in human patients because of difficulty in obtaining samples at early stages of tumorigenesis, insufficient quantities of quality samples, and the heterogeneity of each specimen, especially among genetically diverse ethnic populations with unique environmental surroundings. The majority of human breast cancer studies focus on advanced tumors, with only a few analyses of *in situ* stages of breast carcinogenesis[22-34].

In the last six years, at least two independent laboratories have developed panels for analysis of 14-76 human transcripts predictive of clinical prognosis, based on gene expression analyses, in an attempt to offer more personalized diagnosis and an affordable, less-subjective evaluation of human breast cancer[35-37]. In fact, two of these, OncoTypeDX^{TM}, an RT-PCR platform, and MammaPrint^{R}, a microarray-based platform, are commercially available to clinicians[38-40]. However, markers to predict whether carcinoma *in situ* of the breast will progress to invasive cancer or metastatic disease are lacking. Furthermore, the underlying cause for some women with ductal carcinoma *in situ* to suffer relapse after surgery, while others do not, requires years of investigation and remains largely unknown. For these reasons, identification of early changes

that may be predictive of progression to invasion in mouse mammary tumorigenesis, with follow-up in human carcinogenesis, is the ultimate goal of this thesis research.

1.3. The Experimental Hypothesis

The foundation of this thesis is based on the hypothesis that a reliable early marker of tumorigenesis is epigenetic deregulation of retrotransposons and their robust transcriptional activation within the early stages of tumor progression in mouse and human mammary cancers. To investigate this hypothesis, a comprehensive survey of retrotransposon expression was undertaken using microarray data from a mouse model of mammary cancer, Waptag1. Whole mammary glands and tumors or microdissected epithelial tumor cells from lesions of DCIS and mammary tumors taken from multiparous Waptag1 females were compared with control whole glands or microdissected C57BL/6J normal mammary cells, respectively. Target sequences from all microarray platforms were interrogated for retrotransposons. Additionally, an investigation of retrotransposons in tumors in five mouse models of mammary cancer, as well as in five datasets from human hyperplasia, *in situ* carcinoma and breast cancer, was performed by *in silico* mining of publicly available expression data, to compare with results from Waptag1.

1.4. Mouse Models of Mammary Cancer

Although no one mouse model completely recapitulates all aspects of any subtype of human breast cancer, genetically engineered mice that model important aspects of human cancers are invaluable. Mouse models assist in the elucidation of pathways involved in normal development and disease, discovery of novel markers for eelier detection of cancer and as potential candidates for diagnostic trials, and efficacy testing of new targeted therapeutics. Several laboratories have developed key mouse models to facilitate the study of mammary cancer[41-56]. The bulk of this thesis will detail primarily one mouse model, Waptag1, however, several mouse models were utilized to a lesser extent as well and therefore, they will briefly be discussed below.

FVB/N-Tg(WapMyc)212Bri/J (Myc) hemizygous transgenic mice, containing the myelocytomatosis oncogene (*c-myc*), driven by the whey acidic protein promoter, develop mammary tumors (of several histotypes) at about 6 months of age in mated females[55]. Virgin female Myc mice develop mammary tumors around 250 days, median 8.3 months of age

(Fancher K *et.al.*, future publication), and lung metastases are observed in about 20% of all females[55] (Fancher K *et.al.*, future publication). Among several mammary tumor types observed in Myc mice, the glandular carcinoma with epithelial to mesenchymal transition (EMT) provides a unique mouse model of EMT, a feature observed in human breast cancer as well[57]. Human tumors with EMT show genomic amplification of MYC[58]. Myc transgenic mice provide a system to study late-stage mammary tumor progression in a well characterized mouse model.

FVB.Cg-Tg(Wnt1)1Hev/J, or Wnt1 transgenic mice contain the wingless-related MMTV integration site 1 (*Wnt1* or *int-1*) oncogene driven by the mouse mammary tumor virus (MMTV) promoter. WNT1 is a secreted protein and a ligand for frizzled receptors. Several Wnt signaling pathways have been discovered and are current subjects of active research. *Wnt1* was first discovered as a proto-oncogene when it was the target of MMTV insertion into the genome[59]; the retrovirus MMTV integration enhanced *Wnt1* expression within the mammary gland epithelial cells and subsequently mammary tumors developed[42]. Wnt1 transgenic mice develop early hyperplasia through late tumor progression including metastases[60,61]. Similar to Myc mice, Wnt1 virgin females can develop mammary tumors, but parous females present tumors with a shorter latency[60,62]. Tumors appear as early as 90 or as late as 210 days of age in hemizygous females, with a median age of death at 133 days of age in virgin females[60,62] and (Fancher K *et.al.*, future publication).

Neu transgenic mice [FVB/N-Tg(MMTVneu)202Mul/J], express an unactivated, wild-type rat *Neu* gene in the mammary glands, salivary glands, thymus, spleen, and lungs. *Neu*, also known as *ErbB2* and HER2, is a member of the epidermal growth factor receptor family of type I receptor tyrosine kinases (RTKs). NEU phosphorylation leads to aberrant increased RTK activity, primarily within the mammary glands, which ultimately results in metastatic mammary cancer. The median age of tumor onset is 205 days with 72% of mice showing lung metastases at about 8 months of age[44,45]. Amplification of HER2/ERBB2/NEU occurs in ~25% of human lymph node-positive breast cancer and is correlated with undesirable nuclear grade[63-65].

In *Trp53* mutant mice [FVB.129S2(B6)-*Trp53*^{tm1Tyj}/J], the targeted mutation deletes exons 2-6 of the endogenous mouse *Trp53* gene and replaces it with a neomycin cassette, thereby creating

a null allele[41]. Although *Trp53* null mice develop lymphomas, heterozygous knockout (+/-) females regularly develop mammary tumors at approximately 300 days of age [41]. When *Trp53* (+/-) mice are combined with another transgenic mouse mammary tumor model, double mutant mice exhibit accelerated tumor onset and shortened life span, compared with single transgenics (Fancher K *et.al.*, future publication and[66-72]). TP53 is the most frequently mutated gene in breast cancer and, indeed, in most human cancers[67,73].

PyMT mice [FVB/N-Tg(MMTV-PyVT)634Mul], containing the oncogenic polyoma virus middle T-antigen driven by the MMTV long terminal repeat, also develop early hyperplasia through late tumor progression, including a high percentage of metastatic tumors[43]. Virgins, breeders, and males present with multiple mammary adenocarcinomas as early as 35 days of age[43]. The PyMT transgene enhances phosphatidylinositol 3-kinase pathway signaling, which eventually leads to uncontrolled cell cycling[74].

C57BL/6J-Tg(WapTAg)1Knw (Waptag1) mice, which express the early region of Simian Virus 40 Tumor-antigen (SV40 Tag), under the control of the whey acidic protein (WAP) promoter within the genetically-defined C57BL/6J inbred background[50], were chosen as a model system for this study. The hormone-responsive WAP promoter drives expression during late pregnancy through mid lactation, coincident with endogenous WAP expression[75], in the mammary gland luminal epithelial cells in Waptag1[50]. Waptag1 parity-induced tumors develop late in the life of multiparous females, allowing for evaluation of early stages of cancer progression: atypical cells appear in visually normal mammary tissue and multiple stages of ductal carcinoma *in situ* (DCIS) progress into mammary tumors of one or more histological subtypes. Histologically, Waptag1 early DCIS and papillary adenocarcinomas resemble human DCIS and human papillary breast tumors[50]. Genetically, two of the most common genes mutated in human breast cancer, TP53 and RB1[73], would be predicted to show abrogated function in Waptag1. Similarly, two other human cancer genes, FBW7 a member of the ubiquitination apparatus[76], and a mitotic spindle checkpoint protein, BUB1[77], should show similar altered expression in Waptag1 females. Cytogenetically, genomic instability, including aneuploidy which is frequently seen in human breast cancer is a consequence of SV40 tumorigenesis[78]. Chromosomes 9 and 16 are frequently

deleted in SV40 Tag-induced mouse tumors; these regions are syntenic with human chromosomes 3p21, 3q, 6q12, 15q24, and 22q, several of which undergo deletion in human cancers[79,80].

Extrintistic factors, such as transgene integration site, genetic background, and copy number, can dramatically affect SV40-transgene tumorigenic phenotype. In sharp contrast to homozygous Waptag1 multiparous females, which develop mammary tumors around 12 months of age, hemizygous virgin or mated females generated from the same construct (WAP-SV-T on the outbred NMRI genetic background), show earlier tumor onset[46-48]. In addition, the WAP-SV-T females are unable to suckle or feed their young[46-48]. Female mice from a second founder line from the Knowles laboratory, [C57BL/6J-Tg(WapTAg)3Knw/J, Waptag3], did not suckle their pups, or routinely develop mammary tumors. Instead, Waptag3 males and females developed osteosarcomas of the *os petrosum*, as well as adrenal adenocarcinomas at about 6 months of age[50]. Together these findings suggest the WAP promoter may be leaky and that transgene integration site and environmental dissimilarities can change the rate and location of tumor formation. Additionally, variation in transgene copy number and genetic background undoubtedly play a role in the phenotypic differences among these three strains.

1.5. *In Situ* Carcinoma Studies *in Vivo*

One of the primary goals of these experiments was to characterize ductal carcinoma *in situ*, (DCIS) in mouse models as it has been a limited area of study. Although over-expression of Wnt1 is not a common occurrence in human breast carcinogenesis, the MMTV-Wnt1 mouse model has been used to investigate universal transcriptional changes in DCIS[42,61]. The MMTV-PyMT mouse model develops an early onset DCIS, which rapidly progresses to multiple mammary tumors[43], which may hinder characterization of the early stages. Some transgenic mouse models expressing SV40 Tumor antigen in the mammary gland epithelial cells develop DCIS that progresses to multiple mammary tumors, which resemble stages of human breast carcinogenesis[46-50,78]. Each model, developed independently from the same construct and characterized on different genetic backgrounds, offers unique opportunities for *in vivo* studies of DCIS and progression to invasive cancer. Unfortunately, only one of these models is well

described and publicly available, C57BL/6J-Tg(WapTAg)1Knw. The Waptag1 mouse model (detailed within) has been extensively characterized using histology[50] and microarray analysis. Here, Waptag1 changes during tumor progression have been compared directly with those of human DCIS with progression to invasive cancer (Chapter 4). Finally, although previously used to study metastatic progression, mouse models which offer inducible and reversible oncogene activation may be useful early models in the future as well[81-84].

1.6. SV40 Tumor Antigen (SV40 Tag)

Transgenic mouse models that over express oncogenes, such as Simian Virus 40 Tumor Antigen (SV40 Tag), provide a means to study multiple stages of tumor progression within genetically-defined inbred backgrounds. Numerous mouse models exploit the oncogenic potential of different SV40 Tag regions providing the research community with models of human cancers, including osteosarcomas, insulinomas, hepatocellular carcinoma, kidney and choroid plexus tumors, lymphomas, mesotheliomas, mammary carcinoma, and others (for review[85]).

Simian Virus 40 is a member of the Polyomaviridae family of DNA viruses; its natural host is the rhesus macaque monkey, where it lives in kidney cells for an indefinite length of time. The virus was first discovered as a contaminant of the poliovirus vaccine when 'vacuolar cytopathatic changes' were seen in cultures of African green monkey cells[86]. A short time later, two groups revealed that injection of SV40 produced tumors in newborn hamsters[87,88]. Consequently, SV40 Tag oncoproteins have been shown to induce a wide variety of neoplasms.

The early region of SV40 Tag encodes three unique proteins: Large Tag (708 amino acids [aa]), small tag (174 aa) and 17K, the mini-tag (135 aa). SV40 Tag is capable of transforming host cells through the actions of two of its early proteins, Large Tag and small tag, although the contribution to transformation of 17K tag remains largely undefined. Large Tag and small tag proteins bind a plethora of host cell proteins, altering their normal function, either through inhibition or stimulation. Three SV40 Large Tag domains (J, Rb, and Trp53) and one additional small tag domain (Pp2a), as well as the proteins which bind those SV40 Tag domains, are involved in transformation; the requirement of domains depends upon cell type, environment, and/or genetic background. In some systems, the N-terminal 127 amino acids of SV40 Tag were

sufficient for transformation and progression to hepatic or pancreatic tumorigenesis[89-92], implying that the C-terminal, Trp53 domain was not required in these cell types.

SV40 Tag binds to and stimulates HSPA8 activity. All three SV40 Tumor antigen proteins function as co-chaperone DNA J proteins and bind the DnaK chaperone HSPA8 (formerly HSC70), a homolog of heat shock protein 70[90,93]. Binding of SV40 Tag's N-terminal J-domain to HSPA8 is ATP-dependent and stimulates HSPA8's ATPase activity[94]. Following activation, conformational changes occur (in HSPA8, SV40 Tag and the bound substrate) which permit chaperone activities such as: protein folding/unfolding, protein transport (import/export of host cell proteins across ER and mitochondrial membranes), and dissociation of multimeric protein complexes, such as E2F- RBL2 (formerly p130)[95]. In summary, SV40 Tag's J domain- HSPA8 complex is essential for transformation, viral DNA synthesis, and activation of transcription[90]. The SV40 Tag-HSPA8 complex promotes additional transformation events: 1) association of retinoblastoma family proteins with SV40 Tag[96]; 2) phosphorylation and degradation of substrates[90]; 3) inhibition of apoptosis[97].

SV40 Tag binds the retinoblastoma family proteins to inactivate them. Retinoblastoma (RB) family members, RB1, RBL1 (p107) and RBL2 (p130), normally form a complex with transcription factors such as E2F, MYOD1, PAX3, and ABL1 (c-Abl), repressing their transcriptional activation capabilities. In the presence of ATP, the SV40Tag-RB and SV40Tag-HSPA8 protein complexes disrupt RBL2-E2F, and RBL1-E2F associations. This liberates E2F to activate transcription of target genes which promotes cell cycle progression, DNA synthesis/replication, and nucleotide metabolism[77,95]. In addition, SV40 Tag simultaneously blocks the activity of tumor suppression genes, such as Trp53, at a molecular level.

SV40 Tag complexes with TRP53, inhibiting its normal functions: 1) to act as a gate-keeper and prevent cell cycle progression during conditions of DNA damage, synthesis of viral DNA, or other genotoxic stress; 2) activate genes involved in cell cycle regulation, DNA repair, metabolism and apoptosis. SV40 Tag proteins inhibit Trp53-mediated activities in multiple ways, including several pathways independent from direct TRP53 binding[98,99]. Although the mechanism is

currently unknown, SV40 Tag proteins can inhibit apoptosis and override Trp53-mediated growth suppression[98-101].

CREBBP and EP300 proteins associate with SV40 Tag to either inhibit or stimulate their function. The proteins EP300 (formerly p300) and cyclic AMP-response element binding protein (CREB-Binding Protein, CREBBP or CBP), are structurally and functionally similar; both function as transcriptional co-factors and acetyltransferases; either protein can bind to TRP53 to enhance its transcriptional activation, through acetylation of TRP53 and/or histone acetylation at the promoter of Trp53-mediated genes[77,102,103]. CREBBP also acetylates SV40 Tag at lysine 697 (K697) in a TRP53-independent manner[90,103,104]. Thus, acetylation of SV40 Tag may promote or disrupt its associations with other host proteins thereby affecting protein stability, localization, interactions with other proteins, and/or interactions with DNA during chromatin remodeling.

The cellular consequence of SV40 Tag binding to CREBBP/EP300 proteins is currently an unresolved issue. It is not clear if SV40 Tag binding to CREBBP/EP300 proteins *inhibits* or *stimulates* normal function. Some researchers suggest that the SV40 Tag-EP300 association is inhibitory; Tag alters EP300's phosphorylation state, rendering it inactive, especially its TRP53-mediated co-activation of transcription[90,104,105]. Others suggest that SV40 Tag increases CREBBP's acetyltransferase activity, enhancing it for its own viral needs[103,106], such as increasing E2F-mediated transcriptional activation. SV40 Tag may bind different family members (EP300, CREBBP, or EP400), one to inhibit, another to stimulate; post-translational modifications may influence binding with SV40 Tag with CREBBP/EP300 proteins.

SV40 binds host cell proteins, such as FBW7, CUL7, BUB1, and NBS1 to inhibit their normal function. *Fbw7*, F-box and WD-40 domain protein 7, archipelago homolog (Drosophila), is a human tumor suppressor and a member of the ubiquitination apparatus. When bound by SV40 Tag, FBW7 is relocalized to the nucleoplasm, thereby inhibiting the degradation of proteins needed for cell cycle progression[76]. CUL7/p185/p193 is part of the E3 ubiquitin ligase complex and is part of a scaffold complex that marks no-longer needed proteins for degradation by the proteosome. SV40 Tag inactivates FBW7 and CUL7 to promote growth, probably through increasing the longevity of proteins, resulting in enhanced transformation and tumorigenesis[104].

Bub1, budding uninhibited by benzimidazoles 1 homolog (S. cerevisiae), encodes a mitotic spindle checkpoint protein, and is found mutated in human colorectal cancer and leukemia[77]. The SV40 Tag-BUB1 interaction may contribute to SV40 Tag-induced transformation, aneuploidy, and/or genomic instability[107]. Alongside MRE11A and RAD50, Nijmegen breakage syndrome protein 1 (NBS1) forms the MRN complex involved in double-strand break repair. The SV40 Tag-NBS1 association reinitiates DNA replication regardless of the need for DNA repair, thereby contributing to genomic instability[77,108].

SV40Tag binds SP1, enhancing its activity. *Sp1* encodes stimulating protein 1, a ubiquitously expressed transcription factor that binds the GC-box motif within the promoters of many genes. The SV40 Tag-SP1 complex may enhance Sp1-mediated transcriptional activation[109].

SV40 Tag supports other functions, needed for its own viral replication, which may enhance transformation towards tumorigenesis, such as the following: 1) ATPase activity, which provides energy for DNA unwinding; 2) helicase activity, for unwinding (or rewinding) DNA; 3) transactivation of promoters to supply needed molecules for entry into S phase, such as *Ccna1* and *Ccna2*, as well as those for cell cycle progression; 4) loose transactivation of Pol I, II, and III[77,85].

SV40 small t-antigen (tag) binds to and inhibits PP2A. Protein phosphatase 2a, encoded by the *Pp2a* gene, is a trimeric protein complex whose normal function is to phosphorylate and *inactivate* kinases (and other proteins) involved in cell cycle and mitotic checkpoints, apoptosis, signal transduction, and cytoskeletal organization[110-113]. SV40 tag inhibits PP2A's phosphatase activity, which leaves kinases, such as PRKACA/PKA, MAPK1/ERK, AKT1, PIK3/PI3K, MAP2K1/MEK1, as well as the Na^+/H^+ antiporter SLC9A1, and BCL2 constitutively activated[110,111,113]. The SV40 tag/PP2A complex promotes increased cell migration and cell spreading via increased paxillin in the cell[110,113]. Taken altogether, the SV40 tag-PP2A interaction contributes to tumorigenesis through constitutive activation (via lack of deactivation) of kinases, cell cycle proteins, and apoptotic factors.

SV40 is not naturally found in the human population, to our knowledge. However, some percentage of Americans were accidentally infected with live SV40 virus in their polio vaccinations in the 1960s (for review [85]). Moreover, the effects of SV40 Tag oncoproteins mimic many aspects of human cancers: continued activation of SV40 Tag over time results in at least three major cancer phenotypes. First, loss of cell cycle and mitotic checkpoints (due to SV40 Tag's association with TRP53, RB1, BUB1, NBS1, and PP2A) leads to uninhibited and autonomous growth[77,85,90,102,107,110-115]. Second, loss of initiation of at least one apoptotic pathway and anti-apoptotic signals (through SV40 Tag's interaction with TRP53 and PP2A) results in survival of abnormal cells[97-102,110,111,113,114]. Third, loss of a DNA repair mechanism (via NBS1 and TRP53) leads to genomic instability[107,108]. Finally, sequestering of ubiquitination machinery (CUL7 and FBW7) by SV40 Tag prevents degradation of critical proteins[76,104,116], and recruitment of transcriptional co-activators (HSPA8, SP1, CREBBP/EP300) enhances the aforementioned effects[90,103-106,109]. In summary, the many functions of SV40 Tag proteins ultimately cause cancer with many features that imitate human cancers characteristics.

1.7. Chromatin Remodeling, Retrotransposons and Cancer

Deregulation of gene expression is an early step in neoplastic transformation. Given the close link with chromatin structure and operation, it is not surprising that abnormal expression of many cellular genes goes hand in hand with epigenetic changes, such as DNA methylation and histone modifications. Dysfunction of chromatin remodeling processes coupled with aberrant gene expression and genomic instability are contributing factors in cancer.

Uncharacteristic DNA methylation patterns are frequently associated with malignancy in humans[117]. Both genomic DNA hypomethylation and hypermethylation have been causally linked with carcinogenesis. For example, targeted mutant mice hypomorphic/null for *Dnmt1*, develop thymic lymphomas at 4-8 months of age and display chromosomal aberrations, suggesting that genome-wide hypomethylation can lead to tumor formation[118]. On the other hand, the CpG island within *Sfn*, a gene encoding a cell cycle checkpoint regulator 14-3-3 σ, is hypermethylated and silenced in human breast cancer[119]. Thus, the effect of DNA methylation on oncogenesis can depend on locus-specific factors as well as the global DNA methylation status.

Generally, histone modifications, such as acetylation or demethylation are associated with open chromatin structure, allowing transcriptional apparatus access to the underlying DNA. Conversely, heterochromatin is associated with silenced DNA, due to histone H3 tri-methylation at lysine 9 (tri-meH3K9); moreover, mono-, di-, or tri-methylation at K9 may trigger methylation of the associated DNA, which then may recruit histone deacetylases, resulting in hypoacetylated lysines along histone tails[120]. In mammals, the histone methyltransferases SUV39H1/SUV39H2 are responsible for tri-meH3K9; this modification is recognized by heterochromatin protein chromobox homolog 5 (CBX5/HP1 or other chromobox family member), which binds to SUV39H1/SUV39H2 and recruits DNA methyltransferases, such as DNMT3B, that methylate the cytosine within CpG dinucleotides[121]. Subsequently, methylated cytosines cause the methyl binding protein MECP2 to bind to DNA and together with a DNMT3 family member, recruit histone deacetylases (such as HDAC1) to histones H3 and H4[122].

When considering both active and inactive chromatin modifications, histone tails contain lysines which may be methylated, ubiquitinated, or acetylated and phosphorylatable threonines and serines. In addition to mono-, di-, or tri-methylation at K9, H4 tri-methylation at K20 occurs independently via SUV420H2 methyltransferase, and this epigenetic modification is thought to be directed by CBX5 (formerly HP1), rather than CBX5 being recruited first[123]. Phosphorylation at serine 10 of histone H3 by the mitotic aurora B kinase (AURKB), which occurs during M phase of the cell cycle, results in the dissociation of CBX5 from heterochromatin, but leaves tri-meH3K9 unaltered (see below)[124,125]. Serine phosphorylation-induced blocking of CBX5 binding has also been identified in linker histone variant H1.4, in which methylated K26 binds CBX5, but subsequent phosphorylation of S27 inhibits CBX5 binding[126]. Finally, many lysines are flanked by a serine or threonine, suggesting that this mechanism of control may be even more diverse than previously thought[127,128].

The current model of protein-DNA interactions and histone modifications is as follows: locally, *if* there is no promoter CpG methylation, nor MECP2 family member, HDAC, or inhibitory protein complex bound, *then* transcription factor complexes, containing histone acetyltransferases can bind to promoters (containing their consensus motif), acetylate lysines along histone H3 and

H4 tails, which allows a more open chromatin formation such that the 'holo' RNA Pol II complex may bind and initiate transcription. Genes expressed in one system but repressed in another, such as the estrogen receptors (*Esr1/Esr2*) and *Hoxd4*, provide examples of local active or inactive chromatin states[129,130]. Alternatively, others theorize that initially global opening of chromatin occurs (i.e. within open chromatin fibers) and then local epigenetic modifications, such as histone acetylation, are necessary for gene transcription[131-133].

Repetitive sequences, such as retrotransposons (or transposable elements), are generally located within regions of heterochromatin and therefore normally are transcriptionally silenced by epigenetic mechanisms, such as DNA and histone methylation, and deacetylation of histones[120,134-136]. A few reports detail the methylation status of retrotransposons in cancer, however, evidence for this as a pre-malignant occurrence is lacking. Thus, elevated expression of transposable elements in cancer may reflect alterations in the epigenetic backbone of the genome. Elevated levels of retrotransposon expression are found in a number of human cancers and are often associated with genomic rearrangements. Gene deletions, amplifications, conversions, or translocations as well as disruption of normal gene function may be a consequence of homologous recombination at retrotransposon 'hot spots'[137].

Retrotransposons are a distinct type of very abundant, repetitive genetic elements able to propagate themselves within a genome and are ubiquitous components of genomes in most eukaryotic organisms. In mammals, 40-50% of genomes consist of retrotransposons or their 'remnants', i.e. phylogenetically older insertions no longer capable of propagation[138,139]. Although they are transcribed by the host cell, their mode of propagation is through an RNA-intermediate, which is reverse transcribed and integrated into the genome, thus increasing their copy number.

Approximately one-tenth of retrotransposons resemble retroviruses in that they are flanked by Long Terminal Repeats (LTRs), while the majority of retroelements are non-LTR retrotransposons, Long and Short Interspersed Nuclear Elements (LINEs and SINEs)[139]. LINEs contain two open reading frames and the ability to retro-transpose, unlike SINEs which are 'non-autonomous' and may transpose and integrate using the protein machinery of LINEs, or rarely, other elements. Some LTR retrotransposons, such as endogenous retroviruses, are closely

related to, and originate from infectious retroviruses; however, most LTR retrotransposons have evolved to be non-infectious parasites. One distinction between certain transposable elements is their preference for GC-rich genomic regions (SINEs) or GC-poor locations (LINEs and LTR elements)[140]. Endogenous LTR retroviruses usually contain genes which encode: 1) the viral particle (*gag* gene) in which reverse transcription takes place; 2) the enzymes protease, integrase, and reverse transcriptase (*pol* gene); 3) the envelope (*env*) gene, often mutated or absent in endogenous retroviruses, which facilitates infection of a neighboring cell; as well as additional regions including, a primer binding site, (used by the virus to begin reverse transcription), leader sequence, (non-translated region downstream of the transcription start site, present at the 5' end of all virus mRNAs), polypurine tract, a short run of (~10) A/G residues for initiating (+) strand synthesis during reverse transcription. The internal coding region is flanked by 5' and 3' LTRs: a U3 region, which contain promoters (including transcription factor binding sites) and enhancers (including hormone response elements) to further their expression; a terminally redundant Repeat (R) region; and a U5 region, a unique, non-coding region, first part of the genome to be reverse transcribed, forming the 3' end of the provirus genome. Three different classes of LTR retrotransposons have been described in mice. LTR Class I family members (e.g. the murine leukemia virus, *Mlv1*) contain all three *gag*, *pol* and *env* genes flanked by LTRs. Many members of the *Mlv1* family are defective due to deletions with *pol* and/or *env* genes. Without sequencing of individual insertions, it is difficult to determine whether these LTR Class I retroviruses are defective or not. LTR Class II retrotransposons, such as the intracisternal A particle (*Iap*) contain *gag* and *pol* genes between their 5' and 3' LTR sequences, whereas early transposon element 1 (*Etn1*) genes lack functional enzymatic genes but contain normal LTR sequences. LTR Class III retrotransposons contain normal LTRs and may or may not have *pol* genes; the mammalian apparent non-autonomous LTR retrotransposon (MaLR) family members, specifically MTs or mouse transcripts, are examples of Class III elements. Most transposable elements are able to propagate within the DNA of cells under the appropriate conditions, such as when a cell's means to suppress viruses/double stranded RNA/heterochromatin has been compromised.

Dysfunctional molecules involved in RNA interference might affect unsilencing of transposable elements. RNA interference, RNAi, suppresses target sequences (retrotransposons, repetitive sequences, viruses, and other elements) through formation of double-stranded RNA (dsRNA), which ultimately triggers silencing. Initially, when endogenous retrotransposons are transcribed, it is in both sense and antisense directions, which then results in formation of dsRNA[141]. In mammals, dsRNA is processed to 21-26-nucleotide-long sense and antisense fragments by an endoribonuclease III, DICER1, to produce small interfering RNAs (siRNAs). Next, together with a member of the Argonaute family of translation initiation factors, an RNA-dependent RNA polymerase (RdRP), a helicase, and the rest of RISC, RNA-induced silencing complex, the siRNAs target complementary mRNAs for degradation. RNAi is functional in plants, worms, flies, mice, and humans, although the enzymes involved may differ among organisms.

RNAi machinery operates similarly for microRNAs (miRNAs), functioning as a post-transcriptional mechanism to regulate gene expression. One notable difference is that miRNAs predominantly bind to the 3' untranslated region of mRNAs. Whereas siRNAs are gene-specific, miRNAs are *cis*-acting regulatory sequences that bind to numerous transcripts to regulate entire groups of genes. Like expression of gene transcripts, expression of miRNAs may be cell-type specific and their up-regulation or down-regulation effects the protein levels of their targets. Several key genes aid in miRNA production: RNASEN (formerly DROSHA) and the microprocessor complex with DGCR8 cleave pri-miRNAs to form pre-miRNAs that are exported out to the nucleus by XPO5, exportin 5. Once in the cytoplasm, cleavage by DICER1 results in the functional, mature miRNA, which can bind with the RISC complex (a similar mechanism to siRNAs) to repress translation and/or cleave target mRNA molecules (for review, see[142]). Up- and down-regulation of different miRNAs is associated with human diseases, including cancer and metastases (see[143-147] for reviews of miRNAs in various diseases). Moreover, transcription or repression of miRNAs may be regulated by DNA methylation and/or modifications to histone tails (for review see[148]). Thus, it may be a *combination* of mechanisms by which retrotransposons are transcriptionally activated and associated with open chromatin, genes are deregulated, mutations

occur without repair, cells escape apoptosis, chromosomes exhibit genomic instability, and ultimately uninhibited growth favors progression of tumorigenesis to invasion.

A causal relationship between carcinogenesis and integration of endogenous retroviruses is well established. *De novo* insertions of LTR class I, murine leukemia virus 1 (*Mlv1*), has been shown to activate the myelocytomatosis oncogene (*Myc or c-Myc*), proviral integration site 1 (*Pim1*), or lymphocyte protein tyrosine kinase (*Lck*) genes leading to development of lymphomas in mice[149]. Insertional mutagenesis by Mouse Mammary Tumor Virus (MMTV) and subsequent activation of the following: fibroblast growth factor 3 (*Fgf3*, formerly *Int2*); Notch gene homolog 4 (*Notch4*/*Int3*); eukaryotic translation initiation factor 3, subunit E (*Eif3e*, formerly *Int6*); or wingless-related MMTV integration site 1 (*Wnt1*), leads to mammary carcinoma[149-151]. Insertional gene activation by a LTR class II elements, intracisternal a particle (*Iap*), results in mammary carcinoma, lymphoma and myeloid leukemia in mice[152,153]. To date, no relationship between LTR class III retrotransposons and carcinogenesis has been described, although these elements drive alternative transcripts and expression of many genes in oocytes and cleavage stage embryos[154]. Insertional mutations due to LINE1 elements result in human breast cancer, hemophilia A, and Duchenne muscular dystrophy[155-157]. Similarly, a novel insertion of the SINE element *Alu* results in aberrant splicing of the E74-like factor 3, *Elf3*, gene in human breast cancer and cell lines[158]; this gene has been shown to be over expressed in human DCIS and breast cancer[159]. Recently, increased numbers of SINEs in close proximity of a gene was shown to be proportional to the deregulation of that gene's expression in tumors[160]. Based on these findings and other data, several laboratories are utilizing retroviral insertional mutagenesis strategies as a cancer gene discovery tools and to generate new mouse models[161-165], or for review[166,167].

LTR class II elements, Human Endogenous RetroVirus (HERV)-H, HERV-K, and their splice variants, are up-regulated in breast, ovarian, colon and testicular cancers; leukemia; germ cell tumors; and tumor cells lines when compared with normal tissues[168-171]. In ovarian cancer, HERV-W expression increased four-fold, and LINE1 element increased fifty-fold when compared with non-malignant ovarian controls[172]. In prostate cancer, HERV-E was found to be highly expressed and similarly, HERV-F was overly active in a variety of cancer cell lines[173,174]. In five

human DCIS samples from breast biopsy, HERV-K *env* expression was up-regulated six- to twenty-fold compared with normal controls[168]. Thus, it is possible that over expression of HERV *env* could be used as an early indicator of malignant transformation in early ductal carcinoma *in situ*. A thorough evaluation of transposable elements across many mouse models as well as human breast cancer and carcinoma *in situ* will clarify whether elevated expression of retrotransposons may serve as a useful marker for detection of early alterations in tumorigenesis.

1.8. Statistical Microarray Analysis

The most important aspect of microarray analysis is proper ***experimental design*** to permit accurate, reliable, and consistent data collection that form the basis of scientific interpretation. Sufficient sample size, multiple levels of replication, and randomization techniques *throughout* will simplify analysis of the data by ensuring enough degrees of freedom (see definition below) to perform statistical tests, providing a specified level of confidence in the results, avoiding and/or reducing biases which can lead to skewed interpretation of the data. Another important concept is that statistical tests are performed only on *samples* from the true population. Therefore, all of the estimates are approximations for the population, based on the samples.

There are many sources of error that need to be adequately addressed in the experimental and technical design of microarray experiments. There is *biological variation* (i.e. mouse to mouse differences), *technical variation* due to RNA extraction, cDNA synthesis, labeling, hybridization, as well as scanning and gridding of the array spots with two-color arrays, and *array variation* (array to array and within array variation, such as non-uniformity, spatial heterogeneity, dust, scratches, print-tip problems, etc.) In addition, with two-color microarray experiments, there can be *dye variation* (Cy3 versus Cy5), which can be decreased by using a dye-swap or loop experimental design or eliminated with single-color arrays. Thus, the first task is to decrease and/or stabilize the variance in the microarray experiment as much as possible by utilizing the following criteria: 1) several samples per sample 'group' or 'treatment' (biological replicates); 2) multiple arrays per sample and/or numerous spots for each gene per array; 3) quality control tests and appropriate data transformation. Reducing the variation increases the *power* (i.e. the probability of detecting *true* differential gene expression) in a microarray experiment. In addition

to design considerations, the appropriate statistical test needs to be applied to the dataset to yield useful results.

One statistical test common in microarray experiments is the ***analysis of variance (ANOVA)***. ANOVA is a method of conducting statistical tests to determine differences among samples within and across treatments. It offers a level of confidence in the results that cannot be achieved by direct comparison alone. ANOVA utilizes the inherent variation in the each individual value across samples as well as the variance among samples within each treatment. The ANOVA model separates the total variance into multiple components, each of which can be treated as either fixed or random effects. In a *fixed* ANOVA model, only one source of variation is used to calculate the test statistics. However, if there is more than one source of random variation in the model, then the *mixed* ANOVA model offers a method to treat additional components as *random effects*, (i.e. if the population was sampled again, these components would show similar, but not identical effects)[175].

For microarray data, ***R/MAANOVA*** (MicroArray ANalysis Of VAriance, in R) provides *relative* differential gene expression levels, which are both statistically and biologically significant[175]. R/MAANOVA is a software package used to analyze a wide variety of microarray data. It provides a list of statistically significant genes with an associated level of confidence, rather than merely a fold-change in the expression of a particular gene between two samples. Initially, two models are developed, a null model: no differential gene expression (i.e. the variation in the mean expression (log transformed) of a gene across samples is zero) and an alternative model: there is differential gene expression (i.e. that variation does not equal zero). As an example, a reference dye-swap design can be used to demonstrate the ANOVA model for a *two-color microarray experiment*. A reference dye-swap involves each mouse tumor *sample* and a standard *reference* be put on two arrays. On the first array, one sample is labeled with Cy3, while the reference is labeled with Cy5 and then dyes are swapped for the second array, using another aliquot of the *same* sample and reference. As mentioned above, this design reduces the variation due to dye effects and adds another level of technical replication. The ANOVA model for *two-color array* experiment, in which *each gene* is fit to the following model, is

$$\mathbf{y_{ij} = \mu + A_i + D_j + R_h + M_k + T_l + \varepsilon_{ij}} \quad \text{Equation 1}$$

where, $y_{ij} = \log_2$ of the signal intensity for any given gene, μ= mean expression level for that gene, A_i = array effect, D_j = dye effect, R_h = for the reference, M_k = mouse effect, i.e. the variation from mouse to mouse, T_l = treatment effect for that gene, and ε_{ij} = the residual[175-177]. In a mixed ANOVA model, μ, D_j, R_h, T_l and G_m are treated as fixed effects and A_i, M_k, and ε_{ij} are random effects[175,176]. Random effects are estimated using restricted maximum likelihood (REML, a means to approximate components in a mixed ANOVA linear model)[178] and used to compare treatments.

For *single-color Affymetrix array* experiments, where one mammary tumor sample, for example is placed onto a single array, the following model is fit *for every gene*:

$$\mathbf{y_{ij} = \mu + \alpha_i + \beta_j + T_k + \varepsilon_{ij}} \quad \text{Equation 2}$$

where, $y_{ij} = \log_2$ of the signal intensity for any given gene, μ = mean expression intensity, α_l = effect of the probe (for which there are several probes per gene), β_j = effect of the array, T_k = treatment effect for that gene and ε_{ij} = the residual error[176,179]. Since only one sample per mouse was put onto an array, the array and mouse effect are completely confounded, and no mouse term is included in the model. The fixed ANOVA model is most frequently used for Affymetrix datasets, since, aside from the residual, the only additional random effect would be array (β_j) and that has only a minute influence on the test statistic, so array becomes part of the residual error component[175,176].

As shown in the equations above, R/MAANOVA utilizes the variation *of each gene* across samples (but also within and across treatments) to determine statistically significant expression differences. R/MAANOVA calculates the expression level of each gene in a sample *relative* to its weighted average across all samples. Since microarray data error variances may not be normally distributed, or their distribution may be unknown, for hypothesis testing or to determine significance thresholds and confidence intervals, randomization techniques are employed. Unless the sample size is very large, the method used is shuffling of the *residuals* (the difference between the observed value and the fitted value according to the null model); R/MAANOVA then

computes *p-values* (the probability of getting such an extreme result if the null hypothesis is true) as well as confidence intervals (estimates of the *true* values for the population, that show the variance and/or precision of those estimates). By conducting permutations – 'shuffling the data' - across genes, the magnitude and distribution of gene expression differences can be simulated.

For comparison of two sample groups a ***t-test (test statistic)*** or pairwise comparison is often used because it compares the *mean* expression levels between two treatments, utilizing the variance within each treatment (*within group* variation) separately (unless otherwise stated). In R/MAANOVA, inference methods for the test statistic are based on the t-distribution (assuming a normal distribution of the population from which the samples came) with modified degrees of freedom due to non-equal variances[180]. The degrees of freedom in an experiment depend upon how many 'things' can fluctuate freely in the model[175]; in the simplest situation, if you are only estimating one factor, it is one less than the number of samples. Determining the degrees of freedom becomes more complicated when estimating error across many treatments. Using the log-ratios of the expression levels, *for each gene* (one at a time), the t-test utilizes the inherent *within group* variation, i.e. between mice within one sample group, for both treatments, to estimate the variation within the experiment and calculate significant expression levels. Therefore, if the *within group* variation is high within one treatment, R/MAANOVA may underestimate the number of statistically significant genes between treatments. Since only two treatments are being compared in a t-test, the variation across treatments cannot be utilized, like ANOVA accomplishes, therefore for large datasets, contrasts are preferred over t-tests.

A ***contrast*** is a method used for comparisons ranging from a simple pair-wise comparison to a comparison between two averages of treatment means or a comparison of one treatment mean against the average of all treatment means for the experiment. Furthermore, in some instances weighed averages may also be incorporated[181]. Like ANOVA, contrasts utilize the *within* as well as *across* treatment variation, to determine statistically significant genes. For simple pairwise comparisons, a contrast integrates information across all treatments and samples, while only calculating the expression values and the error associated with the two treatments of interest. For example, for 6 treatments, many different contrasts could be designed to compare them.

However, suppose that treatments 1&2 are similar (but not quite the same), while treatments 4&5 are also similar, but differ from treatments 1&2, and that comparing treatments 3&6 are not relevant. Using planned contrasts, treatments 1&2 can be compared to treatments 4&5, without calculating each pair-wise comparison individually, but while utilizing all the variation. For this contrast, the null hypothesis would be that the average of each of the two groupings (1&2 vs. 4&5) will be same (i.e. the difference between them will be zero), whereas the alternative hypothesis is that the mean averages will be different (i.e. the difference does not equal zero). Contrasts offer more power than a standard t-test comparison.

The R/MAANOVA software offers orthogonal contrasts as part of the function 'matest' in which the user may specify a matrix for comparison of the treatment term. For the above example, zero represents a treatment that is NOT in the contrast, 1 is used for the first half of the comparison and -1 for the second half of the comparison. Therefore, the contrast would be 1,1,0,-1,-1,0 in which treatments 1&2 are given a value of 1 and treatments 4&5 are given a value of -1. In R/MAANOVA contrasts greatly reduce the amount of preparation time needed to set up all pairwise comparisons among treatments. Instead, a contrast matrix can be called to retrieve as many as 14 contrast results with one ANOVA test.

The ***F-test***, a generalization of the t-test statistic, is used for analysis of more than two groups of samples. It detects differences among sample means (and treatments) by evaluating the sample variation within and among different treatments. Inference procedures for F-tests are gene-specific, rely on the models, and are calculated using the residual sum of squares (i.e. the sum of the square of the residuals, see above) and associated degrees of freedom. Significance levels are derived from the F-distribution (or permutation techniques). There are two primary F-statistics for either the fixed or mixed ANOVA model with the common F-test essentially the same as the F1 test below, which follows an F-distribution. The other F-test has an unknown distribution and therefore permutation tests are performed. Using an approximation of the variance within an experiment, the confidence intervals are calculated: F1, which estimates gene-specific error variance; Fs, which uses a James-Stein shrinkage estimator to compute the gene-specific variances and does not make any assumptions about constant error variances[176].

Unlike F1, which uses only gene-specific variances, the Fs test utilizes the data *across* genes as well, in order to attain more powerful estimates[176].

The recent R/MAANOVA algorithms used to compute the Fs statistic have several advantages: the most consistent results, despite permutation test stochasticity, better variance histograms than F1 tests, and very manageable gene lists (due to the shrinkage estimator). Therefore, the Fs statistic is the most suitable test-statistic for microarray analysis[176] and was used for all analyses in which raw data was available.

When thousands of genes are tested, resulting in thousands of hypothesis tests, the ***problem of multiple testing*** becomes evident: there will be a large number of false-positives among the true statistically significant results due to random chance. This yields a situation called a *Type I* error or 'false-positive error', which indicates a gene is considered differentially expressed and rejects the null hypothesis although it is not truly significant. The other error, *Type II* error, occurs when a gene that is truly differentially expressed is not detected as significant, which is also known as a false-negative error. Hypothesis testing in a microarray experiment typically attempts to control for type I error, achieving a certain power (*see above*; also defined as one minus the probability of a type II error), depending on the sample size and experimental design. There are two primary ways to address the problem of multiple testing, Family-Wise Error Rate (FWER) and False-Discovery Rate (FDR) adjustments to the test statistics, both of which increase the level of confidence in detecting true differential gene expression.

Hypothesis testing with a ***family-wise error rate (FWER)*** adjustment gives the probability of producing one or more false-positive (type I) errors based on all of the statistical tests performed[175]. There are multiple FWER adjustments: 1) the Bonferroni correction, simply divides the significance value by the number of hypothesis tests performed; 2) the Westfall and Young one-step adjustment method takes into consideration that test statistics may not be independent[182]. FWER adjustment sets a level of significance for *all* tests in the experiment, regardless of the comparison(s) made; it is very strict and therefore offers a high level of confidence that the list of differentially expressed genes does not contain false-positives.

However, this adjustment may be associated with a loss of statistical power and may be too stringent to detect smaller changes, which may be biologically important. Therefore, FDR is the more sensible way to handle 'multiplicity of testing' when analyzing microarray data.

The ***false discovery rate (FDR)*** modification to the test statistic supplies a post-data level of confidence in the results. FDR adjustment establishes the proportion of type I (false-positive) errors according to the total number of null hypotheses rejected (i.e. differentially expressed genes). Therefore, with FDR control, the number of false-positives *expected* is based on the estimated number of differentially expressed genes, not on the number of tests performed (like FWER). FDR type of control allows a certain number of 'errors of inference' in order to gain better sensitivity and more power than the rigid FWER adjustment[183]. Thus, FDR improves the power, utilizes the number of differentially expressed genes (post ANOVA), and allows detection of smaller changes between treatments, even when the sample size is small[184]. In R/MAANOVA, the FDR-adjustment of the Fs statistic is performed using pooled permutation p-values to produce the q-values[184,185].

1.9. Summary

Putative early markers of DCIS which were predictive of invasion, currently lacking in human carcinogenesis, were identified through this thesis research. Using whole gland and microdissected samples taken throughout the lifespan of the Waptag1 mouse model of human breast cancer, microarray analyses were used to uncover early transcriptional changes associated with tumor progression. Comparison of Waptag1 gene expression profiles with those of human DCIS reveals the following early alterations: genes involved in cell cycle, cell division, and DNA replication are up-regulated predominantly within tumor cells; chromatin modification transcriptional changes coincide with retrotransposon over expression; the most efficient form of energy derivation, electron carrier activity, is dysfunctional. Genes in these categories may provide potential markers of human DCIS, predictive of invasion, to perhaps reduce invasion and metastatic spread in patients with breast cancer.

Chapter 2

MATERIALS AND METHODS

2.1. Mice and Samples

2.1.1. Mice

Multiparous females of C57BL/6J-Tg(WapTAg)1Knw (Waptag1; JAX® Mice stock# 003188, The Jackson Laboratory, Bar Harbor, ME) or age- and parity-matched C57BL/6J controls (B6; JAX® Mice stock# 000664) were housed in cages with pine shavings and free access to water and food (6% fat NIH PMI Mills) on a photoperiod of 12:12 (lights on 06:00) with a male until 6-11 months of age. Females were continuously mated in pairs or trios at 6 weeks of age and first litters were weaned approximately 6 weeks later. Virgins, single and dual pregnancy females failed to develop consistent, predictable mammary tumors. Therefore, only females that sustained a minimum of three pregnancy and lactation cycles were retired at monthly intervals (6-11 months) and group housed until their mammary glands were collected (no sooner than 3 weeks post-wean to insure adequate time for mammary gland involution). Non-lactating glands were collected from Waptag1 females which had not sustained three pregnancies at 4 and 5 months of age; these were Waptag1 'normal' controls. Mice were euthanized by cervical dislocation before their tumors reached ~10mm in size, according to approved Institutional Animal Care and Use Committee protocols. Briefly, each tumor-bearing mouse was monitored three times weekly by checking for changes in tumor size or appearance and the mouse's behavior and mobility. If changes appeared that would result in discomfort to the mouse, it was euthanized. The Jackson Laboratory's Animal Care and Use Committee (ACUC) summary for Dr. Barbara Knowles (#01011, ACUC summary name: Cell Cycle Control), was amended and approved for this research.

2.1.2. Whole mammary tumors and glands

Immediately following euthanization, ~90% of each gland or tumor was removed under aseptic conditions and dissected into 8mm^3 pieces in RNAlater (Ambion, Inc., cat. no. 7021), while the remainder was fixed in Telly's (Fekete's acid-alcohol-formalin) or Bouin's for histology. All samples for RNA were left at room temperature for 3-6 hours to allow the RNAlater to

penetrate the tissue, kept overnight at 4°C, and stored at -20°C until RNA extraction and purification (TRIzol reagents) according to the manufacturer's protocols (Invitrogen Life Technologies, cat. no. 15596018). RNA concentration and purity was determined spectrophotometrically and RNA was stored at -70°C until cDNA/cRNA preparation (First-Stand cDNA Synthesis, Superscript III, Invitrogen). Following reverse transcription with Superscript III (Invitrogen) and an oligo(dT)-T7 primer, *in vitro* transcription with a T7 RNA Polymerase (Ambion) and biotinylated nucleotides (Enzo Diagnostics) created the labeled cRNA that was hybridized to the Affymetrix chips (see below).

2.1.3. Cryopreserved tumors and mammary glands

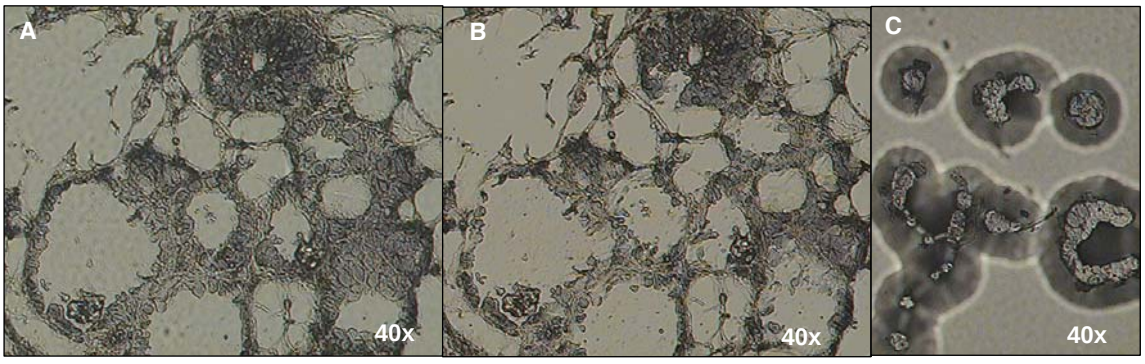
Immediately following euthanization, samples were collected following aseptic technique and cryopreserved in Optimal Cutting Temperature (OCT) compound using 2-methyl butane, chilled by liquid nitrogen. Ten slides of eight micron cryosections were cut from each mammary gland or tumor using a cryostat (kept at–40° to –20°C). Serial cryosections were mounted on ten nuclease-free, plus glass slides. For histological classification, slides one and six were stained with hematoxylin and eosin, while the remaining eight slides were stored at -80°C until laser capture microdissection was performed.

2.1.4. Laser Capture Microdissected Samples

For each mammary cryoblock, serial cryosections were cut (at eight to twelve microns thick) and mounted onto ten nuclease-free, plus glass slides. For histological classification, slides one and six were stained with Hematoxylin and Eosin (H&E), while the remaining eight slides were stored below -80°C until time of laser capture. H&E was only used for histological identification of stages of tumor progression, as this was not optimal for staining tissues when downstream applications required good quality RNA. Although it did not allow identification of cells of atypia (see Figure 3.1), a toluididine blue-based nuclear stain was chosen to facilitate identification of epithelial cells from stromal, muscle, or fat cells within the cryosections. The resolution of the microscope was not optimal (at 40x) to distinguish subtle differences, such as atypical versus normal epithelial cells; however, training and practice allowed recognition of the appropriate cells of normal B6, ductal carcinoma *in situ* (DCIS) and tumors (as shown in Figure 2.1). Variations in

thickness of the cryosections, high fat cell content of the whole mammary glands, and faulty LCM caps further complicated cryosectioning, staining, and/or isolation of specific epithelial cells. Even with these obstacles, sufficient amplified RNA, from laser-captured cells across three biological replicates per stage of tumor progression was obtained. For each sample histologically defined as advanced DCIS (AdvDCIS), cells were captured across multiple slides, but only from one area of AdvDCIS per sample, i.e. if one region of DCIS was used from slide 2, then that same AdvDCIS would be used from slides 3 and 4, to obtain the ~100 cells desired. Since regions of early DCIS (EDCIS) were substantially smaller and dispersed more thinly than in AdvDCIS glands, multiple regions of EDCIS were collected per slide in order to achieve sufficient numbers of cells, within the time constraint. Figure 2.1 shows representative sections from each of the stages of Waptag1 tumor progression: areas of EDCIS (Figure 2.1.A-E), AdvDCIS nodules (Figure 2.1.F-I), papillary adenocarcinomas (Figure 2.1.J-L), and solid, invasive tumors (Figure 2.1.M, N), as they appeared before, during and/or after LCM. Because LCM only picks up the top layer of cells, thicker sections showed cells remaining after initial laser capture (Figure 2.1.D). In these instances, the slide was moved slightly and the region with cells remaining was pulsed with the laser a second time to remove all the cells of DCIS.

Figure 2.1. Laser Capture Microdissection (LCM) Images. A) Early DCIS lesion (Female 3991, R4 mammary gland) before LCM; **B)** same Early DCIS after initial collection of cells and cap was removed; **C)** LCM cap containing about 15 cells of Early DCIS; **D)** H&E stained section of Advanced DCIS lesion, Female 6460, L2 mammary gland; **E)** same Advanced DCIS before LCM; **F)** same Advanced DCIS after cap was removed; **G)** LCM cap containing about 30 cells of Advanced DCIS; **H)** H&E stained section of Papillary adenocarcinoma, Female 4549, L2 mammary tumor; **I)** same Papillary tumor, before LCM; **J)** LCM cap containing about 25 Papillary tumor cells; **K)** Solid, invasive tumor, Female 4545, R1 mammary tumor; **L)** same Solid tumor, after LCM, with cells removed; **M)** Isolation of single tumor cells, under textbook conditions: ideal thickness of the tissue section for model staining and capture; an optimal LCM cap, which rested perfectly on the surface of the glass slide; laser with the precise focus and voltage pulsed directly on the cells of interest; 9 cells are shown. [Magnification: 40x(A-C, G,I,J, M); ±40x(D-F,K,L); 4x(H). Histogene™, toluidine blue stain (Arcturus Bioscience, Inc.), unless otherwise stated. Abbreviations: H&E, hematoxylin and eosin; DCIS, ductal carcinoma *in situ*.]



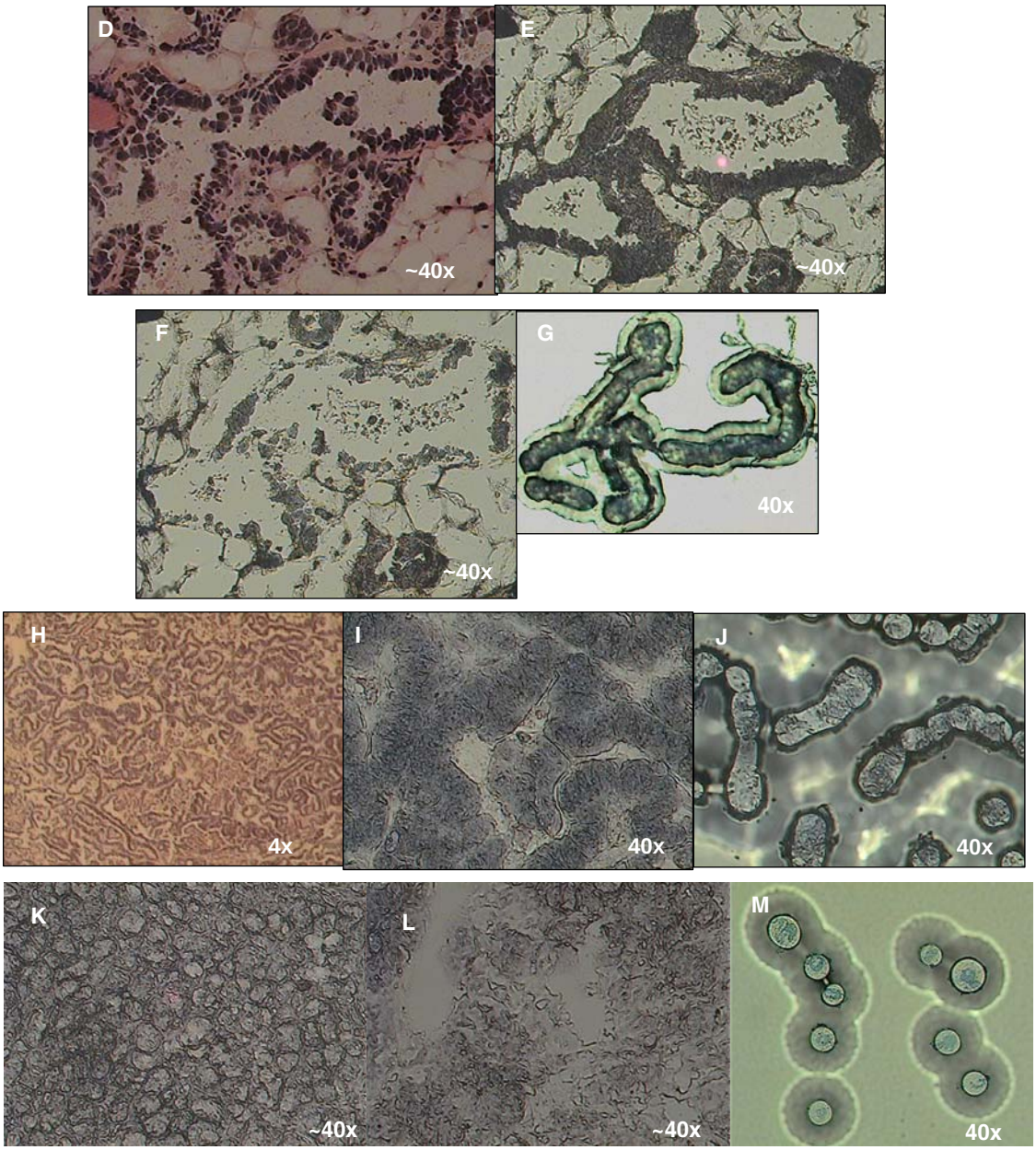


Figure 2.1. Continued.

Single slides were removed from the freezer, one at a time, fixed in 75% ethanol, stained, dehydrated and dried for 15 minutes (Histogene^{TM} LCM Frozen Section Staining Kit, Arcturus) immediately preceding LCM. A longer drying time was used to completely evaporate the xylene. For isolation of RNA, no sample was extended more than a 90-minute window of time at room temperature, from removal from the freezer through to extraction of the cells for RNA. Approximately 100 epithelial cells were isolated per sample using the PixCell^{R} IIe LCM System at the following settings: laser spot size, 7.5µm; pulse power, 80-100mW; pulse width, 0.80ms; and

threshold voltage, 180-220mV, depending upon thickness of the section. Following LCM, the film was peeled from the cap and cells were lysed in 50 microliters extraction buffer (XB) for 20 minutes at 42°C (PicoPure™ RNA Extraction Kit, Arcturus). After incubation, the film was swirled around in the tube using forceps, flicked to remove any remaining buffer, and then discarded; extracted cells were stored in extraction buffer at -70°C. RNA was extracted, purified, and linearly amplified according to protocols (PicoPure™ RNA Extraction Kit and RiboAmp™ RNA amplification Kit, Arcturus). Following amplification, RNA quality was assessed using a 2100 Bioanalyzer instrument and RNA 6000 Nano LabChip assay (Agilent Technologies); good quality amplified RNA was hybridized to microarray chips.

2.2. Microarrays

Two color microarrays were initially used, primarily for detection of retrotransposons, within unknown and/or unannotated clones. However, single-color Affymetrix (Affy) arrays offer many advantages over traditional two-color glass slide arrays. Affy arrays are uniquely synthesized *in silico* using photolithographic masks. Each target sequence (gene) on the array is represented by a 'probe set', consisting of more than 3 different 25-mer oligonucleotides per gene (perfect matches) and 1 mismatch per perfect match. The advantage of this technology is that a single gene is being interrogated by a 'probe set', thereby reducing sequence-dependent factors, non-specific background effects, error and variance.

2.2.1. Mouse

Two-color microarrays (Toronto, Ontario, Canada) were hybridized with cDNA from Waptag1 whole tumors or whole gland samples, or control C57BL/6J glands. These NIA microarrays feature 60mer oligonucleotide probes which encompass most of the NIA 15K clone set, allowing for discovery of unknown elements (http://lgsun.grc.nia.nih.gov/cDNA/15k.html)[186]. The experimental design was a reference dye-swap[177]; each cDNA sample was hybridized against the Stratagene Universal Mouse Reference using Cy5 (sample) and Cy3 (reference), then a second aliquot of the same sample & reference were labeled with dyes swapped. The cDNA hybridization, scanning, & gridding, were performed as described[187]. Briefly, samples were hybridized in pairs at 42°C for about 18 hours, then washed and dried before scanning with

GenePix 4000B (Axon Instruments) scanner. Although raw mean-intensity data was adjusted with and without background subtraction, the background- subtracted median-intensity values were used for subsequent data analysis.

Single-color Affymetrix GeneChip^{R} Mouse Genome 430 v2.0 (Affy) microarrays were used for the majority of the experiments. The Affy 430 v2.0 arrays include alternative transcripts and unknown sequences as part of their repertoire for a total of 16K to 39K transcripts, depending on pre-processing (see below). Biotinylated cRNA from one sample was hybridized per Affy chip according to manufacturer's protocols (Affymetrix, Inc.) A streptavidin-phycoerythrin biotinylated anti-streptavidin antibody tagged the sample and slides were washed, stained and dried prior to scanning at high resolution (GeneChip^{R} Scanner 3000, Affymetrix) for detection of a single fluorophore. The raw image data (.CEL files) were obtained for statistical analysis.

2.2.2. Human

Human Dataset1 breast cancer data from samples enriched in DCIS (27 samples) or Invasive Breast Cancer (IBC; 24 samples) was received from Drs. D. Craig Allred and Sangjun Lee (Washington University, St. Louis, MO) as .CEL files; samples were run on the Affymetrix (Affy) U95Av2 human gene chips. Normal human breast data, Human *Dataset2* was downloaded from the Gene Expression Omnibus [GEO; http://www.ncbi.nlm.nih.gov/geo/; Acc# GDS2250] as .CEL files; seven samples were run on Affy U133plus2.0 gene chips[188]. *Human Datasets3-5,* were downloaded from GEO for independent analysis of the prognostic value of genes mutual to early stages of tumor progression in Waptag1 and human[37,189,190].

2.3. Statistical Experimental Design and Analysis

Numerous distinct experiments, three mouse and eight human datasets were analyzed during this thesis research and will be described briefly below. At least three biological replicates were used for every stage of Waptag1 tumorigenesis, each mouse mammary tumor model, or controls, to capture, utilize, and potentially reduce the inherent biological mouse variation. For the NIA 15K and the Compugen two-color microarrays, a reference dye-swap design was utilized. Since Affy arrays are single color, no dye-swap design was available, without using two arrays for every sample; therefore, technical variation was integrated into the overall error in the model.

2.3.1. NIA 15K two-color microarrays (Waptag1 only)

To reduce error in the model as much as possible, four data quality checks were performed in R statistical program, which, depending on the quality of the raw data, led to removal of 'bad' spots. A natural log transformation of the data insured that the sources of variation (error) would be additive. Using either a fixed model, which embraces the variation between biological replicates, or a mixed model, which treats that variation as a random effect, we fit the data to the model and completed the statistical analysis. Using the pooled permutation p-values from the Fs statistic of the fixed model (unless otherwise stated), we determined which genes were statistically significant at $p \leq 0.05$. The Fs test adds a shrinkage estimation to the variance that takes into account both gene-specific error and common error across all genes[176]. The result is a list of statistically significant genes. Although a multiple test correction was used for ANOVA analysis across all stages of tumorigenesis ($q \leq 0.01$), due to the small differences between *adjacent* stages of Waptag1 tumor progression (as well as no genes significant at the less stringent $q \leq 0.1$ threshold), we chose no FDR-adjustment ($p \leq 0.05$), with the understanding that false positives were not taken into account in the pair-wise comparisons, however, subtle gene expression changes could be uncovered.

2.3.2. Affy one-color microarrays (Waptag1 or human)

For each of the three different Affy arrays (430 v2.0 mouse, U95Av2 human, and U133plus2.0 human) the raw image data (.CEL files) were imported into R/MAANOVA for separate preprocessing. Pre-processing and normalization (including mean-centering) of the data was performed using RMA, Robust Multivariate Analysis[191], and the appropriate *cdf* file (see details below). With many sequences interrogating each gene, we collapsed the probe set data using the University of Michigan Brain Array method (UMBA cdf)[192] resulting in just over 16,000 unique genes. For evaluation of retrotransposons and other repetitive elements only, the Affy annotation {http://www.affymetrix.com/support/technical/byproduct.affx?cat=arrays} method (Affy cdf) was used for pre-processing. Using both the Affy and UMBA cdf files to collapse the probe set data resulted in two gene lists for every analysis. Data was next filtered and control spots

(except GAPDH) were removed prior to analysis. For the two different human microarrays, only the subset of common genes was used for further analysis.

Using R/MAANOVA[177] we fit the data to the fixed model and completed the statistical analysis, using the pooled permutation p-values from the modified Fs test-statistic[184]. The resulting lists included statistically significant genes, with false discovery rate multiple test adjustment ($q \leq 0.1$) as well as without it ($p \leq 0.05$). In addition to the statistical cutoffs mentioned above, any transcript with a fold change of >2x in any pairwise contrast was also included in the gene list, regardless of statistical significance[193].

For mouse Affy cdf analyses, pair-wise comparisons were made, similar to the standard t-test; however, with UMBA cdf analyses, pair-wise contrasts were run instead to utilize the across tumor type variation, rather than just the variation between the two groups. In retrospect, contrast matrices should have been used for both analyses, because with a contrast, information from other samples throughout all stages of tumor progression (i.e. samples not used in the contrast) can be intergraded, whereas the t-test can only use the information within and between the two tumor types. However, since only the retrotransposons were evaluated using the Affy cdf method, the previously run pair-wise comparison results were used.

2.3.3. Survival Analysis

All survival analyses were performed using R/SURVIVAL. Kaplan-Meier log-rank test was used to assess statistical significance between our 63-gene signature, extracted from the commonalities between Waptag1 and human DCIS for prediction of poor prognosis (see Table 4.3), and 3 independent human breast cancer datasets[37,189,190]. The clinical endpoints for evaluation of reduced survival were either time to metastasis[37,189] or time to death[190] with follow-up limited to 10 years. Cox-hazard regression analysis was used to identify individual genes whose expression levels showed significant association with tumor free survival ($p \leq 0.01$) in three human breast cancer cohorts[37,189,190].

2.4. Retrotransposons

2.4.1. Datasets

2.4.1.1. Mouse Microarrays

Mouse Dataset1 consisted of forty-six single transgenic lactating and non-lactating mammary tumors (representing four transgenics: Myc, Neu, Wnt1, and Notch4) as well as FVB/NJ control mammary glands, with at a minimum of three-fold biological and two-fold technical replication, run on ninety-two oligonucleotide arrays. Arrays were spotted from the 22K Mouse Release 2.0 Oligo Library (Compugen, San Jose, CA) and labeled cDNA from each sample was hybridized in pairs, with the Universal Mouse Reference (Stratagene, La Jolla, CA), using a reference dye-swap design to reduce the bias due to the dye. Using the pooled permutation p-values from the Fs statistic of the mixed model, we determined which genes were statistically significant, using a multiple test correction for ANOVA analysis ($q \leq 0.01$). Only the ANOVA results across all transgenic tumors (compared with normal controls) will be discussed here. *Mouse Dataset2* consisted of Waptag1 data across all stages of tumor progression from NIA arrays (see above). *Mouse Dataset3* was comprised of the Affymetrix array results throughout all stages of Waptag1 tumor progression (see above).

2.4.1.2. Human Two Color Microarray Data

Human Dataset6 was obtained from publicly available data at GEO, the Gene Expression Omnibus [http://www.ncbi.nlm.nih.gov/geo/]. It featured 50 human breast tumors (32 luminal, 10 basal, and 9 ERBB2-positive) and four normal breast tissue samples. Luminal tumors are estrogen receptor positive (ER+) and androgen receptor positive (AR+); basal tumors are ER negative (ER-) and AR negative (AR-). Samples were run on cDNA microarrays harboring over 8,000 human genes including expressed sequence tags (GEO accession GDS84)[4,5]. Median-centered data was used; statistical significance was determined for each tumor type (luminal, basal, or ERBB2+) compared against normal breast using a two-tailed t-test ($p \leq 0.05$) analysis in GEO as the complete raw data was unavailable.

2.4.1.3. Human Affymetrix Data

Independent data from four human datasets was analyzed for comparison with Waptag1. Raw data was retrieved from GEO, unless otherwise stated. *Human Dataset7* featured 49 breast tumors (27 luminal, 16 basal, and 6 apocrine). Apocrine tumors are ER- and AR+; as mentioned above, luminal tumors are ER+ and AR+; basal tumors are ER- and AR- (GEO accession GDS1329)[6]. Samples were hybridized to Affymetrix GeneChip^R Human Genome U133 Array Set (Affy HG-U133A) chips and .CEL files were normalized with RMA[6]. Statistical significance was determined for each tumor type (luminal, basal, or apocrine) compared against each other using a two-tailed t-test ($p \leq 0.01$) analysis in GEO as the complete raw data was unavailable. *Human Dataset8* featured four fresh Atypical Ductal Hyperplasia (ADH) samples from women with a history of cancer compared against four samples from women with NO history of cancer (GEO accession GDS1250)[3]. Affy HG-U133A arrays were used; significance was determined for using a two-tailed t-test ($p \leq 0.05$) analysis in GEO as the complete raw data was unavailable. As mentioned above, *Human Dataset1* was received as rawdata (.CEL files) from collaborators, Dr. D. Craig Allred and Dr. Sanjun Lee, Washington University School of Medicine (St. Louis, MO). This dataset featured 27 samples enriched in DCIS and 24 Invasive Breast Cancer (IBC) samples run on Affy HG-U95Av2 arrays. *Human Dataset2* was downloaded as rawdata (.CEL files) from GEO; among other samples, it featured 7 normal breast samples run on Affy HG-U133plus2 arrays. Human Dataset1 and Dataset2 were combined after RMA pre-processing (using Affy cdf for pre-processing), as previously described (see above) and ANOVA results of statistically significant genes from the permuted Fs statistic, based on FDR multiple test adjustment ($q \leq 0.01$), were searched for retrotransposons.

2.4.2. Classification

All target sequences for the entire Affymetrix chip (either human or Waptag1) were run through RepeatMasker (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker); for the two color arrays (either human or mouse), only unknown or un-annotated sequences were queried against RepeatMasker database to obtain proper annotation of repeat sequences.

The main drawback of using microarray data to look for retrotransposons is that the target sequences must be queried; the actual array probes may or may not contain the repeat. However, our hypothesis was that retrotransposons are over-expressed early in the tumorigenic process, specifically within the tumor cells; the identification of exactly which elements was not our primary objective, but rather the similarity of mechanism of retrotransposon expression within early tumorigenesis.

2.5. Cross-Species Comparisons

Official mouse gene symbols and human / mouse gene orthology was determined using Mouse Genome Informatics (MGI) database (http://www.informatics.jax.org/), which features curated data with evidence codes and/or references for each orthologous match between mouse and human. Among the 16,026 unique mouse genes on the Affy 430 v2.0 arrays, 1689 had no human ortholog and 7392 were not found on the human arrays, whereas of 8065 unique human genes (common between the two human arrays), 513 had no human ortholog and 711 were not on the Affy 430 v2.0 arrays using the UMBA cdf. In total, 6945 genes were orthologous and represented on all microarrays, regardless of species. ANOVA pair-wise contrast results of human or Waptag1 DCIS or tumors (versus the appropriate species-specific normal controls) were compared between species. Similarities and differences among joined results were determined three ways: 1) genes that were common to both human and mouse; 2) genes that were only expressed in Waptag1 stages of tumor progression; 3) genes which were only expressed in human DCIS or IBC.

2.6. Gene Ontology and Pathway Analysis

2.6.1. VisuaL Annotation Display (http://proto.informatics.jax.org/prototypes/vlad-1.02/)

VisuaL Annotation Display (VLAD) was used to classify genes according to their Gene Ontology (GO) terminology[194] and (Richardson J and Bult C, unpublished). To determine the statistical significance of a gene list, VLAD uses a hyper-geometric statistical analysis, comparing one gene list to a universal set. Several universal sets were used for all analysis, but the MGI default universe will be described. For mouse datasets, the default is all genes annotated in MGI; for human genes, the default is all genes annotated in GOA_Human. VLAD was the source of all

GO classification results; the threshold set was $p<0.001$, unless otherwise stated; reduced threshold was $p<0.01$.

2.6.2. GenMAPP Energy Pathway Analysis (http://www.genmapp.org)

Using Entrez gene identifiers, statistically significant genes (from R/MAANOVA) were mapped using GenMAPP to specific metabolic pathways, in particular, those suggested by VLAD results as being significantly over-represented. Unlike VLAD, gene lists in GenMAPP were NOT compared against a universal set, as genes were merely mapped to pathways according to expression levels, because they were already determined to be significantly up or down-regulated compared with normal controls. Besides its flexibility and short learning curve, a user can choose and alter any color scheme according to their preferences in GenMAPP software. In addition, GenMAPP borrows curated pathway information from KEGG, providing a level of confidence to the user[195,196].

2.6.3. MouseCyc Pathway Representation (www.informatics.jax.org/pathways)

The Pathway Tools software development kit software (version 10.0) was downloaded from SRI and installed on a Sun Fire X4100 server running SUSE Linux, the dedicated host for the MouseCyc database (http://mousecyc.jax.org, ver. 1.36, June 2008; 19). Briefly, a catalog of mouse genes and their annotations were downloaded from MGI (November 6, 2007). Gene annotations included gene name and symbol, EC numbers, Gene Ontology annotations, genome coordinates and accession identifiers for Entrez Gene, UniProt, and MGI. RNA genes and pseudogenes were excluded. A total of 47 input files were created as input to the PathoLogic algorithm. Annotation files were created for 19 mouse autosomes, 2 sex chromosomes, the mitochondrial genome, and for genes with unknown chromosome location. Following the automated build of MouseCyc, the data-editing tools built into the PathwayTools software system are used for on-going manual refinement and annotation of biochemical pathways and reactions. Using MGI gene identifiers, statistically significant genes were mapped to classical and cellular metabolic pathways. MouseCyc features biochemical pathways curated from the wealth of data available in the Mouse Genome Informatics databases.

Chapter 3

CHARACTERIZATION OF WAPTAG1, A MOUSE MODEL OF DUCTAL CARCINOMA *IN SITU* WITH PROGRESSION TO INVASIVE CANCER

3.1. Waptag1 Tumorigenesis, a Multi-Stage Progressive Disease with Delayed-Onset

C57BL/6J-Tg(WapTAg)1Knw (Waptag1) mice contain a transgene comprised of the entire SV40 Tumor antigen early region (SV40 Tag; coding for all three proteins: Large Tag, small tag, and 17K tag) which is driven by the hormone-responsive whey acidic protein (WAP) promoter[50]. Waptag1-induced tumors develop mid-life in multiparous females, remarkably resembling the latent human condition[50]. Whereas virgins, single and dual pregnancy females failed to develop consistent, predictable mammary tumors, in females that sustained a minimum of three pregnancy and lactation cycles, tumors arose by 12-14 months of age. To visualize the earliest stages of tumorigenesis, several females were sacrificed at monthly intervals, starting at 4 months up until the development of tumors. Histological evaluation of samples revealed no abnormalities until 6-8 months of age when atypical cells were observed (Figure 3.1.A) in females that sustained a minimum of three pregnancy/lactation/involution cycles. Multiple stages of

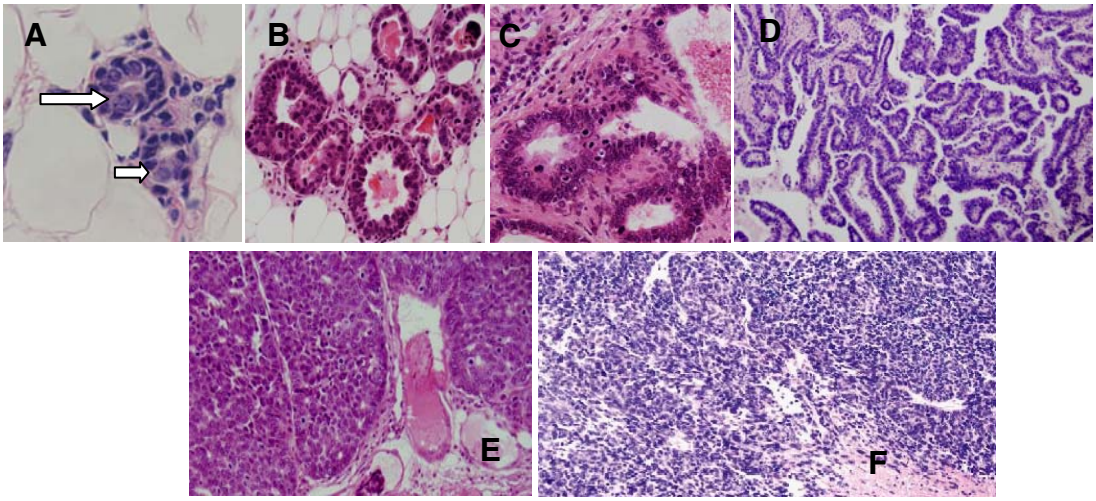


Figure 3.1. Tumor Progression in Waptag1 Mammary Carcinogenesis. In the initial stage of atypia **(A)**, the epithelium lining the mammary ducts contains a few scattered light atypical cells (white arrows) with large hyperchromatic nuclei. In more complex lesions **(B,C)**, ductal carcinoma *in situ* is depicted by intraductal neoplastic proliferation in which malignant cells stack multifocally in a disorderly fashion and occasionally form minute papillae; mild fibrosis and lymphocytic inflammation are observed. Papillary adenocarcinoma **(D)**, the primary tumor in Waptag 1 mice, is characterized by formation of papillary projections lined by cuboidal to columnar malignant epithelial cells, supported by a moderate amount of fibrovascular stroma. Solid / invasive carcinoma **(E)** is comprised of large solid areas with inconspicuous lumen and very little stroma**.** These tumors also exhibit invasion into the surrounding tissues. The myoepitheliomas **(F)** is comprised of solid areas of spindloid to polygonal cells supported by a scant amount of fibrovascular stroma. (Hematoxylin and Eosin stain. Magnification: A,100x; B,C,40x; D,E,F,10x.)

ductal carcinoma *in situ* (DCIS) were evident in females of 8-10 months of age (Figure 3.1.B, 3.1.C). Mammary tumors of one or more histological subtypes developed at 9-14 months: papillary adenocarcinomas (Figure 3.1.D), solid/invasive carcinomas (Figure 3.1.E), and rarely, myoepitheliomas (Figure 3.1.F).

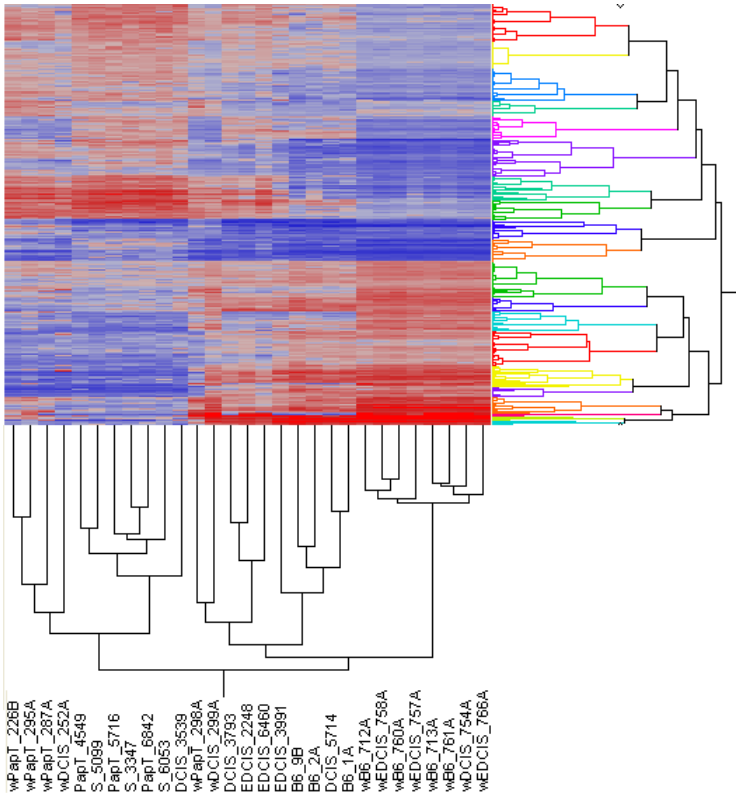


Figure 3.2. Hierarchical Clustering across Stages of Waptag1 Mammary Tumor Progression. **A)** Whole and LCM samples. **Right side:** genes with similar expression across all samples; **bottom:** how the samples cluster across genes, based on ANOVA expression values of the top 844 significant genes ($q \leq 0.00001$). Red, high expression; blue, low or negative expression. [Abbreviations: w, whole gland or tumor sample, if no "w" is listed, the sample is linearly amplified cells isolated using Laser Capture Microdissection (LCM); PapT, Papillary tumor; DCIS, ductal carcinoma *in situ*; S, solid tumor cells; EDCIS, early DCIS; B6, C57BL/6J controls.]

To fully characterize the potential of the Waptag1 mouse model, whole glands or tumors and laser-capture microdissected cells were collected throughout the lifespan of multiparous Waptag1 females. Age- and parity-matched C57BL/6J (B6) mammary glands or epithelial cells were collected as controls. Using these samples and an appropriate experimental design (see methods), we performed microarray analysis to investigate global gene expression changes in Waptag1 tumorigenesis. Figure 3.2 displays the unsupervised clustering of all samples, based on the ANOVA microarray results.

3.2. Waptag1 Whole Glands and Tumors Exhibit Stage-Specific Expression Profiles

Waptag1 whole mammary glands showed much variation *within* certain tumorigenic stages. Since histological classification of each stage was based on less than 1% of the mammary gland

(or tumor), multiple stages of tumor progression, not visible in the histological sample, may have been present in the sample collected for RNA. To better document the extent of within-stage variation, scatter plots were used to depict the mean expression values of one biological replicate plotted against another across all genes on the microarray. For comparison purposes, C57BL/6J samples (Figure 3.3.A) demonstrate how expression data scatter rather tightly around the y=x line, with very few values outside of the threshold ($\log_2$ of the fold changes between the two arrays, at greater than 1.5 times the standard deviation, across all arrays). Waptag1 Normal samples show some variation (Figure 3.3.B), but appear adequate as biological replicates. Conversely, as Figure 3.3.D and E show, the variation among the same tumorigenic stage can be large; the greatest deviation appears in Array2 of Waptag1 advanced DCIS (Figure 3.3.D). However, a sharp 45 degree (y=x) line is found with early DCIS replicates (Figure 3.3.C). Papillary tumor replicates display a moderate amount of variation, which is commonly observed when scattering expression data of one tumor against another within the same transgenic model (Fancher K, personal observation). With a large amount of within-stage variation, it may be difficult to unmask the across-stage genetic differences, even with a relaxed significance threshold. Indeed, this emphasizes the importance of scrutinizing specific cell types within each stage of tumor progression, thereby obtaining results based on only the specific DCIS or papillary tumor cells of interest.

3.2.1. Transcriptional Changes in Whole Glands (Early DCIS vs. C57BL/6J)

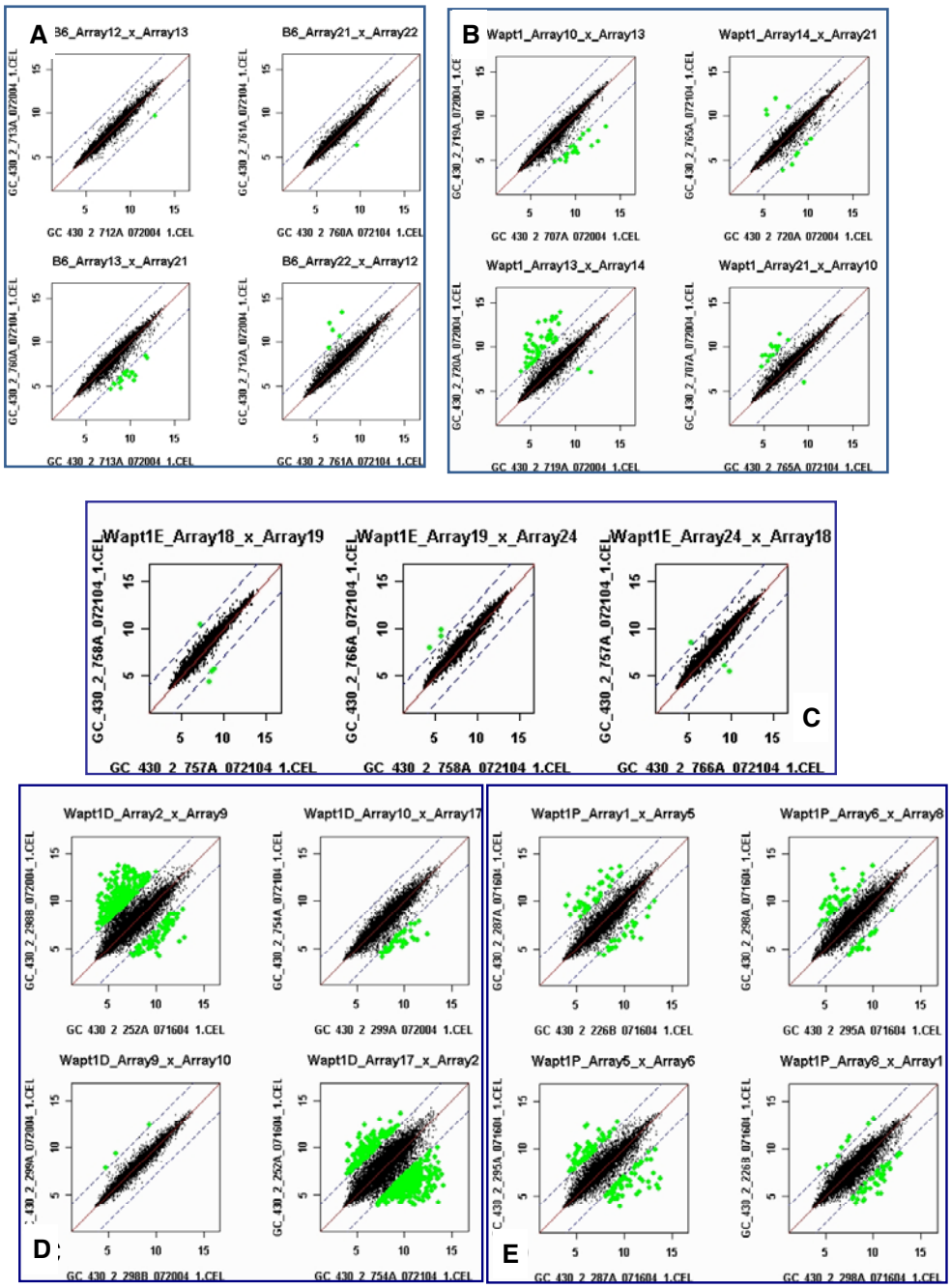
In the pair-wise contrast between whole early DCIS (EDCIS) and Waptag1 normal glands, only 73 genes were expressed at a significantly different level ($p \leq 0.05$). However, when EDCIS glands were compared with control C57BL/6J (B6) glands, 461 genes were significantly differentially expressed ($p \leq 0.05$). *Anti-apoptosis* and *negative regulation of apoptosis* genes were over expressed compared to B6 controls (Table 3.1). Thus, inhibition of apoptosis is one of the earliest changes occurring in Waptag1 tumorigenesis, probably due to dysfunctional TRP53 and PP2A, the consequence of the SV40 Tag transgene. *Regulation of immune system process* and *regulation of immune response* genes were down-regulated in whole glands containing EDCIS (Table 3.1). C57BL/6J inbred mice are normally 'high-responders' to SV40 Tag. Thus,

these findings probably reflect the fact that, like other SV40 Tag transgenic mice, Waptag1 mice are tolerant to SV40 Tag[197-200] and (Knowles BB, unpublished). Genes involved in *organ development*, specifically *organ morphogenesis* were up-regulated in whole EDCIS glands, compared with whole B6 glands, coinciding with the accepted paradigm that genes involved in development are often over expressed in cancer and/or function as oncogenes[201-210].

3.2.2. Transcriptional Changes in Whole Glands (Advanced DCIS vs. Early DCIS)

To determine gene alterations during progression to the next tumorigenic stage, each remaining stage was compared with the stage immediately preceding it, which could mask dominant gene alterations observed throughout tumor progression. In the pair-wise contrast between whole glands containing advanced DCIS (AdvDCIS) and EDCIS, 1323 genes were differentially expressed ($p \leq 0.05$). Genes involved in *cell cycle*, specifically *mitotic cell cycle*, *cell division* and *DNA replication, chromosome segregation* and *response to DNA damage stimulus* (in particular *DNA repair)* were up-regulated in AdvDCIS glands (Tables 3.1 and data not shown). Additionally, when analyzed by Affy cdf, retrotransposons were found in target sequences for 'genes' with the highest fold changes (see Chapter 5). *Monocarboxylic acid, oxidation reduction*, and *fatty acid metabolic processes*, *oxidoreductase activity*, as well as the cellular components *mitochondrion* and *peroxisome* were down-regulated in glands containing AdvDCIS vs. those with EDCIS (Tables 3.1 and Fancher K *et.al.*, future publication). Although defects in mitochondrial energy metabolism have been associated with disease and cancer, (for review[211]) these mutations usually result in generation of reactive oxygen species (ROS), though there are exceptions. Oxidative damage to lipids, proteins, and even DNA occurs under conditions of oxidative stress, hypoxia, and/or production and accumulation of ROS, and cells respond by attempting to neutralize ROS. Were this a response to stress, one would expect up-regulated, rather than down-regulated, expression of genes in these categories. Other genes with down-regulated expression in glands of AdvDCIS compared against EDCIS, fell into the following GO categories: *vasculature development, blood vessel development*, *blood vessel morphogenesis*, and *angiogenesis*. Genes in these categories do not overlap with other GO terms, however, so this finding was a bit perplexing, since one 'Hallmark of Cancer' is maintaining angiogenesis[11].

Figure 3.3. Scatter Plots across Biological Replicates of Control or Waptag1 Mammary Gland and Tumor Samples. A) Control C57BL/6J glands; **B)** Waptag1 'Normal' mammary gland samples; **C)** Waptag1 early DCIS samples; **D)** Waptag1 advanced DCIS samples; **E)** Waptag1 papillary tumors. Each set of scatter plots depicts the variation between any two arrays, based on RMA-preprocessed expression data. X- and Y-axes are normalized mean expression values for the 16,642 genes (no q-value limitation). Upper and lower boundaries (blue dashed lines) depict values 1.5 times the standard deviation, based on all arrays, with outliers (green spots) showing log 2 fold changes (between the two arrays) greater than this threshold.



3.2.3. Transcriptional Changes in Papillary Tumors vs. Whole Glands Containing Advanced DCIS

In the pair-wise contrast of whole papillary tumors versus glands containing AdvDCIS, 2160 genes were differentially expressed. Among 1114 up-regulated genes, *cell cycle/cell division*, *DNA replication, chromosome segregation,* and *DNA repair* genes were further over expressed (Table 3.1), while genes involved in *ATP binding,* and *DNA binding* were also over represented, at a reduced threshold (data not shown). Genes down-regulated in papillary tumors when compared with glands containing AdvDCIS were those in GO categories *carboxylic acid metabolic process, monocarboxylic acid metabolic process, lipid metabolic process, fatty acid metabolic process* and *oxidoreductase activity* (five GO terms); many of which were further down-regulated from the previous pairwise comparison (Table 3.1 and Fancher K *et.al.*, future publication). In addition, overabundance of down-regulated *electron carrier activity* genes in papillary tumors (Table 3.1) suggests that the most efficient method of ATP production is dysfunctional. However, ATP is still being consumed for cell cycle, cell division, and DNA replication processes.

In summary, analyses of entire mammary glands containing early stages of tumor progression are not ideal to understand the genetic changes occurring throughout tumorigenesis in Waptag1. High *within-stage* variation and low *between-stage* genetic changes made the discovery of the true, important stage-specific signatures complicated. However, several distinct and consistent transcriptional alterations seem to define each stage of tumor progression. Initially, in EDCIS mammary glands, anti-apoptotic genes first appear transcriptionally and regulation of immune response genes are down-regulated, compared with control glands. Mammary glands containing these sparsely-dispersed early nodules display one highly characteristic change associated with cancer (i.e. diminished apoptosis) and they appear to be negatively regulating immune responsiveness. Glands containing densely-packed regions of AdvDCIS appear to be in the commitment stage for tumorigenesis, exhibiting up-regulated expression of genes involved in cell cycle/division and DNA replication (Table 3.1), as well as chromatin modification genes, retrotransposons, and histone variants (see Chapter 5). In

addition, transcripts of genes regulating oxidoreductase activity were down-regulated. In papillary tumors, further up-regulation of genes involved in cell cycle/division, DNA replication, and DNA repair was apparent and immune response genes were over expressed. Interestingly, oxidoreductase activity, carboxylic acid metabolism, and lipid metabolic process genes were further down-regulated compared to glands containing AdvDCIS, and genes involved in electron carrier activity were also significantly down-regulated.

Table 3.1. Genes Representative of Stage-Specific Transcriptional Changes within the Most Significant Gene Ontology Classifications in Waptag1 Mammary Glands and Tumors. Each stage of progression exhibits a unique profile. Early DCIS results represent gene changes in glands containing Early DCIS (based on histological classification) compared to C57BL/6J control glands; Advanced DCIS, glands containing Advanced DCIS vs. those of Early DCIS; Papillary tumor, papillary tumor samples vs. glands containing Advanced DCIS. Fold changes determined by pairwise contrasts listed above and results were statistically significant ($p \leq 0.05$, ANOVA; $p \leq 1.0 \times 10^{-08}$, VisuaL Annotation Display). [Abbreviations: UP, Gene Ontology term for which gene expressed was up-regulated in the stage listed; DOWN, down-regulated.]

Stage of Tumorigenesis **GO_Category** / Gene symbol	**Gene Name**	**Significance:** fold change/**p-value**
Early DCIS		
Anti-apoptosis *UP*		**2.64×10^{-7}**
Akt1	thymoma viral proto-oncogene 1	1.78x
Cebpb	CCAAT/enhancer binding protein (C/EBP), beta	1.84x
Dapk1	death associated protein kinase 1	1.35x
Hspa1b	heat shock protein 1B	1.66x
Mapk8ip1	mitogen-activated protein kinase 8 interacting protein 1	1.39x
Spp1	secreted phosphoprotein 1	1.63x
Stat5a	signal transducer and activator of transcription 5A	1.55x
Trp63	transformation related protein 63	1.35x
Regulation of immune system process *DOWN*		**1.71×10^{-18}**
Btla	B and T lymphocyte associated	-5.06x
Cd19, Cd28, Cd4, Cd5, Cd55, Cd79b	CD# antigens	-1.33x to -6.32x
Fcer2a	Fc receptor, IgE, low affinity II, alpha polypeptide	-2.05x
H2-Oa	histocompatibility 2, O region alpha locus	-2.69x
Icosl	icos ligand	-1.56x
Il27ra, Il2rg	interleukin receptors	-2.01x, -1.94x
Lck	lymphocyte protein tyrosine kinase	-2.09x
Prkcq	protein kinase C, theta	-1.74x
Ptpn6, Ptprc	protein tyrosine phosphatases	-2.26x, -1.93x
Tnfrsf13c	tumor necrosis factor receptor superfamily, member 13c	-2.38x

Stage of Tumorigenesis **GO_Category** / Gene symbol	**Gene Name**	**Significance:** fold change / **p-value**
Advanced DCIS		
Mitotic cell cycle *UP*		**4.18 x 10^{-43}**
Aurka, Aurkb	aurora kinases	3.94x, 3.56x
Birc5	baculoviral IAP repeat-containing 5	8.79x
Bub1, Bub1b	budding uninhibited by benzimidazoles 1 homologs (S. cerevisiae)	3.85x, 3.60x
Ccna2, Ccnb1, Ccnb2, Ccnf	cyclins	4.3x, 7.1x, 6.2x, 2.2x
Cdc20, Cdc25b, Cdc25c, Cdc2a, Cdc6, Cdca2, Cdca3, Cdca5, Cdca8	cell division cycle homologs	1.69x to 6.02x
Fbxo5	F-box protein 5	2.52x
Hells	helicase, lymphoid specific	4.83x
Mad2l1	MAD2 (mitotic arrest deficient, homolog)-like 1 (yeast)	3.08x
Mtbp	Mdm2, transformed 3T3 cell double minute p53 binding protein	1.92x
Ncapd2, Ncapg2, Ncaph	non-SMC condensin complex, subunits	3.47x, 2.88x, 3.18x
Ndc80, Nuf2, Spc24, Spc25	NDC80 kinetochore complex component homologs (S. cerevisiae)	2.35x to 4.95x
Nek2	NIMA (never in mitosis gene a)-related expressed kinase 2	3.89x
Nusap1	nucleolar and spindle associated protein 1	5.02x
Plk1	polo-like kinase 1 (Drosophila)	3.39x
Rcc1, Rcc2	regulators of chromosome condensation	1.75x, 2.48x
Smc2, Smc4	structural maintenance of chromosomes	1.98x, 2.64x
Ube2c	ubiquitin-conjugating enzyme E2C	5.19x
DNA Replication *UP*		**2.23 x 10^{-26}**
Blm	Bloom syndrome homolog (human)	2.92x
Cdc45l, Cdc6	cell division cycle homologs (S. cerevisiae)	2.52x, 2.95x
Cdt1	chromatin licensing and DNA replication factor 1	3.5x
Chaf1b	chromatin assembly factor 1, subunit B (p60)	3.68x
Gins1	GINS complex subunit 1 (Psf1 homolog)	2.58x
Mcm2-7, Mcm10	minichromosome maintenance deficient genes	1.99x to 4.51x
Pola1,d1,e,e2	polymerases (DNA directed)	2.54x, 1.77x, 3.2x, 3.1x
Prim1, Prim2	DNA primase, p49 & p58 subunits	3.38x, 2.19x
Rad51, Rad51c, Rad54l	RAD51, 54 homologs (S. cerevisiae)	2.41x, 1.63x, 3.13x
Rfc2, Rfc3, Rfc4	replication factor C (activator 1) genes	1.67x, 2.49x, 3.08x
Rpa1, Rpa2, Rpa3	replication protein A genes	1.8x, 2.04x, 1.78x
Rrm1, Rrm2	ribonucleotide reductase M genes	2.5x, 7.02x
Tk1	thymidine kinase 1	4.04x
Oxidation Reduction & **Oxidoreductase Activity *DOWN***		**9.13 x $10^{-\ }$ &** **2.32x10^{-10}**
Acadl, Acads, Acox1	acyl-Coenzyme A dehydrogenases & oxidase1, palmitoyl	-1.58x, -1.51x, -2.14x
Adh1, Adhfe1, Aldh1a1, Aldh1a7, Aldh1l1, Aldh2, Aldh6a1, Aldh9a1	alcohol & aldehyde dehydrogenases	-1.68x to -8.3x
Aoc3, Aox1, Maob	amine, aldehyde & monoamine oxidases	-6.88x, -3.64x, -4.7x
Cdo1	cysteine dioxygenase 1, cytosolic	-4.85x
Cyp27a1, Cyp2e1, Cyp4b1, Cyp4v3	cytochrome P450, family members	-1.88x to -5.38x
Decr1	2,4-dienoyl CoA reductase 1, mitochondrial	-2.55x
Dpyd, Ehhadh, Gpd1, Hsd11b1, Hsd3b7, Hsdl2, Idh1, Pdha1, Xdh	dehydrogenases (miscellaneous)	-1.6x to -3.95x
Fasn, Nos3	fatty acid & nitric oxide synthases	-2.58x, -1.82x
Fmo1, Fmo5	flavin containing monooxygenase	-2.59x -2.69x
Heph	hephaestin	-4.37x
Loxl1	lysyl oxidase-like 1	-2.06x
Mod1	malic enzyme, supernatant	-2.49x
Por	P450 (cytochrome) oxidoreductase	-2.97x
Sc5d, Scd1	sterol-C5 & stearoyl-CoA desaturases	-2.25x, -3.92x
Sod3	superoxide dismutase 3, extracellular	-3.51x

Table 3.1. Continued.

Stage of Tumorigenesis GO_Category / Gene symbol	**Gene Name**	**Significance:** fold change / **p-value**
PAPILLARY TUMOR		
Immune Response *UP*		**1.67 x 10^{-10}**
B2m	beta-2 microglobulin	1.51x
Bcl3	B-cell leukemia/lymphoma 3	1.65x
C1qb, C1qg	complement component 1, q subcomponents	2.03x, 2.04x
Ccl2,3,4,5, Ccl7,8, Ccl12, Cxcl2, Cxcl10, Cxcl16	chemokine (C-C and C-X-C motif) ligands	1.53x to 5.98x
Fcer1g, Fcgr1	Fc receptor, IgE or IgG, high affinity I	2.09x, 2.65x
H2-Aa, H2-Ab1, H2-D1	histocompatibility 2, class II antigen A or D region locus 1	2.24x, 2.53x, 1.64x
Il1b, Il1rn, Il4ra	interleukin genes	1.81x, 2.5x, 1.66x
Irf7	interferon regulatory factor 7	4.57x
Ly86, Ly96	lymphocyte antigen 86 and 96	3.23x, 1.73x
Mx1, Mx2	myxovirus (influenza virus) resistance genes	3.62x, 1.63x
Pglyrp1	peptidoglycan recognition protein 1	7.5x
Tlr2, Tlr3, Tlr13	toll-like receptors	1.85x, 2x, 1.9x
Cell Division *UP*	***genes listed above not shown, even if further up-regulated***	**1.1 x 10^{-30}**
Ccne2, Cdk2, Cdkn2a	cyclin E2, cyclin-dependent kinase 2 & inhibitor	4.59x, 1.72x, 2.59x
Cdc14a, Cdc45l, Cdca4	cell division cycle homologs	1.83x, 2.63x, 1.58x
Cep55	centrosomal protein 55	2.77x
Dsn1, Nsl1	MIND kinetochore complex components, homologs	2.45x, 2.26x
Kif11	kinesin family member 11	2.66x
Kntc1, Zwilch	kinetochore associated genes	2.16x, 3.85x
Lig1	ligase I, DNA, ATP-dependent	2.49x
Mis12	MIS12 homolog (yeast)	1.88x
Plk1, Prc1	polo-like kinase 1 (Drosophila) & protein regulator of cytokinesis 1	2.83x, 4.37x
Racgap1	Rac GTPase-activating protein 1	4.3x
Sept5, Sept9	septins 5, 9	2.88x, 1.53x
Sgol1, 2	shugoshin-like 1, 2 (S. pombe)	3.2x, 3.41x
Spc24, 25	spindle pole component 24, 25 homologs (S. cerevisiae)	3.66x, 2.9x
Syce2	synaptonemal complex central element protein 2	2.57x
Timeless	timeless homolog (Drosophila)	2.1x
Top2a	topoisomerase (DNA) II alpha	4.27x
Electron Carrier Activity *DOWN*	***genes listed above not shown, even if further down-regulated***	**1.25 x 10^{-08}**
Acad11, Acadsb, Acox1	acyl-Coenzyme A dehydrogenase & oxidase genes	-2.31x, -2.15x, -1.98x
Aifm2	apoptosis-inducing factor, mitochondrion-associated 2	-2.72x
Cox6a2, Cox7a1, Cox7a2l	cytochrome c oxidase, subunits	-3.18x, -2.06x, -1.5x
Cyp1a1, Cyp1b1	cytochrome P450, family 1, subfamilies a,b polypeptide 1	-2.16x, -3.12x
Dhdh, Ivd, Sdha	dehydrogenases (dihydrodiol, isovaleryl Co-A, succinate)	-2.55x, -2.5x, -1.68x
Etfb	electron transferring flavoprotein, beta polypeptide	-1.6x
Glrx5	glutaredoxin 5 homolog (S. cerevisiae)	-1.6x
Ndufs1, Ndufs8, Ndufv2	NADH dehydrogenase (ubiquinone) genes	-1.7x, -1.52x, -1.51x
Uqcrfs1	ubiquinol-cytochrome c reductase, Rieske iron-sulfur polypeptide 1	-1.79x

Table 3.1. Continued.

3.3. Laser Capture Microdissected Tumor Cell Transcriptomes are Nearly Identical, Irrespective of Waptag1 Tumorigenic Stage

In whole gland samples (described above), it is possible that multiple stages of tumor progression, not seen in the histological sample, may have been present in the sample taken for RNA and processed on the microarray since these samples were classified based on less than 1% of the mammary gland (or tumor) that was collected. Variation among biological replicates may have affected the results of our analysis. Therefore, to circumvent this potential problem, we

used laser capture microdissection to isolate specific cell types from each stage of Waptag1 tumor progression. To obtain sufficient quantities of RNA from the laser captured cells, however, linear amplification was performed using the RNA from ~100 cells. Linear amplification was not performed on the whole gland samples; therefore this is a substantial difference in the procedures between whole glands and laser-captured cells prior to their analysis on microarrays.

3.3.1. Laser Capture Microdissection(LCM) of Mammary Samples

Results from one of three biological replicates for each stage of Waptag1 tumor progression are shown in Figure 2.1. Variation was minimal across biological replicates. Unlike the whole mammary glands, laser-captured cells exhibit very little within-stage variation, with scatter plots appearing similar to whole papillary tumor replicates (Figure 3.2. and data not shown). The majority of this variation was due to the cells comprising the DCIS lesion appearing more tumor-like or more B6-like than could be determined from histological classification for the stage of progression. This is illustrated in the hierarchical clustering across all LCM samples (Figure 3.4.), in which DCIS sample #3539 clustered with the LCM cells from papillary and solid tumors, whereas DCIS sample #5714 grouped with the B6 cells.

3.3.2. A Tumor Profile Within LCM Cells of DCIS

Among the very early changes in Waptag1 cells, i.e. comparing LCM cells of EDCIS with LCM control B6 epithelial cells, genes with up-regulated expression were involved in the following processes: *cell cycle*, specifically *mitotic cell cycle* and *cell cycle checkpoint*; *cell division*, *DNA replication*, *DNA repair*, and *chromosome organization and biogenesis*. These findings, though familiar gene categories, were not observed at this early stage in whole glands. The genetic profile of these specific cells within EDCIS is comparable to those obtained from Waptag1 whole tumors. Many of these alterations are due to the functions of SV40 Tag proteins, thus confirming isolation of the appropriate cells during LCM. In addition, retrotransposons and twenty-eight genes categorized under the classification of *chromosome organization and biogenesis* (2.02 x 10-5; 28 of 223 genes) including histone variants *H2afx*, *H2afz*, and *Hist1h4i*, the histone deacetylase, *Hdac10*, and the chromatin remodeling factors *Ezh2*, *Sirt7* and *Smarcd2* were over expressed in LCM EDCIS cells. Thus, deregulation of chromatin maintenance machinery and

transcriptional activation of retrotransposons occur *early* and *within* the specific pre-tumor cells of Waptag1 mammary glands.

Genes with down-regulated expression in LCM EDCIS cells were categorized as energy-related. *Generation of precursor metabolites and energy*, specifically, *aerobic respiration*, *tricarboxylic acid cycle*, and *electron transport* as well as *coenzyme catabolic process, oxidoreductase activity*, *electron carrier activity*, and the related cellular component *mitochondrion*, *mitochondrial envelope*, and *mitochondrial inner membrane* genes were down-regulated, compared with LCM B6 cells. These results suggest that the cancer cell switch *from* normal mitochondrial respiration, ATP production via electron transport chain (which produces 36 ATP/glucose molecule) *to* less efficient glycolysis/fermentation (the breakdown of glucose into

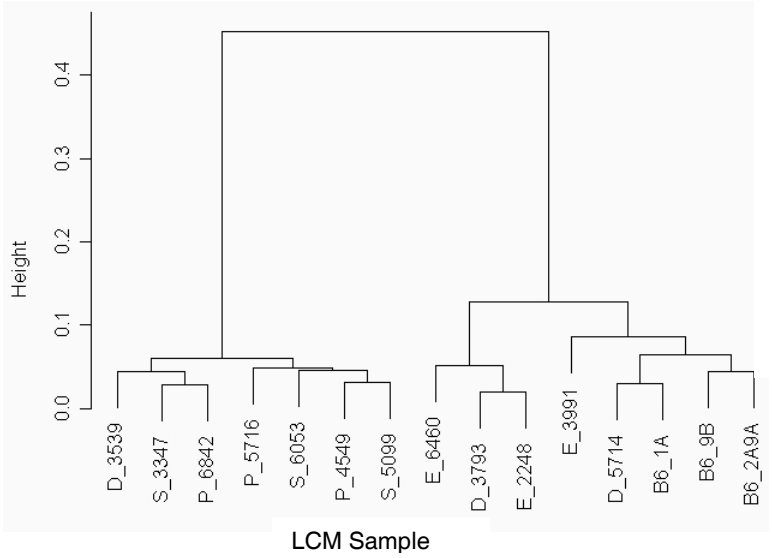


Figure 3.4. Hierarchical Clustering across Laser-Captured Stages of Waptag1 Mammary Tumor Progression. Clustering based on normalized median expression values after RMA (no q-value limitation). [Abbreviations: laser capture microdissection (LCM); P, papillary tumor; D, ductal carcinoma *in situ*; S, solid tumor cells; E, early DCIS; B6, C57BL/6J controls.]

lactic acid, which produces only 2 ATP/glucose) even in the presence of oxygen, as hypothesized by Warburg and described by others[12,211-214] occurs *early* in the tumorigenic process. Indeed, phosphofructokinase *Pfkm*, which controls glycolysis, is up-regulated in LCM EDCIS cells, thereby supporting glycolysis as a means by which these cells obtain ATP when electron transport genes are down-regulated.

Based on the laser-capture early DCIS cell results, once fat, normal epithelium and stromal cells (including lymphocytes and other immune cells) are excluded, LCM EDCIS cells and papillary tumors are quite similar. In fact, transcriptional differences (genes up- or down-

regulated) in LCM EDCIS cells versus B6 control cells were further deregulated in LCM Papillary tumor cells versus LCM EDCIS cells.

In the pair-wise contrast of LCM AdvDCIS with LCM EDCIS cells, only a couple of hundred genes were differentially up- or down-regulated and no GO categories were significantly represented at the specified threshold for VLAD ($p<0.001$). In light of this, the data from LCM AdvDCIS cells was compared against LCM B6 control cells and the GO categories were screened for categories not found in the EDCIS versus B6 comparison described above. In AdvDCIS cells, no additional GO categories were observed among genes up-regulated. Genes down-regulated in LCM AdvDCIS cells, but not in EDCIS cells, were categorized as involved in *lipid* and *fatty acid metabolic processes*, *fatty acid oxidation*, and *FAD binding*. *Fatty acid metabolism* and *oxidoreductase activity* (mentioned above) gene changes are indicative of a cellular stress response. However, the organelles which harbor enzymes to oxidize molecules, catalyze the generation of energy, and neutralize reactive oxygen species, are under-expressed in cells of AdvDCIS: the *mitochondrion* (also introduced above) and the *peroxisome*. Thus, in Waptag1 tumorigenesis, either reactive oxygen species are not being generated or if present, they are not being neutralized.

3.3.3. LCM Papillary Tumor Genetic Changes

In the pairwise contrast of LCM Papillary tumor with LCM AdvDCIS cells, the same GO categories revealed further up- and down-regulation, as both the numbers expressed and the magnitude of expression of genes in these categories were increased. Besides changes observed in EDCIS and AdvDCIS, when LCM papillary tumor cells were compared with B6 control epithelial cells, significant changes were in the GO categories *mRNA metabolic process*, in particular *mRNA processing,* as well as *RNA splicing*, *RNA transport, specifically mRNA transport*, and related cellular component, *spliceosome*. This suggests that aberrant mRNA metabolism / processing may be a factor contributing to tumor progression. Additionally, up-regulation of *DNA-dependent ATPase activity*, *DNA binding, helicase activity,* and *DNA packaging* genes, supports the massive up-regulation of DNA replication previously observed in whole papillary tumors. Considering SV40 Tag proteins contain both ATPase and helicase

activities, however, some of these results may be effects of SV40 Tag expression. Regardless, they undoubtedly play a role during tumorigenesis in Waptag1.

Alongside the aforementioned gene classifications, in LCM papillary tumor cells, *organ morphogenesis* and *vasculature development*, specifically *blood vessel development* and in particular, *blood vessel morphogenesis* genes were down-regulated, similar to a previous finding in whole glands containing AdvDCIS (see above). As stated previously, this is a perplexing finding since increased vasculature is anticipated to promote invasion during carcinogenesis.

3.3.4. LCM Solid, Invasive Tumor Cell-Specific Alterations

Regrettably, whole solid, invasive tumors were not included as part of the original whole glands/tumors put on Affymetrix arrays. However, these solid tumor cells were captured to complete the profile of gene expression throughout all stages of Waptag1 tumor progression. LCM solid, invasive tumor cells were markedly similar to LCM papillary tumor cells: only a few hundred genes, and consequently no GO categories, showed up-regulation in the pair-wise contrast between these two. The only new GO term, in addition to those discussed above, with genes significantly up-regulated in LCM solid tumor cells versus B6 control cells was *RNA localization*, which, when combined with the *mRNA processing, RNA splicing,* and *RNA transport* cited above, suggests extensive RNA alterations are prevalent within these late-stage tumor cells, such as processing of the mRNA and then export out of the nucleus, for example.

The LCM solid, invasive tumor cells, when compared with B6 cells, exhibited down-regulation of genes in the GO categories *glutathione transferase activity*, *lyase activity*, *vitamin metabolism*, in particular, *water-soluble vitamin metabolism*, which happen to fall under the oxidative stress response. Interestingly, vitamins are required for normal metabolism within cells; in addition, antioxidants, such as vitamins A, C & E, can reduce the potential damage to a cell. Therefore, it seems complimentary to find down-regulation of vitamin metabolism alongside energy metabolic process and oxidative stress response genes. *Sulfur metabolic process* genes were decreased as well. Thus, even within the late stage cells of Waptag1 tumorigenesis inhibition is maintained.

These analyses 1) confirmed changes previously uncovered in whole glands, although many transcriptional alterations were realized in advance in LCM EDCIS cells; 2) identified up-

regulation of transposable elements and chromatin modification genes within specific pre-tumor cells of Waptag1 mammary glands; 3) demonstrated the unexpected similarity of pre-tumor and tumor cells when surrounding stroma, fat, etc. are removed, leaving only an oncogene-induced tumor-like profile.

3.4. Stroma-Related Transcriptional Changes

Stromal cells *were* captured at the time of LCM tumor cell collection. Unfortunately, insufficient numbers of cells, inadequate biological replicates, and technical obstacles prohibited satisfactory analysis of stromal cells across all stages of Waptag1 tumor progression. Therefore, to dissect potential stromal changes, which contribute to the 'cancer phenotype' in Waptag1, from whole gland results, we used a subtraction approach. Significant genes from two pairwise contrasts, whole AdvDCIS versus whole B6 glands and LCM AdvDCIS versus LCM B6 cells , were filtered for common genes with the same directionality. The remaining 2179 genes were analyzed for enrichment of specific GO terms; these represent stroma-contributed gene expression changes. Table 3.2 lists the significantly over represented GO categories ($p<0.0001$) among the expression changes of the 2179 'stromal' genes; genes over expressed in whole glands of AdvDCIS as well as under expressed are classified. *Cell adhesion* genes were the most significantly up-regulated in Waptag1 glands containing DCIS compared to B6 glands, but this was not the case in LCM DCIS cells (Table 3.2 and 3.2.1). Additionally, 'stromal' genes showed enrichment in *actin cytoskeleton components*, *extracellular matrix structural constituents*, *collagen* family members, *cytoskeleton organization and biogenesis*, and were part of the *apical junction complex* (Table 3.2). Although these findings might be anticipated, they confirmed the appropriate techniques and methods used for laser-captured cells. Cell to matrix or *extracellular matrix* genes, specifically those increased in whole glands containing AdvDCIS, are listed in Table 3.2.1. The level of significance of *(mono)carboxylic acid metabolism* genes among down-regulated 'stromal' genes (Table 3.2) suggests that some of the energy-related metabolic changes may reside within cells of the stroma, rather than exclusively within pre-tumor and tumor cells. Genes in *blood vessel* and *vasculature developmental* categories were also predominantly down-regulated in AdvDCIS whole gland samples (Table 3.2), compared with whole C57BL/6J

glands, a consistent recurring finding. Finally, *immune system* genes were a significant part of the 'stroma-related' changes (Table 3.2 and 3.2.2) as might be expected when comparing whole glands with laser-captured cells.

Table 3.2. Early Stroma-Related Gene Ontology Classifications. GO categorization of 2179 genes up- or down-regulated in whole glands containing AdvDCIS versus whole C67BL/6J glands ($p \leq 0.0001$, VLAD). Genes were directionally different or absent in LCM AdvDCIS versus LCM B6 cells. [Abbreviations: GO, Gene Ontology; BP, Biological Process, MF, Molecular Function, CC, Cellular Component; VLAD, VisuaL Annotation Display software. A full list of genes and expression values will appear in Fancher K *et.al.* future publication.]

GO Ontology	GO ID	GO Classification	p-value	# genes	# in MGI
BP	GO:0007155	cell adhesion	7.33E-15	123	575
BP	GO:0022610	biological adhesion	7.33E-15	123	575
BP	GO:0065007	biological regulation	3.77E-08	517	4028
BP	GO:0008152	metabolic process	1.01E-07	752	6204
BP	GO:0050789	regulation of biological process	1.28E-07	469	3635
BP	GO:0016043	cellular component organization and biogenesis	7.91E-07	298	2194
BP	GO:0044237	cellular metabolic process	1.15E-06	680	5616
BP	GO:0050794	regulation of cellular process	1.60E-06	428	3344
BP	GO:0044238	primary metabolic process	2.98E-06	670	5557
BP	GO:0048856	anatomical structure development	3.88E-06	247	1799
BP	GO:0050793	regulation of developmental process	4.08E-06	113	708
BP	GO:0032787	monocarboxylic acid metabolic process	6.04E-06	45	216
BP	GO:0048523	negative regulation of cellular process	1.45E-05	132	878
BP	GO:0019752	carboxylic acid metabolic process	1.48E-05	72	415
BP	GO:0006082	organic acid metabolic process	1.61E-05	72	416
BP	GO:0048519	negative regulation of biological process	1.88E-05	141	955
BP	GO:0001568	blood vessel development	3.09E-05	39	189
BP	GO:0048731	system development	3.65E-05	220	1625
BP	GO:0001944	vasculature development	4.47E-05	39	192
BP	GO:0048518	positive regulation of biological process	7.10E-05	135	932
BP	GO:0007010	cytoskeleton organization and biogenesis	7.48E-05	73	443
BP	GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	7.95E-05	32	150
BP	GO:0002376	immune system process	8.59E-05	105	694
BP	GO:0048646	anatomical structure formation	8.66E-05	37	184
MF	GO:0005488	binding	8.42E-19	1258	10373
MF	GO:0005515	protein binding	1.03E-18	685	4943
MF	GO:0001871	pattern binding	1.03E-06	26	89
MF	GO:0030247	polysaccharide binding	1.13E-06	25	84
MF	GO:0005539	glycosaminoglycan binding	2.79E-06	23	77
MF	GO:0008201	heparin binding	3.32E-06	20	62
MF	GO:0005201	extracellular matrix structural constituent	5.97E-06	19	59
MF	GO:0003824	catalytic activity	1.31E-05	609	5016
MF	GO:0030020	extracellular matrix structural constituent conferring tensile strength	1.80E-05	12	29
MF	GO:0005509	calcium ion binding	6.05E-05	120	804
MF	GO:0030246	carbohydrate binding	7.36E-05	50	272

GO Ontology	GO ID	GO Classification	p-value	# genes	# in MGI
CC	GO:0044424	intracellular part	7.49E-08	1041	8607
CC	GO:0015629	actin cytoskeleton	1.78E-07	46	192
CC	GO:0005622	intracellular	1.82E-07	1072	8926
CC	GO:0005737	cytoplasm	2.57E-06	691	5547
CC	GO:0005581	collagen	3.09E-06	15	37
CC	GO:0030017	sarcomere	6.01E-06	22	73
CC	GO:0044449	contractile fiber part	1.24E-05	22	76
CC	GO:0030016	myofibril	1.40E-05	23	82
CC	GO:0032432	actin filament bundle	1.51E-05	10	20
CC	GO:0043292	contractile fiber	2.15E-05	23	84
CC	GO:0043226	organelle	2.33E-05	862	7174
CC	GO:0043229	intracellular organelle	2.63E-05	861	7170
CC	GO:0044420	extracellular matrix part	5.03E-05	24	94
CC	GO:0043296	apical junction complex	7.05E-05	23	90
CC	GO:0001725	stress fiber	7.05E-05	9	19
CC	GO:0005578	proteinaceous extracellular matrix	8.11E-05	52	280

Table 3.2. Continued.

3.4.1. Cell Adhesion Genes

Up-regulation of genes involved in *cell adhesion*, with 123 genes differentially expressed in whole Waptag1 DCIS tissue compared with whole B6 glands (and absent in LCM cells), was the most significant finding (Table 3.2.1). GO terms related to cell adhesion, such as genes that belong to the cellular components *extracellular matrix part* and *collagen* were also significantly expressed (Table 3.2). These findings, though anticipated, confirm and verify the use of the

8430419L09Rik	*Cd97*	*Col5a1*	*Dst*	*Lgals1*	*Pcdhga4*	*Pecam1*	*Stab2*
Adam23	*Cdh1*	*Col5a3*	*Emilin2*	*Lgals7*	*Pcdhga5*	*Perp*	*Sympk*
Amigo2	*Cdh16*	*Col6a1*	*Epdr1*	*Lypd3*	*Pcdhga6*	*Pkd1*	*Tcfap2a*
Barx2	*Celsr1*	*Col6a2*	*F8*	*Mia1*	*Pcdhga7*	*Pkp1*	*Thbs4*
Bcl10	*Ceacam1*	*Col8a1*	*Fat1*	*Mpdz*	*Pcdhga8*	*Pkp3*	*Tnc*
Bcl2l11	*Cfdp1*	*Col8a2*	*Fat4*	*Muc4*	*Pcdhga9*	*Ptk7*	*Tnxb*
Boc	*Cldn1*	*Col9a1*	*Fbln5*	*Mybpc2*	*Pcdhgb1*	*Ptprc*	*Ttyh1*
Cd164	*Cldn3*	*Col9a2*	*Glycam1*	*Nrp1*	*Pcdhgb2*	*Ptprf*	*Tyro3*
Cd2	*Clstn3*	*Cpxm1*	*Hectd1*	*Omd*	*Pcdhgb4*	*Pvrl4*	*Vtn*
Cd209d	*Cntn1*	*Ctgf*	*Igfals*	*Pcdh1*	*Pcdhgb5*	*Pxn*	*Vwf*
Cd22	*Cobl*	*Ctnna1*	*Igsf5*	*Pcdh21*	*Pcdhgb6*	*Ret*	*Wisp1*
Cd24a	*Col11a1*	*Cx3cl1*	*Itga6*	*Pcdhga1*	*Pcdhgb7*	*Rnd1*	*Zyx*
Cd34	*Col12a1*	*Cxadr*	*Itgb8*	*Pcdhga10*	*Pcdhgb8*	*Rnd3*	
Cd44	*Col15a1*	*Cyr61*	*Jam3*	*Pcdhga12*	*Pcdhgc3*	*Sell*	
Cd9	*Col16a1*	*Ddr1*	*Jam4*	*Pcdhga2*	*Pcdhgc4*	*Sorbs1*	
Cd96	*Col4a5*	*Dsg2*	*Lama4*	*Pcdhga3*	*Pcdhgc5*	*Sox9*	

Table 3.2.1. Cell Adhesion Genes, Stroma-Related. Of the 2179 gene changes found in whole AdvDCIS lesions, which were not present in LCM AdvDCIS cells, 123 genes were classified as *cell adhesion* genes ($p=7.33 \times 10^{-15}$, VLAD), the majority of which were up-regulated. Official mouse gene symbols were obtained from the Mouse Genome Informatics (www.informatics.jax.org).

laser-capture microdissection techniques and methods to isolate and analyze control epithelial and tumor cells. Cell to matrix or extracellular matrix genes specifically increased in whole glands containing AdvDCIS were as follows: cadherins, *Chd1*, *Cdh16*, and the related cadherin EGF LAG seven-pass G-type receptor 1, *Celsr1*; catenin (cadherin associated protein), alpha 1, *Ctnna1*; protocadherins, *Pcdh1*, *Pcdh21*; CEA-related cell adhesion molecules, *Ceacam1*, *Ceacam20*; claudins *Cldn1*, *Cldn3*, *Cldn23*; integrins *Itga6*, *Itgb8*; collagens, in particular type IV family members, *Col4a5*, *Col4a6*; fibrillin, *Fbn2*; matrix metallopeptidases, *Mmp12*, *Mmp13*, *Mmp14*; tenascin C, *Tnc*; tight junction proteins, *Tjp2*, *Tjp3*. Additional over expressed genes that appear to be stroma-related include: stromal cell-derived factor 2-like 1, *Sdf2l1*; engulfment and cell motility 3, ced-12 homolog, *Elmo3*; endothelial cell-specific molecule 1, *Esm1*.

3.4.2. Immune System Genes

Immune system genes were significantly over-represented in whole Waptag1 AdvDCIS glands compared with whole B6 samples (Table 3.2). In total, about 100 *immune system process* transcripts were differentially expressed in whole glands that were not present in LCM cells. Several related GO terms were just below the cutoff threshold shown in Table 3.2 including *antigen processing and presentation*, specifically MHC Class II, and *leukocyte migration*. Selected immune genes up-regulated greater than 1.5x in whole glands include the following: B-cell leukemia/lymphoma gene, *Bcl11a*; chemokine (C-C motif) ligand 12, *Ccl12*; CD24a antigen,

2210010L05Rik	*Cblb*	*Cd79b*	*Ercc1*	*Igl-C1*	*Msh6*	*Tap2*
Abcc9	*Cblc*	*Cfb*	*Faim3*	*Igl-C2*	*Myd88*	*Tbx1*
Ada	*Ccl12*	*Cfp*	*Fcgrt*	*Igl-V1*	*Ndrg1*	*Tgfb2*
Akt1	*Ccl24*	*Clec4d*	*G6pdx*	*Il15*	*Nup85*	*Tlr2*
Alas2	*Ccl6*	*Coro1a*	*H2-Aa*	*Il16*	*Pglyrp1*	*Tnfrsf13c*
Bank1	*Ccl5-6*	*Cr2*	*H2-Ab1*	*Il18*	*Plscr1*	*Tnfsf12*
Bcl10	*Ccl9*	*Cx3cl1*	*H2-DMb1*	*Il18r1*	*Prkcd*	*Tnfsf13*
Bcl11a	*Ccr7*	*Cxcl12*	*H2-DMb2*	*Itga6*	*Ptprc*	*Tnfsf9*
Bmp4	*Cd14*	*Cxcl13*	*H2-Eb1*	*Jag2*	*S100a9*	*Trp53*
Btla	*Cd24a*	*Cxcl15*	*H2-Oa*	*Jmjd1a*	*Satb1*	*Trpm4*
C1r	*Cd27*	*Cxcl9*	*H2-Ob*	*Kit*	*Scye1*	*Tshr*
C1s	*Cd300lg*	*Cxcr4*	*Hells*	*Ltb*	*Sh2b2*	*Xrcc6*
C2	*Cd34*	*Dyrk3*	*Id2*	*Lyst*	*Sigirr*	*Zbtb16*
Casp3	*Cd55*	*Egr1*	*Igbp1*	*Ms4a1*	*Sp1*	
Cbfb	*Cd74*	*Epas1*	*Ighg*	*Msh2*	*Tal1*	

Table 3.2.2. Immune System Genes, Stroma-Related. Of the 2179 gene changes found in whole AdvDCIS lesions, which were not present in LCM AdvDCIS cells, 103 genes were classified as involved in the *immune system process* ($p=8.59 \times 10^{-05}$, VLAD). Official mouse gene symbols were obtained from the Mouse Genome Informatics (www.informatics.jax.org).

Cd24a; chemokine (C-X3-C motif) ligand or receptor, *Cx3cl1*, *Cxcr4*; early growth response 1, *Egr1*; integrin alpha 6, helicase, lymphoid specific, *Hells*; *Itga6*; Immunoglobulin heavy chain (gamma polypeptide), *Ighg*; immunoglobulin-like domain containing receptor 1, *Ildr1*; mutS homologs, *Msh2*, *Msh6*; nucleoporin, *Nup85*; S100 calcium binding protein A9 (calgranulin B), *S100a9*; transforming growth factor, beta 2, *Tgfb2*; toll-like receptor 2, *Tlr2*. Unconventional genes, showing substantially increased expression in AdvDCIS whole glands, which did not appear in Table 3.2.2 were E26 avian leukemia oncogene 2, 3' domain, *Ets2*; the T-cell differentiation protein 2, *Mal2*; leukocyte cell derived chemotaxin 1, *Lect1*; lymphocyte antigen 6-related genes, *Ly6e*, *Lypd3*; T-cell, immune regulator, *Tcirg1*; T-cell immunoglobulin and mucin domain containing 2, *Timd2*.

3.5. Summary

Microarray analysis was used as a data-driven, discovery approach to uncover transcriptional changes associated with the early stages of tumorigenesis in a mouse model of human breast cancer, Waptag1. For each stage of tumor progression in Waptag1 whole glands and tumors, a unique transcriptional signature was apparent. In whole glands containing EDCIS, anti-apoptosis genes first appeared transcriptionally, while genes regulating immune response were down-regulated compared with normal glands. In AdvDCIS glands, genes involved in cell cycle/division and DNA replication, as well as retrotransposons and histone variants were over expressed, whereas oxidoreductase activity genes were decreased, when compared with EDCIS. In whole papillary tumors, cell cycle/division, and DNA replication genes were further up-regulated, when compared with AdvDCIS glands, as well as immune response and DNA repair genes; alternatively, oxidoreductase activity genes were further down-regulated than in AdvDCIS glands and carboxylic acid and lipid metabolic process genes were also down-regulated. Contrary to stage-specific gene expression profiles found in whole glands and tumors, only minute step-wise expression changes could be found in laser capture microdissected (LCM) cells. Once fat, normal epithelium and stromal cells (including immune cells) were removed, leaving only transcriptional changes from LCM pre-neoplastic EDCIS or AdvDCIS epithelial cells lining the ducts, cells strikingly resembled whole papillary tumor results. Amplified LCM cells exhibited a

global tumor transcriptional profile, i.e. abundant transcription of genes involved in cell cycle, cell division, DNA replication, and DNA repair as well as transcription of chromatin modification genes and retrotransposons. Gene categories that showed significant down-regulated genes expressed included aerobic respiration, electron transport, and oxidoreductase activity genes, even as prematurely as EDCIS. These findings suggest that at the cellular level, a pre-tumor cell is predominantly a tumor cell; therefore, in the case of SV40 Tag-induced tumors, the surrounding tumor cell microenvironment presents the stage-specific transcriptional alterations. Comparison of whole glands with microdissected samples at the same stage of tumor progression confirmed these theories, exhibiting changes in cell adhesion, immune response, and vasculature genes.

Future comparison of early changes in Waptag1 DCIS and tumors with those of human will elucidate the significance of these findings in mouse, for potential earlier detection of the human disease.

Chapter 4

EARLY TRANSCRIPTIONAL CHANGES IN WAPTAG1 PARALLEL HUMAN BREAST CARCINOGENESIS

To determine, at the molecular level, the full potential of the Waptag1 mouse model in detecting early changes relevant to breast cancer, 58 human breast samples were analyzed, using the same statistical procedures as for the mouse, across three histological classifications: normal breast tissue, DCIS, or invasive breast cancer (IBC). Unsupervised hierarchical clustering suggested multiple stages of DCIS and IBC in the human samples as well as two DCIS outliers. Therefore the DCIS and IBC samples were reclustered (Figure 4.1.A) and the samples were regrouped accordingly: DCIS1, 9 samples; DCIS2, 16 samples; IBC1, 8 samples; IBC2, 16 samples, and 7 normal human samples (not shown in Figure 4.1.A). When clustered alone, the two distinct DCIS groups were recapitulated (Figure 4.1.B). Allred and colleagues independently showed the same clustering for DCIS samples but further defined each cluster as a specific subtype[27]. The ERBB2+ or basal samples have been grouped together here as 'DCIS1' and the luminal/mixed samples are designated 'DCIS2'. The pairwise contrast of human DCIS1 versus normal breast samples resulted in 6771 differences ($q \leq 0.05$), of which 396 had no mouse ortholog and 588 were not represented on the mouse microarray chip, although mouse orthologs were known.

Waptag1 samples were regrouped according to clustering (Figure 3.4) for consistency and to determine the greatest commonalities between species. Unsupervised hierarchical clustering of the 15 Waptag1 LCM samples revealed two DCIS samples clustered tightly with the C57BL/6J control cells (D_5714 and E_3991, Figure 3.4). After their removal, the remaining four DCIS samples were grouped together. The pairwise contrast between Waptag1 LCM DCIS cells and whole C57BL/6J samples returned 10,393 differences ($q \leq 0.05$), of which 922 had no human ortholog and 4794 were not represented on the human microarrays.

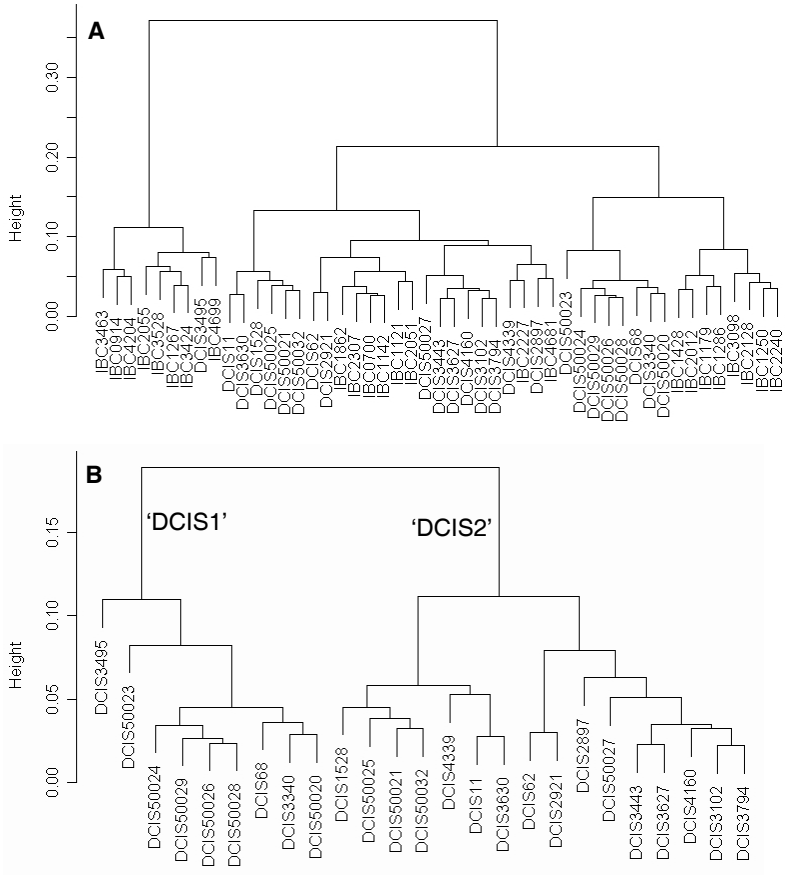


Figure 4.1. Hierarchical Clustering across Human Samples. **A)** Unsupervised clustering of human samples enriched in IBC or DCIS; **B)** DCIS samples cluster into two main subgroups, DCIS1 and DCIS2, respectively. [Normalized median expression values after RMA (no q-value limitation, see methods). Abbreviations: DCIS, ductal carcinoma *in situ*; IBC, Invasive Breast Cancer.]

4.1. Similarities between Human and Waptag1 Ductal Carcinoma *In Situ* (DCIS)

Waptag1 mammary tumorigenesis remarkably resembles human breast carcinogenesis, at both the histological, as well as the molecular, levels. Comparison of microarray results from human DCIS1 with those from Waptag1 DCIS revealed 2097 genes commonly differentially expressed, of 6945 genes orthologous and represented on all microarrays (Figure 4.2).

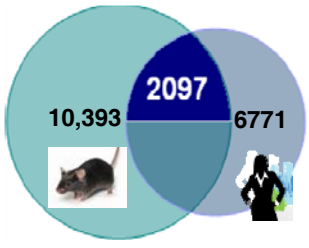


Figure 4.2. Venn Diagram of Common Transcriptional Changes in Waptag1 LCM DCIS and Human DCIS1. Cross-species comparison revealed 2097 genes commonly up- or down-regulated of 6945 orthologous genes. Within the 10,393 LCM Waptag1 DCIS (n=4) versus whole B6 samples (n=4) gene changes ($q \leq 0.05$; Fancher K *et.al.*, future publication), 922 had no human ortholog and 4794 were not represented on the human microarrays. Among 6771 human DCIS1 (n=9) versus normal samples (n=7) transcriptional differences ($q \leq 0.05$; Fancher K *et.al.*, future publication), 396 had no mouse ortholog and 588 were not represented on the mouse chip (see methods for additional details).

Regulation of apoptosis is aberrant in both human and Waptag1 DCIS. Although apoptosis-related Gene Ontology terms were not significant ($p \leq 0.0001$), genes categorized as involved in *regulation of apoptosis* and *regulation of programmed cell death* were among the 2097 genes

commonly differentially expressed between these two species, albeit at a reduced threshold. However, species-specific differences revealed *apoptosis*, *regulation of apoptosis,* and specifically *positive regulation of apoptosis* genes down-regulated in Waptag1 DCIS (see above). Similarly, in human DCIS1, genes categorized under *negative regulation of apoptosis* and *anti-apoptosis* exhibited up-regulated expression, whereas *apoptosis*, particularly *positive regulation of apoptosis* genes were down-regulated (see below).

Genes in the GO categories *developmental process, system development,* specifically *nervous system development* and *cellular developmental process*, *cell-cell signaling* and *neurotransmitter transport* genes were up-regulated in Waptag1 and human DCIS (Table 4.1).

4.1.1. Cells of Waptag1 DCIS and Human DCIS1 are Dividing and Mitotically Active

Waptag1 DCIS cells show abundant mitotic figures (Figure 3.1.B,C) and the 969 genes commonly up-regulated and orthologous in human and Waptag1 DCIS were enriched in the GO terms *DNA replication*, *cell cycle*, specifically *M phase of the mitotic cell cycle*, as well as the cellular component *nucleus* (Table 4.1). Figure 4.3 depicts predominantly genes over expressed in the cell cycle, from among all 2097 common differentially expressed genes in Waptag1 mouse and human DCIS. Of interest, *Trp53*/TP53 is down-regulated among the plethora of genes up-regulated (Figure 4.3). Additional key genes commonly over expressed in Waptag1 and human DCIS, including *Aurka*/AURKA, *Birc5*/BIRC5, *Ccnb1*/CCNB1, *Erbb2*/ERBB2, *Grb7*/GRB7, *Mki67*/MKI67, *Mmp11*/MMP11 (Table 4.3 and Fancher K *et.al.*, future publication), are used as clinical prognostic indicators for human breast cancer in the commercial OncoTypeDX^{TM} platform. Increased expression of these genes is associated with high risk of recurrence in human breast cancer.

These findings address two previous uncertainties: first, the appropriate use of microarrays for cross-species comparison of global gene expression changes; second, Waptag1 is an approximate mouse model for the study of human DCIS early changes, predictive of invasive breast cancer.

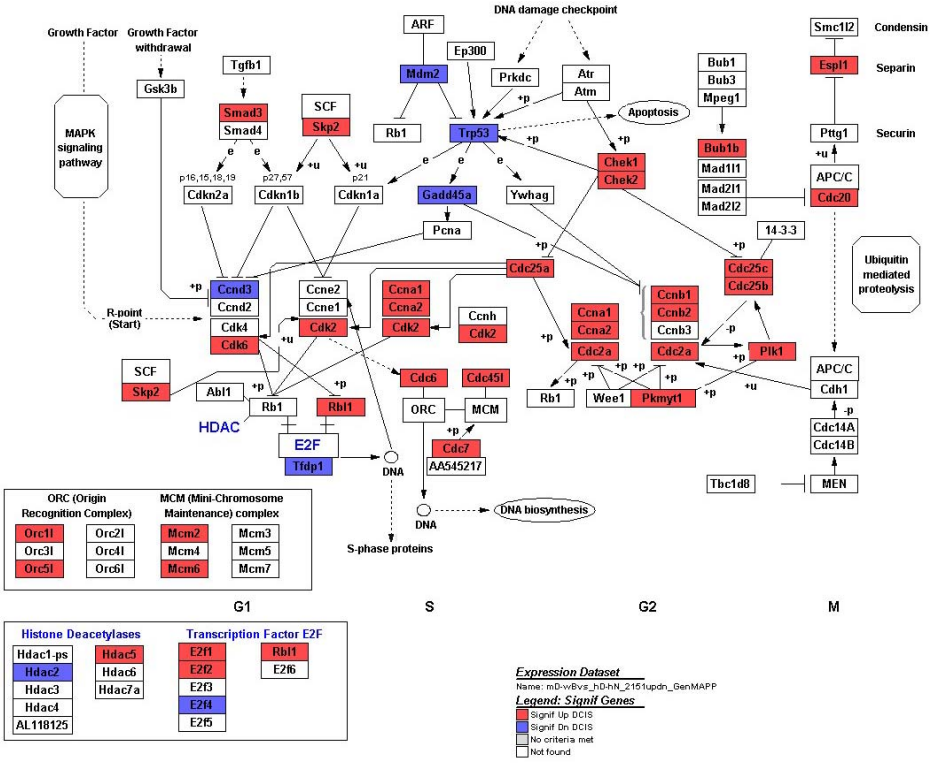


Figure 4.3. Cell Cycle Gene Changes Mutual to Waptag1 LCM DCIS and Human DCIS1. Among the 2097 significant common transcriptional changes, cell cycle genes were predominantly up-regulated (red) in both species [blue, down-regulated compared with species-specific controls; mouse GenMAPPv2 pathway and software.]

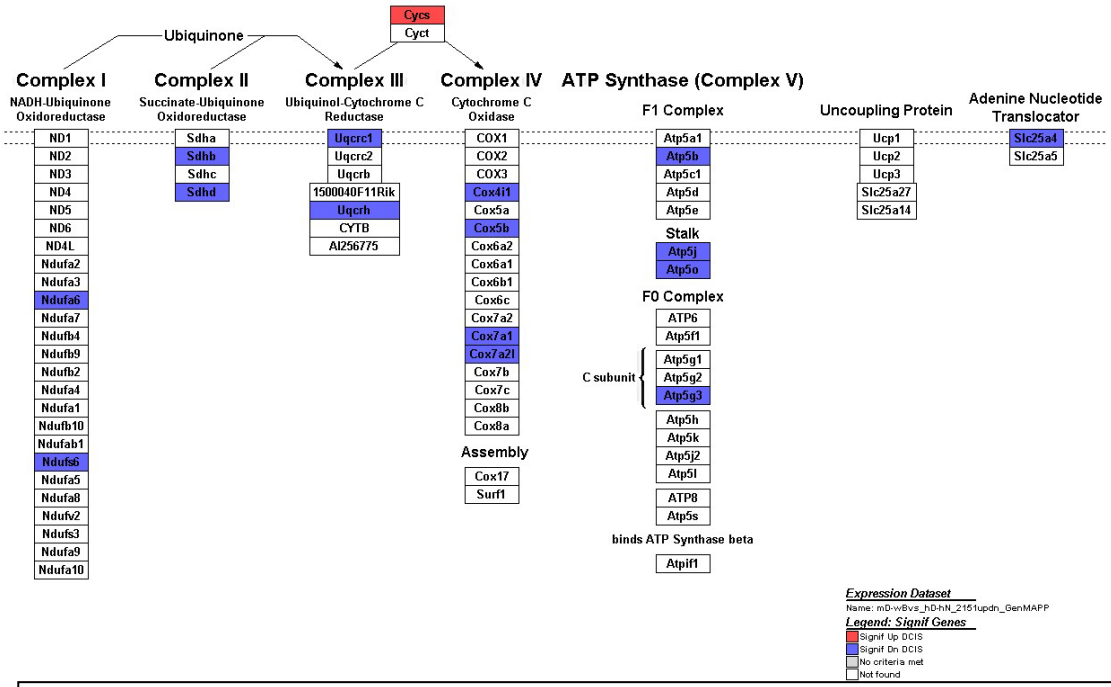


Figure 4.4. Electron Transport Chain Transcriptional Alterations Common to Waptag1 LCM DCIS and Human DCIS1. A classical aerobic energy-derivation pathway is predominantly down-regulated (blue shading) in both species based on the 2097 common genes. [red, up-regulated in DCIS compared with species-specific controls; mouse GenMAPPv2 pathway and software.]

Table 4.1. Significant Gene Ontology Classifications Among the 2097 Genes Commonly Differentially Expressed in Waptag1 DCIS and Human DCIS1. The most over-represented Gene Ontology terms in **A)** the 969 genes up-regulated, **B)** the 1128 genes down-regulated ($p \leq 0.0001$, VLAD, mouse). [Abbreviations: GO, Gene Ontology; BP, Biological Process; MF, Molecular Function; CC, Cellular Component. A full list of genes and expression values will appear in Fancher K *et.al.*, future publication.]

A GO	GO_ID	GO_Classifications of Up-regulated Genes	p-value
BP	GO:0032502	developmental process	6.47E-11
BP	GO:0030154	cell differentiation	1.38E-10
BP	GO:0048869	cellular developmental process	1.38E-10
BP	GO:0006259	DNA metabolic process	4.90E-10
BP	GO:0007049	cell cycle	8.24E-10
BP	GO:0006260	DNA replication	1.80E-09
BP	GO:0048856	anatomical structure development	1.09E-08
BP	GO:0043283	biopolymer metabolic process	1.64E-08
BP	GO:0009653	anatomical structure morphogenesis	1.93E-08
BP	GO:0044237	cellular metabolic process	3.23E-08
BP	GO:0044238	primary metabolic process	5.62E-08
BP	GO:0048731	system development	9.30E-08
BP	GO:0007275	multicellular organismal development	1.13E-07
BP	GO:0051301	cell division	1.55E-07
BP	GO:0022403	cell cycle phase	1.65E-07
BP	GO:0009987	cellular process	1.67E-07
BP	GO:0065007	biological regulation	1.77E-07
BP	GO:0022402	cell cycle process	2.66E-07
BP	GO:0007399	nervous system development	3.88E-07
BP	GO:0000278	mitotic cell cycle	5.05E-07
BP	GO:0000279	M phase	6.85E-07
BP	GO:0006468	protein amino acid phosphorylation	1.57E-06
BP	GO:0007267	cell-cell signaling	1.86E-06
BP	GO:0009719	response to endogenous stimulus	2.08E-06
BP	GO:0021983	pituitary gland development	2.59E-06
BP	GO:0008152	metabolic process	2.60E-06
BP	GO:0050789	regulation of biological process	3.78E-06
BP	GO:0007268	synaptic transmission	3.88E-06
BP	GO:0043170	macromolecule metabolic process	4.08E-06
BP	GO:0007067	mitosis	4.09E-06
BP	GO:0000087	M phase of mitotic cell cycle	4.57E-06
BP	GO:0007417	central nervous system development	4.87E-06
BP	GO:0048699	generation of neurons	6.53E-06
BP	GO:0021536	diencephalon development	6.74E-06
BP	GO:0006974	response to DNA damage stimulus	1.04E-05
BP	GO:0006836	neurotransmitter transport	1.26E-05
BP	GO:0022008	neurogenesis	1.38E-05
BP	GO:0016310	phosphorylation	1.91E-05
BP	GO:0030182	neuron differentiation	1.98E-05
BP	GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	2.02E-05
BP	GO:0050794	regulation of cellular process	2.07E-05
BP	GO:0045944	positive regulation of transcription from RNA polymerase II promoter	2.42E-05
BP	GO:0051179	localization	2.59E-05
BP	GO:0048513	organ development	2.72E-05
BP	GO:0007420	brain development	2.83E-05
BP	GO:0048666	neuron development	3.11E-05
BP	GO:0000902	cell morphogenesis	3.25E-05
BP	GO:0032989	cellular structure morphogenesis	3.25E-05
BP	GO:0019226	transmission of nerve impulse	3.47E-05
BP	GO:0001505	regulation of neurotransmitter levels	4.35E-05

GO	GO_ID	Up-regulated Genes GO_Classification	p-value
BP	GO:0030030	cell projection organization and biogenesis	5.12E-05
BP	GO:0048858	cell projection morphogenesis	5.12E-05
BP	GO:0032990	cell part morphogenesis	5.12E-05
BP	GO:0045941	positive regulation of transcription	5.30E-05
BP	GO:0006793	phosphorus metabolic process	5.46E-05
BP	GO:0006796	phosphate metabolic process	5.46E-05
BP	GO:0006281	DNA repair	5.55E-05
BP	GO:0032508	DNA duplex unwinding	5.85E-05
BP	GO:0032392	DNA geometric change	5.85E-05
BP	GO:0006268	DNA unwinding during replication	5.85E-05
BP	GO:0000904	cellular morphogenesis during differentiation	9.32E-05
MF	GO:0005524	ATP binding	5.92E-12
MF	GO:0032559	adenyl ribonucleotide binding	1.41E-11
MF	GO:0005488	binding	1.62E-10
MF	GO:0030554	adenyl nucleotide binding	1.64E-10
MF	GO:0032555	purine ribonucleotide binding	1.78E-09
MF	GO:0032553	ribonucleotide binding	1.78E-09
MF	GO:0017076	purine nucleotide binding	1.21E-08
MF	GO:0005515	protein binding	5.56E-08
MF	GO:0000166	nucleotide binding	7.27E-08
MF	GO:0004674	protein serine/threonine kinase activity	2.60E-07
MF	GO:0004672	protein kinase activity	5.25E-07
MF	GO:0030594	neurotransmitter receptor activity	8.63E-07
MF	GO:0042165	neurotransmitter binding	8.63E-07
MF	GO:0016773	phosphotransferase activity, alcohol group as acceptor	3.05E-06
MF	GO:0016772	transferase activity, transferring phosphorus-containing groups	4.18E-06
MF	GO:0015370	solute; sodium symporter activity	7.59E-06
MF	GO:0008504	monoamine transmembrane transporter activity	8.45E-06
MF	GO:0022857	transmembrane transporter activity	3.53E-05
MF	GO:0022836	gated channel activity	4.26E-05
MF	GO:0003824	catalytic activity	4.71E-05
MF	GO:0016301	kinase activity	7.04E-05
MF	GO:0022891	substrate-specific transmembrane transporter activity	8.67E-05
CC	GO:0044421	extracellular region part	5.00E-07
CC	GO:0005615	extracellular space	1.63E-06
CC	GO:0045202	synapse	2.08E-06
CC	GO:0005622	intracellular	7.55E-06
CC	GO:0044424	intracellular part	9.60E-06
CC	GO:0005634	nucleus	1.55E-05
CC	GO:0000793	condensed chromosome	3.09E-05
CC	GO:0044456	synapse part	3.57E-05
CC	GO:0005576	extracellular region	9.12E-05

Table 4.1. Continued.

B GO	GO_ID	Down-regulated Genes GO_Classification	p-value
BP	GO:0008152	metabolic process	4.71E-16
BP	GO:0044237	cellular metabolic process	5.06E-14
BP	GO:0044238	primary metabolic process	3.17E-11
BP	GO:0048518	positive regulation of biological process	7.82E-09
BP	GO:0019538	protein metabolic process	9.96E-09
BP	GO:0044267	cellular protein metabolic process	1.07E-08
BP	GO:0001568	blood vessel development	1.51E-08
BP	GO:0001944	vasculature development	2.29E-08
BP	GO:0048514	blood vessel morphogenesis	8.41E-08
BP	GO:0043170	macromolecule metabolic process	1.56E-07
BP	GO:0048522	positive regulation of cellular process	1.61E-07
BP	GO:0044260	cellular macromolecule metabolic process	1.90E-07
BP	GO:0009987	cellular process	4.76E-07
BP	GO:0006629	lipid metabolic process	5.88E-07
BP	GO:0006631	fatty acid metabolic process	7.20E-07
BP	GO:0032787	monocarboxylic acid metabolic process	2.10E-06
BP	GO:0006091	generation of precursor metabolites and energy	2.28E-06
BP	GO:0015980	energy derivation by oxidation of organic compounds	2.35E-06
BP	GO:0006464	protein modification process	3.99E-06
BP	GO:0044255	cellular lipid metabolic process	1.08E-05
BP	GO:0006793	phosphorus metabolic process	1.15E-05
BP	GO:0006796	phosphate metabolic process	1.15E-05
BP	GO:0019221	cytokine and chemokine mediated signaling pathway	1.17E-05
BP	GO:0043412	biopolymer modification	1.45E-05
BP	GO:0042127	regulation of cell proliferation	1.47E-05
BP	GO:0048856	anatomical structure development	1.62E-05
BP	GO:0007243	protein kinase cascade	1.66E-05
BP	GO:0016310	phosphorylation	1.77E-05
BP	GO:0032103	positive regulation of response to external stimulus	1.86E-05
BP	GO:0048513	organ development	2.05E-05
BP	GO:0043283	biopolymer metabolic process	2.47E-05
BP	GO:0002260	lymphocyte homeostasis	3.77E-05
BP	GO:0006084	acetyl-CoA metabolic process	3.77E-05
BP	GO:0048869	cellular developmental process	4.15E-05
BP	GO:0019752	carboxylic acid metabolic process	4.81E-05
BP	GO:0006082	organic acid metabolic process	5.12E-05
BP	GO:0045859	regulation of protein kinase activity	7.58E-05
BP	GO:0048731	system development	9.01E-05
BP	GO:0009887	organ morphogenesis	9.43E-05
MF	GO:0005515	protein binding	7.97E-21
MF	GO:0005488	binding	3.64E-20
MF	GO:0003824	catalytic activity	1.14E-12
MF	GO:0004713	protein tyrosine kinase activity	2.46E-06
MF	GO:0016301	kinase activity	3.21E-06
MF	GO:0050662	coenzyme binding	5.81E-06
MF	GO:0017076	purine nucleotide binding	7.57E-06
MF	GO:0030554	adenyl nucleotide binding	8.53E-06
MF	GO:0016563	transcription activator activity	8.85E-06
MF	GO:0016740	transferase activity	1.05E-05
MF	GO:0016773	phosphotransferase activity, alcohol group as acceptor	1.75E-05
MF	GO:0000166	nucleotide binding	1.90E-05
MF	GO:0016627	oxidoreductase activity, acting on the CH-CH group of donors	2.42E-05
MF	GO:0050660	FAD binding	4.98E-05
MF	GO:0048037	cofactor binding	5.52E-05
MF	GO:0016772	transferase activity, transferring phosphorus-containing groups	6.14E-05
MF	GO:0004672	protein kinase activity	9.81E-05

Table 4.1. Continued.

GO	GO_ID	Down-regulated Genes GO_Classification	p-value
CC	GO:0005737	cytoplasm	1.61E-38
CC	GO:0005622	intracellular	1.02E-31
CC	GO:0044424	intracellular part	1.48E-31
CC	GO:0043227	membrane-bounded organelle	9.67E-23
CC	GO:0044444	cytoplasmic part	1.50E-22
CC	GO:0043231	intracellular membrane-bounded organelle	2.75E-22
CC	GO:0043226	organelle	1.21E-19
CC	GO:0043229	intracellular organelle	5.81E-19
CC	GO:0031090	organelle membrane	1.61E-13
CC	GO:0005739	mitochondrion	9.76E-11
CC	GO:0005783	endoplasmic reticulum	2.44E-07
CC	GO:0019866	organelle inner membrane	2.48E-07
CC	GO:0031967	organelle envelope	3.38E-07
CC	GO:0031975	envelope	3.66E-07
CC	GO:0012505	endomembrane system	7.57E-07
CC	GO:0005740	mitochondrial envelope	7.85E-07
CC	GO:0044429	mitochondrial part	9.30E-07
CC	GO:0044422	organelle part	9.67E-07
CC	GO:0005743	mitochondrial inner membrane	1.05E-06
CC	GO:0043234	protein complex	1.45E-06
CC	GO:0031966	mitochondrial membrane	1.50E-06
CC	GO:0044446	intracellular organelle part	1.85E-06
CC	GO:0044432	endoplasmic reticulum part	2.63E-06
CC	GO:0005578	proteinaceous extracellular matrix	4.27E-06
CC	GO:0031300	intrinsic to organelle membrane	4.98E-06
CC	GO:0031012	extracellular matrix	6.06E-06
CC	GO:0044421	extracellular region part	7.50E-06
CC	GO:0005623	cell	7.98E-06
CC	GO:0044464	cell part	7.98E-06
CC	GO:0031227	intrinsic to endoplasmic reticulum membrane	1.23E-05
CC	GO:0005789	endoplasmic reticulum membrane	3.42E-05
CC	GO:0042175	nuclear envelope-endoplasmic reticulum network	5.65E-05
CC	GO:0000267	cell fraction	8.08E-05
CC	GO:0005746	mitochondrial respiratory chain	8.81E-05

Table 4.1. Continued.

4.1.2. Inhibition of Normal Energy Derivation

The 1128 orthologous genes whose expression was commonly down-regulated in Waptag1 DCIS and human DCIS1 were categorized under the following GO terms: *generation of precursor metabolites and energy*, specifically *energy derivation by oxidation of organic compounds*, *lipid metabolic process*, and the cellular components *mitochondrion*, specifically *mitochondrial membrane* and *mitochondrial respiratory chain* (Table 4.1.B). At a reduced threshold, genes down-regulated in both species were involved in *oxidoreductase activity*, *aerobic respiration*, particularly *tricarboxylic acid cycle*, *electron carrier activity*, *cofactor binding*, specifically *coenzyme binding*, and lastly, *ATP binding*. Figure 4.4 depicts the electron carrier activity genes as visualized on the electron transport chain pathway diagram. Genes in each complex of the respiratory pathway are commonly down-regulated. Figure 4.5 shows genes

under expressed within the aerobic respiration pathway. Ubiquinol-cytochrome c reductase genes, *Uqcrh*, and *Uqcrfs1* show the most substantial down-regulation in mouse. Additional genes in these GO categories were down-regulated in either human or Waptag1 as well (though not both), suggesting that some genes may have evolved different functions over time in each species.

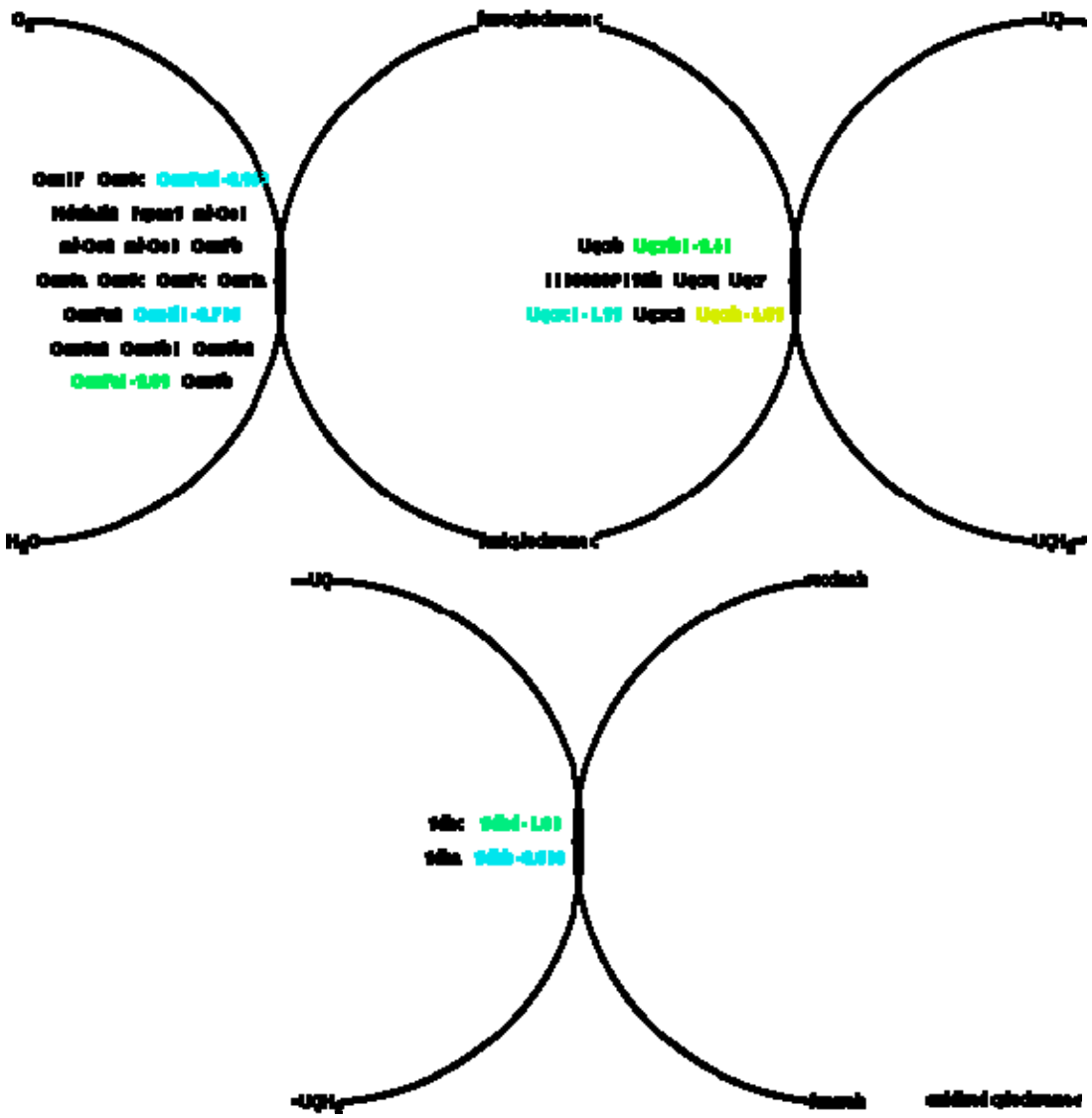


Figure 4.5. Aerobic Respiration, Electron Donor II Pathway Commonly Down-Regulated in Waptag1 LCM DCIS and Human DCIS1. Another classical aerobic energy-derivation pathway is predominantly down-regulated (yellow to green to blue spectrum) in DCIS in both species compared with species-specific controls. [Expression data from mouse is shown. Color Codes: yellow, down-regulated more than 4x (log2); green, beyond 1.8x (log2); blue, more than 0.6-fold (log2) down; MouseCyc software.]

One important finding, presented here in an *in vivo* model, is the down-regulation of the most efficient form of energy production, *early* in the tumorigenic process, in both mice and humans. When oxidative phosphorylation and generation of energy via the electron transport chain are predominantly inhibited, glycolysis is likely the means of ATP production. In support of this, numerous genes in the *glycolysis* pathway (for example, aldolase B, fructose-bisphosphate, *Aldob*/ALDOB, enolase 2, *Eno2*/ENO2, and hexokinase 3 *Hk3*/HK3) were commonly over expressed in DCIS (although the GO term itself was not significant at the cutoff threshold). Additionally, lactate dehydrogenases, *Ldhb* and *Ldhc*, in Waptag1, and ENO1, ENO3, glyceraldehyde-3-phosphate dehydrogenase, GAPDH, phosphoglycerate mutase 2 (PGAM2) and phosphofructokinases PFKP and PFKL in human DCIS were up-regulated compared with controls. (Of note, *Gapdh* and *Eno1* were not represented in the annotation for the mouse array.) Finally, the transcriptional coactivators and enhancers of oxidative phosphorylation, peroxisome proliferative activated receptors *Ppargc1a* and *Ppargc1b*, are both down-regulated in laser-captured cells of Waptag1 DCIS compared with B6 controls. (However, these genes were not present on the U95Av2 human array.)

4.1.3. Chromatin Remodeling, an Early Pre-tumor Modification

In Waptag1 DCIS, chromatin remodeling was among the early transcriptional changes, accompanied by over expression of retrotransposons, histones, and histone variants. Similarly, among genes commonly up-regulated in Waptag1 and Human DCIS, the following GO categories were significantly over-represented: *positive regulation of transcription*, specifically, *positive regulation of transcription from RNA polymerase II promoter* (Table 4.1.A). At a reduced threshold, *chromatin binding*, *histone modification*, and the cellular components, *chromosome*, in particular *condensed chromosome* categories were enriched in both human and Waptag1 DCIS.

In Waptag1 and human DCIS, epigenetic modifications and transcriptional deregulation of the genome appear to be an early event in tumorigenesis. Alongside the up-regulated common histone variants (*H2afx* and *H2afy*) and chromatin assembly, organization, modifying/remodeling factors: *Actl6a*/ACTL6A, *Bmi1*/BMI1, *Cenpa*/CENPA, *Chaf1b*/CHAF1B, *Ezh2*/EZH2, *Mcm2*/MCM2, *Rad54l*/RAD54L, and *Smarcd2*/SMARCD2 (Table 4.2), *chromatin binding* was a

significant GO category in the list of 2097 genes (at a reduced threshold). Table 4.2 depicts genes categorized predominantly under the GO terms *chromatin modification* and/or *chromatin organization & biogenesis*, however known and/or predicted chromatin-associated factors, including those that activate, repress, or may regulate chromatin have been included as well (Table 4.2). Of note, chromobox genes *Cbx4*/CBX4 and *Cbx7*/CBX7, suppressor of zest, *Suz12*/SUZ12, and methyl CpG binding protein, *Mecp2*/MECP2, genes were down-regulated in DCIS in both species (Table 4.2). Since the MECP2 protein binds methylated cytosines to maintain silenced DNA, it further supports the alteration of epigenetic suppression of chromatin.

AURKB/*Aurkb* expression was up-regulated in human and Waptag1 DCIS (Tables 4.2 and 4.3). Phosphorylation of histone H3 by mitotic AURKB, which occurs during M phase of the cell cycle, results in the liberation of Hp1, or another chromobox (CBX) protein, from heterochromatin[124,125]. Thus, up-regulation of AURKB/*Aurkb* may be a potential key early player because unsilencing of heterochromatin is associated with chromatin modifications. Perhaps *Aurkb* is a harbinger of changes that are to come.

Table 4.2. Chromosome Modification or Organization Factors in Waptag1 DCIS and Human DCIS1. Genes involved in chromatin modification/organization were selected to detail the expression levels of chromatin-related gene changes. Fold changes shown are statistically significant according to Fs statistic (ANOVA, $q \leq 0.01$). [Abbreviations: MGI, Mouse Genome Informatics; Wapt1, Waptag1; LCM, laser-capture microdissection; DCIS, ductal carcinoma *in situ*; whB6, whole glands from C57BL/6J; N/A, no mouse or human ortholog; NOA, not on array; NS, not significant, though the gene was present on the array.]

MGI_symbol	Gene_Name	Fold Change Waptag1 LCM DCIS vs. whB6	Human Symbol	FoldChange Human DCIS1 vs. Norm
Acd	adrenocortical dysplasia	NS	**ACD**	2.37
Actl6a	actin-like 6A	2.92	**ACTL6A**	5.19
Actl6b	actin-like 6B	1.19	**ACTL6B**	4.18
Aebp2	AE binding protein 2	NOA	**AEBP2**	NOA
Akap8	A kinase (PRKA) anchor protein 8	NS	**AKAP8**	-1.20
Aof2	amine oxidase (flavin containing) domain 2	3.13	**AOF2**	-1.45
Ard1	N-acetyltransferase ARD1 homolog (S. cerevisiae)	-3.42	**ARD1A**	-1.23
Arid1a	AT rich interactive domain 1A (Swi1 like)	-2.08	**ARID1A**	-5.45
Arid4b	AT rich interactive domain 4B (Rbp1 like)	-1.27	**ARID4B**	NOA
Asf1a	ASF1 anti-silencing function 1 homolog A (S. cerevisiae)	2.39	**ASF1A**	-6.28
Asf1b	ASF1 anti-silencing function 1 homolog B (S. cerevisiae)	4.77	**ASF1B**	NOA
Ash1l	ash1 (absent, small, or homeotic)-like (Drosophila)	NS	**ASH1L**	NOA
Aurkb	aurora kinase B	14.53	**AURKB**	14.02

MGI_symbol	Gene_Name	Fold Change Waptag1 LCM DCIS vs. whB6	Human Symbol	FoldChange Human DCIS1 vs. Norm
Banp	Btg3 associated nuclear protein	1.68	**BANP**	NOA
Baz1a	bromodomain adjacent to zinc finger domain 1A	NS	**BAZ1A**	1.28
Baz1b	bromodomain adjacent to zinc finger domain, 1B	-1.76	**BAZ1B**	-2.88
Bcor	Bcl6 interacting corepressor	NS	**BCOR**	NOA
Bcorl1	BCL6 co-repressor-like 1	1.30	**BCORL1**	NOA
Blm	Bloom syndrome homolog (human)	6.76	**BLM**	1.97
Bmi1	Bmi1 polycomb ring finger oncogene	2.18	**BMI1**	1.93
Bptf	bromodomain PHD finger transcription factor	1.43	**BPTF**	NOA
Brca2	breast cancer 2	1.46	**BRCA2**	NS
Brd8	bromodomain containing 8	1.32	**BRD8**	1.46
Brdt	bromodomain, testis-specific	1.22	**BRDT**	-1.39
Bub3	budding uninhibited by benzimidazoles 3 homolog	NS	**BUB3**	-1.40
Carm1	coactivator-associated arginine methyltransferase 1	-1.31	**CARM1**	3.43
Cbx1	chromobox homolog 1 (Drosophila HP1 beta)	1.67	**CBX1**	-1.35
Cbx2	chromobox homolog 2 (Drosophila Pc class)	NS	**CBX2**	1.76
Cbx3	chromobox homolog 3 (Drosophila HP1 gamma)	NOA	**CBX3**	NS
Cbx4	chromobox homolog 4 (Drosophila Pc class)	-1.80	**CBX4**	-2.06
Cbx5	chromobox homolog 5 (Drosophila HP1a)	NS	**CBX5**	NOA
Cbx6	chromobox homolog 6	NS	**CBX6**	NOA
Cbx7	chromobox homolog 7	-1.53	**CBX7**	-1.71
Cbx8	chromobox homolog 8 (Drosophila Pc class)	NS	**CBX8**	NOA
Ccnb1ip1	cyclin B1 interacting protein 1	NS	**CCNB1IP1**	NOA
Cdyl	chromodomain protein, Y chromosome-like	-1.29	**CDYL**	1.88
Cebpg	CCAAT/enhancer binding protein (C/EBP), gamma	1.30	**CEBPG**	-1.85
Cenpa	centromere protein A	11.55	**CENPA**	3.04
Cenph	centromere protein H	2.91	**CENPH**	NOA
Chaf1b	chromatin assembly factor 1, subunit B (p60)	4.47	**CHAF1B**	1.56
Chd1	chromodomain helicase DNA binding protein 1	-2.76	**CHD1**	1.28
Chd3	chromodomain helicase DNA binding protein 3	1.23	**CHD3**	2.10
Chd4	chromodomain helicase DNA binding protein 4	-2.10	**CHD4**	4.12
Chd6	chromodomain helicase DNA binding protein 6	2.35	**CHD6**	NOA
Chd7	chromodomain helicase DNA binding protein 7	NS	**CHCHD7**	NOA
Chd9	chromodomain helicase DNA binding protein 9	NS	**CHD9**	NOA
Cit	citron	NOA	**CIT**	-2.08
Cpeb1	cytoplasmic polyadenylation element binding protein 1	-1.85	**CPEB1**	NOA
Crebbp	CREB binding protein	-1.94	**CREBBP**	-1.63
Dapk3	death-associated protein kinase 3	-4.15	**DAPK3**	6.11
Dclre1c	DNA cross-link repair 1C, PSO2 homolog (S. cerevisiae)	1.37	**DCLRE1C**	NOA
Dmap1	DNA methyltransferase 1-associated protein 1	1.97	**DMAP1**	NOA
Dnmt1	DNA methyltransferase (cytosine-5) 1	1.36	**DNMT1**	NS
Dnmt3a	DNA methyltransferase 3A	1.90	**DNMT3A**	NOA
Dnmt3b	DNA methyltransferase 3B	1.61	**DNMT3AB**	NOA
Eed	embryonic ectoderm development	1.24	**EED**	1.87
Ehmt2	euchromatic histone lysine N-methyltransferase 2	-2.28	**EHMT2**	NOA
Ep400	E1A binding protein p400	NS	**EP400**	-1.63
Epc1	enhancer of polycomb homolog 1 (Drosophila)	1.89	**EPC1**	NOA
Epc2	enhancer of polycomb homolog 2 (Drosophila)	3.19	**EPC2**	NOA
Ercc1	excision repair cross-complementing rodent repair deficiency, c	NS	**ERCC1**	1.62
Ezh2	enhancer of zeste homolog 2 (Drosophila)	4.71	**EZH2**	5.86
Fancd2	Fanconi anemia, complementation group D2	1.64	**FANCD2**	NOA
Fbxl10	F-box and leucine-rich repeat protein 10	NOA	**FBXL10**	NOA
Fbxl11	F-box and leucine-rich repeat protein 11	NOA	**FBXL11**	1.45
Gpx4	glutathione peroxidase 4	NOA	**GPX4**	1.40
Gsg2	germ cell-specific gene 2	4.55	**GSG2**	NOA
H1f0	H1 histone family, member 0	NS	**H1F0**	-13.79
H2afv	H2A histone family, member V	NS	**H2AFV**	NS
H2afx	H2A histone family, member X	2.81	**H2AFX**	1.45
H2afy	H2A histone family, member Y	1.46	**H2AFY**	1.89
H2afy2	H2A histone family, member Y2	-1.42	**H2AFY2**	NOA
H2afz	H2A histone family, member Z	2.32	**H2AFZ**	-2.61
H3f3a	H3 histone, family 3A	NOA	**H3F3A**	NOA
H3f3b	H3 histone, family 3B	1.53	**H3F3B**	-6.40
Hdac1	histone deacetylase 1	NOA	**HDAC1**	-1.78
Hdac10	histone deacetylase 10	1.58	**HDAC10**	NOA
Hdac11	histone deacetylase 11	1.98	**HDAC11**	NOA
Hdac2	histone deacetylase 2	-1.61	**HDAC2**	NS
Hdac3	histone deacetylase 3	1.32	**HDAC3**	-4.06
Hdac4	histone deacetylase 4	NS	**HDAC4**	1.64
Hdac5	histone deacetylase 5	1.53	**HDAC5**	1.80
Hdac6	histone deacetylase 6	-1.21	**HDAC6**	NS
Hdac7	histone deacetylase 7	-1.57	**HDAC7**	NOA
Hdac8	histone deacetylase 8	NS	**HDAC8**	NOA
Hdac9	histone deacetylase 9	-1.35	**HDAC9**	1.53
Hells	helicase, lymphoid specific	3.62	**HELLS**	NOA
Hils1	histone H1-like protein in spermatids 1	1.53	**HILS1**	NOA
Hira	histone cell cycle regulation defective homolog A	NS	**HIRA**	2.63

Table 4.2. Continued.

MGI_symbol	Gene_Name	Fold Change Waptag1 LCM DCIS vs. whB6	Human Symbol	FoldChange Human DCIS1 vs. Norm
Hist1h1b	histone cluster 1, H1b	NOA	**HIST1H1B**	3.30
Hist1h1c	histone cluster 1, H1c	1.55	**HIST1H1C**	-2.08
Hist1h1d	histone cluster 1, H1d	NOA	**HIST1H1D**	3.57
Hist1h1e	histone cluster 1, H1e	NS	**HIST1H1E**	1.72
Hist1h1t	histone cluster 1, H1t	1.51	**HIST1H1T**	NS
Hist1h2ad	histone cluster 1, H2ad	NOA	**HIST1H2AD**	17.67
Hist1h2ae	histone cluster 1, H2ae	NOA	**HIST1H2AE**	2.19
Hist1h2ai	histone cluster 1, H2ai	NOA	**HIST1H2AI**	-2.33
Hist1h2ak	histone cluster 1, H2ak	NOA	**HIST1H2AK**	-1.20
Hist1h2be	histone cluster 1, H2be	NOA	**HIST1H2BE**	2.24
Hist1h2bf	histone cluster 1, H2bf	NOA	**HIST1H2BF**	-1.49
Hist1h2bk	histone cluster 1, H2bk	NOA	**HIST1H2BK**	-1.83
Hist1h2bl	histone cluster 1, H2bl	NOA	**HIST1H2BL**	2.32
Hist1h2bm	histone cluster 1, H2bm	NOA	**HIST1H2BM**	6.60
Hist1h3f	histone cluster 1, H3f	2.31	**HIST1H3F**	NOA
Hist1h4b	histone cluster 1, H4b	NOA	**HIST1H4B**	1.28
Hist1h4c	histone cluster 1, H4c	NOA	**HIST1H4C**	-2.04
Hist1h4h	histone cluster 1, H4h	-7.34	**HIST1H4H**	NOA
Hist1h4i	histone cluster 1, H4i	3.86	**HIST1H4I**	NOA
Hist2h2be	histone cluster 2, H2be	5.47	**HIST2H2BE**	-1.67
Hist2h3c1	histone cluster 2, H3c1	-1.50	**HIST2H3C**	NOA
Hist2h3c2	histone cluster 2, H3c2	3.56	N/A	
Hist3h2a	histone cluster 3, H2a	1.89	**HIST3H2A**	NOA
Hist3h2ba	histone cluster 3, H2ba	1.93	**HIST3H2BA**	NOA
Hltf	helicase-like transcription factor	1.74	**HLTF**	-6.54
Hmg20a	high mobility group 20A	NS	**HMG20A**	NOA
Hmg20b	high mobility group 20 B	-2.92	**HMG20B**	NS
Hmgn1	high mobility group nucleosomal binding domain 1	3.92	**HMGN1**	NOA
Hnf1a	HNF1 homeobox A	NS	**HNF1A**	NS
Hopx	HOP homeobox	-1.95	**HOP**	-1.48
Hp1bp3	heterochromatin protein 1, binding protein 3	1.28	**HP1BP3**	NOA
Hspa1a	heat shock protein 1A	2.51	**HSPA1A**	NOA
Hspa1b	heat shock protein 1B	2.04	**HSPA1B**	NOA
Ing2	inhibitor of growth family, member 2	1.93	**ING2**	-1.98
Ing3	inhibitor of growth family, member 3	NS	**ING3**	-2.45
Jarid1a	jumonji, AT rich interactive domain 1A (Rbp2 like)	NS	**JARID1A**	-1.46
Jarid1b	jumonji, AT rich interactive domain 1B (Rbp2 like)	NS	**JARID1B**	NS
Jarid1c	jumonji, AT rich interactive domain 1C (Rbp2 like)	-1.42	**JARID1C**	1.46
Jarid1d	jumonji, AT rich interactive domain 1D (Rbp2 like)	NS	**JARID1D**	2.25
Jmjd1a	jumonji domain containing 1A	2.06	**JMJD1A**	-2.59
Jmjd1b	jumonji domain containing 1B	1.25	**JMJD1B**	-10.18
Jmjd1c	jumonji domain containing 1C	NS	**JMJD1C**	-8.36
Jmjd2a	jumonji domain containing 2A	NOA	**JMJD2A**	2.70
Jmjd2b	jumonji domain containing 2B	-1.26	**JMJD2B**	1.78
Jmjd2c	jumonji domain containing 2C	-1.58	**JMJD2C**	NOA
Jmjd2d	jumonji domain containing 2D	1.43	**JMJD2D**	NOA
Jmjd3	jumonji domain containing 3	1.77	**JMJD3**	2.51
Kat2b	K(lysine) acetyltransferase 2B	-1.59	**PCAF**	-3.10
Kat5	K(lysine) acetyltransferase 5	-2.83	**HTATIP**	-3.09
Klf1	Kruppel-like factor 1 (erythroid)	NS	**KLF1**	1.40
L3mbtl2	l(3)mbt-like 2 (Drosophila)	NS	**L3MBTL2**	NOA
Lig4	ligase IV, DNA, ATP-dependent	1.68	**LIG4**	-2.35
Mad2l1	MAD2 (mitotic arrest deficient, homolog)-like 1 (yeast)	7.14	**MAD2L1**	NS
Map3k12	mitogen-activated protein kinase kinase kinase 12	1.90	**MAP3K12**	2.25
Mbd1	methyl-CpG binding domain protein 1	1.92	**MBD1**	-4.80
Mbd2	methyl-CpG binding domain protein 2	-1.73	**MBD2**	5.66
Mbd3	methyl-CpG binding domain protein 3	-1.98	**MBD3**	1.71
Mcm2	minichromosome maintenance deficient 2 mitotin	4.37	**MCM2**	3.95
Mecp2	methyl CpG binding protein 2	1.44	**MECP2**	1.57
Men1	multiple endocrine neoplasia 1	-1.50	**MEN1**	NS
Mettl8	methyltransferase like 8	3.99	**METTL8**	NOA
Mlh3	mutL homolog 3 (E coli)	2.84	**MLH3**	NS
Mll1	myeloid/lymphoid or mixed-lineage leukemia 1	NS	**MLL**	2.43
Mll3	myeloid/lymphoid or mixed-lineage leukemia 3	NS	**MLL3**	NOA
Mll5	myeloid/lymphoid or mixed-lineage leukemia 5	NOA	**MLL5**	NOA
Morf4l1	mortality factor 4 like 1	NOA	**MORF4L1**	NOA
Morf4l2	mortality factor 4 like 2	NOA	**MORF4L2**	-2.09
Mphosph8	M-phase phosphoprotein 8	1.37	**HSMPP8**	-3.02
Msh4	mutS homolog 4 (E. coli)	NS	**MSH4**	-1.74
Msh5	mutS homolog 5 (E. coli)	1.71	**MSH5**	NOA
Msl31	male-specific lethal-3 homolog 1 (Drosophila)	2.06	**MSL3L1**	-3.40
Msx3	homeobox, msh-like 3	1.18	N/A	
Mta2	metastasis-associated gene family, member 2	-3.02	**MTA2**	NOA
Myst1	MYST histone acetyltransferase 1	-1.34	**MYST1**	2.29
Myst2	MYST histone acetyltransferase 2	-1.21	**MYST2**	-4.39
Myst3	MYST histone acetyltransferase (monocytic leukemia) 3	NS	**MYST3**	2.12
Myst4	MYST histone acetyltransferase monocytic leukemia 4	1.78	**MYST4**	NOA

Table 4.2. Continued.

MGI_symbol	Gene_Name	Fold Change Waptag1 LCM DCIS vs. whB6	Human Symbol	FoldChange Human DCIS1 vs. Norm
Nap1l1	nucleosome assembly protein 1-like 1	NS	**NAP1L1**	-3.48
Nap1l2	nucleosome assembly protein 1-like 2	1.21	**NAP1L2**	NOA
Nap1l3	nucleosome assembly protein 1-like 3	-1.21	**NAP1L3**	-2.00
Nap1l4	nucleosome assembly protein 1-like 4	-1.18	**NAP1L4**	3.41
Ncapd2	non-SMC condensin I complex, subunit D2	NOA	**NCAPD2**	4.11
Ncapd3	non-SMC condensin II complex, subunit D3	NS	**NCAPD3**	-1.25
Ncapg2	non-SMC condensin II complex, subunit G2	3.24	**NCAPG2**	NOA
Ncaph	non-SMC condensin I complex, subunit H	9.01	**NCAPH**	NS
Ncaph2	non-SMC condensin II complex, subunit H2	2.83	**NCAPH2**	NOA
Ncor1	nuclear receptor co-repressor 1	-1.42	**NCOR1**	NOA
Nek2	NIMA (never in mitosis gene a)-related expressed kinase 2	7.82	**NEK2**	-1.49
Noc2l	nucleolar complex associated 2 homolog (S. cerevisiae)	NOA	**NOC2L**	NS
Npm2	nucleophosmin/nucleoplasmin 2	NS	**NPM2**	NOA
Nr3c1	nuclear receptor subfamily 3, group C, member 1	NS	**NR3C1**	-3.18
Nsd1	nuclear receptor-binding SET-domain protein 1	NS	**NSD1**	NOA
Nusap1	nucleolar and spindle associated protein 1	16.50	**NUSAP1**	NOA
Padi4	peptidyl arginine deiminase, type IV	1.67	**PADI4**	NOA
Parp1	poly (ADP-ribose) polymerase family, member 1	NS	**PARP1**	1.52
Parp12	poly (ADP-ribose) polymerase family, member 12	NS	**PARP12**	NOA
Parp14	poly (ADP-ribose) polymerase family, member 14	2.65	**PARP14**	NOA
Parp16	poly (ADP-ribose) polymerase family, member 16	2.38	**PARP16**	NOA
Parp2	poly (ADP-ribose) polymerase family, member 2	1.59	**PARP2**	-1.95
Parp9	poly (ADP-ribose) polymerase family, member 9	NS	**PARP9**	NOA
Pax5	paired box gene 5	-1.25	**PAX5**	3.33
Pcgf2	polycomb group ring finger 2	-1.61	**PCGF2**	4.93
Phf21a	PHD finger protein 21A	1.39	**PHF21A**	3.01
Pot1a	protection of telomeres 1A	1.23	**POT1**	NOA
Prdm9	PR domain containing 9	NS	**PRDM9**	NOA
Prkdc	protein kinase, DNA activated, catalytic polypeptide	2.07	**PRKDC**	-1.39
Prm1	protamine 1	NS	**PRM1**	1.60
Prm2	protamine 2	-1.23	**PRM2**	2.00
Prm3	protamine 3	1.66	**PRM3**	NOA
Prmt5	protein arginine N-methyltransferase 5	-1.58	**PRMT5**	NOA
Pttg1	pituitary tumor-transforming 1	-1.70	**PTTG1**	2.26
Rad54l	RAD54 like (S. cerevisiae)	6.81	**RAD54L**	6.10
Rb1	retinoblastoma 1	3.38	**RB1**	-5.19
Rbbp4	retinoblastoma binding protein 4	-3.80	**RBBP4**	-1.86
Rbbp7	retinoblastoma binding protein 7	2.87	**RBBP7**	NOA
Rbl1	retinoblastoma-like 1 (p107)	2.17	**RBL1**	3.62
Rbl2	retinoblastoma-like 2	-5.87	**RBL2**	-3.68
Rcbtb1	regulator of chromosome condensation (RCC1) and BTB (POZ) domain containing protein 1	-1.78	**RCBTB1**	NOA
Rcor1	REST corepressor 1	NS	**RCOR1**	NOA
Recql4	RecQ protein-like 4	3.31	**RECQL4**	2.02
Rere	arginine glutamic acid dipeptide (RE) repeats	NS	**RERE**	6.59
Ring1	ring finger protein 1	-1.93	**RING1**	NOA
Rnf2	ring finger protein 2	-2.26	**RNF2**	3.76
Rnf20	ring finger protein 20	NS	**RNF20**	NOA
Rnf40	ring finger protein 40	-1.40	**RNF40**	NS
Rtel1	regulator of telomere elongation helicase 1	2.08	**RTEL1**	3.60
Ruvbl1	RuvB-like protein 1	NOA	**RUVBL1**	1.28
Ruvbl2	RuvB-like protein 2	-1.44	**RUVBL2**	1.42
Sap18	Sin3-associated polypeptide 18	1.32	**SAP18**	NS
Satb1	special AT-rich sequence binding protein 1	-6.55	**SATB1**	-2.19
Set	SET translocation	NOA	**SET**	2.27
Setd1b	SET domain containing 1B	NS	**SETD1B**	-3.10
Setd7	SET domain containing (lysine methyltransferase) 7	NS	**SETD7**	NOA
Setd8	SET domain containing (lysine methyltransferase) 8	-1.36	**SETD8**	NOA
Setdb1	SET domain, bifurcated 1	NS	**SETDB1**	1.35
Setdb2	SET domain, bifurcated 2	1.93	**SETDB2**	NOA
Sgol2	shugoshin-like 2 (S. pombe)	5.81	**SGOL2**	NOA
Shprh	SNF2 histone linker PHD RING helicase	NS	**SHPRH**	NOA
Sirt1	sirtuin 1 (silent mating type information regulation 2, homolog) 1 (S. cerevisiae)	-2.05	**SIRT1**	NOA
Sirt2	sirtuin 2	-1.51	**SIRT2**	NOA
Sirt3	sirtuin 3	1.26	**SIRT3**	NOA
Sirt4	sirtuin 4	-1.51	**SIRT4**	NOA
Sirt5	sirtuin 5	-1.24	**SIRT5**	NOA
Sirt6	sirtuin 6	1.23	**SIRT6**	NOA
Sirt7	sirtuin 7	1.45	**SIRT7**	NOA

Table 4.2. Continued.

MGI_symbol	Gene_Name	Fold Change Waptag1 LCM DCIS vs. whB6	Human Symbol	FoldChange Human DCIS1 vs. Norm
Smarca1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 1	NS	SMARCA1	NS
Smarca2	SWI/SNF related, matrix associated, regulator of chromatin	NS	SMARCA2	NS
Smarca4	SWI/SNF related, matrix associated, regulator of chromatin	NS	SMARCA4	1.79
Smarca5	SWI/SNF related, matrix associated, regulator of chromatin	-3.21	N/A	
Smarcal1	SWI/SNF related, matrix associated, regulator of chromatin	-1.34	SMARCAL1	NOA
Smarcb1	SWI/SNF related, matrix associated, regulator of chromatin	NS	SMARCB1	1.64
Smarcc1	SWI/SNF related, matrix associated, regulator of chromatin	NS	SMARCC1	2.29
Smarcc2	SWI/SNF related, matrix associated, regulator of chromatin	-1.45	SMARCC2	-2.10
Smarcd1	SWI/SNF related, matrix associated, regulator of chromatin	-1.80	SMARCD1	-1.38
Smarcd2	SWI/SNF related, matrix associated, regulator of chromatin	4.53	SMARCD2	1.90
Smarcd3	SWI/SNF related, matrix associated, regulator of chromatin	NS	SMARCD3	NS
Smarce1	SWI/SNF related, matrix associated, regulator of chromatin	-1.66	SMARCE1	1.55
Smc1a	structural maintenance of chromosomes 1A	NS	SMC1A	2.18
Smc1b	structural maintenance of chromosomes 1B	NS	SMC1B	NOA
Smc2	structural maintenance of chromosomes 2	1.69	SMC2	NS
Smc3	structural maintenace of chromosomes 3	-1.98	SMC3	-1.32
Smc4	structural maintenance of chromosomes 4	1.92	SMC4	-4.50
Smchd1	SMC hinge domain containing 1	NOA	SMCHD1	-5.33
Smyd1	SET and MYND domain containing 1	-1.18	SMYD1	NOA
Smyd3	SET and MYND domain containing 3	NS	SMYD3	NOA
Sox1	SRY-box containing gene 1	1.52	SOX1	3.79
Sox3	SRY-box containing gene 3	NS	SOX3	5.02
Stag3	stromal antigen 3	NS	STAG3	NOA
Stra8	stimulated by retinoic acid gene 8	1.58	STRA8	NOA
Suds3	suppressor of defective silencing 3 homolog (S. cerevisiae)	NS	SUDS3	NOA
Supt4h1	suppressor of Ty 4 homolog 1 (S. cerevisiae)	NOA	SUPT4H1	-1.84
Supt4h2	suppressor of Ty 4 homolog 2 (S. cerevisiae)	NOA	N/A	NOA
Suv39h1	suppressor of variegation 3-9 homolog 1 (Drosophila)	1.39	SUV39H1	NOA
Suv39h2	suppressor of variegation 3-9 homolog 2 (Drosophila)	1.61	SUV39H2	NOA
Suv420h1	suppressor of variegation 4-20 homolog 1 (Drosophila)	NOA	SUV420H1	NOA
Suv420h2	suppressor of variegation 4-20 homolog 2 (Drosophila)	-1.19	SUV420H2	NOA
Suz12	suppressor of zeste 12 homolog (Drosophila)	-2.49	SUZ12	-4.74
Syce2	synaptonemal complex central element protein 2	6.41	SYCE2	NOA
Tbl1xr1	transducin (beta)-like 1X-linked receptor 1	NOA	TBL1XR1	NOA
Terf1	telomeric repeat binding factor 1	NS	TERF1	-2.18
Terf2	telomeric repeat binding factor 2	NS	TERF2	-4.69
Tinf2	Terf1 (TRF1)-interacting nuclear factor 2	1.95	TINF2	NOA
Tlk1	tousled-like kinase 1	NS	TLK1	8.47
Tlk2	tousled-like kinase 2 (Arabidopsis)	NS	TLK2	1.52
Tnp2	transition protein 2	1.28	TNP2	-2.06
Top2a	topoisomerase (DNA) II alpha	7.05	TOP2A	6.34
Trrap	transformation/transcription domain-associated protein	-1.54	TRRAP	NS
Tspyl1	testis-specific protein, Y-encoded-like 1	-1.91	TSPYL1	-13.58
Tspyl2	TSPY-like 2	NS	TSPYL2	-2.11
Tspyl4	TSPY-like 4	1.68	TSPYL4	-1.53
Tspyl5	testis-specific protein, Y-encoded-like 5	NS	TSPYL5	-2.05
Ttf1	transcription termination factor 1	NOA	TTF1	-3.13
Utp3	UTP3, small subunit processome component, homolog	NS	SAS10	-1.55
Utx	ubiquitously transcribed tetratricopeptide repeat gene, X chromosome	-1.84	UTX	-5.54
Uty	ubiquitously transcribed tetratricopeptide repeat gene, Y chromosome	NS	UTY	NS
Vps72	vacuolar protein sorting 72 (yeast)	-3.82	VPS72	2.07
Wbp7	WW domain binding protein 7	1.34	MLL4	3.28
Whsc1	Wolf-Hirschhorn syndrome candidate 1 (human)	1.37	WHSC1	3.47
Wrn	Werner syndrome homolog (human)	1.39	WRN	-1.35
Xrn1	5'-3' exoribonuclease 1	1.38	XRN1	NOA
Yeats4	YEATS domain containing 4	1.25	YEATS4	NOA
1600027N09Rik	RIKEN cDNA 1600027N09 gene	-1.33	C20orf20	NOA
1700054O13Rik	RIKEN cDNA 1700054O13 gene	NS	CXorf27	-1.38
2210018M11Rik	RIKEN cDNA 2210018M11 gene	1.53	C11orf30	NOA
2410016O06Rik	RIKEN cDNA 2410016O06 gene	1.92	C14orf169	NOA
2610028A01Rik	RIKEN cDNA 2610028A01 gene	4.58	PINX1	NOA

Table 4.2. Continued.

4.1.4. Transcriptional Activation and Expression of Retrotransposons

Retrotransposons, often found in heterochromatic regions, are abundantly expressed in both Waptag1 and human DCIS. Although there is little sequence similarity among human and mouse

retrotransposons, the similarity of mechanisms underlying activation of these elements early in tumorigenesis in both human and mouse is very likely, given that these repetitive elements are generally *not* expressed in normal cells. To query for retrotransposon expression, specifically in the datasets described here, the target sequences for each of the Affymetrix probes on the mouse or human arrays were processed through RepeatMasker[215]. (These data are the only ones for which the UMBA cdf was not used for pre-processing.) Numerous retrotransposons were among the top transcripts with the highest changes in expression, irrespective of the species (Table 5.4, Chapter 5). Long terminal repeat (LTR) Class I, II, and III retrotransposons, as well as LINEs and SINEs, were over-expressed in Waptag1 and human DCIS. The LTR retrotransposons that may be genetic predictors of committed transformation to tumorigenesis are discussed in Chapter 5.

Since transcriptional activation of retrotransposons is associated with epigenetic changes and aberrant gene expression, these molecules may prove useful indicators of early stages of mammary tumorigenesis. Data presented in the next chapter test the hypothesis: there may be a common mechanism by which repetitive elements are activated early in the tumorigenic process in both mouse and human mammary cancer.

4.1.5. Early Transcripts in Waptag1 DCIS and Human DCIS1 Predictive of Reduced Survival in Human Breast Cancer

To find the *earliest* gene changes in LCM Waptag1 DCIS that could be putative predictive markers of an invasive phenotype in human and indicative of progression from DCIS1 to invasive breast cancer and metastasis, the data were reanalyzed. LCM Waptag1 DCIS samples were regrouped , to remove D_3539, the LCM DCIS sample that clustered closest to the tumor samples (Figure 3.4), leaving five DCIS samples that were the most similar to B6. The pairwise contrast with whole B6 resulted in 8238 significantly expressed genes ($q \leq 0.05$). Comparison of the mouse genes from the LCM DCIS samples to the 6771 human DCIS1-Norm (see above) genes, revealed an overlap of 1488 differentially expressed genes, with the same directionality. Since DCIS2 was predicted to be less likely to be invasive[27], these 1488 genes were then compared with the human DCIS1-DCIS2 contrast (data not shown). These subtractions resulted

in 63 up-regulated genes, which predicted progression to invasion in three independent human breast cancer datasets (Kaplan-Meier log-rank test; $p \leq 0.017$; Figure 4.6; Table 4.3). Although a subset of the 65 down-regulated genes was not prognostic (data not shown), among the 128 total commonly up- and down-regulated genes, 14 of the 63 up-regulated showed significant enrichment for genes indicative of poor breast cancer survival in human, a much greater proportion than would be expected by chance alone ($p=1.1 \times 10^{-19}$; Figure 4.6.B; Table 4.3). Within the 2097 transcriptional changes most common between species (detailed throughout this chapter), expression levels of all 63 up-regulated genes were significant, showing substantial fold changes in human DCIS1 and Waptag1 DCIS, when compared with normal controls (Table 4.3 and Fancher K *et.al.*, future publication).

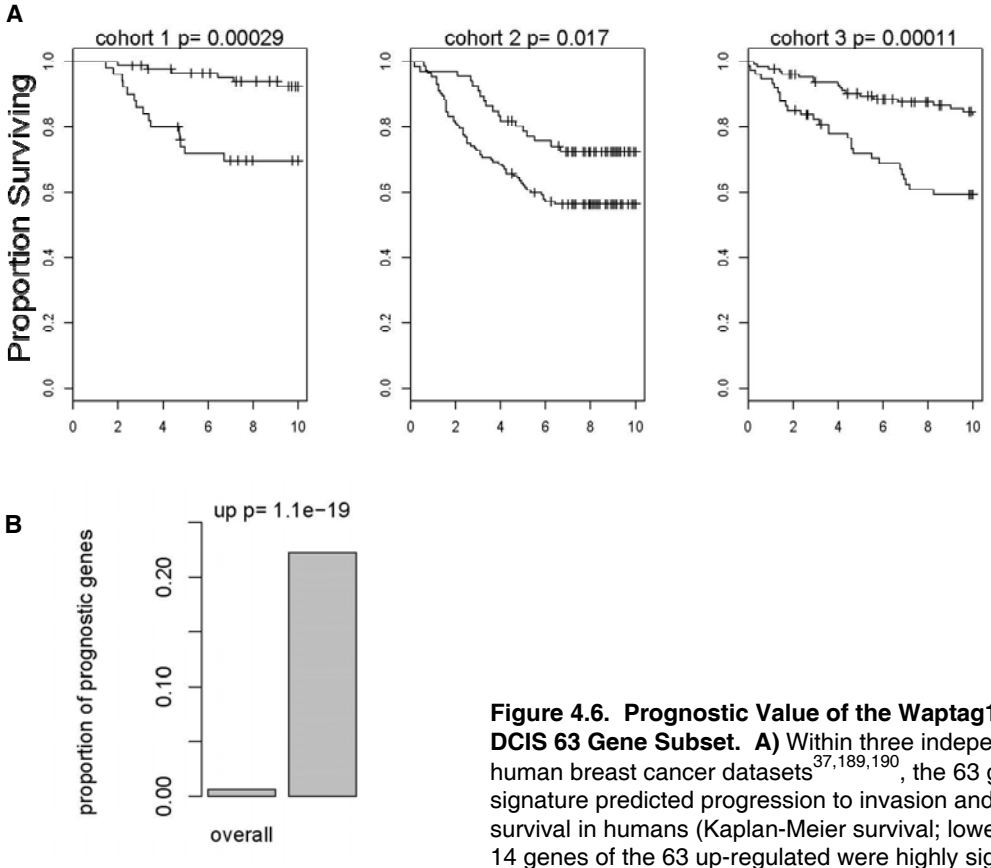


Figure 4.6. Prognostic Value of the Waptag1 *Early* DCIS 63 Gene Subset. A) Within three independent human breast cancer datasets[37,189,190], the 63 gene signature predicted progression to invasion and decreased survival in humans (Kaplan-Meier survival; lower lines); **B)** 14 genes of the 63 up-regulated were highly significant ($p=1.1 \times 10^{-19}$; Cox-hazard regression analysis). [Abbreviations: cohorts1-3, human datasets 6-8; p, p-value; overall, combined data, 63 up- and 65 down-regulated genes.]

MGI Mouse Symbol	FoldChange Wapt1 LCM DCIS-whB6	Human Symbol	Fold Change Human DCIS1-Norm	MGI Name
Aurkb	14.53	AURKB	14.02	aurora kinase B
Faim2	1.85	FAIM2	9.13	Fas apoptotic inhibitory molecule 2
Cdc45l	5.02	CDC45L	8.45	cell division cycle 45 homolog (S. cerevisiae)-like
Mcm6	10.96	MCM6	11.05	minichromosome maintenance deficient 6(S. cerevisiae)
Abp1	1.98	ABP1	7.84	amiloride binding protein 1 (amine oxidase, copper-containing)
Tpx2	11.44	TPX2	7.78	TPX2, microtubule-associated protein homolog
Dmbt1	1.44	DMBT1	6.39	deleted in malignant brain tumors 1
*2810417H13Rik**	5.87	KIAA0101 *	7.58	RIKEN cDNA 2810417H13 gene
*Aurka **	3.23	AURKA *	6.68	aurora kinase A
Ccna2	7.37	CCNA2	5.96	cyclin A2
Lad1	6.93	LAD1	5.78	ladinin
*Kif11 **	6.38	KIF11 *	5.02	kinesin family member 11
*Top2a **	7.05	TOP2A *	6.34	topoisomerase (DNA) II alpha
Krt6a	1.62	KRT6A	4.44	keratin 6A
*Hmmr **	2.63	HMMR *	3.75	hyaluronan mediated motility receptor (RHAMM)
*Cdc20 **	13.48	CDC20 *	4.55	cell division cycle 20 homolog (S. cerevisiae)
Nudt1	2.53	NUDT1	3.47	nudix (nucleoside diphosphate linked moiety X)-type motif 1
Mcm2	4.37	MCM2	3.95	minichromosome maintenance deficient 2 mitotin
Pik3cb	2.92	PIK3CB	3.05	phosphatidylinositol 3-kinase, catalytic, beta polypeptide
Tk1	5.64	TK1	3.60	thymidine kinase 1
Krt86	1.40	KRT86	3.20	keratin 86
Gpr161	1.20	GPR161	2.87	G protein-coupled receptor 161
Trip13	2.12	TRIP13	3.55	thyroid hormone receptor interactor 13
Slc7a6	1.24	SLC7A6	2.97	solute carrier family 7 (cationic amino acid transporter), 6
St14	5.12	ST14	2.95	suppression of tumorigenicity 14 (colon carcinoma)
Abtb2	3.80	ABTB2	3.26	ankyrin repeat and BTB (POZ) domain containing 2
*Rrm2 **	19.12	RRM2 *	3.21	ribonucleotide reductase M2
*Cenpa **	11.55	CENPA *	3.04	centromere protein A
Cdc6	8.23	CDC6	4.69	cell division cycle 6 homolog (S. cerevisiae)
Fanca	2.45	FANCA	2.51	Fanconi anemia, complementation group A
*Ube2c **	4.05	UBE2C *	3.12	ubiquitin-conjugating enzyme E2C
B830045N13Rik	1.37	FAM5C	2.52	RIKEN cDNA B830045N13 gene
Ccnb1	35.29	CCNB1	2.88	cyclin B1
Nfrkb	1.58	NFRKB	2.38	nuclear factor related to kappa B binding protein
Ada	2.45	ADA	2.39	adenosine deaminase
Fgb	1.95	FGB	3.58	fibrinogen, B beta polypeptide
Marveld3	4.00	MARVELD3	3.53	MARVEL (membrane-associating) domain containing 3
*Bub1b **	11.97	BUB1B *	2.74	budding uninhibited by benzimidazoles 1 homolog, beta
Amd1	1.46	AMD1	2.82	S-adenosylmethionine decarboxylase 1
Stmn1	8.17	STMN1	2.17	stathmin 1
Fadd	1.20	FADD	2.00	Fas (TNFRSF6)-associated via death domain
*Cdkn3 **	3.27	CDKN3 *	2.30	cyclin-dependent kinase inhibitor 3
Arhgef4	1.83	ARHGEF4	2.05	Rho guanine nucleotide exchange factor (GEF) 4
Srpk3	1.77	SRPK3	1.87	serine/arginine-rich protein specific kinase 3
Stk10	1.31	STK10	2.04	serine/threonine kinase 10
Col13a1	1.38	COL13A1	1.83	collagen, type XIII, alpha 1
Mall	1.59	MALL	2.15	mal, T-cell differentiation protein-like
Nptx2	1.59	NPTX2	2.37	neuronal pentraxin 2
*Cdc2a **	3.87	CDC2 *	1.75	cell division cycle 2 homolog A (S. pombe)
Smarcd2	4.53	SMARCD2	1.90	SWI/SNF related, matrix associated, regulator of chromatin, subf
BC039210	1.86	FAM38A	2.03	cDNA sequence BC039210
Srd5a1	1.82	SRD5A1	2.19	steroid 5 alpha-reductase 1
Cdk8	2.38	CDK8	1.63	cyclin-dependent kinase 8
Tram2	1.54	TRAM2	1.73	translocating chain-associating membrane protein 2
Gpx2	8.08	GPX2	1.54	glutathione peroxidase 2
*Spag5 **	2.73	SPAG5 *	1.95	sperm associated antigen 5
Traf3ip2	2.06	TRAF3IP2	1.77	Traf3 interacting protein 2
Myo1e	2.23	MYO1E	1.66	myosin IE
Orc5l	1.77	ORC5L	1.52	origin recognition complex, subunit 5-like (S. cerevisiae)
Camsap1	1.45	CAMSAP1	1.45	calmodulin regulated spectrin-associated protein 1
Tagln2	2.37	TAGLN2	1.78	transgelin 2
*Ccnb2 **	8.43	CCNB2 *	1.91	cyclin B2
Prpf19	1.91	PRPF19	1.96	PRP19/PSO4 pre-mRNA processing factor 19 homolog

Table 4.3. Waptag1 *Early* DCIS Genes Predictive of Poor Prognosis in Human. These sixty-three genes were over expressed in Waptag1 early DCIS and Human DCIS1, but not in human DCIS2 lesions and may be predictors of progression, invasion, metastases and/or reduced survival in human breast cancer. Genes were statistically significant in each species-specific contrast ($q \leq 0.05$). [Abbreviations: MGI, Mouse Genome Informatics; Wapt1, Waptag1; LCM, laser-capture microdissection; DCIS, ductal carcinoma *in situ*; whB6, whole glands from C57BL/6J; *, indicates the 14 most significant prognostic genes represented in Figure 4.6.B]

4.2. Cross-Species Differences Found in Waptag1 Tumorigenesis

It is critical to consider divergence when investigating cross-species comparisons. Species-specific gene alterations may fall into the same or similar ontologies, thereby further supporting the commonalities among mouse and human expression changes during tumorigenesis. Indeed, deregulation of apoptosis is one example where changes in different genes, but with similar GO categories are found. Aside from gene expression changes common between species, Waptag1 DCIS and tumors revealed up-regulation of supplemental *cell cycle*, *cell division*, and *DNA replication* genes, as well as the associated *nucleic acid binding* and cellular components *nucleus* and *chromosome, pericentric region*. Based on the effects of the SV40 Tag transgene, these findings seem appropriate for this model. Complementary to the chromatin remodeling changes and rampant expression of retrotransposons mentioned above, aberrant gene expression abounds in Waptag1 DCIS. The GO categories *gene expression, regulation of gene expression, transcription, regulation of transcription* were enriched among genes up-regulated in Waptag1 DCIS. Waptag1 exhibits even further down-regulation of genes involved in energy derivation than was common with human DCIS [i.e. *generation of precursor metabolites and energy* (40 additional genes) *oxidoreductase activity* (148 more genes), *fatty acid metabolism / lipid metabolism* (50 and 129 genes, respectively), and *cofactor binding* (50 extra genes), specifically *coenzyme binding* (32 other genes) as well as the cellular components *mitochondrial part* (103 more genes), specifically *mitochondrial inner membrane* (73 extra genes)]. However, these findings could be due to orthological differences, as suggested by the further down-regulation of energy-related genes in the human results, not common to Waptag1 (see below). Mice exhibit a faster basal metabolic rate than humans and may account for some additional down-regulation of genes involved in metabolism; alternatively, this could be a consequence of the SV40 Tag oncogene. *Vitamin binding* (33 genes) was another GO category of genes down-regulated in Waptag1 DCIS versus B6 controls that was not discovered among human results, perhaps due to lack of a human ortholog for several of these genes. Vitamin binding is required to regulate normal metabolism within cells especially antioxidants, such as vitamins A, C & E, can reduce the potential direct oxidative stress damage to a cell. Interestingly, five of the genes exhibiting

significant down-regulation normally bind vitamin C (ascorbic acid). Therefore, it seems complimentary to find down-regulated expression of genes involved in vitamin binding alongside genes down-regulated in various metabolic processes as well as those involved in oxidative stress response.

4.2.1. DNA Repair Mechanisms in Waptag1 cells of DCIS

One striking difference among the global gene expression changes in DCIS between species was the up-regulated expression of genes in the GO biological process *DNA repair*, specifically *double-strand break repair* in LCM Waptag1 DCIS. Since transgene-induced SV40 Tag proteins bind to and inhibit NBS1, Nijmegen breakage syndrome protein 1, a protein involved in double-strand break repair, and the mitotic checkpoint protein BUB1(budding uninhibited by benzimidazoles 1 homolog), genomic instability results, without the need for down-regulation of DNA repair. Conversely, genes in the GO categories *regulation of DNA repair*, in particular *positive regulation of DNA repair* genes were down-regulated at the cutoff threshold in human DCIS1. The only two common DNA repair genes with alternate directionality (up-regulated in Waptag1 and down-regulated in human) are the familial breast cancer prognostic, Brca1/BRCA1 and the ubiquitin molecule, Ube2n/UBE2N. Although the particular genes may vary between mouse and human DCIS, both species deregulate DNA repair mechanisms to promote tumor progression. However, this brings up a technical flaw in these comparisons: of 80 *DNA repair* genes significantly over expressed in Waptag1 DCIS, 44 were not represented on the human microarray chips.

4.3. Cross-Species Divergence Discovered in Human Carcinogenesis

The most significant GO term represented amid all up-regulated genes in human DCIS was *developmental process*. Related GO categories include *anatomical structure development, multicellular organismal development, system development, organ development, nervous system development*. Given that developmental genes were previously found over represented among the 969 genes up-regulated in the 2097 list, this suggests that genes in these categories show further up-regulation (of gene numbers and expression levels) than is common between species. Up-regulated expression of ‘developmental’ genes is frequently associated with cancer. For

example, *Pthrp*, parathyroid hormone-related protein, is normally expressed during (and is essential for) development of the mammary gland[201]. Over-expression of *Pthrp* has been detected in human breast cancer and is usually associated with invasive properties, bone metastases, and a poor prognosis[202-207]. Additionally, the Notch, Wnt, and Fgf families of genes are also included in these GO categories and these genes are not only important for development, but play significant roles in signal transduction in cancer. When *Notch1*, *Notch4*, *Wnt1* or *Fgf3* are aberrantly over expressed in mice, either by integration of the mouse mammary tumor virus or by using an MMTV promoter/enhancer, mammary cancer results (for review[208-210]). *Cell-cell signaling* and *signal transduction* genes are over-represented in human DCIS1 as well, though some of these genes could represent differences between laser-captured Waptag1 DCIS cells and the DCIS-enriched human samples.

Human DCIS1 samples showed further down-regulated expression of genes involved in energy metabolism. Genes in the following categories showed additional down-regulation: *generation of precursor metabolites and energy* (32 extra genes), *lipid metabolism* (71 more) genes, specifically (59) *cellular lipid metabolic process* genes, *oxidoreductase activity* (18 genes), in human DCIS1 samples compared with normal breast controls. Undoubtedly, there will be tissue-, environment-, species-specific, and even individual differences within these altered energy pathways. Nonetheless, in both Waptag1 and human DCIS, normal aerobic respiration and energy production via the electron transport chain is inhibited early in the tumorigenic process; some of the genes are orthologous and similarly down-regulated, whereas other genes are uniquely down-regulated in each species, but both show further down-regulation with increasing severity of the disease.

4.3.1. Defense Response Genes Over Represented in Human DCIS1

One striking distinction between the human DCIS1 and the laser-captured DCIS mouse results is the up-regulated expression of genes involved in *response to stimulus*. In total, 64 *defense response* genes, specifically 35 genes involved in *response to biotic stimulus*, and fourteen *cellular defense response* genes were found. At least 11 of these genes either did not have mouse orthologs or were represented in the mouse annotation. Interestingly, *response to*

stress genes were both up- and down-regulated among human DCIS results. Genes categorized in the *T cell activation*, *response to virus*, and *lymphocyte / leukocyte activation* GO categories were significantly over-represented among genes up-regulated in human DCIS1, but not in laser-capture microdissected Waptag1 DCIS cells. Of course it is intriguing to find immune system genes up-regulated early in human carcinogenesis as was found previously in Waptag1 whole tumors (though not in the laser-capture results). A few stromal-related gene changes were found in the human DCIS1 samples, but these were obviously not in the laser-captured Waptag1 DCIS samples used for these cross-species comparisons. Previous preliminary analyses suggested that human samples 'enriched' in DCIS and linearly amplified showed greatest similarity with linearly-amplified laser-captured Waptag1 DCIS cells (data not shown). However, the findings mentioned above as well as the enrichment of genes involved in *locomotor behavior*, *localization of cell*, and *cell motility* (also found among unique human DCIS results) suggest that there may be deregulation of analogous stromal genes between human and mouse DCIS.

4.4. Summary

By comparing human DCIS transcriptional changes with genes differentially expressed in Waptag1 DCIS, over two thousand genes were found to be commonly up- or down-regulated across species. Genes involved in regulating the cell cycle, cell division, DNA replication, and chromatin remodeling were up-regulated (compared with appropriate normal controls), and in addition, retrotransposons were over expressed. A sixty-three gene signature (common to mouse and human DCIS) exhibited significant prognostic ability and was associated with decreased survival in three independent human breast cancer datasets. Genes commonly down-regulated in DCIS across species were classified under the GO categories generation of precursor metabolites and energy, specifically fatty acid metabolism, electron transport, and oxidoreductase activity, suggesting that ATP is generated via an alternative energy pathway, such as glycolysis, *early* in tumor progression.

As one of the 'Hallmarks of Cancer' the finding that escape from cell cycle regulation in order to attain uncontrolled growth occurs early in the tumorigenic process, though intriguing, is not novel[11]. However, this result coupled with over expression of chromatin remodeling genes and

retrotransposons and a metabolic switch in the bioenergetics of the DCIS lesions presents a unique cross-species portrait of the early stages of mammary tumor progression *in vivo*. To complete this evaluation of transcriptional changes in DCIS, however, the full contribution of the stroma in the human DCIS microenvironment is essential, to identify potential stroma-related early markers of progression to invasion. Likewise, isolation of paired tumor cells and stromal components within lesions of Waptag1 DCIS and tumors could expose the real contribution of the stroma to mammary tumor progression.

Chapter 5

UP-REGULATION OF RETROTRANSPOSONS IN MOUSE AND HUMAN CARCINOGENESIS

Abundant expression of retrotransposons was a common finding among ten microarray datasets, containing either mouse or human mammary cancer data. This finding was shared among all mouse mammary tumors tested, regardless of genetic background, and in seven subtypes of human breast cancer. To date, six different classes of retrotransposable elements have been described: Long Terminal Repeat (LTR) Class I, II, & III family members and Long and Short Interspersed Nuclear Elements (LINEs and SINEs). As mentioned previously, although specific retrotransposons may differ among tumor models or human breast cancer subtypes, the commonality of the causes and mechanisms underlying transcriptional activation of these normally silenced elements is the focus of interest.

5.1. Retrotransposons are Over Expressed in Five Independent Transgenic Mouse Mammary Cancer Models

Retrotransposons were initially discovered to be over expressed in Waptag1 DCIS and tumors by uncovering the origin of the most significantly up-regulated unknown/unannotated clones on the NIA arrays, when compared to normal controls. To confirm these original findings independent Waptag1 samples were run on a different microarray platform (Affy). Representatives from all classes of retrotransposons were up-regulated in early DCIS, advanced DCIS, and papillary tumors when compared with normal C57BL/6J mammary glands (Figure 5.1.A). Using microarray data across multiple platforms and among several transgenic mammary tumors, we discovered over expression of retrotransposons was not confined to Waptag1 or the C57BL/6 genetic background (Figure 5.1). Rather, mice of a different genetic background and transgenic for *Neu*, *Int3*, *Myc*, or *Wnt1* showed up-regulation of multiple classes of retrotransposons in their mammary tumors, compared with their appropriate normal FVB/NJ mammary glands (Figure 5.1.B).

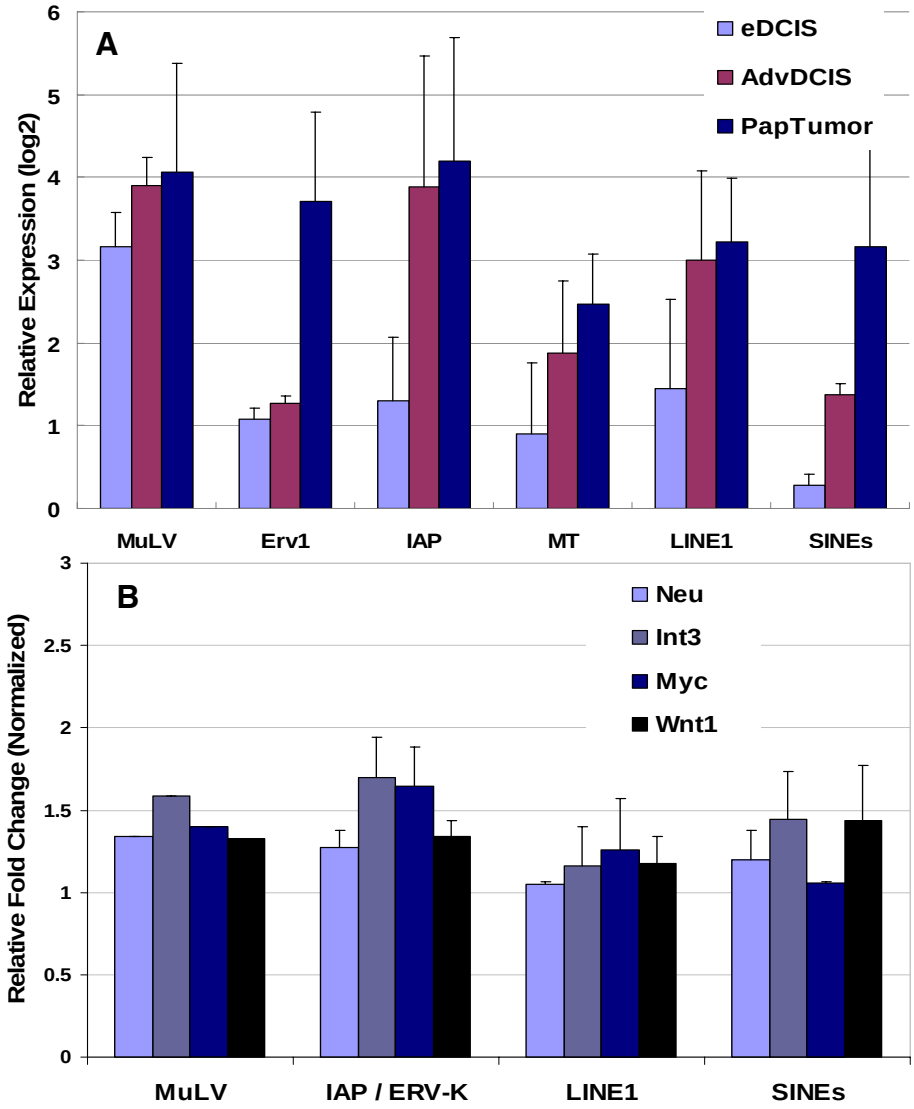
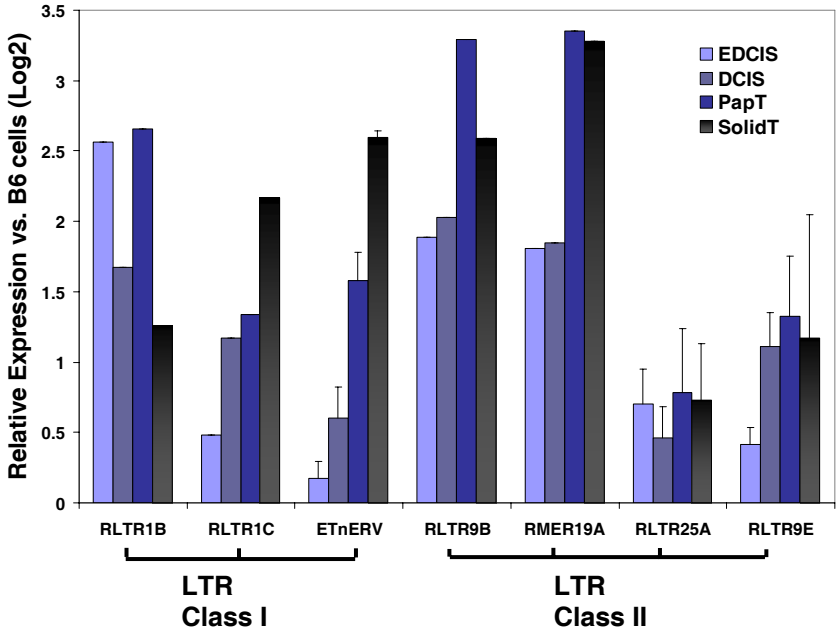


Figure 5.1. Up-Regulation of Retrotransposons in Five Mouse Models of Mammary Cancer. **A)** Expression of all classes of retrotransposons in ***Waptag 1*** mammary glands increases with tumor progression. Combined data from NIA 15K clone set cDNA arrays (NIA) and GeneChip^{R} Mouse Genome 430 2.0 Affymetrix Arrays (Affy) are shown. Representative retrotransposons from: LTR Class I, three MuLV clones(NIA) & two Edv1(Affy); LTR Class II, two IAP(NIA); LTR Class III, two MT clones(NIA); Non-LTR, two LINEs(NIA) & two SINEs(Affy). **Legend:** eDCIS, early Ductal Carcinoma *in situ*; AdvDCIS, advanced DCIS; PapTumor, papillary adenocarcinoma **B)** Over-expression of repetitive elements in mamamary carcinomas in FVB/NJ mice transgenic for ***Neu***, ***Int3***, ***Myc***, or ***Wnt1***. Data from the 22K Compugen Mouse Release 2.0 oligonucleotide probe microarrays was normalized due to spatial variation across some arrays. [All expression values are relative to normal non-lactating mammary glands and are statistically significant (ANOVA; $p < 0.05$). Error bars: standard deviation across clones of the same Class/Family.]

To determine whether retrotransposon over expression was contributed by the tumor cells or the stroma, specific tumor cells in Waptag1 glands were investigated. Using laser capture microdissection on frozen sections of mammary glands and tumors taken throughout the lifespan of Waptag1, we collected pre-tumor cells of early DCIS and advanced DCIS, papillary tumor cells,

solid, invasive tumor cells, and normal C57BL/6J control epithelial cells. After linear amplification and subsequent microarray analysis, we found that indeed expression of retrotransposons occurs within the tumor cells, specifically (Figure 5.2). Although only MaLR, MLT1K and MT-int LTR Class III elements were found consistently up-regulated in both laser captured cells and whole glands/tumors, we did find several LTR Class I and II as well as many LINEs and SINEs over-expressed compared with respective normal controls (Figure 5.2. and Table 5.4). This may suggest that different retrotransposons may be expressed in tumor versus surrounding stromal cells. Perhaps, in response to tumor cells, stromal cells likewise become aberrant.

Figure 5.2. Up-Regulation of Retrotransposons in Laser Capture Microdissected (LCM) Cells Across Stages of Waptag1 Tumorigenesis. Log2 of the expression levels of LCM, linearly amplified cells at each stage of Waptag1 tumor progression relative to LCM control C57BL/6J (B6) epithelial cells. Data from Affy arrays. Representative LTR Class I and II retrotransposons: two probesets ETnERV; three probesets each RLTR25A and RLTR9E; Legend: eDCIS, LCM early ductal carcinoma *in situ* cells; DCIS, LCM advanced DCIS cells; PapT, LCM papillary adenocarcinoma cells; SolidT, LCM solid / invasive tumor cells. [All expression values are statistically significant (ANOVA; $q<0.05$); Error bars: standard deviation of several clones of the same Class/Family.]



5.2. Retrotransposons are Abundant in Human Breast Cancer, DCIS, and Atypical Ductal Hyperplasia

To determine whether retrotransposon expression is a general early feature of human breast carcinogenesis, publicly available databases were searched for human breast tumor or DCIS (or other early pre-tumor stage) data and queried for retrotransposon sequences (see methods).

We initially found over-expression of multiple classes of retrotransposons in three different human breast cancers, luminal, basal, and ERBB2-positive tumors[4,5] when compared with normal breast tissue (Table 5.1).

After manually surveying the results of two datasets, a few human breast cancer-specific retrotransposons were found (Table 5.2). These preliminary results prompted us to investigate whether retrotransposons found early in human carcinogenesis might be predictive of cancer recurrence. Indeed, in atypical ductal hyperplasia (ADH) samples, from patients with a previous cancer history, retrotransposons from several different classes was up-regulated (Table 5.3), when compared to ADH samples from patients without former cancer incident.

	Family / Class	(Range of Expression)
LTR	ERV1 / MER57	(9.0 – 31.2)
LTR	HERV-K / LTR5B	(2.2)
LTR	ERV-L / HERV-L	(2.1 – 2.6)
	MaLR / MLT1	(1.4 – 3.5)
Non-LTR	LINEs / L1, L2, L3, L4	(1.5 – 5.6)
	SINEs / Alu	(1.8 - 8.2)

Table 5.1. Retrotransposon Over-Expression in Three *Human* Breast Cancer Subtypes. Range of expression shown is relative fold change, tumors (luminal, basal or ERBB2+) compared with normal breast tissue[4,5]. If no range, only one repeat was expressed. All comparisons are significant (two-tailed t-test as raw data was unavailable; $p \leq 0.05$).

R/ MAANOVA analysis of human microarray data, from samples enriched in human DCIS and IBC, invasive breast cancer (kindly provided by Drs. Li and Allred, Washington University, St. Louis, MO) and normal breast controls (from public databases), was performed, as with the mouse data. In both human DCIS and IBC, multiple probe sets exhibited over-expression of several classes of retrotransposons, when compared with normal breast tissue (Figure 5.3). The variety and level of over expression of retrotransposons in Waptag1 tumor progression (A) and in human DCIS1 (B) is found in Table 5.4. These results suggest that retrotransposon activation and expression early in the tumorigenic process may be a common feature to both mouse and human mammary carcinogenesis.

As mentioned previously, chromatin maintenance and modification genes are also over-expressed early in tumor progression, suggesting that chromatin remodeling may be a common underlying cause for this retrotransposon expression. Although yet unclear, the cause of retrotransposon unsilencing, chromatin remodeling and aberrant gene expression in pre-tumor

cells may be a response to stress, the result of epigenetic alterations, due to accumulation of un-repaired DNA double strand breaks, or a combination of these factors[216-218].

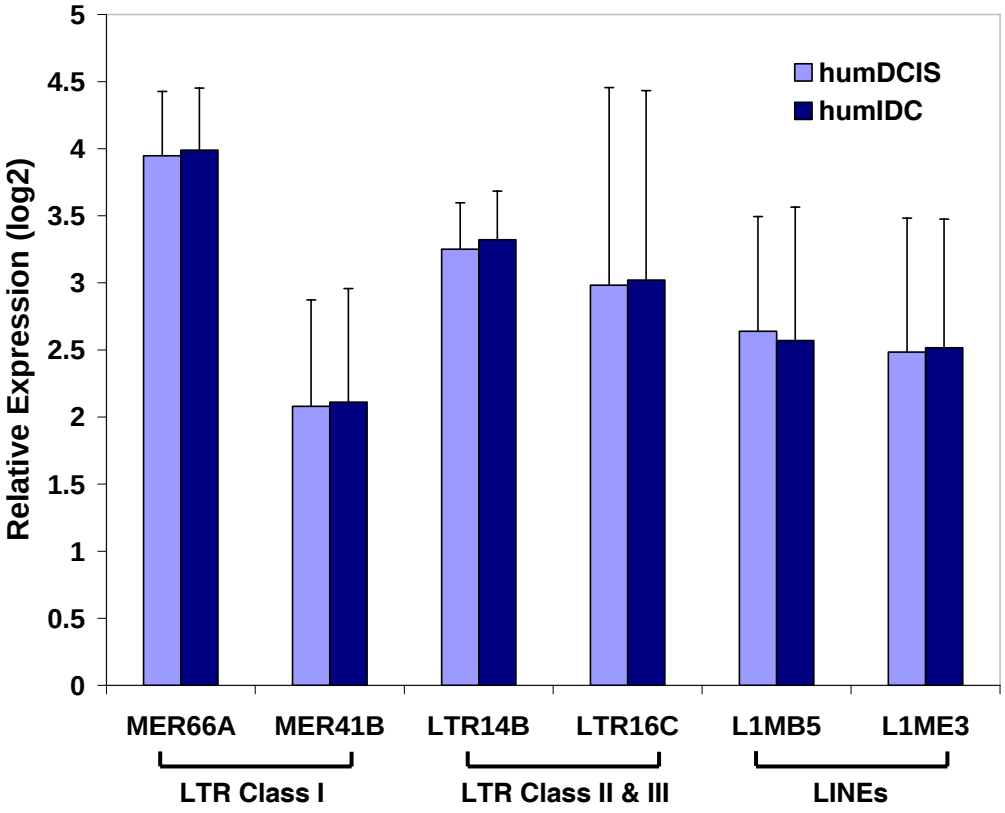
Table 5.2. Over-expression of Retrotransposons in Human Breast Tumors. Results are combined from two datasets: 'dataset1' compared/contrasted 32 luminal, 10 basal, 9 ERBB2+ tumors with 4 normal breast tissues[4,5]; 'dataset2' compared/contrasted 27 luminal, 16 basal, and 6 apocrine tumors[6]. Retrotransposon candidates as markers for *specific* tumor types as shown; tumors were compared with normal glands (dataset1) or other tumor types (dataset2); a consensus of both comparisons is shown for luminal and basal. [All comparisons are statistically significant (two-tailed t-test as the raw data was not available; $p \leq 0.05$). Blank spaces indicate no elements could be found.]

	ERBB2+ (dataset1 only)	Luminal	Basal	Apocrine (dataset2 only)
LTR	ERV1 / *MER57-int*			ERV1 / *LTR12C & MER65B*
LTR	ERVL / *HERV-L* ERVK / *LTR5B*	MaLR / *MLT1K*		
Non LTR	LINEs / *L1M5* SINEs / *AluJb, Jo & AluSg/x*	LINEs / *CR1 L3/ L1MC5*	LINE / *L1ME3A* SINE / *Alu: FAM*	SINE / *FLAM_A*

		ADH, with Cancer		
		Repeat family	Retrotransposon	Fold change
LTR	Class I	*ERV1*	MER110	2.65
			LTR19-int	2.16
LTR	Class II	*HERV-K*	LTR13A	2.81
		HERV-L	LTR69	2.25
Non LTR	Non-LTR	*LINEs*	L1MA6	2.83
			L1MC4	2.06
			L1PA5	2.58
		SINEs	AluSc	2.1 – 2.58
			AluSg	1.9 - 2.38

Table 5.3. Retrotransposon Markers of Progression in Human Atypical Ductal Hyperplasia. Retrotransposon overexpression was investigated in breast samples containing atypical ductal hyperplasia (ADH); four samples from patients *with* a history of *cancer* and four from patients with *no* history of *cancer*[3]. [Fold change is: ADH, with cancer history compared to ADH, *no* cancer; a range of fold changes reveals expression of *multiple* clones of the *same* retrotransposon (two-tailed t-test as the raw data was unavailable; $p \leq 0.05$).]

Figure 5.3. Retrotransposons Over-Expressed in Human Breast Carcinogenesis. Twenty-seven samples enriched in ductal carcinoma *in situ*, **DCIS**, and twenty-four tumors of invasive breast cancer, **IBC** (run on U95Av2 Affy Arrays) were compared to seven normal breast tissue samples (run on U133plus2 Affy Arrays). Representative LTR Class I, II, III and non-LTR retrotransposons are shown as log2 of the mean expression levels of human DCIS or IBC compared to normal breast; three probesets each MER66A, LTR14B, and L1ME3; two, MER41B; four each, LTR16C and L1MB5. [All expression values are statistically significant (ANOVA; $q<0.01$); Error bars: standard deviation across clones of the same Class/Family after outliers were removed.]



5.3. Summary

Though many questions remain, regarding the cause of retrotransposon up-regulation early in tumor progression, the findings presented here suggest that unsilencing of retrotransposons is associated with the initial stages of cellular transformation. Across five different mouse models of mammary cancer on two different genetic backgrounds, five human breast cancer subtypes, human atypical hyperplasia, and DCIS in Waptag1 and human, retrotransposons were consistently among the most differentially over expressed 'genes' in all comparisons. Indeed,

these may be useful markers in the future for detection of early stages of human breast carcinogenesis.

A Array_ID	Repeat Class	Repeat Family	Strand	BPs Left in Query	Fold Change Waptag1 LCM DCIS vs. wB6	Fold Change Waptag1 wPapT-wB6	Fold Change Waptag1 LCM PapT-LCM B6	Fold Change Waptag1 LCM SolidT- LCM B6
1425065_at	MER21B	LTR/ERV1	+	-139	2.27	4.59	4.17	2.93
1428763_at	MER21B	LTR/ERV1	+	-211	1.87	1.64	1.43	1.21
1424717_at	MER31A	LTR/ERV1	C	0	1.21	2.20	3.48	3.39
1424118_a_at	MER34A1	LTR/ERV1	+	-8	6.23	10.70	14.32	17.03
1435628_x_at	RLTR1B	LTR/ERV1	+	0	3.16	1.20	6.32	2.39
1456340_at	RLTR1C	LTR/ERV1	C	-38	2.99	1.39	2.53	4.50
1446714_x_at	ETnERV	LTR/ERVK	+	-5	1.61	1.16	3.29	6.19
1459452_at	ETnERV	LTR/ERVK	+	-5	1.31	1.14	2.71	5.90
1433361_at	MMERVK10C	LTR/ERVK	+	-24	1.30	1.39	1.18	2.28
1455581_x_at	MMERVK10C	LTR/ERVK	+	-309	4.03	1.04	1.67	1.21
1434280_at	MMETn-int	LTR/ERVK	+	0	4.92	7.41	5.98	9.13
1434279_at	MMETn-int	LTR/ERVK	+	0	-1.31	3.71	1.22	1.14
1454904_at	MMETn-int	LTR/ERVK	+	0	-2.04	3.34	1.39	1.54
1457626_at	MYSERV6	LTR/ERVK	+	-5	6.77	2.08	2.04	2.97
1419773_at	RLTR10	LTR/ERVK	+	-137	6.96	1.11	2.60	2.91
1450165_at	RLTR13B4	LTR/ERVK	+	-81	1.77	2.10	2.68	1.96
1422719_s_at	RLTR25A	LTR/ERVK	C	-344	-1.22	1.94	2.45	2.28
1446521_at	RLTR25A	LTR/ERVK	+	-274	3.07	1.16	1.57	1.41
1432155_at	RLTR25A	LTR/ERVK	C	-168	1.32	1.05	1.32	1.40
1419658_at	RLTR9B	LTR/ERVK	+	-21	7.31	2.64	9.78	6.02
1442192_at	RLTR9E	LTR/ERVK	C	-149	2.01	2.77	3.25	4.38
1450165_at	RLTR9E	LTR/ERVK	+	-15	1.77	2.10	2.68	1.96
1417277_at	RLTR9E	LTR/ERVK	+	0	2.25	1.40	1.80	1.34
1421701_at	RLTR9E	LTR/ERVK	+	-41	1.27	1.04	1.18	1.22
1451905_a_at	RMER19A	LTR/ERVK	C	0	4.44	4.35	10.20	9.71
1428105_at	RMER20A	LTR/ERVK	+	-4	2.31	8.51	8.22	10.41
1455674_at	RMER4A	LTR/ERVK	C	-271	1.09	2.30	2.87	3.29
1456803_at	RMER4A	LTR/ERVK	+	-164	1.48	1.20	1.51	1.73
1435973_at	RMER4A	LTR/ERVK	+	-186	3.32	-1.07	2.16	1.60
1430651_s_at	MLT1K	LTR/MaLR	+	-15	4.06	1.72	2.62	2.38
1450156_a_at	MT-int	LTR/MaLR	C	-13	5.06	13.00	11.31	9.99
1427541_x_at	MT-int	LTR/MaLR	C	-1	3.53	4.89	8.75	7.41
1452304_a_at	L3	LINE/CR1	+	-45	1.84	2.75	2.22	1.54
1434043_a_at	L3	LINE/CR1	C	-6	3.03	1.23	1.69	1.72
1439907_at	L1MB4	LINE/L1	C	-11	2.11	1.39	2.20	3.78
1457067_at	L1MB4	LINE/L1	C	0	5.24	1.04	2.14	2.35
1430339_at	L1MC3	LINE/L1	+	-160	2.35	3.46	6.19	4.82
1436032_at	L1MC3	LINE/L1	C	0	2.79	1.64	4.00	2.53
1458130_at	L1ME1	LINE/L1	C	-236	4.41	2.39	3.53	6.02
1458130_at	L1ME1	LINE/L1	C	-20	4.41	2.39	3.53	6.02
1428911_at	L1ME1	LINE/L1	+	0	2.89	1.66	1.71	1.71
1444157_a_at	L1ME1	LINE/L1	C	-39	2.14	1.60	1.14	1.28
1456027_at	L1ME1	LINE/L1	+	-24	1.92	1.38	1.39	1.92
1444158_at	L1ME1	LINE/L1	C	-214	2.36	1.27	1.27	1.65
1454785_at	L1ME1	LINE/L1	C	-4	1.78	1.22	1.27	1.13
1459900_at	L1ME1	LINE/L1	C	-262	2.04	1.15	1.87	1.72
1459900_at	L1ME1	LINE/L1	C	-3	2.04	1.15	1.87	1.72
1437550_at	L1ME1	LINE/L1	C	-31	2.08	1.03	1.30	1.41
1437570_at	L1ME3B	LINE/L1	+	-144	4.47	2.57	2.66	3.68
1423103_at	L1ME3B	LINE/L1	C	-63	2.50	-1.17	1.27	1.24
1436472_at	Lx	LINE/L1	+	0	1.67	6.54	4.66	4.41
1440866_at	Lx	LINE/L1	+	0	-1.79	3.81	1.92	1.46
1442134_at	FLAM_A	SINE/Alu	+	0	2.41	2.03	5.66	7.31
1457636_x_at	FLAM_A	SINE/Alu	+	-3	3.73	1.21	1.72	1.61

Table 5.4. Individual Clones Representing Retrotransposons Expressed during Tumorigenesis in Waptag1 (A) or Human DCIS (B). Fold changes shown are significant (q<0.01). Multiple clone target sequences align to the same retrotransposon, thereby supporting true over-expression of that element. Repeat Class/Family as determined by RepeatMasker (see methods). (Any transposon represented by multiple clones that showed equal over- *and* under-expression was removed from the list.) [Abbreviations: DCIS, ductal carcinoma *in situ*; BPs, basepairs; LCM, laser capture microdissection; wB6, whole C57BL/6J glands; PapT, papillary tumor; SolidT, solid tumor; Norm, normal human breast.]

B Array_ID	Repeat Class	Repeat Family	Strand	BPs Left in Query	Fold Change Human DCIS1 vs. Norm
34702_f_at, 34703_f_at, 34704_r_at	HERVH	LTR/ERV1	+	0	**5.44, 3.14, 1.62**
32250_at	HUERS-P1	LTR/ERV1	+	0	**5.93**
1992_at	LTR26E	LTR/ERV1	+	0	**2.11**
37709_at	LTR37A	LTR/ERV1	+	-64	**3.30**
40700_at	LTR48	LTR/ERV1	+	-112	**3.31**
36105_at	LTR54B	LTR/ERV1	+	0	**8.77**
34842_at	LTR56	LTR/ERV1	+	-60	**47.77**
33983_at	LTR72	LTR/ERV1	+	-1	**1.91**
1748_s_at	LTR7B	LTR/ERV1	+	-328	**3.19**
31442_at & 1582_at	MER31-int	LTR/ERV1	+	0	**1.45 & 2.48**
33559_at	MER34B-int	LTR/ERV1	+	-509	**2.46**
31324_at	MER41B	LTR/ERV1	+	-100	**2.76**
37940_f_at	MER49	LTR/ERV1	+	0	**2.42**
34429_at	MER4A1	LTR/ERV1	+	0	**17.63**
41025_r_at & 41026_f_at	MER51E	LTR/ERV1	C	-7 & 0	**4.73 & 1.55**
36105_at	MER57A-int	LTR/ERV1	+	-255	**8.77**
35956_s_at	MER65-int	LTR/ERV1	+	-279	**11.51**
33503_at	MER66B	LTR/ERV1	+	-79	**4.32**
36286_at	MER89-int	LTR/ERV1	+	-205	**2.85**
1391_s_at & 35412_at	LTR14B	LTR/ERVK	+	0 & -124	**8.01 & 12.04**
35955_at	LTR5B	LTR/ERVK	+	0	**4.84**
33453_at	LTR16C	LTR/ERVL	+	-338	**28.12**
35915_at	LTR33	LTR/ERVL	+	-38	**9.56**
33453_at	LTR40b	LTR/ERVL	+	0	**28.12**
39059_at	LTR41B	LTR/ERVL	+	-11	**3.34**
31650_g_at	LTR69	LTR/ERVL	+	-2	**3.41**
34137_at	LTR75	LTR/ERVL	+	0	**2.34**
33077_at & 32949_at	LTR79	LTR/ERVL	+	-213 & -445	**3.46 & 1.78**
32949_at	LTR79	LTR/ERVL	+	-445	**1.78**
38162_at	MLT2B3	LTR/ERVL	C	-14	**4.60**
32198_at	MLT2B4	LTR/ERVL	C	-214	**17.17**
31652_at	MLT2F	LTR/ERVL	+	-1	**3.68**
34549_g_at, 1411_at, 1412_g_at	MLT1A0	LTR/MaLR	C	-212	**11.46, 4.72, 2.81**
31801_at	MLT1A0	LTR/MaLR	C	-52	**2.38**
34549_g_at, 1411_at, 1412_g_at	MLT1A0-int	LTR/MaLR	C	-4	**11.46, 4.72, 2.81**
41169_at	MLT1G	LTR/MaLR	+	-61	**5.26**
39217_at	MLT1G3	LTR/MaLR	C	-329	**6.56**
41522_at & 32203_at	MLT1J	LTR/MaLR	+ & C	0 & -146	**1.47 & 7.44**
33007_at	MLT1K	LTR/MaLR	C	-66	**4.40**
35117_at	MLT1L	LTR/MaLR	C	0	**2.56**
178_f_at & 1875_f_at	MSTA	LTR/MaLR	+	-64 & 0	**2.50 & 2.04**
32690_s_at	MSTB2	LTR/MaLR	+	-11	**2.01**
32321_at	MST-int	LTR/MaLR	+	-352	**31.47**
882_at	L3	LINE/CR1	C	-151	**48.64**
32539_at	L3	LINE/CR1	C	-40	**9.99**
35519_at	L3	LINE/CR1	C	-201	**1.45**
37880_at & 39953_i_at	L1M5	LINE/L1	C	-4 & 0	**20.21 & 5.76**
35526_at	L1MA4A	LINE/L1	+	-191	**6.27**
1032_at	L1MB4	LINE/L1	+	-66	**13.02**
36720_at	L1MB4	LINE/L1	+	0	**9.93**
1033_g_at	L1MB4	LINE/L1	+	-66	**4.94**
36718_s_at	L1MB4	LINE/L1	+	0	**2.56**
1305_s_at	L1MB5	LINE/L1	+	0	**13.50**
38716_at	L1MB8	LINE/L1	C	-3	**13.86**
32198_at	L1MC4	LINE/L1	+	-457	**17.17**
39673_i_at	L1MC4a	LINE/L1	+	-256	**5.41**
32198_at	L1MDa	LINE/L1	+	-109	**17.17**
31667_r_at & 31366_at	L1ME3	LINE/L1	+	-3 & 0	**7.72 & 2.63**
32689_s_at & 32691_s_at	L1ME3A	LINE/L1	C	0	**6.76 & 5.86**
35736_at	L1ME3B	LINE/L1	C	-126	**8.45**
40765_at	L1MEg	LINE/L1	+	-82	**6.95**
41624_r_at & 35602_at	L2	LINE/L2	C & +	-238 & -226	**6.06 & 5.65**
37569_at, 36702_at, 40507_at	L2a	LINE/L2	C, +, +	-31, -1, -1	**6.11, 5.96, 6.41**
1667_s_at & 34161_at	L2b	LINE/L2	C & +	-201 & -57	**11.71 & 8.46**
33568_at	L2c	LINE/L2	+	-85	**31.63**
1668_s_at	L2c	LINE/L2	+	-335	**12.46**
35602_at	L2c	LINE/L2	+	-129	**5.65**

Table 5.4. Continued.

Chapter 6

DISCUSSION AND CONCLUSIONS

Mouse models of human breast cancer remain important research tools: mouse models, destined to develop cancer, provide samples throughout the tumorigenic process, from the earliest changes through late stage progression; by utilizing different mouse models of mammary tumorigenesis and combining their histological and genetic information with that of human, progress can be made toward a more thorough understanding of mammary cancer; genetically engineered mice that model important elements of human cancers facilitate the elucidation of the molecules and pathways regulating normal development and disease, uncover *de novo* markers for detection of early alterations to be used for earlier diagnosis, and expedite testing of new therapeutic agents. Through comparison of mutual *early* mouse and human changes, predictive of progression to invasion, new prognostic markers may be identified, to more clearly differentiate the lesions of DCIS that will advance to invasive breast cancer from those that will not.

The research presented here directly compares global gene expression changes between a mouse model of DCIS, Waptag1, and human DCIS. Among 6945 orthologous genes represented on all microarrays, over two thousand genes were commonly differentially expressed in both Waptag1 and human DCIS, when compared with species-specific controls. Microarray, ontology, and pathway analyses revealed several novel findings characteristic of the early stages of mammary tumor progression. First, at the cellular level, pre-tumor cells mirror the transcriptional profile of advanced tumor cells, which implies that within lesions of DCIS, the surrounding tumor microenvironment defines the stage of tumor progression. Second, down-regulation of metabolic pathway genes involved in normal aerobic energy derivation supports the theory of an alternative energy pathway utilized throughout tumorigenesis. Third, up-regulation of chromatin modification genes and retrotransposons in DCIS in both species suggest that the epigenetic mechanisms that silence retrotransposons are dysfunctional. In addition to these novel findings, DCIS samples in human and Waptag1 exhibit uncontrolled cell growth, replication, and division as well as deregulation of apoptosis, allowing damaged cells to cycle and divide without potential death, thereby presenting several hallmarks of cancer at early stages of tumor progression.

Dissection of tumor cell versus stromal changes during Waptag1 tumor progression revealed the importance of identifying early pre-tumor cells amidst the overwhelming population of normal cells. For earlier detection and treatment, it is critical to evaluate and target the early pre-tumor cells, independently, especially if stromal changes are a secondary response due to cues from nearby tumor cells. If Waptag1 early DCIS is any indication, the uninhibited growth, massive down-regulation of normal aerobic respiration, and/or the bloom of retrotransposon expression can *best* be detected and at the earliest possible stage *without* the excessive 'noise' found in the surrounding cells of the mammary gland. Alternatively, stromal cell alterations are proving to be important for tumor progression, as harbingers of transformation and invasion, such as those characterized by epithelial to mesenchymal transitions during development and cancer[219-224].

Breast cancer treatments to eliminate tumor cells simultaneously destroy normal, healthy cells; thus, targeting only the tumor cells would be advantageous. Interestingly, an endogenous photosensitizer, protoporphyrin IX, accumulates in pre-tumor and tumor cells of Waptag1 (and other mouse and human tumors), especially after exongenous administration of a pre-cursor, 5-aminolevulinic acid, one that appears to have no ill side effects. Upon accumulation of protoporphyrin IX, pre-tumor cells can *either* be fluorescently detected at 380nm (for identification of tumor margins during surgery, for example) *or* excited with 635nm wavelength of light, to produce reactive oxygen species, ultimately resulting in cell death. Since a severe inhibition of oxidative response and/or neutralization of reactive oxygen species has been detected through analyses listed above *and* Waptag1 cells escape one apoptotic pathway, it would be interesting to use the Waptag1 model to attempt excitation-induced early tumor cell death. The current challenge is proximity of light source to the pre-tumor cells for excitation of the protoporphyrin IX; 635nm of light does not penetrate the skin. However, micro computed tomography (Micro CT), capable of imaging in three dimensions and exciting cells of interest, may in the future enable experiments such as these in mice. Interestingly, FVB/N-Tg(MMTVneu)202Mul/J mice which harbor an inactivated form of rat *Neu*, a receptor tyrosine kinase in the epidermal growth factor receptor family, under the control of the mouse mammary tumor virus (MMTV) promoter (MMTV-Neu[43]) present opposing mitochondrial energetics from that of Waptag1 (or human). Moreover,

MMTV-Neu mice may lend further insight into the mechanisms involved in mitochondrial changes, such as oxidative stress, down-regulation of electron transport, or increased glycolysis.

What roles do mitochondrial bioenergetics play in tumorigenesis? A few studies have shown, by combinatorial transforming manipulations *in vitro*, that from a single transforming 'hit' on upward, pre-tumorigenic cells undergo an early switch in energy metabolism[212,225]. However, to my knowledge, this phenomenon has not been described, as an entire process, throughout tumor progression *within a living system*. Others have published data (in supplemental gene lists) supporting the down-regulation of genes involved in normal aerobic respiration, similar to findings discussed here, yet these changes were not mentioned, even in the discussion, within the text of their publications[61,78]. Glycolysis is most likely the means of energy production in pre-tumor and tumor cells[18-21,211,213,214,226-228], when oxidative phosphorylation and generation of ATP via the electron transport chain are predominantly inhibited. By switching to glycolysis, tumor cells would gain several selective advantages: 1) energy independence, through utilization of an abundant nutrient, glucose, to generate ATP with fewer limitations (no need for oxygen, for example); 2) quicker energy, as the cancer cell produces more ATP through glycolysis than oxidative phosphorylation, given the same amount of time[229-231]; 3) production of biosynthesis pathway metabolites, via glucose breakdown, that are available for synthesis of nucleotides, lipids, etc. for rapid growth. Thus, cancer cells gain increased invasive and metastatic potential.

Clearly, energy-related transcriptional changes are prevalent throughout all stages of tumor progression in Waptag1 and human, but several questions still remain. Why are genes involved in oxidative phosphorylation and mitochondrial respiration down-regulated? Wouldn't it be more efficient to supplement normal electron-transport-produced ATP thru glycolysis? Although defects in mitochondrial energy metabolism have been associated with disease and cancer, (for review[211]), these mutations usually result in generation of reactive oxygen species (ROS), though there are exceptions. In Waptag1 and human tumorigenesis, either ROS are not being generated or if present, they are not being dealt with through normal mechanisms. Could mutations in key metabolic genes (or in their expression), as a result of un-repaired DNA breaks, improper joins, chromosomal abnormalities, or altered epigenetic mechanisms be another means by which tumor

cells promote tumorigenesis? In contrast to Waptag1 and human mammary cancer, the MMTV-Neu mouse model seems to launch a full-blown oxidative response (transcriptionally), exhibiting over-expression of lipid metabolism genes and cellular lipid metabolic processes, such as fatty acid oxidation, whereas transcription of genes from these categories is down-regulated in Waptag1 mouse or human samples (Fancher K *et.al.*, future publication). Up-regulation of many oxidative enzymes (*Lao1, Ddo, Cox7a1 & 8b, Acox2, Suox, Noxo1, Loxl3, Aoc2 & Aof1, Duox1* and *Duoxa1*) suggests that ROS are being produced in Neu tumors; many of these genes were down-regulated in Waptag1 and human. Expression of genes indicative of oxidative stress response and/or neutralization of ROS: the antioxidant superoxide dismutase 3 (*Sod3*), ROS storage and transport genes ceruloplasmin (*Cp*), and transferrin (*Trf*); oxidoreductase enzymes (*Txnip, Txndc1*); and carbonic anhydrase genes (*Car 9*, an hypoxia-related gene, as well as *Car2, 5b 6, 8,* and *12*), implies that cells are attempting to thwart ROS-induced cell injury, unlike Waptag1 mouse or human cancer cells. Lastly, the cellular components over represented in MMTV-Neu tumors are lysosome, endoplasmic retriculum, mitochondrion and peroxisome, all membrane-bounded organelles with enzymes to oxidize molecules, produce ROS, and/or neutralize ROS. These results are antithetical to those in Waptag1 and human. Further comparison of mouse models with human tumors may offer new insight into the mitochondrial bioenergetics of mammary carcinogenesis and identify how different cancers grow and invade uniquely, seemingly with or without the stress response.

What initiates pre-tumor and tumor cells to make the glycolytic switch? Indeed, this is a question over eight decades old[214] and, to my knowledge, remains unresolved. Do all tumor cells undergo this conversion? If Wallace and colleagues are correct and reactive oxygen species are being produced, without neutralizing them, unscavenged free radicals could be the source of initial or supplemental DNA-damage. For example, oxidative damage to DNA could be the original early cause, initiating the cascade of secondary effects which enable and/or promote the tumorigenic process. Alternatively, perhaps it is the dysfunction of the *Trp53*/TP53 tumor suppressor that drives the shift in energy metabolism within tumor cells[16,225,232-234]. Studies of Trp53 knockout mice and mouse or human cell lines confirmed that reduced oxygen

consumption, higher levels of lactic acid, and increased production of ATP from glycolysis (versus aerobic respiration) correlated with Trp53 deficiency[225]. Hypoxia-inducible factor, *Hif1a*/HIF1, can increase glycolytic enzyme levels and therefore may be an initiating factor in the cancer cell's switch to glycolysis[235,236]. The phosphatidylinositol 3-kinase (PI3K) signaling pathway, primarily through activation of *Akt1*/AKT1, increases glycolysis, independently of *Hif1a*/HIF1[237] (for review[14,238]). Additionally, crosstalk between Trp53 and Hif1 pathways complicates identification of the origin of the metabolic switch[239,240].

An early step in the malignant transformation of a cell involves deregulation of gene expression. Given that gene expression is closely linked to chromatin organization and function, it stands to reason that epigenetic changes, such as DNA methylation and histone modifications, coincide with altered expression of many cellular genes and retrotransposons within the genome. Retrotransposon expression may serve as an *indicator* of chromatin changes, rather than a *driver*, based on the following three points: 1) the evidence that retrotransposons and repetitive elements are normally silenced (for review[120]); 2) the multitude of seemingly redundant mechanisms that maintain silencing, which are conserved across most organisms[120,123-128,134-136,241-248]; 3) deletion of key molecules involved in retrotransposon/heterochromatin silencing results in dramatic results[118,249-261]. From rice to worms to mice to humans, transposons are silenced through mechanisms such as cytosine methylation, epigenetic modifications to histone tails, and/or via RNAi machinery. Thus, unsilencing of retrotransposons must occur as a result of epigenetic changes favoring the expression of these elements. Although yet unclear, the cause(s) of retrotransposon unsilencing, chromatin remodeling and aberrant gene expression within early stages of tumor progression may be a response to stress, the result of epigenetic alterations, due to accumulation of unrepaired DNA double strand breaks, or a combination of these factors[216-218].

What is the mechanism by which retrotransposon silencing is relieved? Whether genomic changes that coincide with unsilencing of retrotransposons are local or global mechanisms, or a combination thereof, remains uncertain. Certainly, genes expressed in one system may be repressed in another, such as the estrogen receptor (*Esr1*) or *Hoxd4* genes. These examples

point out local active or inactive chromatin states[129,130]. Alternatively, others theorize that initially global opening of chromatin occurs (i.e. within open chromatin fibers) and then local epigenetic modifications, such as histone acetylation, are necessary for gene transcription[131-133].

Additionally, it remains unresolved what cause initiates which effect in tumorigenesis. Does global histone acetylation via aberrant activity of histone acetyltransferases alter chromatin structure, which leads to subsequent, unrelated unsilencing of retrotransposons, or does the unsilencing of retrotransposons occur first, locally, via a malfunction in the RNAi pathways during tumorigenesis? Do kinases, like aurora B, phosphorylate serines and threonines locally at heterochromatic regions, without neutralizing protein phosphatases due to their abnormal function (globally), which consequently suppresses CBX5 chromobox protein binding as well as formation of heterochromatin, leaving the associated DNA unsilenced? Are there alterations to (or switches in) the histone code (locally) associated with retrotransposons, such as meH3K4, a hallmark of active chromatin, instead of meH3K9? Finally, genomic instability, *if* it occurs early in the tumorigenic process, may contribute to local (altered DNA sequence, for example) and global (i.e. chromosomal gains, losses, or translocations) genome changes which either initiate or promote mammary tumor progression in Waptag1 and human. (Certainly, the inhibition of *Trp53*/TP53 found in DCIS in both species supports the theory that genomic instability is an early occurrence.) High-resolution array CGH analysis of DNA from laser-captured tumor cells throughout all stages of Waptag1 tumor progression could help resolve questions regarding genomic stability changes during tumorigenesis.

What transcriptional changes found early in Waptag1 tumor progression could aid the understanding of a potential mechanism of aberrant expression of retrotransposons? Early in tumor progression, a substantial down-regulation of stress responses and block of normal aerobic respiration, lipid metabolism, and oxidoreductase activity occurs. The event(s) which initiate this cascade of effects remain(s) largely unknown. However, speculation, based on available data and the published literature may provide insight or offer hypotheses for future investigation of the mechanisms underlying retrotransposon expression in early tumor progression.

DNA damage, specifically double-strand breaks that remain un-repaired, can instigate expression and transposition of Short Interspersed Nuclear Elements (SINEs), which ultimately results in genomic instability[262]. However, since SINEs have no reverse transcriptase enzymatic activity, they depend upon Long Interspersed Nuclear Elements (LINEs) and/or other autonomous retrotransposons for transposition. Indeed, results presented here demonstrate that LTR and Non-LTR retrotransposons are actively transcribed. In addition, inhibition of *Trp53*/TP53 markedly accentuates this expression and transposition following DNA damage[262]. Our mouse and human results both exhibit deregulated *Trp53*/TP53 activity. Furthermore, with over ten percent of the human or mouse genomes composed of SINEs (at more than one million copies), over 15% comprised of LINEs (at roughly 500,000 copies), alongside all the other retrotransposons which, when combined constitute over one-third of the species DNA[138,139,263], perhaps it is a global mechanism by which SINEs are activated. Interestingly, a LINE1 element knocked down in human melanoma cells showed subsequent down-regulation of human endogenous retrovirus-K, implying that repetitive elements maintain a hierarchy, even during tumorigenesis[264]. Finally, LINE1 retrotransposition flourishes in cancer cells, and others which are dividing uncontrollably[265]. Hence, the up-regulation of cell cycle, DNA replication, etc. as well as the loss of checkpoints provides an additional role in tumorigenesis by contributing to genomic instability and LINE1 transposition. As shown in Table 5.4, LINE1 elements exhibit the highest fold changes in Waptag1 mouse cells and human DCIS.

Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3 (APOBEC3) genes exhibit an inhibitory effect on mouse mammary tumor virus (MMTV) infection[266], and therefore their down-regulation may initiate or aid the unsilencing of retrotransposons. There are eight members of the APOBEC3 human gene family (APOBEC3A-H), whereas mice only have one gene (*Apobec3*). In mice, APOBEC3 (protein) binds to the MMTV capsid and is packaged into virons; APOBEC3 decreases effective viral replication, largely through its cytidine deamination activity[266]. In fact, Apobec3-/- mice show increased MMTV infection compared with littermates harboring wild-type *Apobec3*[266]. In addition, work in HeLa and 293T cells showed that APOBEC3A, 3B, 3F or 3G could inhibit LINE1 activity, including transposition[267]. At the transcript

level, *Apobec3* was up-regulated in samples throughout most stages of Waptag1 tumor progression, as was APOBEC3B & 3C in the human samples. However, APOBEC3G was substantially down-regulated in human DCIS as was APOBEC3F, to a lesser extent, compared with normal breast tissue.

Adenosine deaminase enzymes are involved in gene silencing/unsilencing and function on double stranded RNA structures. Abnormal RNA editing may contribute to alternative splicing and subsequent multiple protein isoforms, increased variety of miRNAs, or suppression of RNAi mechanisms and genomic instability[268,269]. Expression of the deaminase gene, *Ada*/ADA, is significantly up-regulated within the specific cells of DCIS and tumors in Waptag1 and in human DCIS1 and tumors. In fact, Ada/ADA is among the 63 genes predictive of poor prognosis in human breast cancer (Table 4.3).

Alongside the retrotransposon transcriptional changes investigated at the global level here, characterization of individual elements at the DNA level throughout all stages of tumor progression would offer additional understanding of these elements. Specifically, is there new transposition early in tumor progression that increases the level of retrotransposon gene expression or is it simply an increase in the transcription of a finite number of elements within the genome? Alternatively, it may be both. Since the copy number of Class I endogenous retroviruses may be as low as a single genomic copy within the normal cell, quantitative polymerase chain reaction (qPCR) probes and primers are being designed to investigate these elements. Specifically, the copy number of Class I LTR elements *MER34A1* and *RLTR1C* retrotransposons will be analyzed by qPCR using DNA from tumor cells compared against that from normal cells. A thorough quantitative analysis of the copy number changes of other classes of retrotransposons, within cells of DCIS and tumors, compared against normal cells, will aid in the understanding of these elements and help resolve these uncertainties.

In summary, besides putative retrotransposons (Table 5.4), several key genes, identified through these analyses are significantly up-regulated in both species and may be suitable candidates for future studies as early markers of DCIS, including the mitotic cell cycle/division genes, *Aurka*/AURKA, *Aurkb*/AURKB, *Bub1b*/BUB1B, *Cdc20*/CDC20, and/or *Kif11*/KIF11, the

DNA topoisomerase, *Top2a*/TOP2A, or the little known PCNA-associated factor, *2810417H13Rik*/KIAA0101, the chromatin organization gene, *Cenpa*/CENPA, the ubiquitin-conjugating enzyme, *Ube2c*/UBE2C, the receptor for hyaluronan-mediated motility, *Hmmr*/HMMR, or the induced during growth factor-mediated cell survival gene, *Spag5*/SPAG5 (Table 4.3). Independent validation of these potential early markers, predictive of progression to invasion, suggests that they are consistently up-regulated in human breast cancer and associated with decreased survival (Figure 4.6). Since over two thousand transcriptional changes are shared between human DCIS and the mouse model evaluated here, perhaps Waptag1 mice may serve as an appropriate model for future studies aimed at detection of DCIS and/or therapeutic intervention.

REFERENCES

1. Nowell PC. (1976) The clonal evolution of tumor cell populations. *Science.* 194, 23-28.

2. Mikaelian I, Fancher K, Knowles BB, and Churchill GA. (*et. al.*) Epithelial to Mesenchymal Transition: Acquisition of the Mesenchymal Phenotype is Associated with Decreased Cell Proliferation and Apoptosis.

3. Poola I, DeWitty RL, Marshalleck JJ, Bhatnagar R, Abraham J, and Leffall LD. (2005) Identification of MMP-1 as a putative breast cancer predictive marker by global gene expression analysis. *Nat Med.* 11, 481-483.

4. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, and Botstein D. (2000) Molecular portraits of human breast tumours. *Nature* 406, 747-752.

5. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lonning PE, Brown PO, Borresen-Dale AL, and Botstein D. (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 100, 8418-8423.

6. Farmer P, Bonnefoi H, Becette V, Tubiana-Hulin M, Fumoleau P, Larsimont D, Macgrogan G, Bergh J, Cameron D, Goldstein D, Duss S, Nicoulaz AL, Brisken C, Fiche M, Delorenzi M, and Iggo R. (2005) Identification of molecular apocrine breast tumours by microarray analysis. *Oncogene* 24, 4660-4671.

7. Kinzler KW, and Vogelstein B. (1996) Lessons from hereditary colorectal cancer. *Cell.* 87, 159-170.

8. Sidransky D, Mikkelsen T, Schwechheimer K, Rosenblum ML, Cavanee W, and Vogelstein B. (1992) Clonal expansion of p53 mutant cells is associated with brain tumour progression. *Nature.* 355, 846-847.

9. Beerenwinkel N, Antal T, Dingli D, Traulsen A, Kinzler KW, Velculescu VE, Vogelstein B, and Nowak MA. (2007) Genetic progression and the waiting time to cancer. *PLoS Comput Biol.* 3, e225.

10. Vogelstein B, and Kinzler KW. (2004) Cancer genes and the pathways they control. *Nat Med.* 10, 789-799.

11. Hanahan D, and Weinberg RA. (2000) The hallmarks of cancer. *Cell.* 100, 57-70.

12. Garber K. (2006) Energy deregulation: licensing tumors to grow. *Science.* 312, 1158-1159.

13. Kondoh H. (2008) Cellular life span and the Warburg effect. *Exp Cell Res.* 314, 1923-1928. Epub 2008 Mar 1920.

14. DeBerardinis RJ, Lum JJ, Hatzivassiliou G, and Thompson CB. (2008) The biology of cancer: metabolic reprogramming fuels cell growth and proliferation. *Cell Metab.* 7, 11-20.

15. Wu M, Neilson A, Swift AL, Moran R, Tamagnine J, Parslow D, Armistead S, Lemire K, Orrell J, Teich J, Chomicz S, and Ferrick DA. (2007) Multiparameter metabolic analysis reveals a close link between attenuated mitochondrial bioenergetic function and enhanced glycolysis dependency in human tumor cells. *Am J Physiol Cell Physiol.* 292, C125-136.

16. Vousden KH, and Lane DP. (2007) p53 in health and disease. *Nat Rev Mol Cell Biol.* 8, 275-283.

17. Moreno-Sanchez R, Rodriguez-Enriquez S, Marin-Hernandez A, and Saavedra E. (2007) Energy metabolism in tumor cells. *Febs J.* 274, 1393-1418.

18. Ristow M. (2006) Oxidative metabolism in cancer growth. *Curr Opin Clin Nutr Metab Care.* 9, 339-345.

19. Fantin VR, St-Pierre J, and Leder P. (2006) Attenuation of LDH-A expression uncovers a link between glycolysis, mitochondrial physiology, and tumor maintenance. *Cancer Cell.* 9, 425-434.

20. Bui T, and Thompson CB. (2006) Cancer's sweet tooth. *Cancer Cell.* 9, 419-420.

21. Shaw RJ. (2006) Glucose metabolism and cancer. *Curr Opin Cell Biol.* 18, 598-608.

22. Winston JS, Asch HL, Zhang PJ, Edge SB, Hyland A, and Asch BB. (2001) Downregulation of gelsolin correlates with the progression to breast carcinoma. *Breast Cancer Res Treat* 65, 11-21.

23. Wulfkuhle JD, Sgroi DC, Krutzsch H, McLean K, McGarvey K, Knowlton M, Chen S, Shu H, Sahin A, Kurek R, Wallwiener D, Merino MJ, Petricoin EF, 3rd, Zhao Y, and Steeg PS. (2002) Proteomics of human breast ductal carcinoma in situ. *Cancer Res* 62, 6740-6749.

24. Farabegoli F, Champeme MH, Bieche I, Santini D, Ceccarelli C, Derenzini M, and Lidereau R. (2002) Genetic pathways in the evolution of breast ductal carcinoma in situ. *J Pathol* 196, 280-286.

25. Yao J, Weremowicz S, Feng B, Gentleman RC, Marks JR, Gelman R, Brennan C, and Polyak K. (2006) Combined cDNA array comparative genomic hybridization and serial analysis of gene expression analysis of breast tumor progression. *Cancer Res.* 66, 4065-4078.

26. Schuetz CS, Bonin M, Clare SE, Nieselt K, Sotlar K, Walter M, Fehm T, Solomayer E, Riess O, Wallwiener D, Kurek R, and Neubauer HJ. (2006) Progression-specific genes identified by expression profiling of matched ductal carcinomas in situ and invasive breast tumors, combining laser capture microdissection and oligonucleotide microarray analysis. *Cancer Res.* 66, 5278-5286.

27. Allred DC, Wu Y, Mao S, Nagtegaal ID, Lee S, Perou CM, Mohsin SK, O'Connell P, Tsimelzon A, and Medina D. (2008) Ductal Carcinoma In situ and the Emergence of Diversity during Breast Cancer Evolution. *Clin Cancer Res.* 14, 370-378.

28. Dalgin GS, Alexe G, Scanfeld D, Tamayo P, Mesirov JP, Ganesan S, DeLisi C, and Bhanot G. (2007) Portraits of breast cancer progression. *BMC Bioinformatics.* 8, 291.

29. Valenzuela M, and Julian TB. (2007) Ductal carcinoma in situ: biology, diagnosis, and new therapies. *Clin Breast Cancer.* 7, 676-681.

30. Alexe G, Dalgin GS, Ganesan S, Delisi C, and Bhanot G. (2007) Analysis of breast cancer progression using principal component analysis and clustering. *J Biosci.* 32, 1027-1039.

31. Bates GJ, Fox SB, Han C, Leek RD, Garcia JF, Harris AL, and Banham AH. (2006) Quantification of regulatory T cells enables the identification of high-risk breast cancer patients and those at risk of late relapse. *J Clin Oncol.* 24, 5373-5380.

32. Hannemann J, Velds A, Halfwerk JB, Kreike B, Peterse JL, and van de Vijver MJ. (2006) Classification of ductal carcinoma in situ by gene expression profiling. *Breast Cancer Res.* 8, R61.

33. Seth A, Kitching R, Landberg G, Xu J, Zubovits J, and Burger AM. (2003) Gene expression profiling of ductal carcinomas in situ and invasive breast tumors. *Anticancer Res.* 23, 2043-2051.

34. Adeyinka A, Emberley E, Niu Y, Snell L, Murphy LC, Sowter H, Wykoff CC, Harris AL, and Watson PH. (2002) Analysis of gene expression in ductal carcinoma in situ of the breast. *Clin Cancer Res.* 8, 3788-3795.

35. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, and Friend SH. (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature.* 415, 530-536.

36. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, and Bernards R. (2002) A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med.* 347, 1999-2009.

37. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, Talantov D, Timmermans M, Meijer-van Gelder ME, Yu J, Jatkoe T, Berns EM, Atkins D, and Foekens JA. (2005) Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet.* 365, 671-679.

38. Koscielny S. (2008) Critical review of microarray-based prognostic tests and trials in breast cancer. *Curr Opin Obstet Gynecol.* 20, 47-50.

39. Taube SE, Jacobson JW, and Lively TG. (2005) Cancer diagnostics: decision criteria for marker utilization in the clinic. *Am J Pharmacogenomics.* 5, 357-364.

40. Glas AM, Floore A, Delahaye LJ, Witteveen AT, Pover RC, Bakx N, Lahti-Domenici JS, Bruinsma TJ, Warmoes MO, Bernards R, Wessels LF, and Van't Veer LJ. (2006) Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics.* 7, 278.

41. Jacks T, Remington L, Williams BO, Schmitt EM, Halachmi S, Bronson RT, and Weinberg RA. (1994) Tumor spectrum analysis in p53-mutant mice. *Curr Biol.* 4, 1-7.

42. Li Y, Hively WP, and Varmus HE. (2000) Use of MMTV-Wnt-1 transgenic mice for studying the genetic basis of breast cancer. *Oncogene* 19, 1002-1009.

43. Guy CT, Cardiff RD, and Muller WJ. (1992) Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. *Mol Cell Biol* 12, 954-961.

44. Guy CT, Webster MA, Schaller M, Parsons TJ, Cardiff RD, and Muller WJ. (1992) Expression of the neu protooncogene in the mammary epithelium of transgenic mice induces metastatic disease. *Proc Natl Acad Sci U S A.* 89, 10578-10582.

45. Muller WJ, Arteaga CL, Muthuswamy SK, Siegel PM, Webster MA, Cardiff RD, Meise KS, Li F, Halter SA, and Coffey RJ. (1996) Synergistic interaction of the Neu proto-oncogene product and transforming growth factor alpha in the mammary epithelium of transgenic mice. *Mol Cell Biol.* 16, 5726-5736.

46. Tzeng YJ, Gottlob K, Santarelli R, and Graessmann A. (1996) The SV40 T-antigen induces premature apoptotic mammary gland involution during late pregnancy in transgenic mice. *FEBS Lett.* 380, 215-218.

47. Santarelli R, Tzeng YJ, Zimmermann C, Guhl E, and Graessmann A. (1996) SV40 T-antigen induces breast cancer formation with a high efficiency in lactating and virgin WAP-SV-T transgenic animals but with a low efficiency in ovariectomized animals. *Oncogene.* 12, 495-505.

48. Tzeng YJ, Guhl E, Graessmann M, and Graessmann A. (1993) Breast cancer formation in transgenic animals induced by the whey acidic protein SV40 T antigen (WAP-SV-T) hybrid gene. *Oncogene.* 8, 1965-1971.

49. Schulze-Garg C, Lohler J, Gocht A, and Deppert W. (2000) A transgenic mouse model for the ductal carcinoma in situ (DCIS) of the mammary gland. *Oncogene.* 19, 1028-1037.

50. Hüsler MR, Kotopoulis KA, Sundberg JP, Tennent BJ, Kunig SV, and Knowles BB. (1998) Lactation-induced WAP-SV40 Tag transgene expression in C57BL/6J mice leads to mammary carcinoma. *Transgenic Research* 7, 253-263.

51. Shima N, Alcaraz A, Liachko I, Buske TR, Andrews CA, Munroe RJ, Hartford SA, Tye BK, and Schimenti JC. (2007) A viable allele of Mcm4 causes chromosome instability and mammary adenocarcinomas in mice. *Nat Genet.* 39, 93-98.

52. Sinn E, Muller W, Pattengale P, Tepler I, Wallace R, and Leder P. (1987) Coexpression of MMTV/v-Ha-ras and MMTV/c-myc genes in transgenic mice: synergistic action of oncogenes in vivo. *Cell.* 49, 465-475.

53. Muller WJ, Sinn E, Pattengale PK, Wallace R, and Leder P. (1988) Single-step induction of mammary adenocarcinoma in transgenic mice bearing the activated c-neu oncogene. *Cell.* 54, 105-115.

54. Matsui Y, Halter SA, Holt JT, Hogan BL, and Coffey RJ. (1990) Development of mammary hyperplasia and neoplasia in MMTV-TGF alpha transgenic mice. *Cell.* 61, 1147-1155.

55. Sandgren EP, Schroeder JA, Qui TH, Palmiter RD, Brinster RL, and Lee DC. (1995) Inhibition of mammary gland involution is associated with transforming growth factor alpha but not c-myc-induced tumorigenesis in transgenic mice. *Cancer Res.* 55, 3915-3927.

56. Stewart TA, Pattengale PK, and Leder P. (1984) Spontaneous mammary adenocarcinomas in transgenic mice that carry and express MTV/myc fusion genes. *Cell.* 38, 627-637.

57. Dandachi N, Hauser-Kronberger C, More E, Wiesener B, Hacker GW, Dietze O, and Wirl G. (2001) Co-expression of tenascin-C and vimentin in human breast cancer cells indicates phenotypic transdifferentiation during tumour progression: correlation with histopathological parameters, hormone receptors, and oncoproteins. *J Pathol* 193, 181-189.

58. Trimboli AJ, Fukino K, de Bruin A, Wei G, Shen L, Tanner SM, Creasap N, Rosol TJ, Robinson ML, Eng C, Ostrowski MC, and Leone G. (2008) Direct evidence for epithelial-mesenchymal transitions in breast cancer. *Cancer Res.* 68, 937-945.

59. Nusse R, and Varmus HE. (1982) Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. *Cell* 31, 99-109.

60. Tsukamoto AS, Grosschedl R, Guzman RC, Parslow T, and Varmus HE. (1988) Expression of the int-1 gene in transgenic mice is associated with mammary gland hyperplasia and adenocarcinomas in male and female mice. *Cell* 55, 619-625.

61. Huang S, Li Y, Chen Y, Podsypanina K, Chamorro M, Olshen AB, Desai KV, Tann A, Petersen D, Green JE, and Varmus HE. (2005) Changes in gene expression during the development of mammary tumors in MMTV-Wnt-1 transgenic mice. *Genome Biol* 6, R84.

62. Shackleford GM, MacArthur CA, Kwan HC, and Varmus HE. (1993) Mouse mammary tumor virus infection accelerates mammary carcinogenesis in Wnt-1 transgenic mice by insertional activation of int-2/Fgf-3 and hst/Fgf-4. *Proc Natl Acad Sci U S A* 90, 740-744.

63. Richner J, Gerber HA, Locher GW, Goldhirsch A, Gelber RD, Gullick WJ, Berger MS, Groner B, and Hynes NE. (1990) c-erbB-2 protein expression in node negative breast cancer. *Ann Oncol.* 1, 263-268.

64. Berger MS, Locher GW, Saurer S, Gullick WJ, Waterfield MD, Groner B, and Hynes NE. (1988) Correlation of c-erbB-2 gene amplification and protein expression in human breast carcinoma with nodal status and nuclear grading. *Cancer Res.* 48, 1238-1243.

65. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, and McGuire WL. (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science.* 235, 177-182.

66. Attardi LD, and Donehower LA. (2005) Probing p53 biological functions through the use of genetically engineered mouse models. *Mutat Res.* 576, 4-21.

67. Donehower LA, Harvey M, Slagle BL, McArthur MJ, Montgomery CA, Jr., Butel JS, and Bradley A. (1992) Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature.* 356, 215-221.

68. Elson A, Deng C, Campos-Torres J, Donehower LA, and Leder P. (1995) The MMTV/c-myc transgene and p53 null alleles collaborate to induce T-cell lymphomas, but not mammary carcinomas in transgenic mice. *Oncogene.* 11, 181-190.

69. Garcia-Espana A, Salazar E, Sun TT, Wu XR, and Pellicer A. (2005) Differential expression of cell cycle regulators in phenotypic variants of transgenically induced bladder tumors: implications for tumor behavior. *Cancer Res.* 65, 1150-1157

70. Hosokawa Y, Papanikolaou A, Cardiff RD, Yoshimoto K, Bernstein M, Wang TC, Schmidt EV, and Arnold A. (2001) In vivo analysis of mammary and non-mammary tumorigenesis in MMTV-cyclin D1 transgenic mice deficient in p53. *Transgenic Res* 10, 471-478.

71. Kountouras J, Zavos C, and Chatzopoulos D. (2005) New concepts of molecular biology on gastric carcinogenesis. *Hepatogastroenterology.* 52, 1305-1312.

72. MacGrogan D, and Bookstein R. (1997) Tumour suppressor genes in prostate cancer. *Semin Cancer Biol.* 8, 11-19.

73. Elledge RM, and Allred DC. (1998) Prognostic and predictive value of p53 and p21 in breast cancer. *Breast Cancer Res Treat* 52, 79-98.

74. Maglione JE, McGoldrick ET, Young LJ, Namba R, Gregg JP, Liu L, Moghanaki D, Ellies LG, Borowsky AD, Cardiff RD, and MacLeod CL. (2004) Polyomavirus middle T-induced mammary intraepithelial neoplasia outgrowths: single origin, divergent evolution, and multiple outcomes. *Mol Cancer Ther* 3, 941-953.

75. Pittius CW, Sankaran L, Topper YJ, and Hennighausen L. (1988) Comparison of the regulation of the whey acidic protein gene with that of a hybrid gene containing the whey acidic protein gene promoter in transgenic mice. *Mol Endocrinol.* 2, 1027-1032.

76. Welcker M, and Clurman BE. (2005) The SV40 large T antigen contains a decoy phosphodegron that mediates its interactions with Fbw7/hCdc4. *J Biol Chem.* 280, 7654-7658.

77. Ahuja D, Saenz-Robles MT, and Pipas JM. (2005) SV40 large T antigen targets multiple cellular pathways to elicit cellular transformation. *Oncogene.* 24, 7729-7745.

78. Klein A, Guhl E, Zollinger R, Tzeng YJ, Wessel R, Hummel M, Graessmann M, and Graessmann A. (2005) Gene expression profiling: cell cycle deregulation and aneuploidy do not cause breast cancer formation in WAP-SVT/t transgenic animals. *J Mol Med* 83, 362-376.

79. Dietrich WF, Radany EH, Smith JS, Bishop JM, Hanahan D, and Lander ES. (1994) Genome-wide search for loss of heterozygosity in transgenic mouse tumors reveals candidate tumor suppressor genes on chromosomes 9 and 16. *Proc Natl Acad Sci U S A.* 91, 9451-9455.

80. Hicks J, Krasnitz A, Lakshmi B, Navin NE, Riggs M, Leibu E, Esposito D, Alexander J, Troge J, Grubor V, Yoon S, Wigler M, Ye K, Borresen-Dale AL, Naume B, Schlicting E, Norton L, Hagerstrom T, Skoog L, Auer G, Maner S, Lundin P, and Zetterberg A. (2006) Novel patterns of genome rearrangement and their association with survival in breast cancer. *Genome Res.* 16, 1465-1479.

81. Podsypanina K, Politi K, Beverly LJ, and Varmus HE. (2008) Oncogene cooperation in tumor maintenance and tumor recurrence in mouse mammary tumors induced by Myc and mutant Kras. *Proc Natl Acad Sci U S A* 20, 20.

82. Sarkisian CJ, Keister BA, Stairs DB, Boxer RB, Moody SE, and Chodosh LA. (2007) Dose-dependent oncogene-induced senescence in vivo and its evasion during mammary tumorigenesis. *Nat Cell Biol.* 9, 493-505.

83. Du Z, Podsypanina K, Huang S, McGrath A, Toneff MJ, Bogoslovskaia E, Zhang X, Moraes RC, Fluck M, Allred DC, Lewis MT, Varmus HE, and Li Y. (2006) Introduction of oncogenes into mammary glands in vivo with an avian retroviral vector initiates and promotes carcinogenesis in mouse models. *Proc Natl Acad Sci U S A.* 103, 17396-17401.

84. Wijnhoven SW, Zwart E, Speksnijder EN, Beems RB, Olive KP, Tuveson DA, Jonkers J, Schaap MM, van den Berg J, Jacks T, van Steeg H, and de Vries A. (2005) Mice expressing a mammary gland-specific R270H mutation in the p53 tumor suppressor gene mimic human breast cancer development. *Cancer Res.* 65, 8166-8173.

85. Furth PA (1998) *Simian Virus 40 (SV40): A Possible Human Polyomavirus*. S. Krager AG.

86. Sweet BH, and Hilleman MR. (1960) The vacuolating virus, S.V. 40. *Proc Soc Exp Biol Med.* 105, 420-427.

87. Eddy BE, Borman GS, Berkeley WH, and Young RD. (1961) Tumors induced in hamsters by injection of rhesus monkey kidney cell extracts. *Proc Soc Exp Biol Med.* 107, 191-197.

88. Girardi AJ, Sweet BH, Slotnick VB, and Hilleman MR. (1962) Development of tumors in hamsters inoculated in the neonatal period with vacuolating virus, SV-40. *Proc Soc Exp Biol Med.* 109, 649-660.

89. Bennoun M, Grimber G, Couton D, Seye A, Molina T, Briand P, and Joulin V. (1998) The amino-terminal region of SV40 large T antigen is sufficient to induce hepatic tumours in mice. *Oncogene.* 17, 1253-1259.

90. Sullivan CS, and Pipas JM. (2002) T antigens of simian virus 40: molecular chaperones for viral replication and tumorigenesis. *Microbiol Mol Biol Rev.* 66, 179-202.

91. Tevethia MJ, Bonneau RH, Griffith JW, and Mylin L. (1997) A simian virus 40 large T-antigen segment containing amino acids 1 to 127 and expressed under the control of the rat elastase-1 promoter produces pancreatic acinar carcinomas in transgenic mice. *J Virol.* 71, 8157-8166.

92. Tevethia MJ, Lacko HA, Kierstead TD, and Thompson DL. (1997) Adding an Rb-binding site to an N-terminally truncated simian virus 40 T antigen restores growth to high cell density, and the T common region in trans provides anchorage-independent growth and rapid growth in low serum concentrations. *J Virol.* 71, 1888-1896.

93. Zalvide J, Stubdal H, and DeCaprio JA. (1998) The J domain of simian virus 40 large T antigen is required to functionally inactivate RB family proteins. *Mol Cell Biol.* 18, 1408-1415.

94. Srinivasan A, McClellan AJ, Vartikar J, Marks I, Cantalupo P, Li Y, Whyte P, Rundell K, Brodsky JL, and Pipas JM. (1997) The amino-terminal transforming region of simian virus 40 large T and small t antigens functions as a J domain. *Mol Cell Biol.* 17, 4761-4773.

95. Sullivan CS, Cantalupo P, and Pipas JM. (2000) The molecular chaperone activity of simian virus 40 large T antigen is required to disrupt Rb-E2F family complexes by an ATP-dependent mechanism. *Mol Cell Biol.* 20, 6233-6243.

96. Kim HY, Ahn BY, and Cho Y. (2001) Structural basis for the inactivation of retinoblastoma tumor suppressor by SV40 large T antigen. *Embo J.* 20, 295-304.

97. Gabai VL, Mabuchi K, Mosser DD, and Sherman MY. (2002) Hsp72 and stress kinase c-jun N-terminal kinase regulate the bid-dependent pathway in tumor necrosis factor-induced apoptosis. *Mol Cell Biol.* 22, 3415-3424.

98. Quartin RS, Cole CN, Pipas JM, and Levine AJ. (1994) The amino-terminal functions of the simian virus 40 large T antigen are required to overcome wild-type p53-mediated growth arrest of cells. *J Virol.* 68, 1334-1341.

99. Rushton JJ, Jiang D, Srinivasan A, Pipas JM, and Robbins PD. (1997) Simian virus 40 T antigen can regulate p53-mediated transcription independent of binding p53. *J Virol.* 71, 5620-5623.

100. Pipas JM, and Levine AJ. (2001) Role of T antigen interactions with p53 in tumorigenesis. *Semin Cancer Biol.* 11, 23-30.

101. Slinskey A, Barnes D, and Pipas JM. (1999) Simian virus 40 large T antigen J domain and Rb-binding motif are sufficient to block apoptosis induced by growth factor withdrawal in a neural stem cell line. *J Virol.* 73, 6791-6799.

102. Ali SH, and DeCaprio JA. (2001) Cellular transformation by SV40 large T antigen: interaction with host proteins. *Semin Cancer Biol.* 11, 15-23.

103. Poulin DL, Kung AL, and DeCaprio JA. (2004) p53 targets simian virus 40 large T antigen for acetylation by CBP. *J Virol.* 78, 8245-8253.

104. Kasper JS, Kuwabara H, Arai T, Ali SH, and DeCaprio JA. (2005) Simian virus 40 large T antigen's association with the CUL7 SCF complex contributes to cellular transformation. *J Virol.* 79, 11685-11692.

105. Baluchamy S, Rajabi HN, Thimmapaya R, Navaraj A, and Thimmapaya B. (2003) Repression of c-Myc and inhibition of G1 exit in cells conditionally overexpressing p300 that is not dependent on its histone acetyltransferase activity. *Proc Natl Acad Sci U S A.* 100, 9524-9529.

106. Valls E, de la Cruz X, and Martinez-Balbas MA. (2003) The SV40 T antigen modulates CBP histone acetyltransferase activity. *Nucleic Acids Res.* 31, 3114-3122.

107. Cotsiki M, Lock RL, Cheng Y, Williams GL, Zhao J, Perera D, Freire R, Entwistle A, Golemis EA, Roberts TM, Jat PS, and Gjoerup OV. (2004) Simian virus 40 large T antigen targets the spindle assembly checkpoint protein Bub1. *Proc Natl Acad Sci U S A.* 101, 947-952.

108. Wu X, Avni D, Chiba T, Yan F, Zhao Q, Lin Y, Heng H, and Livingston D. (2004) SV40 T antigen interacts with Nbs1 to disrupt DNA replication control. *Genes Dev.* 18, 1305-1316.

109. Moens U, Seternes OM, Johansen B, and Rekvig OP. (1997) Mechanisms of transcriptional regulation of cellular genes by SV40 large T- and small T-antigens. *Virus Genes.* 15, 135-154.

110. Janssens V, Goris J, and Van Hoof C. (2005) PP2A: the expected tumor suppressor. *Curr Opin Genet Dev.* 15, 34-41.

111. Rundell K, and Parakati R. (2001) The role of the SV40 ST antigen in cell growth promotion and transformation. *Semin Cancer Biol.* 11, 5-13.

112. Van Hoof C, and Goris J. (2003) Phosphatases in apoptosis: to be or not to be, PP2A is in the heart of the question. *Biochim Biophys Acta.* 1640, 97-104.

113. Janssens V, and Goris J. (2001) Protein phosphatase 2A: a highly regulated family of serine/threonine phosphatases implicated in cell growth and signalling. *Biochem J.* 353, 417-439.

114. Lill NL, Grossman SR, Ginsberg D, DeCaprio J, and Livingston DM. (1997) Binding and modulation of p53 by p300/CBP coactivators. *Nature.* 387, 823-827.

115. Whalen B, Laffin J, Friedrich TD, and Lehman JM. (1999) SV40 small T antigen enhances progression to >G2 during lytic infection. *Exp Cell Res.* 251, 121-127.

116. Ali SH, Kasper JS, Arai T, and DeCaprio JA. (2004) Cul7/p185/p193 binding to simian virus 40 large T antigen has a role in cellular transformation. *J Virol.* 78, 2749-2757.

117. Jones PA. (1996) DNA methylation errors and cancer. *Cancer Res* 56, 2463-2467.

118. Gaudet F, Hodgson JG, Eden A, Jackson-Grusby L, Dausman J, Gray JW, Leonhardt H, and Jaenisch R. (2003) Induction of tumors in mice by genomic hypomethylation. *Science* 300, 489-492.

119. Ferguson AT, Evron E, Umbricht CB, Pandita TK, Chan TA, Hermeking H, Marks JR, Lambers AR, Futreal PA, Stampfer MR, and Sukumar S. (2000) High frequency of

hypermethylation at the 14-3-3 sigma locus leads to gene silencing in breast cancer. *Proc Natl Acad Sci U S A* 97, 6049-6054.

120. Mattei M-G, and Luciani J (2003) Heterochromatin, from chromosome to protein. Atlas Genet Cytogenet Oncol Haematol http://www.infobiogen.fr/services/chromcancer/Deep/HeterochromatinDeep.html.

121. Lehnertz B, Ueda Y, Derijck AA, Braunschweig U, Perez-Burgos L, Kubicek S, Chen T, Li E, Jenuwein T, and Peters AH. (2003) Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr Biol.* 13, 1192-1200.

122. Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, and Kouzarides T. (2003) The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J Biol Chem.* 278, 4035-4040.

123. Kourmouli N, Sun YM, van der Sar S, Singh PB, and Brown JP. (2005) Epigenetic regulation of mammalian pericentric heterochromatin in vivo by HP1. *Biochem Biophys Res Commun.* 337, 901-907.

124. Fischle W, Tseng BS, Dormann HL, Ueberheide BM, Garcia BA, Shabanowitz J, Hunt DF, Funabiki H, and Allis CD. (2005) Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature.* 438, 1116-1122.

125. Hirota T, Lipp JJ, Toh BH, and Peters JM. (2005) Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. *Nature.* 438, 1176-1180.

126. Daujat S, Zeissler U, Waldmann T, Happel N, and Schneider R. (2005) HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation blocks HP1 binding. *J Biol Chem.* 280, 38090-38095.

127. Fischle W, Wang Y, and Allis CD. (2003) Binary switches and modification cassettes in histone biology and beyond. *Nature.* 425, 475-479.

128. Flanagan JF, Mi LZ, Chruszcz M, Cymborowski M, Clines KL, Kim Y, Minor W, Rastinejad F, and Khorasanizadeh S. (2005) Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature.* 438, 1181-1185.

129. Sharma D, Blum J, Yang X, Beaulieu N, Macleod AR, and Davidson NE. (2005) Release of methyl CpG binding proteins and histone deacetylase 1 from the Estrogen receptor alpha (ER) promoter upon reactivation in ER-negative human breast cancer cells. *Mol Endocrinol.* 19, 1740-1751.

130. Rastegar M, Kobrossy L, Kovacs EN, Rambaldi I, and Featherstone M. (2004) Sequential histone modifications at Hoxd4 regulatory regions distinguish anterior from posterior embryonic compartments. *Mol Cell Biol.* 24, 8090-8103.

131. Gilbert N, and Ramsahoye B. (2005) The relationship between chromatin structure and transcriptional activity in mammalian genomes. *Brief Funct Genomic Proteomic.* 4, 129-142.

132. Hovhannisyan H, Cho B, Mitra P, Montecino M, Stein GS, Van Wijnen AJ, and Stein JL. (2003) Maintenance of open chromatin and selective genomic occupancy at the cell cycle-regulated histone H4 promoter during differentiation of HL-60 promyelocytic leukemia cells. *Mol Cell Biol.* 23, 1460-1469.

133. Schubeler D, Francastel C, Cimbora DM, Reik A, Martin DI, and Groudine M. (2000) Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus. *Genes Dev.* 14, 940-950.

134. Walsh CP, Chaillet JR, and Bestor TH. (1998) Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* 20, 116-117.

135. Yoder JA, Walsh CP, and Bestor TH. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* 13, 335-340.

136. Lynch CA, Tycko B, Bestor TH, and Walsh CP. (2002) Reactivation of a silenced H19 gene in human rhabdomyosarcoma by demethylation of DNA but not by histone hyperacetylation. *Mol Cancer* 1, 2.

137. Asch BB. (1996) Tumor viruses and endogenous retrotransposons in mammary tumorigenesis. *J Mammary Gland Biol Neoplasia* 1, 49-60.

138. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Chen YJ, and al. e. (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.

139. Waterston RH, Lindblad-Toh K, Lander ES, and et.al. (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520-562.

140. Medstrand P, van de Lagemaat LN, and Mager DL. (2002) Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res.* 12, 1483-1495.

141. Horn PJ, and Peterson CL. (2006) Heterochromatin assembly: A new twist on an old model. *Chromosome Res.* 14, 83-94.

142. Zhang B, Wang Q, and Pan X. (2007) MicroRNAs and their regulatory roles in animals and plants. *J Cell Physiol.* 210, 279-289.

143. Kuss AW, and Chen W. (2008) MicroRNAs in brain function and disease. *Curr Neurol Neurosci Rep.* 8, 190-197.

144. Corney DC, and Nikitin AY. (2008) MicroRNA and ovarian cancer. *Histol Histopathol.* 23, 1161-1169.

145. Mills K. (2008) Gene expression profiling for the diagnosis and prognosis of acute myeloid leukaemia. *Front Biosci.* 13, 4605-4616.

146. Kanellopoulou C, and Monticelli S. (2008) A role for microRNAs in the development of the immune system and in the pathogenesis of cancer. *Semin Cancer Biol.* 18, 79-88.

147. Ma L, and Weinberg RA. (2008) MicroRNAs in malignant progression. *Cell Cycle.* 7, 570-572.

148. Yang N, Coukos G, and Zhang L. (2008) MicroRNA epigenetic alterations in human cancer: one step forward in diagnosis and treatment. *Int J Cancer.* 122, 963-968.

149. van Lohuizen M, and Berns A. (1990) Tumorigenesis by slow-transforming retroviruses--an update. *Biochim Biophys Acta* 1032, 213-235.

150. Nusse R. (1991) Insertional mutagenesis in mouse mammary tumorigenesis. *Curr Top Microbiol Immunol* 171, 43-65.

151. Marchetti A, Buttitta F, Miyazaki S, Gallahan D, Smith GH, and Callahan R. (1995) Int-6, a highly conserved, widely expressed gene, is mutated by mouse mammary tumor virus in mammary preneoplasia. *J Virol* 69, 1932-1938.

152. Kordon EC, Smith GH, Callahan R, and Gallahan D. (1995) A novel non-mouse mammary tumor virus activation of the Int-3 gene in a spontaneous mouse mammary tumor. *J Virol* 69, 8066-8069.

153. Kuff EL. (1990) Intracisternal A particles in mouse neoplasia. *Cancer Cells* 2, 398-400.

154. Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, Solter D, and Knowles BB. (2004) Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev Cell* 7, 597-606.

155. Morse B, Rotherg PG, South VJ, Spandorfer JM, and Astrin SM. (1988) Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma. *Nature* 333, 87-90.

156. Kazazian HH, Jr., Wong C, Youssoufian H, Scott AF, Phillips DG, and Antonarakis SE. (1988) Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* 332, 164-166.

157. Narita N, Nishio H, Kitoh Y, Ishikawa Y, Ishikawa Y, Minami R, Nakamura H, and Matsuo M. (1993) Insertion of a 5' truncated L1 element into the 3' end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. *J Clin Invest* 91, 1862-1867.

158. Kaplan MH, Wang XP, Xu HP, and Dosik MH. (2004) Partially unspliced and fully spliced ELF3 mRNA, including a new Alu element in human breast cancer. *Breast Cancer Res Treat* 83, 171-187.

159. Chang CH, Scott GK, Kuo WL, Xiong X, Suzdaltseva Y, Park JW, Sayre P, Erny K, Collins C, Gray JW, and Benz CC. (1997) ESX: a structurally unique Ets overexpressed early during human breast tumorigenesis. *Oncogene* 14, 1617-1622.

160. Lerat E, and Semon M. (2007) Influence of the transposable element neighborhood on human gene expression in normal and tumor tissues. *Gene.* 396, 303-311.

161. Akagi K, Suzuki T, Stephens RM, Jenkins NA, and Copeland NG. (2004) RTCGD: retroviral tagged cancer gene database. *Nucleic Acids Res* 32, D523-527.

162. Horie K, Yusa K, Yae K, Odajima J, Fischer SE, Keng VW, Hayakawa T, Mizuno S, Kondoh G, Ijiri T, Matsuda Y, Plasterk RH, and Takeda J. (2003) Characterization of Sleeping Beauty transposition and its application to genetic screening in mice. *Mol Cell Biol.* 23, 9189-9207.

163. Ivics Z, Hackett PB, Plasterk RH, and Izsvak Z. (1997) Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell.* 91, 501-510.

164. Takeda J, Izsvak Z, and Ivics Z. (2008) Insertional mutagenesis of the mouse germline with Sleeping Beauty transposition. *Methods Mol Biol.* 435, 109-125.

165. Dasgupta M, Agarwal MK, Varley P, Lu T, Stark GR, and Kandel ES. (2008) Transposon-based mutagenesis identifies short RIP1 as an activator of NFkappaB. *Cell Cycle.* 7, 2249-2256.

166. Dupuy AJ, Jenkins NA, and Copeland NG. (2006) Sleeping beauty: a novel cancer gene discovery tool. *Hum Mol Genet.* 15, R75-79.

167. Collier LS, and Largaespada DA. (2006) Transforming science: cancer gene identification. *Curr Opin Genet Dev.* 16, 23-29.

168. Wang-Johanning F, Frost AR, Jian B, Epp L, Lu DW, and Johanning GL. (2003) Quantitation of HERV-K env gene expression and splicing in human breast cancer. *Oncogene* 22, 1528-1535.

169. Stauffer Y, Theiler G, Sperisen P, Lebedev Y, and Jongeneel CV. (2004) Digital expression profiles of human endogenous retroviral families in normal and cancerous tissues. *Cancer Immun* 4, 2.

170. Armbruester V, Sauter M, Krautkraemer E, Meese E, Kleiman A, Best B, Roemer K, and Mueller-Lantzsch N. (2002) A novel gene from the human endogenous retrovirus K expressed in transformed cells. *Clin Cancer Res* 8, 1800-1807.

171. Armbruester V, Sauter M, Roemer K, Best B, Hahn S, Nty A, Schmid A, Philipp S, Mueller A, and Mueller-Lantzsch N. (2004) Np9 protein of human endogenous retrovirus K interacts with ligand of numb protein X. *J Virol* 78, 10310-10319.

172. Menendez L, Benigno BB, and McDonald JF. (2004) L1 and HERV-W retrotransposons are hypomethylated in human ovarian carcinomas. *Mol Cancer* 3, 12.

173. Wang-Johanning F, Frost AR, Jian B, Azerou R, Lu DW, Chen DT, and Johanning GL. (2003) Detecting the expression of human endogenous retrovirus E envelope transcripts in human prostate adenocarcinoma. *Cancer* 98, 187-197.

174. Yi JM, Kim TH, Huh JW, Park KS, Jang SB, Kim HM, and Kim HS. (2004) Human endogenous retroviral elements belonging to the HERV-S family from human tissues, cancer cells, and primates: expression, structure, phylogeny and evolution. *Gene* 342, 283-292.

175. Cui X, and Churchill GA. (2003) Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol* 4, 210.

176. Cui X, Hwang JT, Qiu J, Blades NJ, and Churchill GA. (2005) Improved statistical tests for differential gene expression by shrinking variance components estimates. *Biostatistics* 6, 59-75.

177. Wu H, Kerr MK, Cui X, and Churchill GA (2002) *MAANOVA: A software package for the analysis of spotted cDNA microarray experiments.* Springer.

178. Searle S, Casella G, and McCulloch C (1992) *Variance Components*. John Wiley and Sons, Inc.

179. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, and Speed TP. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* 31, e15.

180. Welch BL. (1938) The significance of the difference between two means when the population variances are unequal. *Biometrika* 29, 350-362.

181. Neter J, Kutner MH, and Nachtsheim W (1996) *Applied Linear Statistical Models*. WCB McGraw Hill.

182. Westfall PH, and Young SS (1993) *Resampling-based multiple testing : examples and methods for P-value adjustment*. Wiley.

183. Keselman HJ, Cribbie R, and Holland B. (1999) The pairwise multiple comparison multiplicity problem: an alternative approach to familywise and comparisonwise type I error control. *Psychol Methods* 4, 58-69.

184. Yang H, and Churchill G. (2007) Estimating p-values in small microarray experiments. *Bioinformatics.* 23, 38-43.

185. Storey JD, and Tibshirani R. (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A.* 100, 9440-9445.

186. Tanaka TS, Jaradat SA, Lim MK, Kargul GJ, Wang X, Grahovac MJ, Pantano S, Sano Y, Piao Y, Nagaraja R, Doi H, Wood WH, Becker KG, and Ko MS. (2000) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray. *Proc Natl Acad Sci U S A* 97, 9127-9132.

187. Woo Y, Affourtit J, Daigle S, Viale A, Johnson K, Naggert J, and Churchill G. (2004) A comparison of cDNA, oligonucleotide, and Affymetrix GeneChip gene expression microarray platforms. *J Biomol Tech* 15, 276-284.

188. Ge X, Yamamoto S, Tsutsumi S, Midorikawa Y, Ihara S, Wang SM, and Aburatani H. (2005) Interpreting expression profiles of cancers by genome-wide survey of breadth of expression in normal tissues. *Genomics.* 86, 127-141.

189. Desmedt C, Piette F, Loi S, Wang Y, Lallemand F, Haibe-Kains B, Viale G, Delorenzi M, Zhang Y, d'Assignies MS, Bergh J, Lidereau R, Ellis P, Harris AL, Klijn JG, Foekens JA, Cardoso F, Piccart MJ, Buyse M, and Sotiriou C. (2007) Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin Cancer Res.* 13, 3207-3214.

190. Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A, Pawitan Y, Hall P, Klaar S, Liu ET, and Bergh J. (2005) An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci U S A.* 102, 13550-13555.

191. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, and Speed TP. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics.* 4, 249-264.

192. Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, Bunney WE, Myers RM, Speed TP, Akil H, Watson SJ, and Meng F. (2005) Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.* 33, e175.

193. Higdon R, van Belle G, and Kolker E. (2008) A note on the false discovery rate and inconsistent comparisons between experiments. *Bioinformatics.* 24, 1225-1228.

194. Ashburner M, Lewis S, and Reese MG. (2000) Annotating eukaryote genomes. *Curr. Opin. Struct. Biol.* 10, 349-354.

195. Doniger SW, Salomonis N, Dahlquist KD, Vranizan K, Lawlor SC, and Conklin BR. (2003) MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biol.* 4, R7.

196. Dahlquist KD, Salomonis N, Vranizan K, Lawlor SC, and Conklin BR. (2002) GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat Genet.* 31, 19-20.

197. Otahal P, Schell TD, Hutchinson SC, Knowles BB, and Tevethia SS. (2006) Early immunization induces persistent tumor-infiltrating CD8+ T cells against an immunodominant epitope and promotes lifelong control of pancreatic tumor progression in SV40 tumor antigen transgenic mice. *J Immunol.* 177, 3089-3099.

198. Schell TD, Mylin LM, Georgoff I, Teresky AK, Levine AJ, and Tevethia SS. (1999) Cytotoxic T-lymphocyte epitope immunodominance in the control of choroid plexus tumors in simian virus 40 large T antigen transgenic mice. *J Virol.* 73, 5981-5993.

199. Ye X, McCarrick J, Jewett L, and Knowles BB. (1994) Timely immunization subverts the development of peripheral nonresponsiveness and suppresses tumor development in simian virus 40 tumor antigen-transgenic mice. *Proc Natl Acad Sci U S A.* 91, 3916-3920.

200. Faas SJ, Pan S, Pinkert CA, Brinster RL, and Knowles BB. (1987) Simian virus 40 (SV40)-transgenic mice that develop tumors are specifically tolerant to SV40 T antigen. *J Exp Med.* 165, 417-427.

201. Wysolmerski JJ, Philbrick WM, Dunbar ME, Lanske B, Kronenberg H, and Broadus AE. (1998) Rescue of the parathyroid hormone-related protein knockout mouse demonstrates that parathyroid hormone-related protein is essential for mammary gland development. *Development.* 125, 1285-1294.

202. Sirchia R, and Luparello C. (2007) Mid-region parathyroid hormone-related protein (PTHrP) and gene expression of MDA-MB231 breast cancer cells. *Biol Chem.* 388, 457-465.

203. Luparello C, Sirchia R, and Pupello D. (2003) PTHrP [67-86] regulates the expression of stress proteins in breast cancer cells inducing modifications in urokinase-plasminogen activator and MMP-1 expression. *J Cell Sci.* 116, 2421-2430.

204. Luparello C, Romanotto R, Tipa A, Sirchia R, Olmo N, Lopez de Silanes I, Turnay J, Lizarbe MA, and Stewart AF. (2001) Midregion parathyroid hormone-related protein inhibits growth and invasion in vitro and tumorigenesis in vivo of human breast cancer cells. *J Bone Miner Res.* 16, 2173-2181.

205. Yoshida A, Nakamura Y, Shimizu A, Harada M, Kameda Y, Nagano A, Inaba M, and Asaga T. (2000) Significance of the parathyroid hormone-related protein expression in breast carcinoma. *Breast Cancer.* 7, 215-220.

206. Guise TA, Yin JJ, Taylor SD, Kumagai Y, Dallas M, Boyce BF, Yoneda T, and Mundy GR. (1996) Evidence for a causal role of parathyroid hormone-related protein in the pathogenesis of human breast cancer-mediated osteolysis. *J Clin Invest.* 98, 1544-1549.

207. Guise TA, and Mundy GR. (1996) Physiological and pathological roles of parathyroid hormone-related peptide. *Curr Opin Nephrol Hypertens.* 5, 307-315.

208. Bolos V, Grego-Bessa J, and de la Pompa JL. (2007) Notch signaling in development and cancer. *Endocr Rev.* 28, 339-363.

209. Johnson ML, and Rajamannan N. (2006) Diseases of Wnt signaling. *Rev Endocr Metab Disord.* 7, 41-49.

210. Hennighausen L (2007) Biology of the Mammary Gland. (http://mammary.nih.gov/index.html).

211. Wallace DC. (2005) Mitochondria and cancer: Warburg addressed. *Cold Spring Harb Symp Quant Biol.* 70, 363-374.

212. Ramanathan A, Wang C, and Schreiber SL. (2005) Perturbational profiling of a cell-line model of tumorigenesis by using metabolic measurements. *Proc Natl Acad Sci U S A.* 102, 5992-5997.

213. Warburg O. (1956) On the origin of cancer cells. *Science.* 123, 309-314.

214. Warburg O, Posener K, and Negelein E. (1924) The metabolism of cancer cells. *Biochem Z* .

215. Smit AFA, Hubley R, and Green P unpublished http://www.repeatmasker.org, http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker.

216. Wallace DC. (2001) Mouse models for mitochondrial disease. *Am J Med Genet* 106, 71-93.

217. Baylin S, and Bestor TH. (2002) Altered methylation patterns in cancer cell genomes: cause or consequence? *Cancer Cell* 1, 299-305.

218. Widschwendter M, and Jones PA. (2002) DNA methylation and breast carcinogenesis. *Oncogene* 21, 5462-5482.

219. Bonneton C, Sibarita JB, and Thiery JP. (1999) Relationship between cell migration and cell cycle during the initiation of epithelial to fibroblastoid transition. *Cell Motil Cytoskeleton* 43, 288-295.

220. Savagner P. (2001) Leaving the neighborhood: molecular mechanisms involved during epithelial-mesenchymal transition. *Bioessays* 23, 912-923.

221. Thiery JP. (2003) Epithelial-mesenchymal transitions in development and pathologies. *Curr Opin Cell Biol* 15, 740-746.

222. Vincent-Salomon A, and Thiery JP. (2003) Host microenvironment in breast cancer development: epithelial-mesenchymal transition in breast cancer development. *Breast Cancer Res* 5, 101-106.

223. Blumenberg M, Gao S, Dickman K, Grollman AP, Bottinger EP, and Zavadil J. (2007) Chromatin structure regulation in transforming growth factor-beta-directed epithelial-mesenchymal transition. *Cells Tissues Organs.* 185, 162-174.

224. Turley EA, Veiseh M, Radisky DC, and Bissell MJ. (2008) Mechanisms of Disease: epithelial-mesenchymal transition-does cellular plasticity fuel neoplastic progression? *Nat Clin Pract Oncol* 18, 18.

225. Matoba S, Kang JG, Patino WD, Wragg A, Boehm M, Gavrilova O, Hurley PJ, Bunz F, and Hwang PM. (2006) p53 regulates mitochondrial respiration. *Science.* 312, 1650-1653.

226. Kondoh H, Lleonart ME, Bernard D, and Gil J. (2007) Protection from oxidative stress by enhanced glycolysis; a possible mechanism of cellular immortalization. *Histol Histopathol.* 22, 85-90.

227. Kondoh H, Lleonart ME, Gil J, Beach D, and Peters G. (2005) Glycolysis and cellular immortalization. *Drug Discovery Today: Disease Mechanisms* 2, 263-267.

228. Garber K. (2004) Energy boost: the Warburg effect returns in a new theory of cancer. *J Natl Cancer Inst.* 96, 1805-1806.

229. Pfeiffer T, Schuster S, and Bonhoeffer S. (2001) Cooperation and competition in the evolution of ATP-producing pathways. *Science.* 292, 504-507.

230. Guppy M, Leedman P, Zu X, and Russell V. (2002) Contribution by different fuels and metabolic pathways to the total ATP turnover of proliferating MCF-7 breast cancer cells. *Biochem J.* 364, 309-315.

231. Zu XL, and Guppy M. (2004) Cancer metabolism: facts, fantasy, and fiction. *Biochem Biophys Res Commun.* 313, 459-465.

232. Assaily W, and Benchimol S. (2006) Differential utilization of two ATP-generating pathways is regulated by p53. *Cancer Cell.* 10, 4-6.

233. Bensaad K, and Vousden KH. (2007) p53: new roles in metabolism. *Trends Cell Biol.* 17, 286-291.

234. Ma W, Sung HJ, Park JY, Matoba S, and Hwang PM. (2007) A pivotal role for p53: balancing aerobic respiration and glycolysis. *J Bioenerg Biomembr.* 39, 243-246.

235. Semenza GL. (2002) Involvement of hypoxia-inducible factor 1 in human cancer. *Intern Med.* 41, 79-83.

236. Kim JW, Tchernyshyov I, Semenza GL, and Dang CV. (2006) HIF-1-mediated expression of pyruvate dehydrogenase kinase: a metabolic switch required for cellular adaptation to hypoxia. *Cell Metab.* 3, 177-185.

237. Plas DR, and Thompson CB. (2005) Akt-dependent transformation: there is more to growth than just surviving. *Oncogene.* 24, 7435-7442.

238. Shaw RJ, and Cantley LC. (2006) Ras, PI(3)K and mTOR signalling controls tumour cell growth. *Nature.* 441, 424-430.

239. Sanchez-Puig N, Veprintsev DB, and Fersht AR. (2005) Binding of natively unfolded HIF-1alpha ODD domain to p53. *Mol Cell.* 17, 11-21.

240. Ravi R, Mookerjee B, Bhujwalla ZM, Sutter CH, Artemov D, Zeng Q, Dillehay LE, Madan A, Semenza GL, and Bedi A. (2000) Regulation of tumor angiogenesis by p53-induced degradation of hypoxia-inducible factor 1alpha. *Genes Dev.* 14, 34-44.

241. Grewal SIS, and Elgin SCR. (2002) Heterochromatin: new possibilities for the inheritance of structure. *Current Opinion in Genetics & Development* 12, 178-187.

242. Eissenberg JC, and Elgin SC. (2005) Molecular biology: antagonizing the neighbours. *Nature.* 438, 1090-1091.

243. Yamane K, Tateishi K, Klose RJ, Fang J, Fabrizio LA, Erdjument-Bromage H, Taylor-Papadimitriou J, Tempst P, and Zhang Y. (2007) PLU-1 is an H3K4 demethylase involved in transcriptional repression and breast cancer cell proliferation. *Mol Cell.* 25, 801-812.

244. Tahiliani M, Mei P, Fang R, Leonor T, Rutenberg M, Shimizu F, Li J, Rao A, and Shi Y. (2007) The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation. *Nature* 29, 29.

245. Reinhart BJ, and Bartel DP. (2002) Small RNAs correspond to centromere heterochromatic repeats. *Science.* 297, 1831.

246. Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, and Martienssen RA. (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science.* 297, 1833-1837.

247. Finnegan EJ, and Matzke MA. (2003) The small RNA world. *J Cell Sci.* 116, 4689-4693.

248. Mathieu O, and Bender J. (2004) RNA-directed DNA methylation. *J Cell Sci.* 117, 4881-4888.

249. Peters AH, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schofer C, Weipoltshammer K, Pagani M, Lachner M, Kohlmaier A, Opravil S, Doyle M, Sibilia M, and Jenuwein T. (2001) Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell.* 107, 323-337.

250. Grewal SI, and Rice JC. (2004) Regulation of heterochromatin by histone methylation and small RNAs. *Curr Opin Cell Biol.* 16, 230-238.

251. Tachibana M, Sugimoto K, Nozaki M, Ueda J, Ohta T, Ohki M, Fukuda M, Takeda N, Niida H, Kato H, and Shinkai Y. (2002) G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev.* 16, 1779-1791.

252. Li E, Bestor TH, and Jaenisch R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell.* 69, 915-926.

253. Okano M, Bell DW, Haber DA, and Li E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell.* 99, 247-257.

254. Dodge JE, Okano M, Dick F, Tsujimoto N, Chen T, Wang S, Ueda Y, Dyson N, and Li E. (2005) Inactivation of Dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. *J Biol Chem.* 280, 17986-17991.

255. Weaving LS, Christodoulou J, Williamson SL, Friend KL, McKenzie OL, Archer H, Evans J, Clarke A, Pelka GJ, Tam PP, Watson C, Lahooti H, Ellaway CJ, Bennetts B, Leonard H, and Gecz J. (2004) Mutations of CDKL5 cause a severe neurodevelopmental disorder with infantile spasms and mental retardation. *Am J Hum Genet.* 75, 1079-1093.

256. Pelka GJ, Watson CM, Radziewic T, Hayward M, Lahooti H, Christodoulou J, and Tam PP. (2006) Mecp2 deficiency is associated with learning and cognitive deficits and altered gene activity in the hippocampal region of mice. *Brain* 8, 8.

257. Lagger G, O'Carroll D, Rembold M, Khier H, Tischler J, Weitzer G, Schuettengruber B, Hauser C, Brunmeir R, Jenuwein T, and Seiser C. (2002) Essential function of histone deacetylase 1 in proliferation control and CDK inhibitor repression. *Embo J.* 21, 2672-2681.

258. Murchison EP, Partridge JF, Tam OH, Cheloufi S, and Hannon GJ. (2005) Characterization of Dicer-deficient murine embryonic stem cells. *Proc Natl Acad Sci U S A.* 102, 12135-12140.

259. Kanellopoulou C, Muljo SA, Kung AL, Ganesan S, Drapkin R, Jenuwein T, Livingston DM, and Rajewsky K. (2005) Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev.* 19, 489-501.

260. Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, Song JJ, Hammond SM, Joshua-Tor L, and Hannon GJ. (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science.* 305, 1437-1441.

261. Song JJ, Smith SK, Hannon GJ, and Joshua-Tor L. (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. *Science.* 305, 1434-1437.

262. Hagan CR, and Rudin CM. (2007) DNA cleavage and Trp53 differentially affect SINE transcription. *Genes Chromosomes Cancer.* 46, 248-260.

263. Wilson AS, Power BE, and Molloy PL. (2007) DNA hypomethylation and human diseases. *Biochim Biophys Acta.* 1775, 138-162.

264. Oricchio E, Sciamanna I, Beraldi R, Tolstonog GV, Schumann GG, and Spadafora C. (2007) Distinct roles for LINE-1 and HERV-K retroelements in cell proliferation, differentiation and tumor progression. *Oncogene.* 26, 4226-4233.

265. Shi X, Seluanov A, and Gorbunova V. (2007) Cell divisions are required for L1 retrotransposition. *Mol Cell Biol.* 27, 1264-1270.

266. Okeoma CM, Lovsin N, Peterlin BM, and Ross SR. (2007) APOBEC3 inhibits mouse mammary tumour virus replication in vivo. *Nature.* 445, 927-930.

267. Kinomoto M, Kanno T, Shimura M, Ishizaka Y, Kojima A, Kurata T, Sata T, and Tokunaga K. (2007) All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. *Nucleic Acids Res.* 35, 2955-2964.

268. Nishikura K. (2006) Editor meets silencer: crosstalk between RNA editing and RNA interference. *Nat Rev Mol Cell Biol.* 7, 919-931.

269. Ohman M. (2007) A-to-I editing challenger or ally to the microRNA process. *Biochimie* 8, 8

270. Thiery JP, and Chopin D. (1999) Epithelial cell plasticity in development and tumor progression. *Cancer Metastasis Rev* 18, 31-42

271. Guarino M, Micheli P, Pallotti F, and Giordano F. (1999) Pathological relevance of epithelial and mesenchymal phenotype plasticity. *Pathol Res Pract* 195, 379-389

272. Gilles C, Polette M, Piette J, Delvigne AC, Thompson EW, Foidart JM, and Birembaut P. (1996) Vimentin expression in cervical carcinomas: association with invasive and migratory potential. *J Pathol.* 180, 175-180.

273. Hendrix MJ, Seftor EA, Seftor RE, and Trevor KT. (1997) Experimental co-expression of vimentin and keratin intermediate filaments in human breast cancer cells results in phenotypic interconversion and increased invasive behavior. *Am J Pathol.* 150, 483-495.

274. Xue C, Plieth D, Venkov C, Xu C, and Neilson EG. (2003) The gatekeeper effect of epithelial-mesenchymal transition regulates the frequency of breast cancer metastasis. *Cancer Res* 63, 3386-3394.

275. Boyer B, Tucker GC, Valles AM, Gavrilovic J, and Thiery JP. (1989) Reversible transition towards a fibroblastic phenotype in a rat carcinoma cell line. *Int J Cancer Suppl.* 4, 69-75.

276. Medici D, Hay ED, and Goodenough DA. (2006) Cooperation between snail and LEF-1 transcription factors is essential for TGF-beta1-induced epithelial-mesenchymal transition. *Mol Biol Cell.* 17, 1871-1879.

277. Peinado H, Olmeda D, and Cano A. (2007) Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nat Rev Cancer.* 7, 415-428.

278. Mikaelian I, Nanney LB, Parman KS, Kusewitt DF, Ward JM, Naf D, Krupke DM, Eppig JT, Bult CJ, Seymour R, Ichiki T, and Sundberg JP. (2004) Antibodies that label paraffin-embedded mouse tissues: a collaborative endeavor. *Toxicol Pathol* 32, 181-191.

279. Damonte P, Gregg JP, Borowsky AD, Keister BA, and Cardiff RD. (2007) EMT tumorigenesis in the mouse mammary gland. *Lab Invest.* 87, 1218-1226.

280. Wu H, Kerr MK, Cui X, and Churchill GA (2003) *MAANOVA: A software package for the analysis of spotted cDNA microarray experiments.* Springer.

281. Acevedo VD, Gangula RD, Freeman KW, Li R, Zhang Y, Wang F, Ayala GE, Peterson LE, Ittmann M, and Spencer DM. (2007) Inducible FGFR-1 activation leads to irreversible prostate adenocarcinoma and an epithelial-to-mesenchymal transition. *Cancer Cell.* 12, 559-571.

282. Thiery JP. (2002) Epithelial-mesenchymal transitions in tumour progression. *Nat Rev Cancer* 2, 442-454.

283. Declerck Y. (2004) Focus on the cell membrane: the need for dissociation and detachment in tumoral invasion. *Cancer Biol Ther* 3, 632-633.

284. DeClerck YA, Mercurio AM, Stack MS, Chapman HA, Zutter MM, Muschel RJ, Raz A, Matrisian LM, Sloane BF, Noel A, Hendrix MJ, Coussens L, and Padarathsingh M. (2004) Proteases, extracellular matrix, and cancer: a workshop of the path B study section. *Am J Pathol* 164, 1131-1139.

285. Larue L, and Bellacosa A. (2005) Epithelial-mesenchymal transition in development and cancer: role of phosphatidylinositol 3' kinase/AKT pathways. *Oncogene.* 24, 7443-7454.

286. Pujuguet P, Simian M, Liaw J, Timpl R, Werb Z, and Bissell MJ. (2000) Nidogen-1 regulates laminin-1-dependent mammary-specific gene expression. *J Cell Sci* 113, 849-858

287. Yamada KM, Pankov R, and Cukierman E. (2003) Dimensions and dynamics in integrin function. *Braz J Med Biol Res* 36, 959-966.

288. Derenzini M, Trere D, Pession A, Govoni M, Sirri V, and Chieco P. (2000) Nucleolar size indicates the rapidity of cell proliferation in cancer tissues. *J Pathol* 191, 181-186.

289. Diehl JA, Yang W, Rimerman RA, Xiao H, and Emili A. (2003) Hsc70 regulates accumulation of cyclin D1 and cyclin D1-dependent protein kinase. *Mol Cell Biol.* 23, 1764-1774.

290. Malumbres M, and Barbacid M. (2001) To cycle or not to cycle: a critical decision in cancer. *Nat Rev Cancer* 1, 222-231.

291. Sergeev IN. (2005) Calcium signaling in cancer and vitamin D. *J Steroid Biochem Mol Biol* 1, 1.

292. Huber MA, Kraut N, and Beug H. (2005) Molecular requirements for epithelial-mesenchymal transition during tumor progression. *Curr Opin Cell Biol.* 17, 548-558.

293. Andarawewa KL, Erickson AC, Chou WS, Costes SV, Gascard P, Mott JD, Bissell MJ, and Barcellos-Hoff MH. (2007) Ionizing radiation predisposes nonmalignant human mammary epithelial cells to undergo transforming growth factor beta induced epithelial to mesenchymal transition. *Cancer Res.* 67, 8662-8670.

294. Valcourt U, Kowanetz M, Niimi H, Heldin CH, and Moustakas A. (2005) TGF-beta and the Smad signaling pathway support transcriptomic reprogramming during epithelial-mesenchymal cell transition. *Mol Biol Cell.* 16, 1987-2002.

295. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S, Backlund MG, Yin Y, Khramtsov AI, Bastein R, Quackenbush J, Glazer RI, Brown PH, Green JE, Kopelovich L, Furth PA, Palazzo JP, Olopade OI, Bernard PS, Churchill GA, Van Dyke T, and Perou CM. (2007) Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol.* 8, R76.

296. Zhang H, Qian DZ, Tan YS, Lee K, Gao P, Ren YR, Rey S, Hammers H, Chang D, Pili R, Dang CV, Liu JO, and Semenza GL. (2008) Inaugural Article: Digoxin and other cardiac glycosides inhibit HIF-1{alpha} synthesis and block tumor growth. *Proc Natl Acad Sci USA* 25, 25.

297. Korpal M, Lee ES, Hu G, and Kang Y. (2008) The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2. *J Biol Chem.* 283, 14910-14914.

298. Park SM, Gaur AB, Lengyel E, and Peter ME. (2008) The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes Dev.* 22, 894-907.

APPENDIX

Epithelial to Mesenchymal Transition is Associated with Decreased Cell Proliferation, Apoptosis and Metastatic Potential

Igor Mikaelian[1,*] , Karen S. Fancher[2,*] , Jason I. Herschkowitz[3] , Karen Rasmussen[4], Qian Li[2], Yong H. Woo[2], Charles M. Perou[3] , Barbara B. Knowles[2,5] and Gary A. Churchill[2¶]

[1]Hoffmann La-Roche Inc., Nutley, NJ 07110; [2]The Jackson Laboratory, Bar Harbor, ME 04609; [3]Lineberger Cancer Center, CB #7295, University of North Carolina, Chapel Hill, NC 27599; and [4]Maine Medical Center, Portland, ME 04102; [5]Institute of Medical Biology, A*STAR, Singapore, 138648; *I. Mikaelian and K. Fancher contributed equally to this paper.

Abstract

Epithelial to Mesenchymal Transition (EMT) is a common tumor phenotype in the FVB/N-Tg(WapMyc) 212Bri/J [Myc] mouse model of mammary carcinogenesis. In this study, histopathology, immunohistochemistry and gene microarray analysis were used to characterize EMT in this model. When compared with epithelial tumors exhibiting <1% EMT, genes over-expressed during early EMT (primary tumors with 13-50% EMT) showed significant enrichment in cell-matrix adhesion molecules, extracellular matrix homeostasis, and angiogenesis, all classic characteristics of EMT. Tumors with early EMT also exhibited decreased apoptosis, a feature commonly associated with cancer; however, our findings of a cell cycle block and expression of myoepithelial markers in early EMT are novel.

To better understand the transition toward an exclusively mesenchymal phenotype, *in vivo* passage was used to generate epithelial cell-derived pure mesenchymal tumors. Analysis of these spindle cell carcinomas (SCCs) confirmed an invasive molecular profile, including improved vasculature and reduced apoptosis similar to early EMT. SCCs had a lower metastatic rate than the epithelial carcinomas from which they were derived, a previously undescribed feature of EMT. The reduced metastatic capability of SCC cells appears to result from their inability to access the

blood stream since SCCs readily colonized the lungs when injected in the tail vein. Finally, SCCs lacked expression of myoepithelial markers and the archetypal master regulators of EMT (*Twist1*, *Twist2,* and *Snai2*), but had partially conquered the cell cycle block of early EMT, indicating that early EMT and SCCs represent two distinct steps in cancer progression.

Introduction

Acquisition of the mesenchymal phenotype is a key stage in development and tumor progression[219,270,271]. This process, termed epithelial to mesenchymal transition (EMT), is coupled with invasion. Morphologic changes during EMT include loss of the cuboidal/polygonal shaped epithelial cells and of cell-cell adhesions, gain of a spindloid morphology, and transformation of the actin cytoskeleton and extracellular matrix (ECM) to enable motility. Molecularly, EMT is associated with de novo expression of markers of mesenchymal cells such as vimentin (*Vim*)[272,273], markers of invasion such as *S100a4, S100a6,* and metalloproteinases[274], alongside concurrent loss of cell-cell adhesion markers such as *Cdh1* (formerly E-cadherin)[275,276]. Additional EMT-associated markers include members of the *Snail, Twist,* and *Zeb* families of transcription factors, which directly inhibit *Cdh1* transcription to promote EMT (reviewed in[277]).

At least three major signaling pathways have been implicated in EMT: the receptor tyrosine kinase (RTK), transforming growth factor β (TGFB), and the WNT/β-catenin (CTNNB1) pathways (reviewed in[220,222]). How these and other pathways may collaborate during EMT remains under investigation[276,277].

Transgenesis of the mammary gland–targeted myelocytomatosis oncogene (*Myc*) is amongst the first and best described mouse models of carcinogenesis[52,55,56], characterized by the development of glandular and solid mammary adenocarcinomas. In recent years, tumors in this model were found to be particularly prone to undergo EMT[278,279]. The value of Myc transgenic models of mammary carcinogenesis is supported by the observation that human tumors with EMT have genomic amplification of MYC[58].

In the present study, Myc mammary carcinomas were examined using histology, microarrays, and immunohistochemistry to characterize multiple stages of EMT. First, primary Myc tumors with

EMT were compared to epithelial ones. Second, to further our knowledge of tumor progression beyond EMT, *in vivo* passage was used to generate tumors comprised exclusively of either mesenchymal cells (spindle cell carcinomas or SCCs), or pure epithelial tumors. Finally, tail vein injection experiments evaluated the true metastatic ability of in-vivo passage tumors. Our results indicate that tumorigenesis associated with EMT is a multistep process; several characteristic markers of EMT are expressed only in early EMT but not in SCCs; and EMT hinders metastasis in this model.

Methods

Primary Tumors. Hemizygous FVB/N-Tg(WapMyc)212Bri/J [Myc] mice[55] were obtained from The Jackson Laboratory (Bar Harbor, ME; stock #: 002677). Mice were aged (90-443 days) until palpable mammary masses developed. Half of each tumor was collected in RNAlater (Ambion, Austin, TX) and the remaining portion was fixed in Fekete's acid-alcohol-formalin for histology. The proportion of mesenchymal cells in 73 tumors from 41 mice was evaluated on H&E stained sections using Photoshop 6.0 (Adobe, San Jose, CA). On the basis of histopathology, three groups, each containing five carcinomas, were evaluated: the first group contained neoplastic cells with an epithelial phenotype (5 tumors without EMT); the second group featured <1% mesenchymal (spindloid) neoplastic cells (5 tumors); the third group, designated "tumors with early EMT", contained 13-50% neoplastic cells harboring a mesenchymal phenotype (5 tumors with >12% EMT) in which four of these five tumors displayed large multifocal coalescing areas of spindloid cells.

Mitotic rate was assessed in ten randomly-selected high power microscopic fields on whole tumors and nucleolar size was estimated in 45-102 nuclei; in tumors with >12% EMT, spindloid and epithelial portions were evaluated separately. Paraffin-embedded sections (5-6 μm thick) were immunolabeled and graded for αSMA (Sigma, St. Louis, MO), COL4 (Chemicon, Temecula, CA), K1, K5, K6, K10, K14 (BabCo, Richmond, CA), K17 (P. A. Coulombe, The Johns Hopkins University, Baltimore, MD), K8/18 (Progen, Heidelberg, Germany), VIM (Biomeda, Foster City, CA), cleaved caspase 3 (Cell Signaling Technology, Beverly, MA), and CCND1

(Dakocytomation, Carpinteria, CA) as described previously[278]. A tumor was considered negative for a marker if immunolabeling was detected in less than 10 cells within the tumor.

Generation of Tumors with a Pure Phenotype. Portions of tumors (n=18) arising from Myc mice were collected aseptically to prepare cell suspensions while the remainder of the tumor and the lungs were prepared for histology. For single cell suspensions, each tumor portion was finely chopped in sterile phosphate buffered saline (PBS), incubated with 0.1% collagenase (Sigma, St. Louis, MO) for 2-3 hours at room temperature, rinsed three times in PBS, and filtered through Sefar Nitex 112 µm mesh (03-112/40; Sefar Holding, Rüschlikon, Switzerland). Single cell suspensions (5,000-25,000 cells?) were injected in the gastrocnemius muscle of FVB.Cg-Tg(ACTB-EGFP)B5Nagy/J (GFPU; The Jackson Laboratory; stock #: 003516) mice (n=2 mice/single cell suspension preparation) anesthetized with xylazine (19 mg/kg; Phoenix Pharmaceutical Inc., St. Joseph, MO) and ketamine hydrochloride (95 mg/kg; Fort Dodge Animal Health, Fort Dodge, IA). Mice were monitored bi-weekly until a tumor developed at the site of the injection or for 2 months. Portions of these tumors were harvested, new single cell suspensions prepared (as described above), and this process was repeated until three types of tumors with pure phenotypes were obtained: glandular/trabecular, solid/comedo, and pure spindloid tumors (designated as SCCs). Selected tumors were genotyped to ensure the passaged tissue contained the Myc transgene. Tumor invasiveness was evaluated on H&E stained sections by assessing entrapment of skeletal muscle fibers, nerves and adipocytes in randomly selected SCCs (n=35) and solid/comedo carcinomas (n=35). Tumors selected for microarrays were characterized by immunohistochemistry for the same markers as for primary tumors plus tropomyosin (Chemicon).

***In Vivo* Tail Vein Metastasis Analysis.** Single cell suspensions (5,000-25,000 cells) of SCCs (n=58) and solid/comedo carcinomas (n=16; positive control) in a volume of 0.1 mL 0.9% saline were injected intravenously (tail vein) or intra-muscularly (gastrocnemius muscle) into anesthetized GFPU mice (2 mice for intravenous injections and 1 mouse for intramuscular injections). The mice were monitored bi-weekly and euthanized when the mouse injected

intramuscularly had a 10 mm in diameter tumor. Histology was performed on the intramuscular tumor and on the lungs of all mice.

Microarrays. Treatment of tissue samples, RNA isolation, cDNA synthesis and labeling, and hybridizations were performed as previously described[187]. Labeled samples were hybridized in pairs, with the Universal Mouse Reference (Stratagene, La Jolla, CA), using a reference dye-swap method[280] onto Compugen 22K mouse oligonucleotide microarrays. Mean intensity values were used in subsequent statistical analysis. Microarrays were generated for 14 primary tumors representing three phenotypes (4 without EMT, 5 with <1% EMT, and 5 with >12% EMT) and separately for 15 tumors with a pure phenotype (5 glandular, 7 solid/comedo, 3 SCCs). Microarray results for primary tumors were confirmed by quantitative reverse-transcriptase polymerase chain reaction (QRT-PCR) for 12 genes representing various Gene Ontology (GO) categories (Fig. 3C; Fig. 4).

Statistical Microarray Analysis. R/MAANOVA (http://research.jax.org/faculty/churchill/) was used for analysis[280]. The mean channel intensities were imported into R and normalized using lowess transformation. The Analysis of Variance (ANOVA) method was used to decompose the channel intensities into an expression estimate for each tumor sample, after accounting for Array and Dye effects[280]. For each experiment, differential expression among tumor groups was tested by ANOVA using a fixed model. (For the in-vivo passage experiment, pairwise contrasts were also performed to compare between any two groups.) For both experiments, pooled permutation p-values of the Fs statistic were used[176,281] to generate gene lists using a false discovery rate ($q \leq 0.05$) adjustment for multiple testing and a fold change of $\geq$ 1.2. Additionally, any gene with a relative fold change greater than 2.0 was included, regardless of q-value[193]. Hierarchical clustering was carried out in JMP software (SAS Institute, Cary, NC). Gene Ontology categories associated with significant genes were identified using Visual Annotation Display (http://proto.informatics.jax.org/prototypes/vlad-1.0.3/). Pathway analyses were performed using Ingenuity Pathway Analysis (Ingenuity Systems, Redwood City, CA) or GenMAPP 2.1 (Gladstone Institutes, San Francisco, CA). The data have been deposited in the

Gene Expression Omnibus, http://www.ncbi.nlm.nih.gov/geo/, under the accession GSE1498 (primary tumors) and GSE1533 (tumors with a pure phenotype).

Quantitative Reverse Transcriptase-Polymerase Chain Reaction (QRT-PCR). Total RNA (the same RNA isolated and analyzed on the microarrays) from each sample was DNAse treated (DNA-free™, Ambion) prior to cDNA synthesis. Using both random hexamers and oligo dT primers, cDNA was synthesized from 1μg aliquots as described for First-Stand cDNA Synthesis using random primers (Superscript™ III, Invitrogen Corp., Carlsbad, CA). A *Ctnnb1* PCR (with primers spanning intron/exon boundaries) was used to test for genomic contamination. For each gene, Assays-on-Demand™ Gene Expression Products (Applied Biosystems, Foster City, CA) were used to compare cDNA from tumors with >15% EMT (experimental) versus tumors without EMT (control). For each QRT-PCR of each gene, cDNA from the equivalent of 2ng total RNA (prior to DNAse treatment) was used. Accurate expression changes were determined using the Δ, ΔCT method (User Bulletin #2, http://www.appliedbiosystems.com) and *H3f3b* as the endogenous control (normalizer), which showed constant expression across all samples. (*H3f3b* was not significant via microarray results either; in tumors with early EMT, *H3f3b* showed a fold change of one, compared to tumors with no EMT.) Delta, ΔCT values were averaged from two independent experiments (Figure A4), run on 7700 and 7500 Sequence Detection Systems, with three replicates for each gene and each tumor per run.

Results

Histological and Immunohistochemical Characterization of Spindle Cells of Early EMT. Acquisition of the mesenchymal phenotype was accompanied by the loss of glandular/trabecular organization and the formation of bundles and fascicles (Fig. A1, D-F versus A-C). Tumor cells with an epithelial phenotype were polygonal with a basophilic cytoplasm, with round nuclei containing 1-5 large nucleoli. Tumor cells with a mesenchymal phenotype were spindloid with acidophilic cytoplasm and their nuclei were oval to elongated containing 1-3 small nucleoli. Spindloid cells accounted for 32% ± 19% of the five tumors studied (>12% EMT; Fig. A1D-F).

Primary epithelial tumors expressed little or no markers of myoepithelial cells: keratin 5, K5 (0/4), K6 (0/4), K8/18 (0/4), K14 (1/4), and K17 (0/4) (data not shown). In contrast, spindloid cells in tumors with early EMT expressed keratins 5 (3/5), 6 (2/5), 8/18 (5/5), 14 (4/5), 17 (4/5), α-smooth muscle actin (4/5), vimentin (4/5), and were surrounded by collagen type IV (Fig. A1G-M). Thus, early EMT is associated with expression of terminal differentiation markers typical of myoepithelial cells, de novo expression of vimentin, a hallmark of mesenchymal cells, and increased ECM as assessed by collagen type IV.

***In Vivo* Selection Identifies Multistep Tumor Progression.** To gain insight into tumor progression changes following EMT, we generated tumors with a pure phenotype through intra-muscular injection, serial passage and selection. Sequential morphologic steps of tumor progression were represented: (1) glandular carcinomas, composed of epithelial cells that formed trabeculae, cords and glands (Fig. A2A); (2) solid/comedo carcinomas, composed of epithelial cells that formed large solid areas with central necrosis, which arose from glandular carcinomas (Fig. A2B); (3) SCCs which arose from solid/comedo tumors and were composed of spindloid cells of epithelial origin that formed bundles and fascicles (Fig. A2C). Unlike early EMT tumors, neoplastic cells in SCCs showed no expression of epithelial or myoepithelial markers; indeed, among these, only vimentin was expressed (immunohistochemistry data not shown).

Early EMT Tumors and SCCs Are Molecularly Distinct from Epithelial Tumors. Hierarchical clustering of ANOVA expression estimates adequately identified the histological type of all but two primary tumors, Samples 12 and 13 (Fig. A3A). Sample 12, with >12% EMT, showed EMT in only one of its lobes and clustered with tumors without EMT, suggesting that an epithelial portion of this tumor may have been processed for RNA; EMT was multifocal in the other four tumors with >12% EMT. Similarly, Sample 13, without EMT, clustered with tumors showing <1% EMT (Fig. A3A). In light of these findings, we classified tumors into two groups (5 with >12% EMT and 9 with <1% EMT); the pairwise comparison between these identified 526 unique genes of 536 clones differentially expressed (Fig. A3B; Table S1[2]).

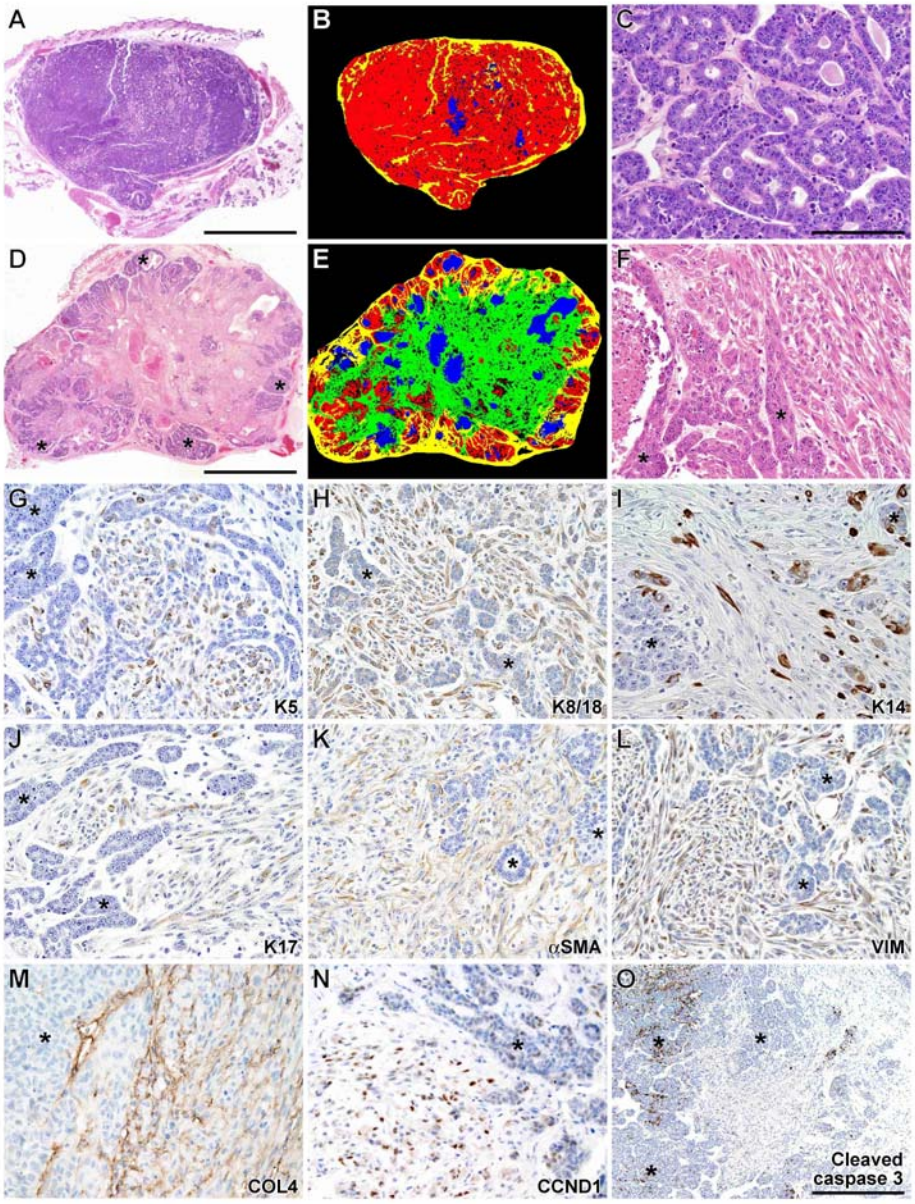


Figure A1. Tumors Arising in Mice Transgenic for Myc generally have an epithelial phenotype (***A-C***) but may also comprise large areas of spindloid neoplastic cells (***D-F***; hematoxylin & eosin). The ratio of epithelial (red) and spindloid neoplastic cells (green) to stroma (yellow) and the areas of necrosis (blue) was evaluated by image analysis (***B, E***). Compared to epithelial neoplastic cells (*), spindloid neoplastic cells show stronger expression of: keratins 5 (***G***), 8/18 (***H***), 14 (***I***), and 17 (***J***); α-smooth muscle actin (***K***); vimentin (***L***); and cyclin D1 (***N***). Areas composed of epithelial cells contain less collagen type IV (***M***) and show more apoptosis (***O***) than the spindloid portions of the tumors. Bar is 4.25 mm (*A-B*), 107 μm (C,*F-N*), 3.9 mm (*D-E*), and 215 μm (*O*).

When compared with epithelial tumors, EMT tumors displayed over-expression of classical markers of EMT: transcriptions factors *Snai2, Twist1, Twist2, Zeb1* and *Zeb2*[277]; and *Vim*[272,273] (Fig. A4; Table S1[2]).

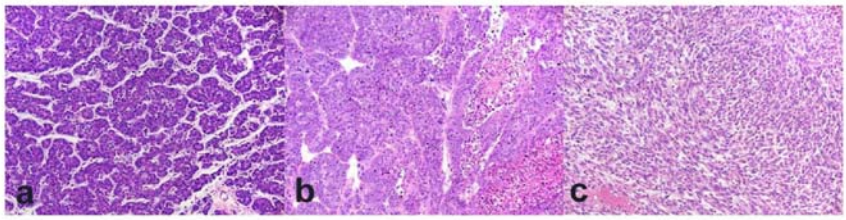


Figure A2. Tumors with a Pure Phenotype Generated Through Serial *In Vivo* Passage from epithelial cell suspensions. Representative histological phenotypes for glandular (*a*), solid/comedo (*b*), or spindloid (*c*) tumors are shown. [Hematoxylin & eosin stain; scale is 10x.]

In SCCs, 3,258 unique genes of 3,360 clones were differentially expressed in the pairwise contrast with epithelial tumors (Table S4[2]). Conventional markers of EMT (*Snai1, Zeb1, Zeb2* and *Vim*) were abundantly over-expressed while *Cdh1* was down-regulated, as expected (Table S1, S4[2]). However, some other classical EMT markers observed in early EMT were not differentially expressed in SCCs, specifically *Twist1, Twist2* and *Snai2*.

EMT is Associated with Altered Expression of Adhesion Molecules and ECM Homeostasis Genes. Extracellular matrix production and remodeling are hallmark features of EMT and suggest progression towards an invasive phenotype[220,274,282-284]. Indeed, the GO categories related to cell adhesion, the ECM, basement membrane, actin cytoskeleton, and collagen were enriched in early EMT tumors and SCCs (Fig. A3C; Fig. A4; Table S1, S2, S4[2]). Tumors with early EMT and SCCs over expressed genes with metalloproteinase activity (*Mmp2, 3, 9, 13,* and *14,* and *Bmp1*) as well as *Timp1, Timp2, Timp3, Fbln1, Fbln2,* and *Efemp2* (Fig. A4; Table S1, S4[2]), demonstrating the importance of maintaining homeostasis of the ECM; up-regulation of some of these is associated with cell migration in EMT[285]. Contributing to cell motility[283,284], tumors with early EMT and SCCs over-expressed numerous ECM genes or proteins and their cell surface receptor (Fig. A1M; Fig. A4; Table S1[2]): membrane receptor nidogen 1 and its ECM binding partners laminin 1 and collagen type IV[286]; ECM constituent fibronectin (FN1) with its receptor the heterodimer ITGB1/ITGA5[287]; fibronectin, biglycan, and decorin, which may bind collagen and/or thrombospondin; and *Thbs1* and *Thbs2*, that are also markers of invasion[287].

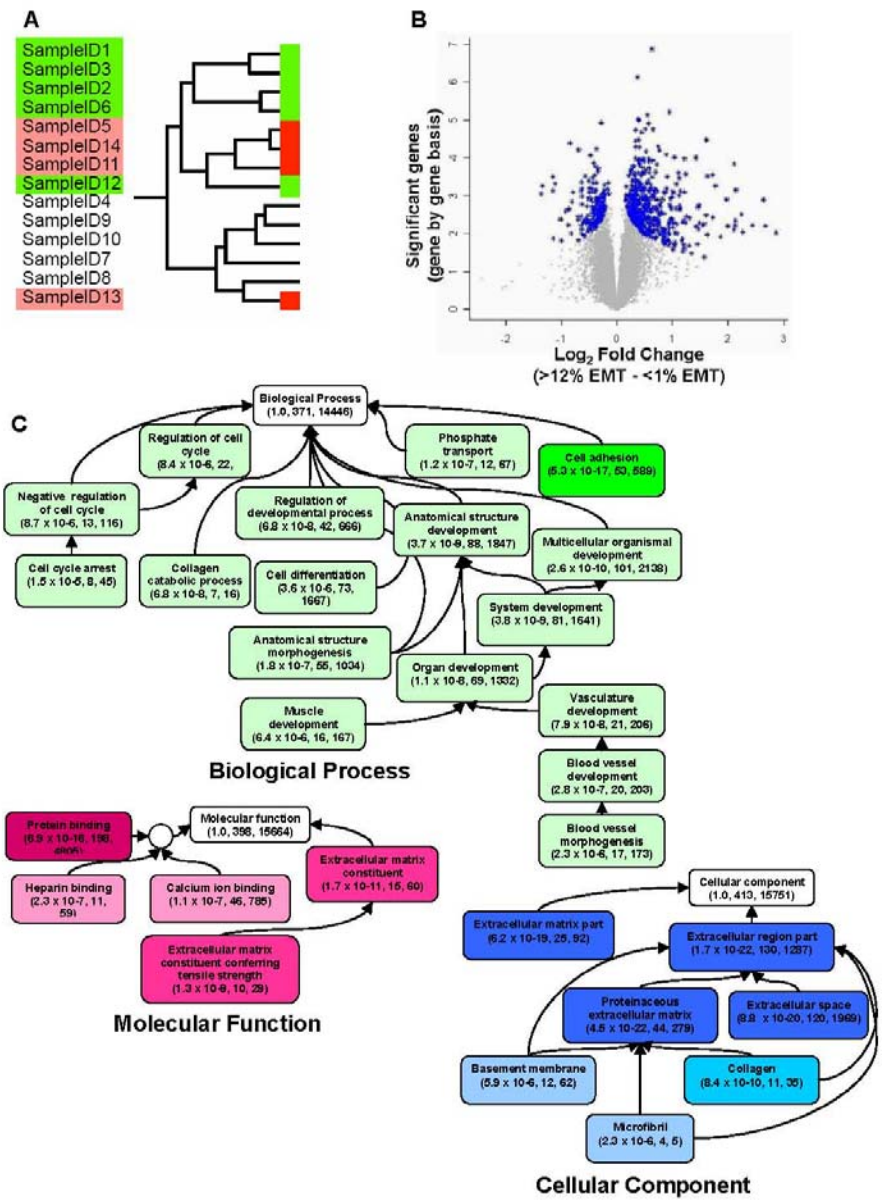


Figure A3. Microarray Results Across 14 Early Epithelial to Mesenchymal Transition (EMT) Myc Tumors. (***A***) Hierarchical clustering of ANOVA estimates. Tumors branched into three clusters: one group was comprised of only tumors with >12% EMT (green highlight); light red highlighting identifies exclusively epithelial cell tumors; no highlighting designates the tumors with minimal EMT; (***B***) When organized into two groups (tumors with >12% EMT and tumors with <1% EMT), 526 differentially expressed genes were identified (blue marks; $q \leq 0.05$); (***C***) Gene Ontology analysis of Biological Process, Molecular Function, and Cellular Component categories over represented in the 526 gene list compared to all genes in Mouse Genome Informatics (www.informatics.org). [Darker color rectangles indicate greater significance. Shown in parenthesis: p-value, # of genes analyzed this experiment, # of GO annotated genes background set. The complete list of statistically significant GO categories is presented in Supplemental Table S2[2].]

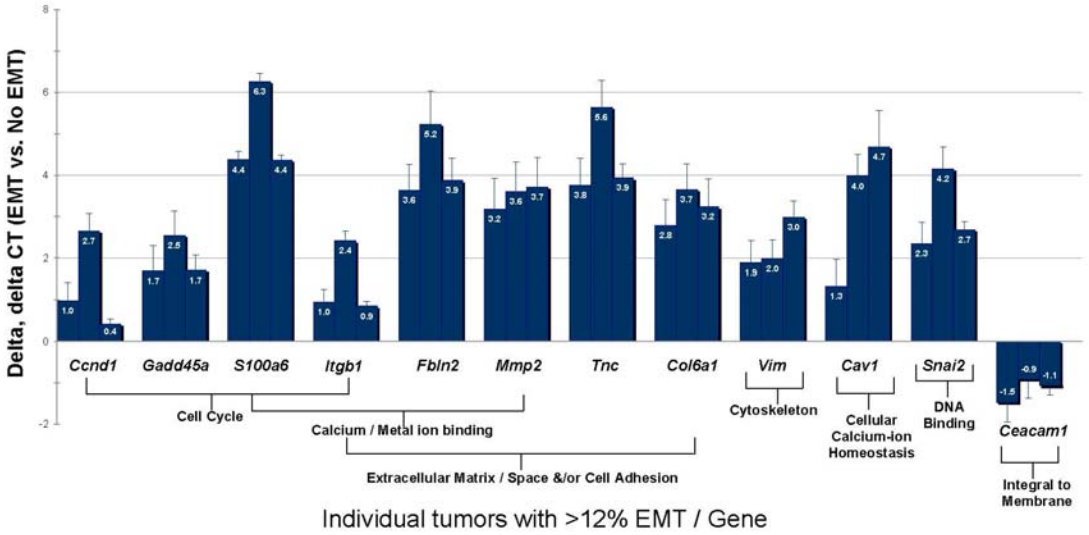
Other reported markers of cell motility/invasiveness, such as *Cd97, Lox, Loxl1, Loxl2, Plaur, S100a6, Tnc,* and *Sparc* were up-regulated during early EMT and in SCCs (Fig. A4; Table S1, S4[2]). Additional GO categories significantly enriched in SCCs, when compared to pure epithelial tumors, indicated increased invasiveness and ECM remodeling (Table S2[2]). In SCCs only, the epithelial adhesion molecules *Cdh1* and claudins *Cldn1, Cldn3,* and *Cldn7* were down-regulated (Table S3[2]), while mesenchymal adhesion molecules, such as *Spp1*, and a marker of invasion *S100a4*, were up-regulated (Table S4[2]).

Spindloid Cells are Less Proliferative than Epithelial Cells. Tumors with >12% EMT showed a lower mitotic rate (116 ± 44 [average ± standard deviation] mitoses per 10 high power fields) than tumors with <1% EMT (173 ± 62; t-test, p=0.03).

This difference in mitotic rate was directly attributable to a low mitotic rate in spindloid neoplastic cells (9 ± 6) compared to the epithelial cells (190 ± 47; t-test, p<0.001). Consistent with a higher mitotic rate[288], epithelial neoplastic cells had a larger nucleolar area (8.3±1.9 μm2) than spindloid neoplastic cells (3.7±1.2 μm2; t-test, p=0.03).

In total, 32 cell cycle genes were differentially expressed in early EMT tumors, thirteen of which were negative regulators of the cell cycle (Fig. A3C; Fig. A4; Tables S1, S2 and S5[2]).

Figure A4. Quantitative RT-PCR Expression Profile of Early EMT Tumors. Relative expression differences [(-) Log2] in 12 genes across three tumors with epithelial to mesenchymal transition (EMT; 15%, 60%, and 40%), compared to epithelial tumors (No EMT). [See methods for details.]



Hspa8, Cdkn1a, Cdkn2a, Ccnd1, and *Ccnd2*, which control passage through G1[289,290], were up-regulated in early EMT tumors (Fig. A4; Table S1[2]), and expression was confirmed at the protein level (CCND1, Fig. A1N). The mRNA levels of S phase cyclin A were similar in tumors with and without EMT. G2 or G2/M transition genes, such as *Gadd45a* (G2/M arrest), *Plk3* (G2 progression), *Cdkn2a, Cdkn1a*, and/or transcriptional regulators, *Usf1* and *Elk3*, were up-regulated in early EMT compared to epithelial tumors, whereas *Atm* was under expressed (Fig. A4; Tables S1, S5[2]). These data and Ingenuity Pathway Analysis (Table S2[2]) suggest that the cell cycle block may occur in G2/M.

Spindloid cells in SCCs showed a higher mitotic rate (113 ± 35, n=11) than in early EMT primary tumors (9 ± 6, n=5), but a lower rate than pure epithelial tumors (227 ± 43; t-test, $p<0.001$). Molecularly, SCCs showed enrichment of genes involved in cell activation and cell growth (Table S2[2]) and Cdkn1a, a key cell-cycle progression regulator up-regulated in early EMT, was not differentially expressed in SCCs (Table S4[2]). Thus, the mechanism underlying cell cycle arrest in early EMT was partly but not entirely abrogated in SCCs.

EMT is Associated with Increased Vascularization and Decreased Apoptosis. Large areas of epithelial tumors are physiologically avascular due to a basement membrane separating neoplastic cells from the stroma and the piling-up of tumor cells within comedo areas. Indeed, tumors with early EMT and SCCs presented substantially better vascularization than epithelial tumors, histologically and transcriptionally; genes involved in angiogenesis and vasculature development were up-regulated (Fig. A3C; Table S1, S2, S4[2]). Better vascularization in spindloid areas may provide a less hypoxic environment than in epithelial areas, thereby limiting apoptosis in spindloid regions. Histopathological analysis of H&E sections and cleaved caspase 3 expression identified more apoptosis in epithelial regions than spindloid areas of tumors with early EMT (Fig. A1O). The following gene changes supported decreased apoptosis in early EMT: *Cycs*, a caspase activator, was down-regulated, and *Cdkn1a* and *Trp63*, both negative regulators of apoptosis, were up-regulated (Table S1[2]). Additionally, regulation of apoptosis genes were enriched in early EMT tumors and SCCs, when compared to epithelial tumors, and the GO term positive regulation of apoptosis was significantly over-represented among genes up-regulated in

SCCs (Table S1, S2, S4[2]). Calcium ion-binding genes and the calcium-ion homeostasis gene *Cav1*, which regulate apoptosis and angiogenesis[291], may contribute to decreased apoptosis in early EMT and SCCs (Fig. A3C; Fig. A4; Table S1, S2; S4[2]).

TGFB and RTK pathway activation and signaling during EMT. GO, Ingenuity and GenMAPP analyses identified features that were undetectable histologically. First, genes in the TGFB pathway were enriched in SCCs and predominantly up-regulated compared with epithelial tumors (Fig. A5). Over-expression of TGFB downstream targets (*Snai1, Zeb1, Zeb2, Hmga2* and *Tgfbi*) further supported this hypothesis (Table S4[2]). Second, genes with receptor tyrosine activity were over represented in SCCs, according to analysis by both Ingenuity and GO (Table S2, S4[2]). For example, transcription of *Akt1, Angpt1, Epha2, Fgfr1, Lox,* and *Pdgfrb* was significantly up-regulated in SCCs versus epithelial tumors (Table S4[2]). Additionally, *Cdh1*, inhibited by the RTK and TGFB pathways during EMT, was down-regulated in SCCs (Table S4[2])[220l,292]. Finally, lack of significant and critical WNT/CTNNB1 genes argued against WNT/CTNNB1 signaling during EMT in Myc tumorigenesis (Fig. S1, Table S4[2]).

Tumor Metabolism changes in early EMT and SCCs. A dramatic shift in the metabolism of SCCs compared to pure epithelial tumors was identified through microarray analysis: compared to epithelial tumors, genes involved in oxidative phosphorylation, electron carrier activity, carboxylic acid metabolism, amino acid metabolism, ubiquinone biosynthesis and pyruvate metabolism were down-regulated (Tables S2, S4[2]). Some of these GO terms were significant in early EMT as well (Tables S1, S2[2]). Concurrently, there was down-regulation of mitochondrion-related GO categories in SCCs and to a lesser degree in early EMT (Table S2[2]). Taken together, these changes suggest a shift from mitochondrial respiration to aerobic glycolysis.

Spindle Cell Carcinomas Are More Invasive But Less Metastatic Than Solid/Comedo Carcinomas. SCCs were more invasive than solid/comedo carcinomas from which they were derived, as assessed by entrapment of skeletal muscle fibers (32/35 versus 2/35; χ^2 test, $p<0.001$), nerves (20/35 vs. 10/35; $p<0.05$), and adipocytes (21/35 vs. 6/35; $p<0.001$; Fig. S2[2]).

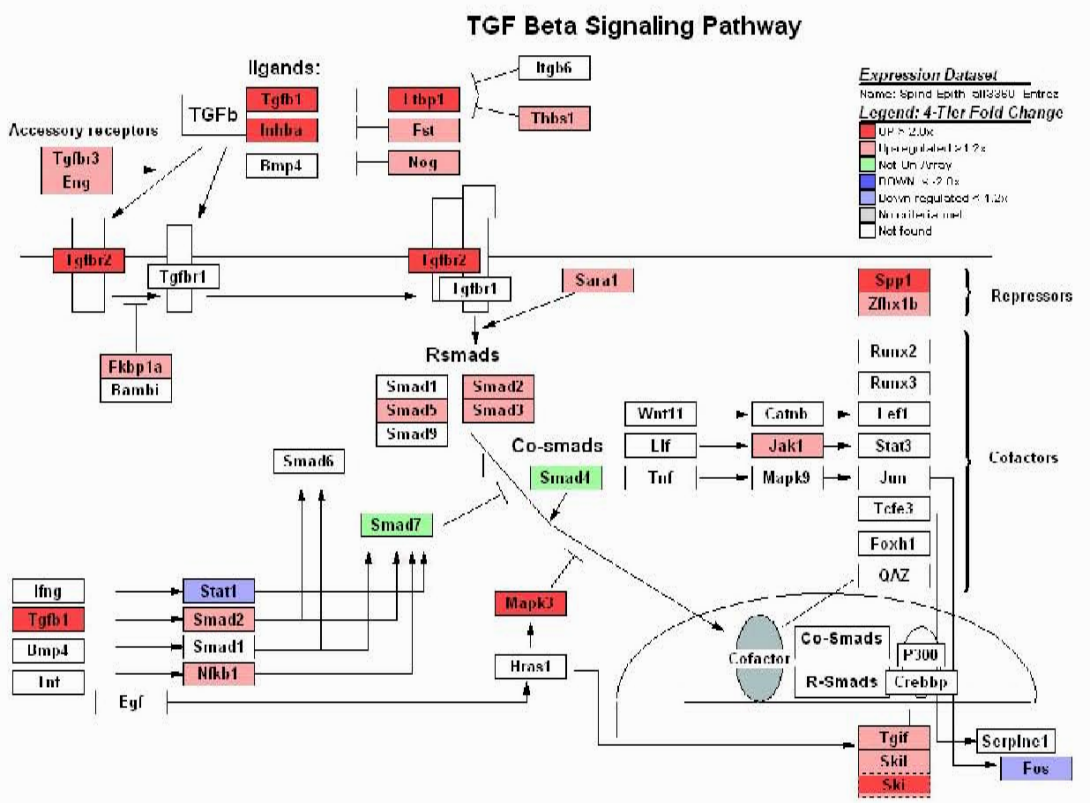


Figure A5. GeneMAPP Visualization of the Transforming Growth Factor β (TGFB) Pathway. Genes primarily up-regulated (red shading) in spindle cell carcinomas compared to epithelial tumors. (Blue shading indicates down-regulation; green shading, genes not present on Compugen microarrays; no shading, not significantly up- or down-regulated.)

However, SCCs (1/170) and glandular carcinomas (0/41) were less metastatic than solid/comedo carcinomas (32/173) as assessed histologically by the presence of pulmonary metastases (or by radiography 0/24; Supplemental Methods). Tail vein injections of single cell suspensions demonstrated that the reduced metastatic rate in SCCs compared with solid/comedo tumors was not due to the lack of metastatic capability, inability to survive in the blood, or failure to develop tumors in the lungs. In fact, intravenous injections of single cell suspensions of SCCs caused the development of tumors in the lungs more frequently (n=41/58) than that of solid/comedo carcinomas (n=7/16, χ^2 test, $p<0.05$).

Discussion

The aim of this study was to characterize EMT in a mouse model of mammary carcinogenesis by assessing early EMT in primary Myc tumors and confirming our findings in SCCs generated by serial *in vivo* passage. Two unanticipated phenomena were found in this

model: first, early EMT is associated with cell cycle arrest; and second, spindle cell carcinomas in late EMT had a lower incidence of metastases than less progressed solid/comedo (epithelial) carcinomas.

Our observation that early EMT is associated with decreased cell cycling is paradoxical because it implies that an important step in tumor progression is unfavorable to cell proliferation. This observation is mirrored by in vitro studies where TGFB-induced EMT in human and mouse epithelial cells induces growth arrest[293,294]. One major cause of cell cycle arrest is DNA damage. However, karyotyping of SCCs failed to identify consistent, reproducible chromosomal abnormalities in these tumors (data not shown). The dramatic reduction in cell cycling observed in early EMT was partially overcome in SCCs. Importantly, *Cdkn1a*, a critical anti-proliferative signal for cyclin dependent kinases, was abundantly over-expressed in early EMT but not significant in SCCs, suggesting it may be a key molecule in recovery from this transitory cell cycle block. Evaluation of other models of EMT is necessary to determine if decreased cell cycling is unique to the model presented here.

SCCs were less metastatic than solid/comedo tumors from which they were derived, in spite of increased vascularization, decreased expression of epithelial cell-cell adhesion genes, and enhanced invasiveness. Tail vein injection experiments, however, established that SCC cells could readily colonize the lungs. Therefore, the lack of metastatic spread in SCCs, growing in the gastrocnemius muscle, is likely due to the inability of neoplastic spindloid cells to access the vasculature and/or to embolize in spite of the rich vascularization of the skeletal muscle. Additionally, an appropriate tumor-cell niche is critical for progression to metastatic invasion. These observations, supported by others[224,292], suggest that EMT and metastasis are uncoupled tumorigenic processes.

Tumors with early EMT did not down-regulate epithelial adhesion molecules, with the exception of Ceacam1, whilst SCCs did (Fig. A4; Table S3[2]). Lack of down-regulation of the prototypical epithelial marker *Cdh1* in early EMT may be a limitation of the Compugen microarray platform because *Cdh1* was down-regulated in subsequent experiments using a different platform[295]. Alternatively, a decrease of *Cdh1* may have been obscured by the epithelial

component of early EMT tumors or by high within-group variation. Indeed, raw ANOVA estimates revealed that three of five primary tumors with >12% EMT showed down-regulation of *Cdh1* when compared to tumors without EMT.

Decreased apoptosis during early EMT and in SCCs was prominent histologically and immunohistochemically. As one of Weinberg and Hanahan's "Hallmarks of Cancer" diversion from apoptotic mechanisms is necessary for successful tumor progression[11]. Increased vasculature in early EMT and SCCs, identified histopathologically and transcriptionally, may limit apoptosis. Additionally, regulation of apoptosis and calcium-ion binding genes were enriched in early EMT and SCCs compared to epithelial tumors (Table S2[2]); the latter group of genes my enhance buffering to decrease calcium-ion mediated apoptosis[291]. Genes differentially expressed and GO terms enriched in early EMT (Fig. A3C; Fig. A4; Table S1[2]) and SCCs (Table S2 and S4[2]) support this hypothesis.

Transcriptional profiling identified striking metabolic differences for tumors with early EMT and SCCs compared to epithelial tumors: genes involved in the generation of precursor metabolites and energy and nuclear mitochondrial genes were down-regulated compared with epithelial tumors. These previously undescribed findings for EMT may be the effect of increased vascularization, a decrease in the number of mitochondria, or evidence of the Warburg effect, a switch in the bioenergetics of tumor cells from normal mitochondrial respiration to aerobic glycolysis/lactic acid fermentation[213]. This shift in energy production even in the presence of oxygen is anticipated to provide a selective advantage to tumor progression[12,211,213]. Drugs which exploit this phenomenon are currently in clinical trials (for review[12,296]).

The RTK, TGFB, and WNT/CTNNB1 pathways have been implicated in EMT. Data presented here supports TGFB and RTK signaling pathways, at the transcriptional level, in SCCs. Although the activation events remain unclear, post-transcriptional or post-translational modifications, such as deregulated microRNAs no longer repressing their targets or epigenetic alterations to histones, may alter master regulatory molecules, thereby initiating one or both of these signal transduction pathways. Indeed, two independent labs recently showed in vitro how the miR-200 family of microRNAs hinders TGFB-induced EMT through targeting *Zeb1* and

Zeb2[297,298]. In vitro exposure of SCC cells to drugs or siRNA, transfection with miRNAs, or *in vivo* treatment with small molecule inhibitors targeting these pathways may uncouple the effects of TGFB from those of RTK signaling to unravel the initiating events of signal transduction in EMT.

In summary, the findings presented here further our understanding of EMT during mammary carcinogenesis in three important ways. First, EMT is a dynamic multi-step process: early EMT is characterized by expression of myoepithelial markers, increased cell motility, invasiveness, and vasculature, decreased apoptosis, and cell cycle arrest. Progression from early EMT to SCCs is accompanied by the loss of myoepithelial markers, partial release from cell cycle block, significant over-expression of cell growth and proliferation genes, enhanced invasion, and cooperation of TGFB and RTK signal transduction. Second, classical markers of EMT may not represent late EMT tumor progression: while over-expressed and highly significant in early EMT, *Twist1, Twist2,* and *Snai2* were not significant in SCCs. Third, in this model, metastasis was hindered by EMT; tumor cell microenvironment and access to the blood stream substantially affected the ability of neoplastic cells to metastasize. Evaluation of other models of EMT is needed to determine whether these characteristics are universal or are unique to the multi-step model of EMT tumorigenesis described herein.

BIOGRAPHY OF THE AUTHOR

Karen Suzanne (Currier) Fancher was born in Great Falls, Montana on December 11, 1972 and then raised in Plymouth, Maine. She graduated from John Bapst Memorial High School, Bangor, Maine in 1991 and from Hartwick College, Oneonta, NY with a Bachelor of Science degree in Biochemistry in 1995. After working in quality control for a small pharmaceutical company in New York, Karen returned to Maine as a research assistant for The Jackson Laboratory in 1998. Karen entered the Interdisciplinary Ph.D. program in Functional Genomics between the University of Maine and The Jackson Laboratory in the fall of 2002. Since July 2007 Karen has worked as a Technical Information Scientist at The Jackson Laboratory. She is a candidate for the Doctor of Philosophy degree Interdisciplinary in Functional Genomics and in Biomedical Sciences from the University of Maine in December 2008.