GENETICALLY ENGINEERED MOUSE MODELS PREDICT ACTIONABLE MUTATIONS IN HUMAN CANCERS

By

Matthew Richard Swiatnicki

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Microbiology and Molecular Genetics – Doctor of Philosophy

2021

ABSTRACT

GENETICALLY ENGINEERED MOUSE MODELS PREDICT ACTIONABLE MUTATIONS IN HUMAN CANCERS

By

Matthew Richard Swiatnicki

In the United States alone, cancer claims the lives of over 600,000 people a year. While progress has been made in understanding this complex set of diseases, more work is needed if we are to end our struggle with cancer. Bioinformatics analysis and genetically engineered mice are important tools for understanding the biological complexities of cancer. When combined, these approaches can be an important avenue to uncover disrupted cellular pathways contributing to cancer formation. While genetically engineered mouse models are important for the study of cancer, genome sequence analysis of many of these models is lacking. Within this work, we sequenced whole genomes of two genetically engineered mouse models of cancer, MMTV-Neu and MMTV-PyMT. Through this sequence data, we have found numerous disruptions to pathways contributing to the metastatic cascade. These include tumor signatures associated with defective mismatch repair, as well as numerous genomic mutations within cell adhesion genes.

More importantly, we have uncovered a conserved V483M missense mutation within the protein tyrosine phosphatase receptor type H (*Ptprh*) gene. Within mice, tumors harboring a *Ptprh* mutation correlate with increased phosphorylation of the epidermal growth factor receptor (EGFR). EGFR is a known oncogene that is mutated in numerous cancers, including non-small cell lung cancer (NSCLC). Lung cancer is the number one cancer cause of death in the United States. Often, prognosis for lung cancer is poor, often due to late diagnosis. NSCLC patients with mutations in *EGFR* typically have a more favorable prognosis, due to treatment with tyrosine kinase inhibitors. More research is needed to improve survival rates of lung cancer patients who do not present with mutations in *EGFR*.

Within NSCLC, 5% of patients have mutations in *PTPRH*, and many of these mutations correlated with increased EGFR activity as well as PI3K/AKT activity. If *PTPRH* mutant patients have increased activation of EGFR and would benefit from TKI therapy, this presents a unique opportunity to treat a large subset of cancer patients with an FDA approved therapy. CRISPR KO of PTPRH within the H23 lung cancer cell line resulted in increased phosphorylation of EGFR and downstream AKT. Furthermore, *PTPRH* mutant NSCLC cell lines H1155 and H2228 respond to the tyrosine kinase inhibitor osimertinib. *In vivo* osimertinib treatment of nude mice injected with H2228 cells also shows partial response, suggesting *PTPRH* mutant patients may benefit from EGFR therapy.

This work is dedicated to my family, especially to my grandmother Rita Swiatnicki
who passed from breast cancer in 2003.
May we one day find a cure to ease the suffering for all those
who toil with the affliction of cancer.

ACKNOWLEDGEMENTS

There are a number of people without whom; I would not have achieved my degree. First I would like to thank my family and friends, especially my soon to be wife Jenna, and my parents. Their support over these last few years has made this possible.

I would like to thank Dr. Eran Andrechek for being a great mentor, and the rest of the Andrechek lab for all their support. I would especially like to thank Dr. Jon Rennhack and Dr. Sean Misek for their help and advice involving my persistent questions with research and experimental design. My thesis committee, including Dr. Kathy Meek, Dr. Susan Conrad, Dr. Hua Xiao, and Dr. Kefei Yu also deserve a large thank you for their support over the years. Especially Kathy, who graciously allowed me to hunt on her property and keep my sanity. Finally, I would like to thank the Microbiology and Molecular Genetics Department for all of their support and numerous funding source overs the years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

APC	Adenomatous polyposis coli
ATRS	A/T-rich sequences
BCR	Breakpoint cluster region protein
C-ABL	Abelson tyrosine kinase
ChIP-seq	Chromatin immunoprecipitation sequencing
CNV	Copy number variant
COSMIC	Catalog of somatic mutations in cancer
CRISPR	Clustered regularly interspaced short palindromic repeats
DCIS	ductal carcinoma in situ
DMBA	7-12,Dimethylbenz[a]anthracene
DSB	Double stranded break repair
EGFR	Epidermal growth factor receptor
ER	Estrogen receptor
FGFR1	Fibroblast growth factor receptor 1
GEF	Guanine nucleotide exchange factor
GEMM	Genetically engineered mouse model
GEO	Gene Expression Omnibus
GFP	Green fluorescent protein
GO	Gene ontology
(SS)GSEA	(single sample) Gene set enrichment analysis
HER2	Human epidermal growth factor receptor
ICGC	International Cancer Genome Consortium
IGF1R	Insulin like growth factor 1 receptor

ILC	Invasive lobular carcinoma
In/del	insertion/deletion
KRAS	Kirsten rat sarcoma
KEGG	Kyoto encyclopedia of genes and genomes
MAPK	Mitogen activated protein kinase
MCA	3-methy1cholanthrene
MIND	Mammary intraductal
MITE-seq	Mutagenesis by integrated tiles sequencing
MMR	Mismatch repair
MMTV	Mouse mammary tumor virus promoter
MNU	N-methyl-N-nitrosourea
NCI	National Cancer Institute
Neu	(Erb-B2) Receptor tyrosine kinase 2
NSCLC	Non-small cell lung cancer
PDX	Patient derived xenograft
PTB	Phosphotyrosine binding
PTP	Protein tyrosine phosphatase
PTPRH	Protein tyrosine phosphatase receptor type H
PyMT	Polyoma middle T antigen
RB	Retinoblastoma
RCAS-TVA	Replication-competent avian sarcoma-leukosis virus – tumor virus A receptor
rPTP	Receptor protein tyrosine phosphatase
RTK	Receptor tyrosine kinase
SB	Sleeping beauty

SC	Small cell lung cancer
SCID	Severe combined immunodeficiency
ScRNA-seq	Single cell RNA sequencing
SNV	Single nucleotide variant
SAP-1	Stomach cancer-associated phosphatase 1 (PTPRH)
TALEN	Transcription activator-like effector nucleases
TCGA	The Cancer Genome Atlas
TKI	Tyrosine kinase inhibitor
VAF	Variant allele frequency
WAP	Whey acidic protein
WES	Whole exome sequencing
WGS	Whole genome sequencing

INTRODUCTION

CANCER AS A GENOMIC DISEASE

Current scientific paradigm surrounding the onset of cancer involves gene mutations leading to dysregulation of cellular pathways controlling proliferation, apoptosis, and cellular maintenance. Often, mutations in a few oncogenes or tumor suppressor genes lead to oncogenic transformation of a cell [1–4]. This is exemplified through the current model of colorectal cancer, which often relies on mutations in the tumor suppressor *APC* (Adenomatous polyposis coli), followed by mutations in the proto-oncogene *KRAS* (Kirsten Rat Sarcoma) to develop a malignancy [5–9]. With recent cost reductions in sequencing technologies, whole genome or whole exome sequencing has been completed on hundreds of thousands of human tumors. This has revealed differing mutation burdens across various forms of cancer, with certain cancers such as glioblastomas harboring few mutations, and others such as colorectal cancers harboring a large number of mutations [4]. Analyzing whole genome sequence data has also revealed the importance of non-exonic mutations within cancer formation. Genetic mutations in regions important for gene regulation, such as gene promoters, can impact tumor formation and growth. Analyzing the impact of non-exonic mutations is a quickly growing area within the cancer field.

Mutations to the genetic code can be broadly classified into two categories. These include small structural changes such as single base pair mutations (SNVs) and small insertions and deletions (in/dels), as well as larger structural changes such as amplification or deletion events (CNVs) and translocations. While the vast majority of single base pair mutations are synonymous, resulting in no changes to protein structure, nonsynonymous and nonsense mutations can lead to amino acid shifts or truncated proteins that alter protein function. Examples of SNVs contributing to cancer formation are L858R *EGFR* mutations in lung cancer, and various amino acid shifting *KRAS* mutations that occur in numerous cancers [10, 11]. Large amplification or deletion events within the genome can result in cancer through disruption of a single gene or multiple genes. This is evidenced within human epidermal growth factor receptor type 2 (HER2) breast cancer patients, where amplification of the *HER2* oncogene contributes to oncogenic

transformation [12]. Translocations are also capable of inducing cancer through a number of mechanisms, including fusing active gene promoters with known oncogenes, or simple truncation of a gene. An example of translocations contributing to cancer formation lies within chronic myelogenous leukemia patients, where a translocation between chromosomes 9 and 22 fuses the *C-ABL* (Abelson tyrosine kinase) oncogene with an active BCR (Breakpoint cluster region protein) promoter [13–15]. Mutations and structural changes affecting gene promoters, splice sites, and other regions important for gene regulation are now becoming more appreciated for their ability to cause cancer [16–18].

With the realization that cancer is largely a disease of underlying genetic mutations, much debate has swirled around the contribution of factors underlying these mutations. There are currently thought to be three mechanisms for the onset of genetic insults, including heritable germline mutations, random mutations originating from DNA replication errors, and errors introduced by environmental mutagens. Heritable germline mutations are perhaps the easiest to trace, but account for the least amount of cancer incidence. Heritable *BRCA1/2* mutant breast cancers account for approximately 25% of all breast cancer cases, while heritable mutations in *RB1* account for 40% of retinoblastoma cases [19, 20]. When analyzing overall cancer rates however, The National Cancer Institute (NCI) estimates germline mutations account for approximately five to ten percent of all cancers.

Approximately 90% of cancers occur due to environmental factors and random chance, however there has been much debate over the contribution of these two factors to cancer incidence. Samuel Epstein's 'The Politics of Cancer' attributes the majority of cancers to increasing environmental pollution via carcinogens [21]. This notion is supported through a slew of epidemiological evidence, such as migrant cancer rates shifting towards rates of their adoptive countries [22], and higher cancer incidence seen in areas located in close proximity to heavy industrial presence [23–25]. However, these statements are more complicated when looking below the surface level. For instance, other studies have shown migrants adopt rates of cancer similar to their adoptive country for only certain cancers, while other cancers

maintain rates to that of their country of origin [26, 27]. The most well known environmental factor contributing to increased cancer rates is smoking. By the 1960s, there were numerous epidemiological and animal studies showing a link between smoking and lung cancer [28–30]. Since then, the data has evolved to show an overwhelming amount of evidence linking smoking to cancer. This includes genomic studies showing differing mutation profiles of lung cancer patients who smoked, versus those who haven't [31, 32].

In a shift from Epstein's line of thinking, some recent evidence suggests a majority of cancers occur by chance, due to random mutations within the genome [33]. This evidence was based on correlations between the number of stem cell divisions occurring within particular tissues, and the cancer incidence within those tissues. This paper has come under fire for a number of reasons, namely, another group was able to show the correlation held in a hypothetical scenario where cancer incidence was high due to environmental effects [34]. A series of letters to *Science* has also pointed out that the original Vogelstein study didn't include breast and prostate cancer in their analysis, two cancers thought to be highly impacted by environmental factors [35–37]. These letters also point out a potential flaw in the Vogelstein statistical analysis, showing that the confidence limits would actually be +/- 30, meaning the rates of incidence could be 30 times less or greater than their predicted value. More recent data from Vogelstein and other groups have reiterated the importance of random DNA replication errors in the formation of cancer [38, 39]. They prudently pointed out however, that these studies don't diminish the impact environmental mutagens have on the formation of cancer.

Questions surrounding the impact of certain gene mutations are important considerations for cancer biologists. While tumors often carry a high mutational burden, it has been traditionally thought that only a few mutations, dubbed 'driver mutations', contribute to tumor progression. A vast majority of the remaining mutations are dubbed 'passenger mutations', and thought to have little impact on tumor progression [4]. More recent data however has suggested that whether collectively or individually,

passenger mutations may have more of an impact on tumor progression than previously thought [40–42]. With the development of MITE-seq (Mutagenesis by integrated tiles), the effect of every possible amino acid substitution within an individual gene can be determined [43, 44]. While this technology shows promise for investigating passenger mutations within individual genes, completing this assay for each gene with the exome remains a tall order. Many important questions surrounding passenger mutations remain. Do these passenger mutations arise within pre-neoplastic tissue, or after tumor formation? Furthermore, how do these mutations contribute to the metastatic process of cancer? Future studies involving single cell, and MITE sequencing may be able to resolve some of these questions.

EFFICACY OF MOUSE MODELS

In cancer research, the use of mouse models is often two-fold. The first includes studying cancer associated oncogenes, pathways, and histology, in other words, studying cancer itself. The second includes utilizing mouse models to study the safety and efficacy of drug treatments. In the two subsections below, I will touch on both of these uses.

I. MICE AS A CANCER MODEL

This subsection of the introduction has previously been published as a review in the Journal of Mammary Gland Biology and Neoplasia titled "How to Choose a Mouse Model of Breast Cancer, a Genomic Perspective". Portions of the review not applicable to this thesis were excluded. While the review focuses specifically on mouse models of breast cancer, the principals can be applied to mouse models of various other cancers.

CARCINOGEN BASED MODELS

A common method for modeling breast cancer is through mouse model systems. Currently there are numerous systems, each with advantages and disadvantages, used to generate different models. Modeling cancer in animals began with the application of coal tar on rabbits and mice, leading to the formation of tumors [45]. Since that point, a wide array of carcinogens employed in mice have been used

to study cancer, including N-methyl-N-nitrosourea (MNU), 3-methy1cholanthrene (MCA), and perhaps the most widely used 7-12,Dimethylbenz[a]anthracene (DMBA) [46, 47]. Tumors in mice treated with carcinogens often express a variety of genomic alterations including mutations in *PTEN*, increased expression of *CCND1* and *MYC*, and the activation of important cellular pathways including NF-κB, Wnt, and PI3K/AKT [48, 49]. Histologically, these tumors vary greatly between models, with MPA treated mice often exhibiting type-B adenocarcinomas, and DMBA treated mice often having tumors of the adenomyoepethelial and myoepithelial histologies [46, 50].

TRANSPLANT MOUSE MODELS

To further study facets of human cancers in a more biologically relevant setting, transplantable mouse models have been developed. These include the mammary intraductal (MIND) model in addition to the previously mentioned cell lined xenografts and patient derived xenograft models. In order to study the progression of human cancers from ductal carcinoma in situ (DCIS), the MIND model mimics human DCIS through the injection of human DCIS cells into the ducts of severe combined immunodeficiency (SCID)-beige mice [51]. Indeed, this method allows for the subtypes of DCIS to be maintained in a mouse model [51, 52]. However, despite their clear strengths, these models are not readily amenable to modification or manipulation to allow quick and easily genetic testing of hypotheses.

GENETICALLY ENGINEERED MOUSE MODELS

The complexity of human cancer may best be modeled through the various forms of genetically engineered mice, including transposon based, transgenic, knock-in, knock-out, and inducible mouse systems. One of their largest advantages these models possess is the acquisition of impactful mutations [53, 54], analogous to the development and progression of human breast cancer.

One method of generating mice with cancer in the mammary glands is through the use of transposable elements [55–57]. These systems are used for germline transmission, as well as generating somatic mutations for the study of cancer [58]. Use of these systems allowed mice to be characterized

with mutations in key genes. As mentioned above, patients with invasive lobular carcinoma (ILC) tend to have loss of E-Cadherin. Using the Sleeping Beauty (SB) transposable system, Kas *et al.* showed the importance of particular genes, including *Myh9*, and *Ppp1r12b*, contributing to tumor formation in mice with ablated E-Cadherin [59].

To study potential oncogenes, transgenic mice are developed to determine whether overexpression of that particular gene results in tumor formation. In these mice, tissue specific promoters direct oncogene expression to a particular organ or tissue. Promoters for the study of breast cancer in mice include the commonly used mouse mammary tumor virus (MMTV) and whey acidic protein (WAP), as well as others including keratins [60–62]. Overexpression of a number of important oncogenes with these promoters has illustrated the importance of key genes, including *C-MYC*, RAS, and *ERBB2* [60, 63, 64]. In addition to the simple overexpression systems, work from the Chodosh lab introduced numerous inducible systems where expression of key oncogenes could be turned on or off in the mammary gland through introduction of doxycycline to the water [53, 65–67]. These systems revealed that while tumors were initially dependent upon the initiating oncogene, they accumulated enough mutations that when expression of the primary driving gene was withdrawn, tumors that initially regressed eventually relapsed. Other studies have used a combination of the inducible and standard transgenic systems to demonstrate oncogene dominance, where only one oncogene in a two oncogene system is needed to maintain tumor viability [68, 69].

In addition to transgenic models with overexpression of various oncogenes, knock-in models have been generated to express oncogenes in their native genomic location. This has allowed for expression of oncogenes under the control of the Rosa26 promoter, resulting in lower levels of transgene expression [70]. Other groups have placed a lox-stop-lox cassette between the endogenous promoter and an oncogene. The advantage of this system is that normal temporal and spatial control of gene expression occurs [71], but depending on timing of the excision event, mice can adapt to oncogene expression [72].

Importantly, with the lox-stop-lox system, erbB2 knock-in mice developed amplification and overexpression of the oncogene, analogous to HER2+ve breast cancer [71]. Numerous other knock-in models have been created to study breast cancer genes, including R273H, R248W, and R175H *Tp53* mutant mice, as well as H1047R *Pik3ca* mutant mice [73, 74].

Alongside overexpression of oncogenes, knock-out mice permit the study of tumor suppressor genes *in vivo*. *TP53*, the most mutated gene in breast cancer, as well as *BRCA1*, which has germline mutations in 5-10 percent of human breast cancer, have been studied extensively through the use of knockout models [75]. The combination of knockout models with transgenic models, where expression of Cre is linked to the transgene, have also allowed the study of specific facets of tumor development while lacking signaling pathways [76, 77].

In addition to standard transgenic and knock-in / knockout systems, engineered nuclease systems, including TALEN (Transcription activator-like effector nucleases) and CRISPR (clustered regularly interspaced short palindromic repeats), are used to generate mouse models. These systems allow for the deletion, addition, and replacement of desired DNA sequences into numerous models, including mice. While TALEN systems are capable of editing genes anywhere in the genome, as opposed to CRISPR needing nearby PAM motifs, CRISPR has become a more widely used tool due to its simplicity and cost effectiveness. Studies utilizing the power of TALEN and CRISPR systems have investigated numerous genes important to breast cancer, including *BRCA1* and *CDH1* [78, 79]. These systems can be employed through manipulation of mouse embryonic cells, or through direct injection of the system components into wildtype mice, and mice containing the CAS9 protein under control of the cre-lox system [80, 81]. Gene specificity is achieved in these systems through the use of guide RNAs. A further review of these systems can be found here [82]. With the recent advent of CRISPR systems easing the transgenic process, it will also be interesting to see whether there is a resurgence in the use of estrogen receptor (ER)+ rat models. Another tool potentially capable of faithfully recapitulating human breast cancer progression is

the replication-competent avian sarcoma-leukosis virus – tumor virus A receptor (RCAS-TVA) system reviewed here [83]. This system can be used for the delivery of oncoproteins and dominant negative tumor suppressors in a timely matter, but is often limited to small insertions into the virus.

With the heterogeneity of human breast cancer and the large number of mouse models available to study the disease, the central question becomes, which model is the best fit for a particular study? This is obviously dependent on the experimental question, but the characterization of the models and their relation to human breast cancer should be considered. This is true on a phenotypic, genomic, and gene expression level.

MOUSE PHENOTYPES

On a phenotypic level, there is a large amount of variation between the various mouse models of breast cancer. In terms of latency, models range from the rapid MMTV-PyMT in the FVB background, to the prolonged GR/J, with tumors appearing at 45 days, and 12 months respectively. Other notable models with strikingly different latency periods include MMTV-NeuNT (Erbb2) transgenics relative to the conditional expression of NeuNT under the control of the endogenous promoter, where tumors appear at 89 days and 15 months respectively [84, 85]. Variation is also observed in the tumor growth rate in various strains. While MMTV-Neu mouse tumors grow to 2500mm^3 from first palpitation in approximately 45 days [86], other models such as MMTV-Myc mice with *Stat3* ablated, can take as long as 109 days to grow to 2500mm^3 from the first palpitation [87]. Fluctuations in tumor latency and growth rate are also context dependent, relying on differentially activated signaling pathways. This is exemplified with ablation of the E2F1 transcription factor in two different mouse models. Loss of E2F1 in the MMTV-Neu mouse model leads to increases in both tumor latency and growth rate, whereas in the MMTV-PyMT model, a decrease in latency and no alteration to growth rate was observed [86, 88]. These differences illustrate the importance of selecting particular models for a study.

Previous research has also shown histological differences between the primary tumors of various mouse models. Genetically engineered mouse models (GEMM) exploring mice harboring specific genome alterations introduced through a number of genome editing techniques, have been important tools for cancer researchers. A review of GEMMs by a panel of experts in 2000 found the majority of genetically engineered mouse tumors to have a set of histological forms unique from non-GEMM tumors such as carcinogen induced models [89]. Some GEM tumors, such as those from models expressing the *neu* and *src* transgenes, have also been found to have histologies similar to those of tumors from human patients [90]. Much like human breast cancer, a large amount of histological variation is seen within certain GEMMs. MMTV-Myc mice have been shown to harbor multiple tumor histologies including papillary, microacinar, and squamous tumors [54]. Similar pathologies were noted in the MMTV-Met mice [91]. In MMTV-PyMT mice, while approximately 40 percent of tumors have a microacinar histology, tumors also display a wide array of histological patterns including adenosquamous, glandular, and those of mixed histology [88]. More recently, certain GEMM tumor histological subtypes have been shown to correlate with particular transcriptional profiles within the model, much like the human disease. In fact, gene expression signatures have been generated that are capable of predicting histological patterns in mouse tumors [92].

The study of metastasis is also heavily reliant on mouse models. While the expression of some oncoproteins such as PyMT and Neu result in a heavy metastatic burden in mice, other transgenic models with potent oncogenes such as WAP-Ras and MMTV-Myc have lower metastatic rates, or fail to metastasize at all [61, 64, 84]. Strain background is also an important consideration in the ability of the primary tumor to metastasize, with expression of PyMT in FVB mice resulting in nearly all tumor bearing mice developing metastasis to the lung. However, the same transgenic line interbred to RF/J, C58/J, and other mouse backgrounds dramatically reduced the metastatic burden [93]. Of GEMMs that metastasize, most result in metastases to the lungs. However, select models have the ability to metastasize to different

organs. MT-Met mice have demonstrated metastasis to the heart and kidney as well as the lung, and tumors from $p53^{fp/fp}$ MMTV-Cre mice are able to metastasize to the liver [94, 95].

GENE EXPRESSION DATA

The advent of microarray and sequencing technologies has made it possible to complete large scale gene analysis on large numbers of samples. In breast cancer, conserved gene expression patterns led to the definition of the intrinsic subtypes of breast cancer [96]. Since the initial work on human breast tumor expression data, numerous studies have applied microarrays to study GEMM mammary tumors. This has been done for individual models [54, 91, 97–103], as well as in a broader survey approach across models.

When examining individual models using array analysis, a surprising amount of molecular heterogeneity has been a recurring finding. Not surprisingly, this heterogeneity was present in tumors with long latency (MMTV-Myc), and correlated with histological subtypes. Predicting that tumors with a short latency would be less heterogeneous would appear to be a logical hypothesis, however, it is notable that tumors with extremely short latency, driven by PyMT, also have a surprising level of heterogeneity from tumor to tumor. Together these studies suggest that both models are dependent upon accumulation of other events for tumor formation and progression. Not all models have extensive heterogeneity, and models such as Wap-Myc, C3(1)Tag, and MMTV-Neu, have less heterogeneity based on gene expression profiles. Comparison of these individual models to human breast cancer has revealed that C3(1)-Tag and Wap-Myc models have expression patterns similar to basal-like human tumors (a highly aggressive molecular subtype of breast tumors), including high expression *CRYAB*, a known human basal-like tumor marker [104]. Expression signatures from other tumor types, such as luminal, do not correlate as well between mouse models and human tumors, although they still share some similar features, like positive staining for the K8/18 marker [104]. While the MMTV-Neu model fails to actually reflect human Her*2*+ breast cancer on a gene expression level, this may simply be due to the altered expression of other

genes within the large HER2 amplicon. A mouse model with amplification of the endogenous erbB2 locus [71] should thus be assayed for similarities to human HER2+ve breast cancer.

In addition to papers that have profiled individual models, there have been several publications that compared various models. Herschkowitz *et al* examined 13 different models of breast cancer, identifying models with similarities to luminal tumors, despite being ER-negative, and having heterogeneous expression patterns. They also identified other GEMMs resembling more basal like tumors. [104]. Hollern *et al* increased the number of samples analyzed (1156) as well as profiling numerous additional models to examine 26 major models with several additional variants (wild type *Myc*, T58A *Myc* etc.). This unsupervised approach demonstrated substantial heterogeneity in the majority of mouse models. Using both a gene expression and a signaling pathway approach, they also noted several similarities between the intrinsic subtypes of human breast cancer, and subsets of various mouse models. Importantly, it was noted that only a portion of tumors from an individual model reflected each of the intrinsic subtypes [105]. Further, Pfefferle et al. examined 356 samples from 27 models to identify 17 distinct mouse mammary tumor intrinsic subtypes, eight of which reflected subtypes in human breast cancer. However, this analysis used an intrinsic approach, a supervised method of clustering that may add bias to the study. Each of these three publications provides an important examination of the diversity of mouse models of breast cancer and are an essential starting point when choosing a mouse model for analysis.

GENOMIC COPY NUMBER ALTERATIONS

In tumor cells, regions of the genome are often deleted or repeated dozens of times, potentially serving to drive tumor formation or modify tumor progression. A prime example of copy number variation (CNV) in cancer is the amplification of human epidermal growth factor receptor type 2 (*HER2*), resulting in uncontrolled activation of downstream signaling cascades, including the mitogen activated protein kinase (MAPK) pathway [12, 106]. While extensive CNV data from mouse tumor models has not been

generated, use of an algorithm that predicts CNV from gene expression data has been generated and validated [107]. Applied to mouse models of breast cancer, the prediction of CNV noted variation across numerous mouse models of breast cancer. However, genes from some CNV regions, such as *Gsn*, are conserved among some models [107]. This same trend was seen within distinct mouse models, whereas some CNV events showed little conservation between mice in a given model, and other events were present in greater than 50 percent of mice in a given model [107]. More interestingly, integrated clustering of CNV events from mouse and human tumors showed conservation of some CNV events between the two species [107], demonstrating that mouse models can be an accurate depiction of human breast tumors in terms of copy number alterations.

PATHWAY ANALYSIS

Research has shown that complex networks of proteins work together in regulatory pathways that control cellular function. These signaling pathways, including the MAPK/ERK and PI3K/AKT pathways, are often dysregulated in cancer [108, 109]. Expression data from the various genes that constitute these pathways and their downstream targets can predict activation or inactivation of particular pathways, making these pathway signatures an important tool for the study of breast cancer. To uncover pathway use, gene expression analysis has been coupled with bioinformatic tools like Gene Set Enrichment Analysis (GSEA), which has been widely applied to many models. Likewise, a Bayesian Regression Pathway signature system [110] has been applied to mouse models of breast cancer to predict cell signaling pathway activity [86–88]. Like differential gene expression data, pathway signatures often vary within GEMMs, the most prominent example of this perhaps being the *Myc* model [105]. In mice, pathway signatures have shown a correlation with histological subtypes, most notable being the microacinar histology associated with amplification events on chromosomes 11 and 15 [107]. Pathway signatures from mouse mammary tumors have also been found to correlate to human breast tumors. A set of highly expressed pathways found in tumors from Myc mice were also found to be highly expressed in Basal-like

human tumors [111]. This trend has been seen in a number of pathway signature sets between mouse and human tumors.

SEQUENCING

Sequencing of human breast cancer samples has led to both the discovery of novel mutations important to breast cancer, such as *FOXP1* [112], as well as further characterization of genes already known to be important to cancer development including *HER2* and *PI3K* [113, 114]. In mouse models, sequencing studies in lung cancer have shown the mutational burden from GEMM tumors to be lower than that of human lung tumors. Tumors from *Kras*, and *Egfr* driven mice carry a mutational burden of ~.05 non-synonymous mutations per mega base, while human tumors harbor a mutational burden of ~4.1 non-synonymous mutations per mega base [115, 116]. While numerous publications have examined gene expression in mouse models of breast cancer, very few models have been examined at the sequence level. Recently, whole genome sequencing (WGS) from mouse mammary tumors (MMTV-Neu and MMTV-PyMT) has also led to the discovery of alterations in genes potentially important to human breast cancer, including *Col1a1* and *Phb* [117]. The potential impacts of these mutations on tumor behavior in such well characterized tumor models underscores the need to complete WGS on mouse models of breast cancer [118].

Researchers are now beginning to appreciate the cellular and genetic heterogeneity of tumors not only between patients, but within single tumors [119]. Intra-tumoral and metastatic site heterogeneity present issues for tumor treatment, as targeted therapies may be effective for only part of the tumor. Single cell RNA sequencing (scRNA-seq) is beginning to confront these challenges through the understanding of the differences present within a primary tumor, and across the metastatic sites. Investigation of copy number alterations in single cell sequencing of two triple negative human breast tumors found four distinct populations of cells, with some shared CNV regions between the cell populations [120]. In mice, scRNA-seq has begun to show the distinct gene expression profiles of

mammary epithelial cells at different developmental stages. In the mammary gland, a shift in gene expression from a basal-like transcriptional profile to a more luminal profile occurs around 5 weeks of age [121]. While more studies are needed using scRNA-seq, key insights into the single cell heterogeneity of cancer should continue to be uncovered as this technology continues to develop.

OTHER CONSIDERATIONS - METABOLOMICS AND PROTEOMICS

While cancer metabolomics is not a new area of study within the field, recent years have seen a surge in metabolic profiling of both human and mouse tumors. A 2018 study from Dai et al. focuses on the metabolic profiles for a number of mouse models, including PyMT, Wnt1, and Neu [122]. This study not only found metabolomic differences between tumor and normal breast tissue for each model, it also found that each oncogene had a unique metabolomic profile. Furthermore, the C3-TAg model was found to have metabolites of prognostic value, illustrating the importance of these studies.

Advances in mass spectrometry have also led to a rise in large scale proteomics analysis. These analyses in breast cancer mouse models have allowed both comparisons to the human disease, as well as enhanced the search for biomarkers capable of early cancer detection. Indeed, proteins found upregulated in the plasma of tumor bearing PyMT mice have been found to coincide with multiple human breast cancer cell lines, including MCF7 and BT474 [123]. In some cases, such as with the conditionally activated Neu mouse model, entire proteomic profiles have been made publically accessible in hopes of enhancing the search for novel cancer biomarkers [124].

CHOOSING A MODEL

Choosing the correct mouse model to investigate human breast cancer is an important experimental decision. As reviewed above, there are numerous categories stratifying the various models. Rather than simply using a model based on availability, investigators should carefully consider the choice of model. First, if the research question is one related to a particular signaling pathway, then this may dictate the choice of model. Numerous models have been profiled in comparison to each other in several

reports [105, 111], and both GSEA and Bayesian pathway predictions have been reported for these models [105]. These data may be downloaded and signaling pathways searched to determine models with high or low activity for a pathway of interest. However, given the gene expression heterogeneity seen in various models [111, 125, 126], the number of tumors with the signaling pathway alterations in question should be considered when calculating the number of experimental subjects required.

If the primary consideration is a phenotype, such as metastatic progression, then the model choice will be constrained by that characteristic. While a majority of studies use the MMTV-PyMT strain for metastatic research, other strains that metastasize are available. The short tumor latency and extensive metastasis are attractive characteristics for the PyMT transgenic mice, but if the gene expression profile and signaling pathways that are of interest do not match, then other strains are available with metastatic properties. Other characteristics, from tumor latency to promoter system can be considered when choosing a mouse model.

For investigators simply looking to ask which mouse model most closely resembles a subtype of human breast cancer, unfortunately there is not an easy answer or single best choice. Examining co-clustering of human and mouse model tumors by gene expression [92] or predicted CNV [107] has revealed that many different models cluster with each of the subtypes of human breast cancer. MMTV-Myc is particularly instructive with varied histological subtypes and gene expression subtypes that individually cluster with most of the major subtypes of human breast cancer [92]. While this confounds the choice of model system, it underscores how sample to sample heterogeneity of gene expression in human breast cancer is reflected in the majority of mouse model systems.

Ultimately, the choice of mouse model system is a multifactorial one. This choice must take into account the initiating oncogene, latency, progression characteristics, gene expression similarities to human cancer, cell signaling pathway use, and whether copy number variation is relevant. Moreover,

once a model is chosen, the resulting tumors must be characterized to determine how the tumor to tumor heterogeneity that is present in the various models has been altered with the experimental manipulations.

DISCUSSION

Numerous genomic perturbations, and a cascade of protein interactions and regulatory pathways all function together to initiate and maintain oncogenic transformation. Given this complexity, the mouse model is highly suited to study breast cancer. The *in vivo* nature of mouse models allows the complexity of cancer to be studied more accurately than cell culture and other *in vitro* experiments alone. Numerous types of mouse models, including carcinogen induced, patient derived xenografts (PDXs), and GEMMs recapitulate certain aspects of the disease. While their usefulness is dependent on the research question, GEMMs are perhaps the most comprehensive due to their ability to closely mimic the initiating oncogenic event that occurs in a number of cancers while maintaining an appropriate tumor microenvironment and functioning immune system.

On an expression and histological level, GEMM tumors are as complex as the human tumors they attempt to mimic. Just as a wide array of histologies are seen within human tumors, tumor histological differences can be seen within single GEMMs. Classifying histological subtypes on their expression profile also shows relevancy to human breast cancer. Since the initial characterization of human breast cancer into intrinsic subtypes, an increasing amount of data has been generated showing mouse subtypes that mimic each. While little whole genome sequencing data has been generated for GEMM tumors, the data available has shown that like human tumors, mouse tumors display a large array of genomic rearrangements, including single nucleotide variants, copy number alterations, and translocations. The histological, expression, and sequencing similarities between human and mouse breast tumors show that when used correctly, genetically engineered mouse models can be an accurate method for studying human breast cancer.

Given the complexity of both human breast cancer and the numerous mouse models used to study it, choosing the correct mouse model is essential for the experimental question. Initial examination of expression based analysis and the human based subtypes that are mimicked through large scale gene expression experiments is critical [96, 105]. Depending on copy number alterations in the gene, it is also beneficial to examine the mouse models for similar changes [107]. Whether through GSEA or a signature based approach, signaling pathways should also be examined [105, 111] to ensure that the appropriate model is used. Recent examples of drug screening in mouse models have taken these parameters into account [127, 128] in important demonstrations of the integration of bioinformatics analysis of mouse models with wet lab experiments.

II. MICE AS MODELS FOR TREATMENT

Clinical trials act as a controlled experiment, allowing researchers and doctors to determine the safety and efficacy of cancer drugs before they are widely prescribed for use. While there are a variety of clinical trials for studying oncology, drug trials are used to study drug safety and efficacy. Typically, drug trials consist of five phases (0-4), with phases 0 and 1 focusing on determining pharmacokinetics and safety respectively [129]. Phases 2 and 3 incorporate a larger number of participants to determine efficacy of the drug, and continue to monitor safety. Finally, phase 4 evaluates long term affects and outcomes of the drug. With the advent of numerous types of mouse models to study oncogenesis at a molecular, cellular, and histological level, there has also been an uptick in the usage of these models as pre-clinical indicators for the safety and efficacy of new cancer drugs. Often, experiments on the safety and efficacy of drugs are completed on mice before a drug can be taken to clinical trials. There is a question however, of whether these models are good indicators of how a drug will perform in the clinic.

Before use of genetically engineered mouse models in pre-clinical trial studies, *in vitro* data and patient derived xenograft models were used widely. Data from the National Cancer Institute (NCI) however showed experiments from these models did not correlate well with phase II clinical trial results

[130]. It was therefore hoped that use of GEMMs would better predict clinical trial results [131]. More recent studies however, have shown a continued failure of mouse models to predict safety and efficacy outcomes within the clinic [132–134]. Elongated telomeres found in laboratory mice may have implications in using mice as models for cancer and clinical studies [135, 136]. It is plausible that long telomeres in lab mice may result in an increased ability to repair tissue and resist toxicity, as well as enhance tumor promotion. However, most carefully designed studies in mice involve normalized controls, which would seemingly circumvent questions surrounding tumorigenesis.

It is also important to note many mouse studies failing to predict drug toxicity and efficacy may result from poorly designed experiments. Careful consideration must be given to the mouse model's histology, gene expression patterns driving that histology, molecular driver, immune microenvironment, and other factors [137, 138]. From available data, it seems the use of mouse models for pre-clinical studies must be reconsidered. With the current paradigm however, they will likely remain a staple for use as preclinical models.

BIOINFORMATICS AS A MEANS TO INVESTIGATE CANCER

The last few decades have seen an explosion of bioinformatics methods used to study cancer. These technologies have vastly improved our understanding of cancer on a molecular and epidemiological level. A large array of new approaches now allows researchers to study gene sequences and expression, cellular pathways, proteins, tumor-stromal interactions, epidemiological trends, and many other facets of carcinogenesis. With these new technologies has also come new hope for improved targeted treatments, and even more recently, a more serious look at pan cancer therapies. The following paragraphs will briefly look at some of the more common technologies and methods that have revolutionized the study cancer biology.

SEQUENCING

Since the initial advent of sanger sequencing in the 1970s and the first draft of the human genome in the early 2000s, sequencing technologies have come a long way. In sequencing the human genome, what initially took 3 billion dollars and 13 years to complete can now be done in a couple days with a few thousand dollars [139]. This is a testament to the newly available next generation sequencing technologies. Sequencing technologies seemingly have the ability to cover most facets of gene regulation. Genome sequencing can uncover large and small changes to the genetic code, chromatin immunoprecipitation (ChIP)-sequencing is capable of discovering protein regulatory changes in promoter regions, and RNA sequencing can determine changes in gene expression. Even more impressive are the advancements in single-cell sequencing, which can be applied at the DNA and RNA level, and has promise to tackle the questions surrounding intra-tumor heterogeneity [140]. On a clinical level, targeted sequencing and exome sequencing have become important for determining course of action for treatment regimes. Utilization of targeted therapies has increased in recent years, but these therapies still rely on genetic information to make sound treatment decisions. This is evident in the treatment of non-small cell lung cancer patients with tyrosine kinase inhibitors (TKI). While TKIs work effectively in patients who harbor activating mutations in the oncogene EGFR, they show no results in patients without the mutations [141]. Targeted sequencing completed on lung tumor biopsies is capable of providing clinicians with the proper information.

GENE EXPRESSION

While sequence analysis plays an important role in research and clinical therapy, gene expression analysis is another necessary piece of the puzzle. Often, gene expression is not affected by mutations to the underlying gene, or gene expression may change without gene mutations. Whether through microarray technology or RNA-sequencing, large scale shifts in gene expression can be determined for a large number of samples relatively simply. While RNA sequencing is now more widely used, microarray

technology is still around, and a more cost effective technology. The main difference between the two technologies is microarray's dependence on transcript specific oligos annealed to a chip, while RNA-seq sequences do not rely on these transcript specific oligos. When comparing the technologies, RNA-seq seems to have an advantage in detecting low level transcripts [142].

Like sequencing, gene expression patterns are often used to study cancer as well as determine the clinical course of action. In the laboratory, gene expression is often used to determine genome wide expression changes across sample groups that are subject to gene knockouts, drug treatment, or other experimental scenarios [105, 143]. In the clinic, gene expression patters are often used to classify patient tumors and determine a course of action for treatment [96, 104].

PATHWAY ANALYSIS

Cellular processes are often organized into complicated pathways and protein networks. A prudent example of this is the Ras/Raf/Mek/Erk signaling pathway stemming from RTK stimulation, and leading to eventual transcription factor activation or repression [144]. With the complicated nature of these pathways and their key role in stimulating and maintaining oncogenesis, researchers have developed a number of tools for their investigation. Often, these tools rely on gene expression data gathered through microarray or RNA-seq technologies. One such tool has been the development of pathway signatures for human breast cancer [110, 145]. Pathway signatures are often developed through overexpression of an oncogene or GFP (green fluorescent protein) control within a particular cell line. Expression data is then gathered from the oncogene or GFP overexpressed line, and a training dataset is developed to allow for classification of future samples. This classification is given as a score that predicts whether the pathway in question is active. Another pathway prediction tool is Gene Set Enrichment Analysis (GSEA) [146]. Briefly, GSEA uses gene expression data to compare two groups of samples in order to determine whether particular gene sets or pathways may be up or downregulated in one sample group compared to the other. This analysis can often be useful when first exploring expression data from two

sample groups, such as drug treated vs. non-treated groups. Overall, these programs can serve as good predictors to which pathways may be activated or repressed within a tumor. This gives researchers the ability to narrow their search when completing lab validation.

DATA ANALYSIS

With the plethora of data that has been generated using the above technologies comes a need for expert data analysis. Over the years, a number of regulations, programs, and analysis methods have been put forth to deal with the large amount of incoming data. To ensure public access to data produced under federal grant money, authors are required to submit datasets to online portals, such as the Gene Expression Omnibus (GEO). Large databases have also been developed to allow for analysis of large datasets by the public. An example of this is The Cancer Genome Atlas (TCGA), a tool used to access genomic mutation data for thousands of human tumors of various cancers.

Hundreds, if not thousands of programs have been generated to deal with the influx of data. Some of these programs are generated by individual labs, while others have been generated through the coordinated effort of multiple groups. For sequence analysis there are programs that "clean and prep" data, programs to align data to reference genomes, and numerous programs to determine genomic variants occurring within the data. Once the initial data processing is complete, there multitudes of other programs to complete specialized analysis, such as determining tumor heterogeneity or tumor mutation signatures. There are also dedicated programs for RNA, ChIP, and single-cell sequencing analysis.

While all of these programs work to achieve the same result, many go about it in a different fashion, making the choice of which program to use dependent on the biological question. For instance, in genome sequence analysis, some programs can uncover rare mutations but also have a higher number of false positives, while other programs have a lower number of false positives but may miss low frequency mutations. Overall, analysis methods have drastically improved to increase statistical power and remove confounding effects. This is exemplified in RNA sequencing data, where data normalization

has improved to remove potential analysis errors including transcript number and length. In many cases these advancements are beneficial, however in some cases they pose even more challenges. For example, microarray technology was the go to for obtaining gene expression data in the early 2000s. Even though microarray is still used, RNA sequencing has become the standard for many labs conducting large gene expression studies. While there is a boon of available data, integrating microarray and RNA-seq datasets is still a challenging endeavor.

THE FUTURE OF CANCER TREATMENT

Cancer therapy has made many strides since the 19th and 20th centuries, however there is still a long way to go. This is evident when examining the treatment regimes and survival rates of breast cancer. Once common place, radical mastectomies are now considered barbaric as less invasive surgeries combined with adjuvant therapy have been found equally effective [147]. Drug treatments have also advanced tremendously, from early mustard gas derivatives [148] to more advanced chemotherapies and targeted therapies [149–151]. These treatments have seen vastly increased 5-year survival rates and decreased observed mortality rate [152]. Late stage metastatic and triple negative breast cancers still carry a poor prognosis, showing the need for improved therapies. Like breast cancer, the overall success for treatment of cancers has varied widely. Some cancers such as breast and skin melanomas are treated with high success, while others, such as lung and pancreatic, yield a poor prognosis [153, 154]. However, just like breast cancer, the overall 5-year survival rates do not tell the whole story. Treatment success can vary widely within certain cancers depending on molecular phenotype, genetic mutations, and stage of diagnosis. Melanoma for instance has an extremely high success rate when caught early, but has a poor prognosis after metastasis has occurred [155].

Current research has focused on characterizing the molecular and histological profiles of tumors in order to develop new therapies. Within the clinic, patients undergo tumor biopsies, which then undergo sequence, molecular, and histological analysis to apply applicable targeted therapies. These

targeted therapies have improved survival rates within the clinic, but resistance mechanisms continue to be a challenging issue. Some clinical trials, such as the ongoing SMMART trial [156], are attempting to circumvent these resistance mechanisms by closely monitoring tumor growth and performing new biopsies once resistance begins to develop. This allows a new treatment regime to begin and a further reduction in tumor volume. While these avenues show a lot of promise, they have issues as well. For instance, some patients cannot be enrolled in the SMMART trial due to a lack of actionable mutations. Furthermore, multiple biopsies can be burdensome on the patient, and unfeasible in certain cancers. Finally, this approach is extremely costly in terms of financial burden and manpower. It is fair to point out these issues may be solved with further research and technology development. A further characterization of cancer genomes and molecular profiles may lead to a greater number of actionable mutations. Improvements in our understanding of, and sequencing extra-cellular vesicles and other biomarkers may eliminate the need for invasive biopsies [157]. Technology advancements may also reduce costs and labor.

The above financial challenges and patient burdens may make it impossible to apply this approach to every cancer patient and thus, other options need to be considered. While the heterogeneous nature of cancer has put finding a 'universal cure' in doubt, a universal cure is an endeavor we should still pursue even if that cure is more akin to a universal process than treatment with a single drug. The biggest obstacle in such an approach would surely be distinguishing tumor cells from cells in normal physiological condition. If this were done however, a number of targeting approaches could foreseeably be taken. One includes treating with already developed drugs that target particular pathways. More intriguing perhaps would be using Crispr technology in conjunction to inhibitors of DNA repair pathways. Hypothetically, this could damage the cancer cells enough to make them undergo cell cycle arrest and apoptosis once they are unable to repair the DNA damage. While these treatments may seem far off, they are surely worth investigation.

CHAPTER 1

ALTERED METASTASIS IN E2F1 KNOCKOUT MODELS OF HUMAN BREAST CANCER

PREFACE

While this chapter is not directly related to the bulk of the work in this thesis, its importance is two-fold. First, this chapter underscores many of the important bioinformatics methods I have learned during my time in the Andrechek lab. These methods are now vital for success as a cancer researcher. Second, the whole genome sequencing completed in this study directly resulted in finding a mutation in the *Ptprh* gene. The characterization of this *PTPRH* mutation and its relevance to human non-small cell lung cancer is the bulk of my thesis work, and illustrates the importance of pan-cancer research.

This chapter is adapted, with additional added data, from a manuscript previously published in Scientific Reports

As: "Metastasis is altered through multiple processes regulated by the E2F1 transcription factor" DOI: 10.1038/s41598-021-88924-y

ABSTRACT

The E2F family of transcription factors is important for many cellular processes, from their canonical role in cell cycle regulation to other roles in angiogenesis and metastasis. Alteration of the Rb/E2F pathway occurs in various forms of cancer, including breast cancer. E2F1 ablation has been shown to significantly decrease metastasis in MMTV-Neu and MMTV-PyMT transgenic mouse models of breast cancer. Here we take a bioinformatics approach to determine the impact of E2F1 loss on the genomic landscape of these tumors, and look specifically at genes related to the metastatic cascade, in both Neu and PyMT models. Through gene expression analysis, we reveal few transcriptome changes in non-metastatic E2F1-/- tumors relative to transgenic tumor controls. However investigation of these models through whole genome sequencing found numerous differences between the models, including differences in the proposed tumor etiology between E2F1-/- and E2F1+/+ tumors induced by Neu or PyMT. For example, loss of E2F1 within the Neu model led to an increased contribution of the inefficient double stranded break repair signature to the proposed etiology of the tumors. While the SNV mutation burden was higher in PyMT mouse tumors than Neu mouse tumors, there was no statistically significant differences between E2F WT and E2F1 KO mice. Investigating mutated genes through gene set analysis also found a significant number of genes mutated in the cell adhesion pathway in E2F1-/- tumors, indicating this may be a route for disruption of metastasis in E2F1-/- tumors. Overall, these findings illustrate the complicated nature of uncovering drivers of the metastatic process.

INTRODUCTION

Breast cancer is the most diagnosed cancer in women. To study genomic events contributing to breast cancer, numerous genetically engineered mouse models have been generated, including MMTV-Neu [158] which recapitulates HER2+ve breast cancer, and MMTV-Polyoma virus Middle T antigen (PyMT) [84]. The PyMT model relies on overexpression of the PyMT oncogene, leading to downstream activation of the SRC and AKT pathways. The PyMT model is highly aggressive, with tumors appearing at 45 days of age. Metastasis to the lung occurs in over 90% of tumor bearing mice, resulting in wide use of PyMT for metastasis studies. Similar to human breast cancers, both Neu and PyMT models have striking heterogeneity at histological and gene expression levels [89, 92, 104, 105], reinforcing the importance of these models as tools for the study of breast cancer.

Previous studies using Neu and PyMT models predicted a key role for the E2F1 transcription factor through a pathway signature analysis, suggesting that mechanisms outside the overexpression of the Neu or PyMT oncogene were contributing to tumor biology [86, 88]. The E2F family of transcription factors is involved in numerous cellular processes, best known for cell cycle control. Usually sequestered by retinoblastoma (Rb), E2F1 is released to act on downstream targets upon Rb phosphorylation [159]. While mutations in E2F1 are not common in human breast cancer, mutations within the E2F pathway occur in over 25% of breast cancer patients, illustrating the importance of the pathway [160–164].

To test the hypothesis that E2F1 regulates key events in Neu and PyMT tumors, E2F1 knockout (KO) mice [163] were interbred with Neu and PyMT models [86, 88]. This resulted in mammary tumors with changes in latency, growth rate, and a significant decrease in metastasis to the lung. Metastasis is the ultimate cause of mortality in cancer, with an estimated 90% of cancer deaths resulting from the spread of cancer cells to distal sites within the body [165]. Typically, cancer cells undergo numerous important steps for completion of the metastatic cascade. These include escape from the primary tumor, intravasation, extravasation, and seeding the distal site [166] as reviewed by Welch [167].

An important component contributing to the metastatic capability of a tumor is its microenvironment. Various collagens and proteins integral to cellular and tissue structure are capable of impacting metastatic potential. Indeed, proteins within the extracellular matrix, including collagen IV, have been found to regulate metastasis within the liver [168]. Collagen IV is a major component of the basement membrane, an important barrier to tumor invasion, and breaching this has been shown to be a critical early step in tumor invasion and metastasis [169, 170]. Interestingly, a previous report demonstrated a decrease in the number of circulating tumor cells within PyMT E2F1$^{-/-}$ mice, suggesting a disruption to the early steps in the metastatic cascade. Other data shows remodeling of the extracellular matrix at pre-metastatic lesion sites to be important for eventual seeding of distant metastasis [171].

Recent advances in bioinformatics methods have facilitated the investigation of cancer biology. Publicly available transcriptomic datasets have allowed for comparisons between primary tumor and distant metastatic lesions [172, 173]. Next generation sequencing has furthered our understanding of cancer genomics. Studies involving the sequencing of human tumors have described the mutation rate of solid tumors [174], and demonstrated that numerous genomic events are required for metastasis [120, 175, 176]. To determine the underlying genomic events behind altered metastatic characteristics in E2F1 KO tumors, gene expression and sequence data was analyzed. Here, we characterize the genome landscape of E2F WT and E2F1 KO tumors from both the Neu and PyMT models and uncover new targets that may be critical to tumor development and progression.

RESULTS

ANALYSIS OF GENE EXPRESSION DATA IN NEU AND PYMT TUMORS

We previously demonstrated altered phenotypic characteristics upon ablation of E2F1 within Neu and PyMT models, including changes in growth rate and tumor latency for the primary tumors (Figure 1.1A). Given the short latency of PyMT mice, it was surprising to observe tumor latency in PyMT mice significantly decreased with E2F1 loss while growth rate remained unaffected. Interestingly, the opposite

effect was seen within Neu E2F1$^{-/-}$ mice, where latency was significantly increased, and growth rate was significantly increased. However, the most striking phenotype was a significant reduction of metastasis with loss of E2F1 in both strains (Figure 1.1B and 1.1C).

To determine whether gene expression differences regulated phenotypic changes in E2F1 knockout tumors, fold change differences were examined. Volcano plots revealed few genes with major gene expression changes when analyzing E2F1 WT and E2F1 KO primary tumors (Figure 1.2A). While there were some genes with a fold change between 1 and 1.5, there were very few genes with a fold change greater than 1.5. To test whether this is recapitulated in human breast cancer, data from The Cancer Genome Atlas (TCGA) was analyzed. E2F1 activity in HER2+ve samples was determined using pathway signature analysis. Samples were stratified into quartiles for E2F1 activity and differential gene expression was determined. As shown in Figure 1.2B, human breast tumors resemble mouse mammary tumors in that low E2F1 activity does not lead to vast gene expression changes. To test for genetic pathways affected by loss of E2F1, Gene Set Enrichment Analysis (GSEA) was completed on Neu and PyMT tumors with and without E2F1. GSEA analysis revealed several differentially regulated pathways, including WNT signaling, and nucleotide excision repair (Figure 1.2C). Importantly, WNT signaling has been shown to regulate the epithelial to mesenchymal transition, a process involved in the metastatic cascade [177, 178].

MUTATION ANALYSIS THROUGH WHOLE GENOME SEQUENCING

Given that the gene expression analysis did not identify a mechanism altering metastatic potential, we examined genomic events occurring in Neu and PyMT tumors with and without E2F1. Whole genome sequencing was completed and single nucleotide variant (SNV) profiles were called for each tumor using TCGA best practices. Initial analysis of the SNV data resulted in an unexpectedly high proportion of SNVs occurring within chromosome 2 of the E2F1 knockout tumors (Figure 1.3A-D). However, E2F1 is located within the qH1 band of chromosome 2 and correlated to where the increased SNVs were observed (Figure 1.3E). While E2F1 knockout mice were backcrossed 12 generations to FVB,

we hypothesized that SNV abundance was called due to residual background strain DNA from the original E2F1 knockout stain. Given that E2F1 mice were generated in the SV129 background, and Neu and PyMT mice are on the FVB background, we filtered SNV calls using a list of SNVs that were generated from comparing the SV129 background against the C57/BL6 background, the standard mouse reference genome (Figure 1.3F). As a result, the majority of chromosome 2 SNV calls were filtered out, and the proportion of SNVs was roughly equal across the 19 autosomal mouse chromosomes in E2F1 WT and E2F1 KO PyMT tumors (Figure 1.3G). This was also the case for E2F1 KO Neu tumors (data not shown). As such, residual background is an important caution when sequencing mouse models.

Interestingly, the SNV mutation burden was higher in PyMT mice as compared to Neu mice (p-value = 0.05), which was surprising due to the brief latency of PyMT tumors (Figure 1.4A). Except for one PyMT E2F1 knockout tumor, the rate of exonic SNVs ranged from .005 to .08 mutations per megabase. This mutation rate is similar to previous rates shown for mouse tumors [179], and is lower than the 1 mutation / megabase exonic mutation rate commonly observed in human breast cancer [174]. Surprisingly, a significant percentage shift of exonic, intronic, and intergenic SNVs occurred when comparing PyMT E2F1 KO tumors to WT tumors (Figure 1.4A). In PyMT WT tumors, the percent of exonic and intronic mutations were approximately 1 and 30 respectively. This is in contrast to E2F1 KO tumors where the percentages were approximately 2 and 38 respectively. The percentage increases (P-value = .05 for exonic and .03 for intronic) seen in E2F1 KO tumors corresponded to percentage decreases (P-value = .03) in the intergenic regions of the tumors. These shifts were not seen in Neu tumors.

MUTATION SIGNATURES GENERATED FROM SNV PROFILES

To analyze distinct types of SNVs occurring within our tumors, and investigate potential mechanisms driving these differences, a mutation signature approach was taken [180]. While trinucleotide signatures showed similarities between Neu and PyMT tumors, there were striking differences, such as T>G mutations occurring almost exclusively in Neu tumors of either E2F1 status

(Figure 1.4B). The signatures for all 12 tumors are shown in (Figure 1.5). Principal component analysis (PCA) completed using mutation signatures from all 12 tumors shows distinct clustering between Neu and PyMT tumors (Figure 1.4C). Furthermore, apart from a single E2F1 KO PyMT tumor, PCA separates E2F1 WT and E2F1 KO tumors into distinct clusters within the Neu and PyMT models. While PyMT E2F1 KO sample 2 has a 6-fold increase in the number of SNVs, this is not reflected within the sample clustering of the principal component analysis. This is due to PCA being completed on the mutation signatures of the samples. For example, if sample X were to have an increased number of SNVs as compared to sample Y, but the overall mutation profile of those SNVs was similar between sample X and Y, they would cluster together.

The contribution of the 30 known COSMIC (catalog of somatic mutations in cancer) signatures to each Neu and PyMT tumor were then determined [180]. While all Neu and PyMT tumors had some contribution from signature 18, there were stark differences in other COSMIC signatures contributing to Neu and PyMT tumors (Figure 1.4D). For example, Neu tumors had contributions from signatures 1 and 3, while PyMT tumors were associated with signatures 4 and 20. Furthermore, there were signature differences when comparing E2F1 WT tumors to E2F1 KO tumors within the Neu and PyMT models. For example, Neu E2F1 WT tumors were associated with signatures 5 and 9, while Neu E2F1 KO tumors lacked these associations. Neu E2F1 KO tumors also had an association with signature 12, while Neu E2F WT tumors lacked this signature. When analyzing the proposed etiology for these signatures, Neu tumor signatures are associated with age, while PyMT tumor signatures have no age association, which correlates with Neu and PyMT tumor latency (Table 1.1). Interestingly, Neu tumors also have an association with inefficient double stranded break repair (DSB), with E2F1 KO tumors being more highly associated than E2F1 WT tumors. E2F1 has been found to recruit DSB processing factors, particularly NBS1, to DSB sites, which serves as a possible explanation for this signature [181]. PyMT E2F1 KO tumor signatures were not associated with DSB, but were highly associated with the smoking signature number

4, and defective DNA mismatch repair (MMR) signature 20. While it may seem counterintuitive that PyMT E2F1 KO tumors would be associated with one MMR signature and not the others (numbers 6, 15, and 26), it is entirely possible for this to occur. Multiple mutational profiles can be associated with a particular etiology, even though the mutational profiles themselves are distinct from each other. Together, these data suggest E2F1 loss drives differences in DNA repair and tumor etiology.

EXAMINING TUMOR CLONALITY

A wealth of evidence has shown tumors to have intra-tumoral heterogeneity on a histological and molecular level [119, 182–186]. Previous research demonstrated a shift in histological heterogeneity within E2F1$^{-/-}$ PyMT mice, where no shift in histology was seen in E2F1$^{-/-}$ Neu mice [86, 88]. To assess the molecular intra-tumoral heterogeneity in PyMT and Neu tumors, variant allele frequencies (VAF) were investigated. Briefly, the VAF is determined by taking a proportion of the number of reads containing a particular SNV mutation versus all of the reads in that location. In a single clone tumor, the VAF for all mutations will be .5 since half of the reads will have the mutation (we assume here that only one copy of the DNA is mutated). When analyzing the clonality of Neu and PyMT tumors, 5 of 6 Neu tumors had two clones, and all PyMT tumors had one clone (Figure 1.6). This is unsurprisingly given the fast growth of PyMT tumors. E2F1 status had no effect on the clonality of Neu or PyMT tumors.

COPY NUMBER AND TRANSLOCATION EVENTS

Multiple programs were also used to determine copy number variants and translocations occurring within Neu and PyMT tumors (Figure 1.7A-D). Based on consensus CNV calls from two programs, over 98% of the copy number events were small in size (under 1 mb), while relatively few larger events (above 1 mb) were observed. Surprisingly, there was a large amount of copy number gene overlap between the E2F WT and E2F1 KO tumors (Figure 1.7E). The large number of shared genes involved in copy number events may indicate E2F1 loss is not a primary driver of these events.

There were also a surprisingly large number of translocations occurring within the Neu and PyMT tumors. When comparing average number of translocations per sample across the genomic models, there were statistically more translocations occurring within Neu tumors than PyMT tumors, regardless of E2F1 status. When comparing E2F1 status within each model, there was no statistically significant difference (Figure 1.7F). To confirm the translocation calls made by Delly and Lumpy, 20 translocations from each tumor were chosen at random and read evidence for these translocations was analyzed using Genome Ribbon [187]. Translocation read data for one tumor is shown in Table 1.2. All tumors had at least 75% of translocations with some read support, with 9 of 12 tumors having at least 85% of translocations with some read support (Table 1.3). Interestingly, all translocation events analyzed had a varying level of wild type reads present. Since care was taken to exclude normal tissue when primary tumor was collected for sequencing, and since the abundance of wild type reads is fairly large for many of the translocation sites, this suggests a large amount of heterogeneity within the tumors. While some normal tissue (vasculature, immune etc.) is present in any tumor, the prevalence of wild type reads is far below that observed for mutations. To verify one of the translocation events from Table 1, PCR was completed with primers flanking the translocation junction. Both translocated and wild type reads were present at the breakpoint, confirming the existence of the translocation (Figure 1.8). Based on this evidence, upwards of 80% of the translocations were predicted to be real events.

ANALYSIS OF DISRUPTED PATHWAYS

To determine whether cancer and metastasis related genes were mutated within E2F1 WT and E2F1 KO tumors, the mutation list was filtered with known cancer genes from COSMIC. This analysis found mutations in a number of cancer associated genes (Table 1.4). While a few of the genes listed in supplemental table 2 have known metastatic implications, they were not consistently mutated within the sample groups, or were mutated exclusively within E2F wildtype tumors. To identify whether an abundance of mutations occurred within particular pathways comparing E2F1 knockout to wildtype

tumors, a database mining approach was taken using Gather [188]. First, genes with potentially impactful mutations were stratified into two gene lists that were distinct in $E2F1^{-/-}$ and $E2F1^{+/+}$ tumors. Potentially impactful mutations included SNVs causing stop gain or nonsynonymous mutations, translocations causing truncated or fusion genes, and copy number segments resulting in the amplification or deletion of genes. These two gene lists were then applied to Gather to determine whether Gene Ontology (GO) lists or KEGG (Kyoto encyclopedia of genes and genomes) pathways were significantly mutated. This analysis determined a number of significant GO lists that were present within the gene list from $E2F1^{-/-}$ tumors, but not $E2F1^{+/+}$ tumors.

In fact, the top three GO pathways associated with E2F1 KO tumors were involved in cell adhesion (GO:0007155 p-value = <.0001, GO:0007156 p-value = <.0001, GO:0016337 p-value = .0001). Genes in those cell adhesion GO annotations included various collagens, integrins, and cadherins (Figure 1.9). Previous research has shown collagens to be important for tumor maintenance, angiogenesis, and metastasis [168]. Collagen IV is the major component of the basement membrane and is comprised of heterogeneous trimers stemming from six COL4A genes. Three collagen IV genes were found mutated in different PyMT E2F1 KO tumors. Other mutations within PyMT E2F1 KO tumors include COL5A2, with collagen V being a component of the interstitial matrix, COL6A1-3, with collagen VI being abundant in the tumor invasive front [168–170] and several integrin and cadherin genes. Interestingly, a closer examination of the gene expression data revealed the integrin pathway was also found to be upregulated within E2F WT tumors, but not E2F1 KO tumors. There was also an abundance of intronic and synonymous mutations within these genes, suggesting they may be hypermutated due to the disruption of E2F1 within the model, although this hasn't been statistically verified. Indeed, of the 64 mutated genes within the cell adhesion Gene Ontology number 0007155, half were noted to have an E2F1 binding motif using TRANSFAC (p-value = .003, data not shown). With E2F1 known to regulate the cell cycle as well as a number of genes involved in DNA repair and adhesion, it is feasible that loss of E2F1 could result in an

abundance of mutations within certain gene profiles through a disruption of the cell's ability to undergo DNA repair during the S phase. E2F1 loss and corresponding disruptions to the cell cycle, especially during S phase could conceivably lead to an increased mutation burden, potentially within E2F regulated genes. E2F1 has also been shown to recruit nucleotide excision repair and double stranded break repair factors to sites of DNA damage [181, 189, 190]. It is possible that loss of recruitment of these factors could lead to inefficient DNA repair, and an increased mutational burden, although this would need to be further explored.

DISCUSSION

Ablation of E2F1 in PyMT and Neu transgenic mice results in a significant decrease in pulmonary metastasis. To determine whether gene expression changes were responsible for altered phenotypes, transcriptomic data was analyzed but showed no large changes in gene expression between $E2F1^{+/+}$ and $E2F1^{-/-}$ tumors. This was recapitulated in human HER2+ breast cancers after separation into E2F1 high/low quartiles. GSEA revealed several pathways differentially regulated between $E2F1^{+/+}$ and $E2F1^{-/-}$ tumors, but without obvious implications in regulating metastasis. To test for genomic alterations impacting metastasis, we completed WGS of $E2F1^{+/+}$ and $E2F1^{-/-}$ tumors in Neu and PyMT models. Mutation trinucleotide signatures showed differences between etiology of Neu and PyMT tumors, as well as between the E2F1 knockout and WT tumors. Neu tumors were more closely associated with double stranded break repair, while PyMT tumors were associated with DNA Mismatch Repair. As noted, Neu E2F1 KO tumors were more closely associated with defective double stranded break repair than Neu E2F wildtype tumors. An interesting question that warrants further investigation would be whether this was due to increased alterations within these genes upon loss of E2F1, or due to some other transcriptional function of E2F1. Analyzing mutated genes for GO and KEGG pathways revealed alterations in cell adhesion. Further analysis of these genes uncovered a role in the basement membrane and interstitial matrix, which could be a potential mechanism for disruption of the metastatic cascade.

Sequencing data from genetically engineered mouse models is largely lacking, with only a few models having been sequenced [115, 179, 191, 192]. SNV mutation rates between previous studies and ours indicate similarities, and small discrepancies may be explained through differences in data processing methods. For copy number variation prior research has shown numerous small copy number events and a few larger events [115], although this was estimated from whole exome sequencing data. This was recapitulated in our data, with the exception that large events were not prevalent after taking the consensus of two structural variant callers. We also noted a substantially greater number of translocations within the mouse tumors as compared to a previous study comparing Neu and PyMT wildtype tumors, while the same trend of Neu tumors having more translocations than PyMT tumors held. This increase in called translocations is likely due to differences in calling methods. Overall, the field would benefit from a large comparison of mouse tumor sequencing data with tumors analyzed under the same parameters.

After analyzing mutated genes using a pathway approach, many genes involved in cell adhesion were found having potentially impactful mutations in E2F1 knockout tumors, but not E2F1 wild type tumors, including various collagens, integrins and cadherins. Of the mutated genes found important to cell adhesion, genes such as *Col4a1* are important components of the basement membrane and are involved in tumor progression. Disruptions to the basement membrane and collagen formation has potential to disrupt the metastatic process. This theory is supported by previous data we generated, which found a significant decrease in circulating tumor cells [88]. Interestingly, we have also previously noted amplification of *Col1a1* in Neu E2F1 WT tumors which impacted the metastatic process [193]. Combined, these data suggest collagens and proteins within the basement membrane are important to the metastatic process in Neu and PyMT tumors.

SNV profiling for human tumors has utility for both discovery and treatment purposes. Sequencing of human breast tumors has revealed larger genomic trends as well as mutation rates for

oncogenes and tumor suppressors [194]. The importance of determining SNVs within mouse models is evidenced by previous research from our lab and others [115, 179]. Potential sources of error when determining SNVs can stem from differing genetic background within mice, even after backcrossing, as well as being too loose or too stringent with the filtering process. Interestingly, our prior work identified and validated a SNV in *Ptprh* in PyMT tumors [179], but this mutation was not present within this sequence analysis. While the initial paper stipulated an SNV call must pass 3 of 4 SNV calling programs, the work herein stipulated a call must pass 3 of 3 programs used, leading to the discrepancy. When analyzing the SNV data for each program used, a *Ptprh* SNV was called from SomaticSniper and Varscan, but not called from Mutect2. This suggests the usage of multiple programs to call SNVs is more applicable for discovery purposes, and that less stringent filtering parameters may be beneficial.

When analyzing copy number alterations and translocations within the models, there were a surprising lack of differences across E2F1 status, suggesting E2F1 loss is not a primary driver of these events. Furthermore, the varying read support seen for confirmed translocations indicates a high amount of tumor heterogeneity occurring in both models, regardless of E2F1 status. While there were numerous COSMIC associated genes mutated within the models, no mutations conserved between E2F1 knockout tumors (within or across models) were immediately apparent as important to the metastatic process.

Analyzing gene expression changes between E2F1 WT and E2F1 KO tumors showed no major changes upon E2F1 loss. This was recapitulated among human HER2+ve breast cancer tumors stratified between low and high E2F1 activity. The lack of large gene expression changes may indicate that numerous small changes result in phenotypic alterations, or that genomic mutations are leading to altered protein function/localization. Interestingly, the gene encoding Transcription Factor AP-2 Beta was significantly upregulated in Neu E2F1 KO mice. This, combined with the data showing a lack of major gene expression changes between E2F1 WT and E2F1 KO tumors, indicates some possible compensation by Transcription Factor AP-2 Beta, as well as other members of the E2F family [86, 105]. The sequencing

data from E2F1$^{-/-}$ Neu and PyMT mice indicate phenotypic changes may be due to an abundance of mutations in particular pathways, in addition to minor expression changes. Taking into consideration that the metastatic process likely originates from a small population of metastatic cells within the primary tumor, the contribution of a few metastatic cells to the bulk tumor gene expression or sequencing data may cause key events to be lost within the noise of the primary tumor. Future work will address these issues through single cell sequencing and gene expression in matched primary and metastatic tumors.

MATERIALS AND METHODS

GENE EXPRESSION ANALYSIS

Gene expression data was described previously [86, 92]. Volcano plots for Neu and PyMT tumors were generated by removing outliers for each sample group using Nowaclean (Holsb, Einar. 2017. "*nowaclean*"), samples greater than 3.0 standard deviations away when constructing PCA plots were removed. Data were log2 transformed, and the mean for each gene was calculated within the four sample groups. Fold change was calculated by subtracting the E2F1 KO mean from the E2F1 WT mean for each gene. P-values were calculated and data plotted using EnhancedVolcano (Blighe, Kevin. 2018. "EnhancedVolcano") in R. Human RSEM normalized RNAseq breast cancer data from TCGA was downloaded from UCSC Xena, filtered to HER2+ samples, and sorted by E2F1 expression. Lower and upper quartiles were kept and data were processed for volcano plots as above. GSEA plots were generated from combining Neu and PyMT gene expression datasets. Datasets were collapsed and combatted to remove batch effects. GSEA was run using GenePattern [195].

WHOLE GENOME SEQUENCING AND PROCESSING

Raw whole genome sequencing data from mouse tumors was previously obtained [28]. Briefly, three samples from each group (total of 12) were used, DNA from flash frozen extracted following manufacture's protocol for Qiagen Genomic-tip 20/G kit. Sequencing was completed at a depth of 40x with paired end, 150 base pair reads. DNA was prepared and sequenced using Illumina TruSeq Nano DNA

library preparation and an Illumina HiSeq 2500. For this study, raw fastq files were assessed for quality control using FASTQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimmed using Trimmomatic [196]. Files were aligned to mm10 mouse reference using BWA MEM [197] with standard parameters. Picard tools ("Picard Toolkit." 2019) was used to add read groups and remove duplicates. Samtools [198] was used to sort and index files.

VARIANT CALLING

Somatic SNVs were called using SomaticSniper [199], Mutect2 [200], and VarScan [201]. Consensus calls were merged using R (R Core Team (2018)) base programming, and mutations were only kept if called by all three programs. SNV calls were filtered using base R to account for differences between the FVB strain and mm10 alignment (C57/BL6), as well as differences between the SV129 strain (original E2F1 mouse background) and C57/BL6. SNVs were annotated using Annovar [202]. CNVs were determined by keeping the consensus of Lumpy [203] and Delly [204]. Consensus was determined using Intansv (Yao W 2019) at a threshold of .2, and events smaller than 10,000 bp were filtered out. Intansv was also used to annotate CNV events. Translocations were called using Lumpy and Delly, and filtered based on read evidence. Lumpy calls were kept if they had at least 20 supporting split end and paired end reads, Delly calls were kept if there was split end and paired end read evidence for the call. WT FVB mouse sequence was used as a normal control.

MUTATION SIGNATURES

Trinucleotide mutation signatures were completed using the Musica [205] shiny app in R. Musica code was altered to allow for the use of the mouse mm10 reference genome.

TUMOR CLONALITY

Clonality for each tumor was determined individually using the MAGOS program in R [206]. An updated R script was acquired through email correspondence with the author. Base R was used to extract

VAFs from the consensus SNV calls, and to prep files for use in MAGOS. VAFs of 0 and 1 were removed as per author's suggestion.

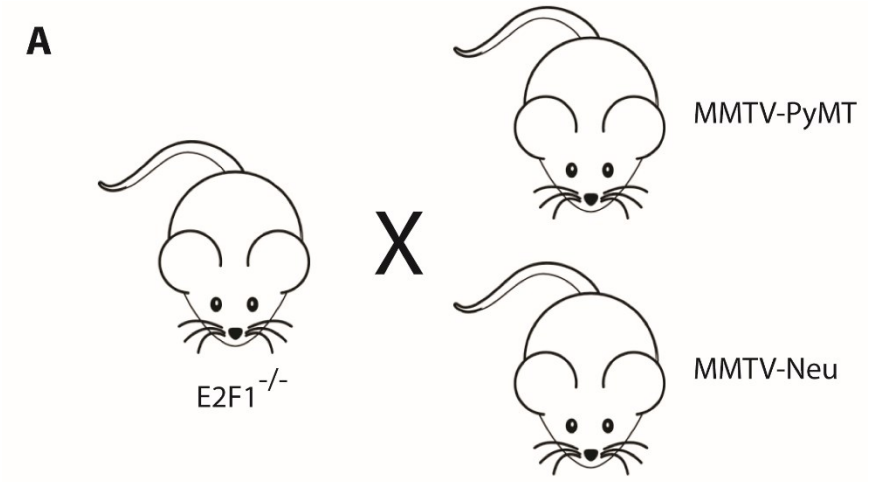
CIRCOS PLOTS

Circos plots were generated for each sample using CIRCOS version .69 [207]. Genetic variants were plotted according to the mm10 reference genome.

TRANSLOCATION VERIFICATION

Read evidence for 20 randomly selected translocations from all 12 sequenced samples was examined using GenomeRibbon [187]. For PCR verification, primers were designed with at least 400 bp flanking the predicted breakpoint.

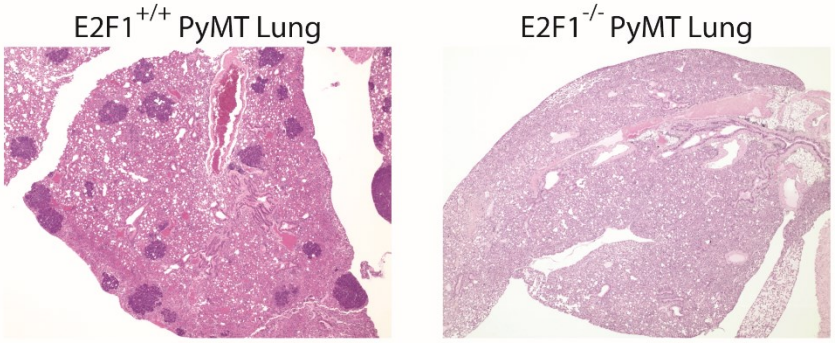
APPENDIX



B

PyMT Phenotypic Changes

E2F1 Status	Tumor Latency	Growth to 2000mm^3 (days)	Average # Lung Metastasis
Wildtype	50% at 42 days	44	87.8
Knockout	50% at 35 days	45	17.4



C

Neu Phenotypic Changes

E2F1 Status	Tumor Latency	Growth to 2500mm^3 (days)	Average # Lung Metastasis
Wildtype	212.7	46.1	6.3
Knockout	287.7	28.3	2.9

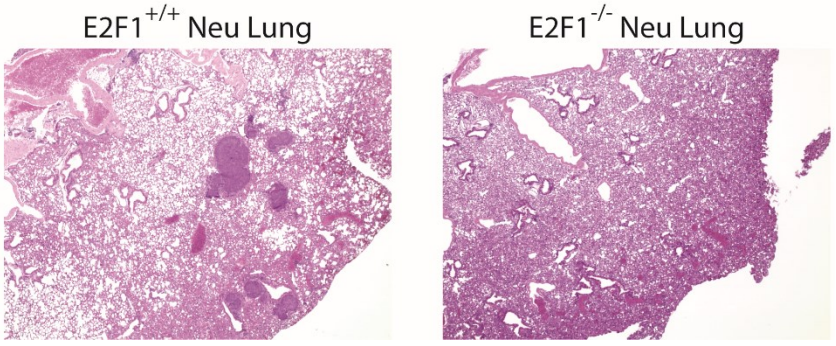


Figure 1.1: Altered phenotypic characteristics in E2F1$^{-/-}$ tumors

A) E2F1$^{-/-}$ mice were crossed with MMTV-Neu and MMTV-PyMT mice on the FVB background to create

Figure 1.1 (cont'd)

E2F1 knockouts in both models. B) Phenotypic changes seen in PyMT E2F1$^{-/-}$ mice and (C) Neu E2F1$^{-/-}$ mice, summarizing changes in latency, growth rate, and number of metastasis. H&E staining of E2F1$^{+/+}$ mouse lung shows a large number of metastasis, while E2F1$^{-/-}$ mice have little to no metastasis. Histology of the lungs was obtained at primary tumor endpoint.

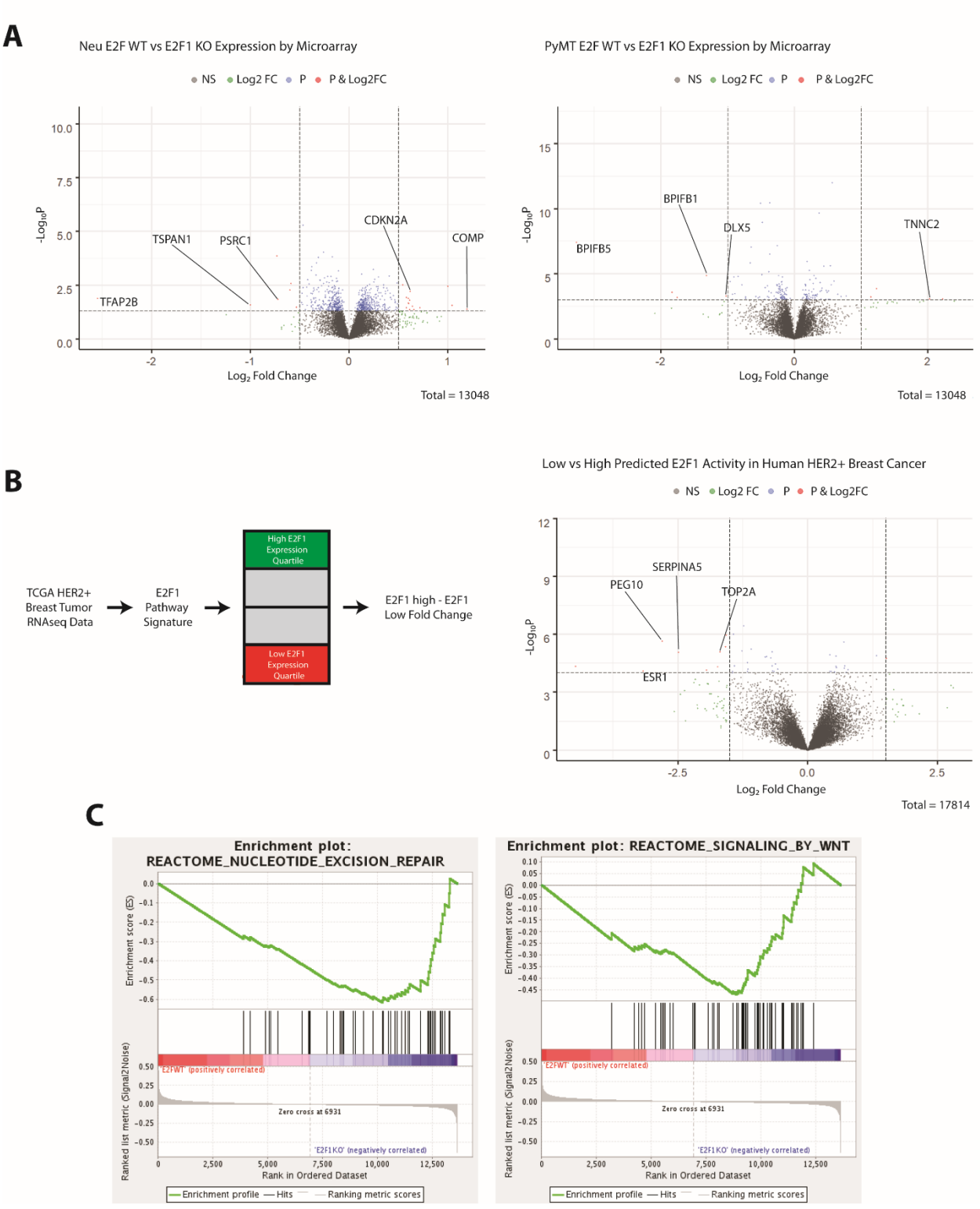


Figure 1.2: Gene expression changes in E2F1$^{-/-}$ mouse tumors, and E2F1 low human breast cancer

A) Two volcano plots show significant fold changes in genes from Neu and PyMT mouse tumors

Figure 1.2 (cont'd)

respectively. Fold change was determined by subtracting the E2F1 KO mean from the E2F1 WT mean for each gene. Fold change and p-value cutoff for Neu tumors was .5, and .05 respectively. Fold change and Pvalue for PyMT tumors was 1.0 and .001 respectively. B) Diagram represents data processing steps for human TCGA data. A volcano plot shows significant fold change genes in E2F1 high vs. E2F1 low human HER2+ve tumors. Fold change was determined by subtracting samples in the lowest E2F1 quartile mean from the highest E2F1 quartile mean for each gene. Fold change cutoff and p-value for human tumors was 2.0, and 10e−60 respectively. C) GSEA plots generated for E2F1 WT vs E2F1 KO tumors (Neu and PyMT combined) show enrichment of Nucleotide excision repair, and WNT signaling pathways in E2F1 KO tumors.

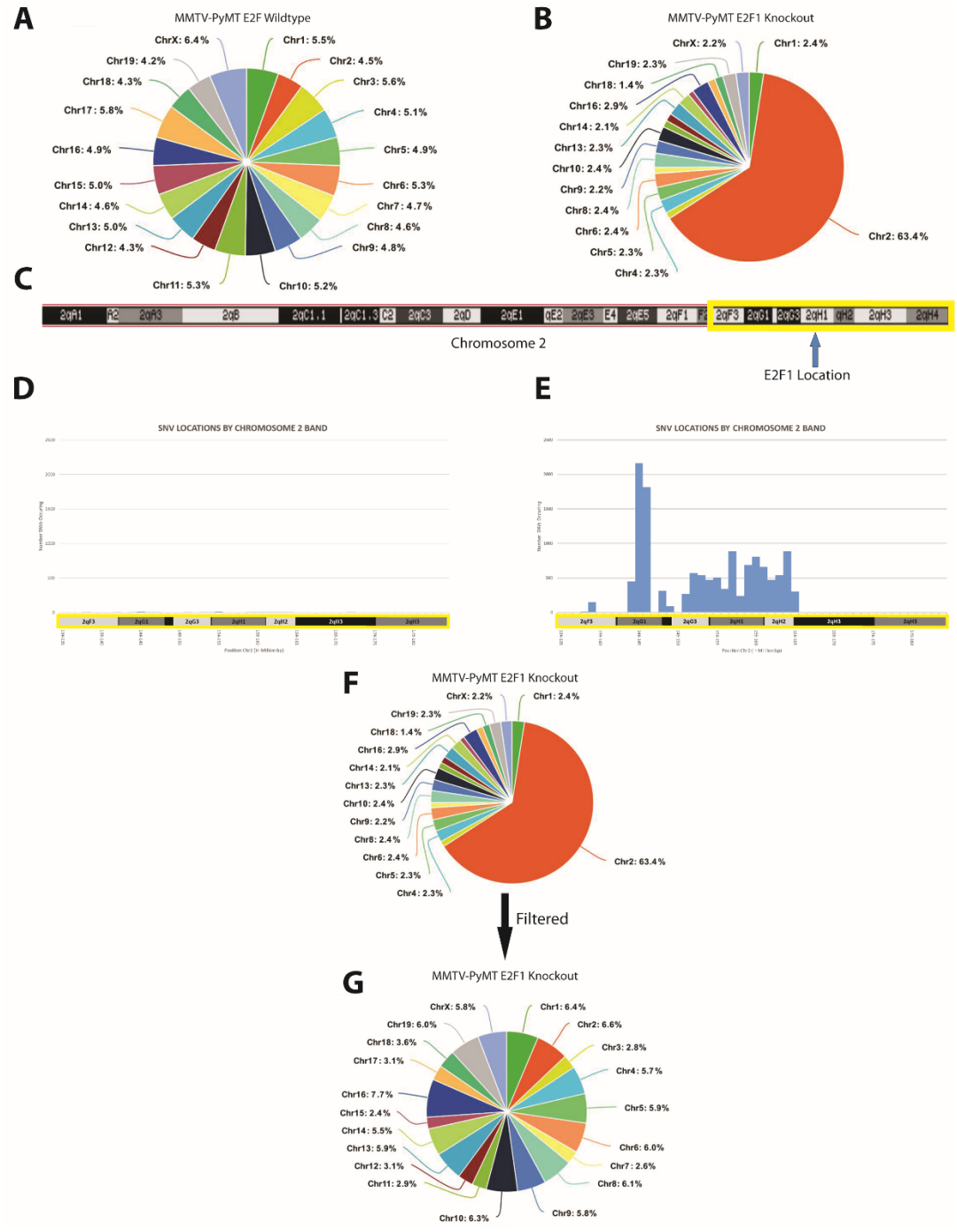


Figure 1.3: Filtering background strain to remove artifacts that have potential to confound analysis

A) Pie chart from an E2F1$^{+/+}$ PyMT tumor represents the normalized (SNVs/Chromosome Size) percentage of SNVs within each chromosome. B) Pie chart from an E2F1$^{-/-}$ PyMT tumor represents the normalized percentage of SNVs within each chromosome. An abundance of SNVs within chromosome 2

Figure 1.3 (cont'd)

is observed. C) The banding pattern of mouse chromosome 2. The arrow highlights the location of E2F1, and the yellow box represents the bands represented in D and E. D) Manhattan plot shows the number of SNVs occurring within the 2qF3-2qH3 bands of chromosome 2, in the $E2F1^{+/+}$ sample from A. E) Manhattan plot shows the number of SNVs occurring within the 2qF3-2qH3 bands of chromosome 2, in the $E2F1^{-/-}$ sample from B. F) Top pie chart is the same as in B. Bottom pie chart represents the percentage of SNVs across each chromosome of the same sample as above, after filtering on the sv129 background.

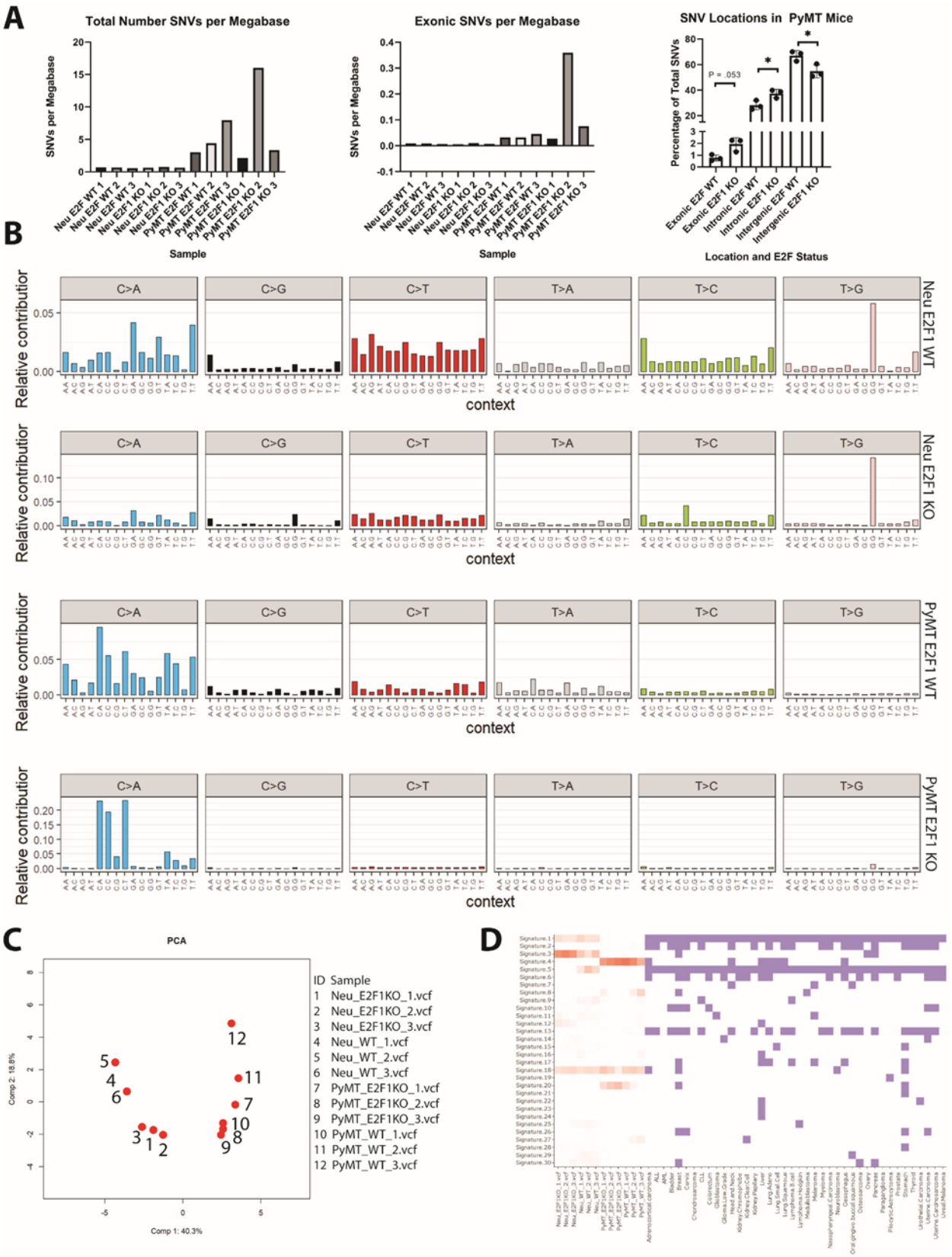


Figure 1.4: SNV mutation burden in Neu and PyMT tumors

Figure 1.4 (cont'd)

A) First two bar graphs represent the number of total or exonic mutations per megabase occurring in all 12 sequenced tumors. Third graph represents the percentage shift of exonic, intronic, and intergenic mutations in $PyMT^{+/+}$ and $PyMT^{-/-}$ tumors. B) Shows representative mutation profiles for each of the four classes of samples sequenced. Mutation profiles are derived from 96 bp trinucleotide signatures originally developed by Alexandrov et. al. Four classes of samples are Neu $E2F1^{+/+}$, Neu $E2F1^{-/-}$, PyMT $E2F1^{+/+}$, PyMT $E2F1^{-/-}$. C) PCA plots derived from trinucleotide signatures show clustering of all 12 samples sequenced. D) The heatmap of cancer signatures for the 12 sequenced tumors, as well as various cancers is shown.

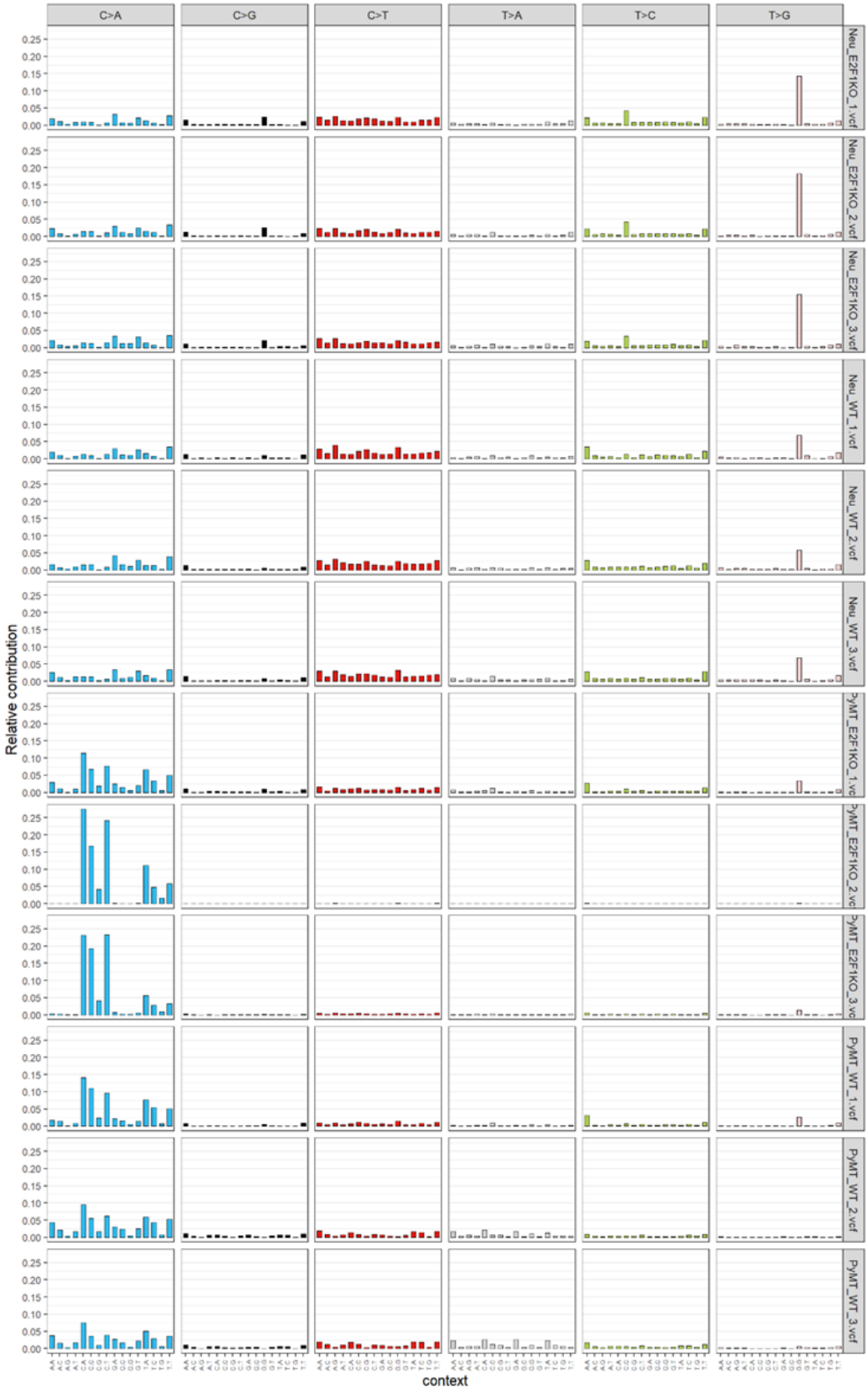


Figure 1.5: Mutation profiles

Figure 1.5 (cont'd)

Mutation profiles for all 12 Neu and PyMT mouse tumors corresponding to four classes in Figure 4B. Mutation profiles derived from 96 bp trinucleotide signatures originally developed by Alexandrov et. al.

Figure 1.6: Clonal heterogeneity in Neu and PyMT tumors

Graphs showing clonal populations in representative Neu and PyMT tumors. Each dot represents a specific mutation, with the Y-axis showing the total number of reads covering that mutation, and the X-axis showing the variant allele frequency of that mutation. Each color represents a different predicted clone. E2F1 status did not affect clonality.

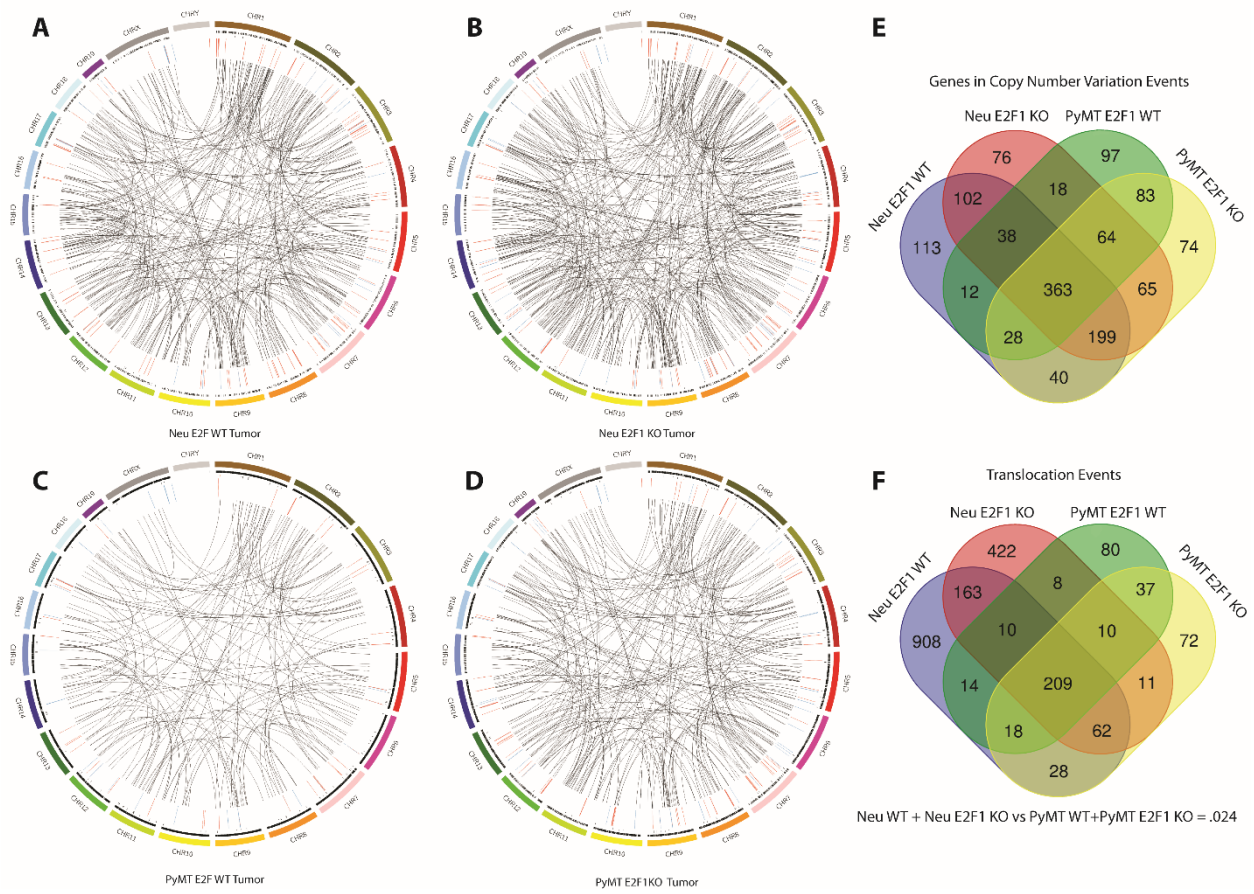


Figure 1.7: Mutation burden in Neu and PyMT tumors

A) Circos plot for a representative Neu E2F1$^{+/+}$ sample. B) Circos plot for a representative Neu E2F1$^{-/-}$ sample. C) Circos plot for a representative PyMT E2F1$^{+/+}$ sample. D) Circos plot for a representative PyMT E2F1$^{-/-}$ sample. For A-D Circos plots, outer most ring represents the mouse chromosomes. Four successive inner rings represent the following mutation types; total SNVs, exonic SNVs, Copy number variation with green being amplification and red being deletion, and translocations. E) Venn diagram showing the overlap of genes within copy number events. Consensus copy number events were generated for each of the three samples within the four sample classes. Genes were then extracted and compared across the sample classes. F). Venn diagram showing the overlap of translocations occurring within the four sample classes. Consensus translocations calls from each of the three samples within each class were generated, and the four classes were then compared.

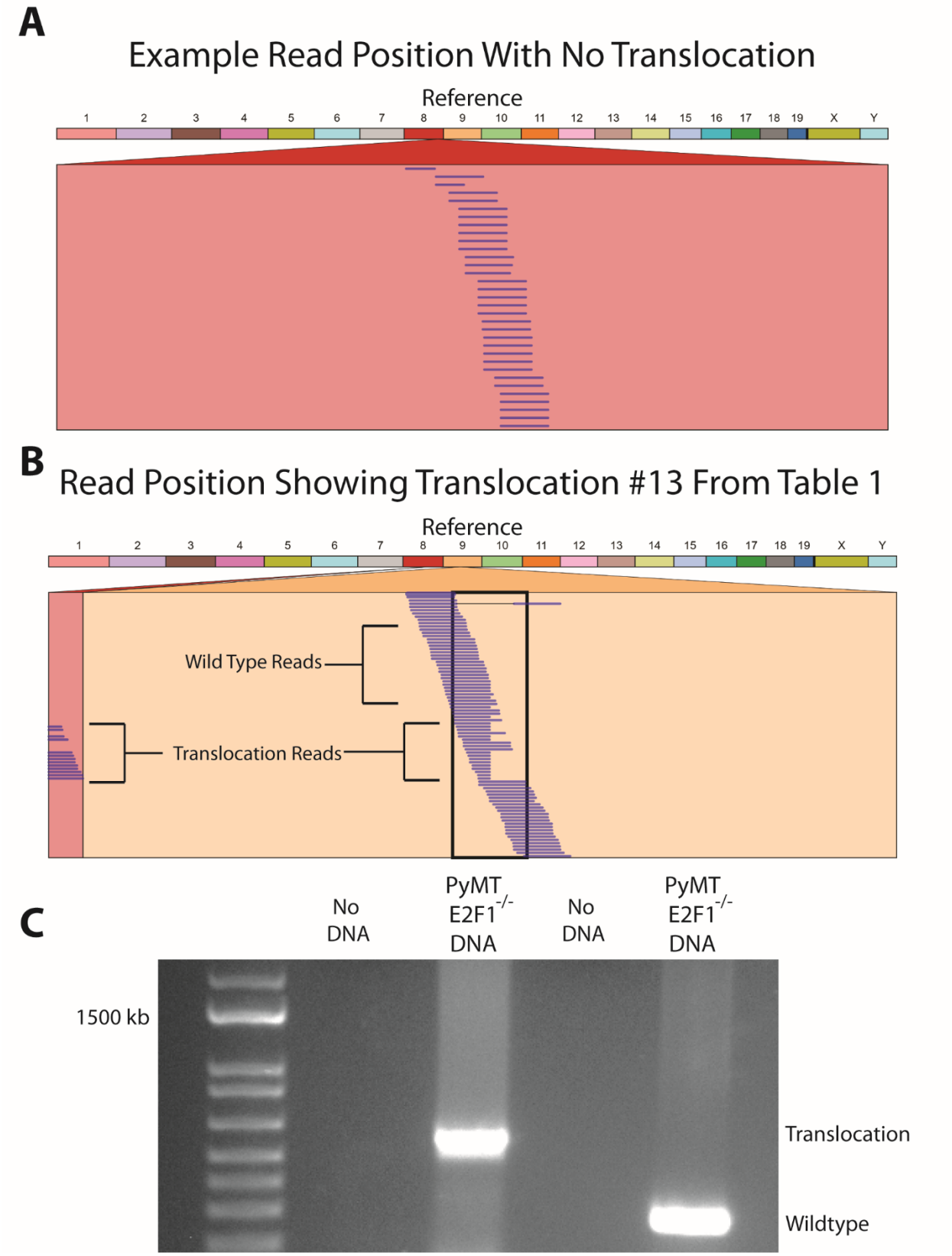


Figure 1.8: Verification of translocation calls

Figure 1.8 (cont'd)

A) Example of a GenomeRibbon plot where no structural variation occurs. The top colored bands represent each chromosome of the mouse, and the red box below represents the location searched within a sample's bam file. Each line within that box represents a different read. B) A GenomeRibbon plot representing translocation number 13 from table 1. Translocated reads are shown between chromosome 9 and chromosome 8. C) Gel image of the chromosome 8/9 translocation from the GenomeRibbon plot above. DNA was from a PyMT E2F1$^{-/-}$ tumor. Both translocation and wild type tumor DNA were amplified. Translocated reads were amplified using a primer set flanking the region where the two translocated ends ligate.

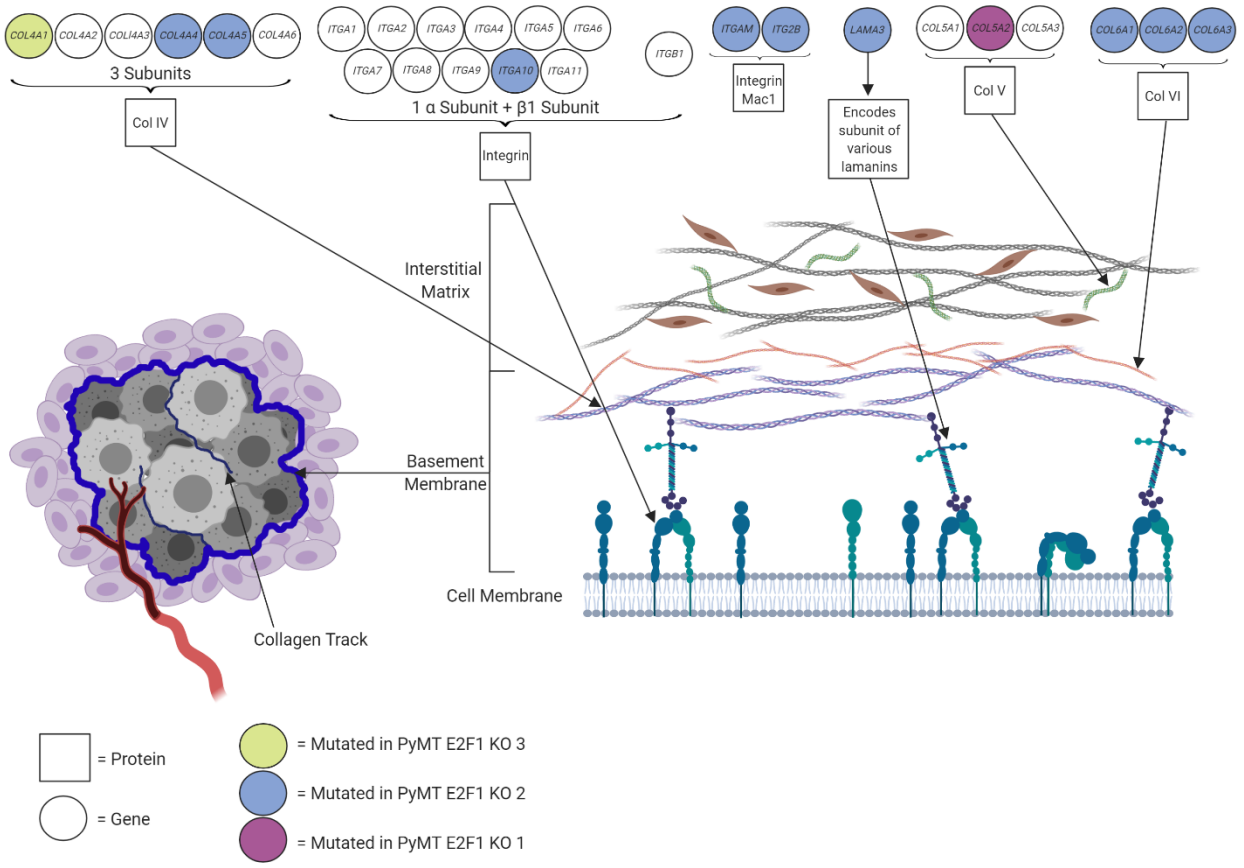


Figure 1.9: Mutations in basement membrane genes

Diagram shows various mutations occurring in genes that code for proteins making up the basement membrane and interstitial matrix. Circles at top indicate genes with colors representing 1 of 3 sequenced E2F1−/− PyMT tumors that has a mutation in that gene. Image on left represents a breast tumor with surrounding basement membrane. Image on right represents the basement membrane and interstitial matrix on the outer edge of a tumor.

Signature	Proposed_Etiology	Neu_E2F1KO_1	Neu_E2F1KO_2	Neu_E2F1KO_3	Neu_WT_1	Neu_WT_2	Neu_WT_3	PyMT_E2F1KO_1	PyMT_E2F1KO_2	PyMT_E2F1KO_3	PyMT_WT_1	PyMT_WT_2	PyMT_WT_3
1	Age	0.148	0.117	0.111	0.185	0.128	0.14	0	0	0	0	0	0
2	APOBEC	0	0	0	0.004	0.014	0.002	0.004	0	0	0	0.019	0.022
3	BRCA1 / BRCA2 (failure of DNA DSBR / large INDELs)	0.518	0.59	0.506	0.254	0.138	0.274	0	0	0	0	0	0
4	Smoking	0	0	0	0	0	0	0.471	0.557	0.616	0.65	0.558	0.368
5	Unknown (all cancer types)	0	0	0	0.107	0.277	0.176	0	0	0	0	0	0
6	Defective DNA MMR / MSI (small INDELs)	0	0	0	0	0	0	0.003	0	0	0	0	0
7	UV light	0	0	0	0	0	0	0	0	0	0	0	0.004
8	Unknown (breast cancer and medulloblastoma)	0	0	0	0	0	0.009	0.074	0	0	0	0.082	0.201
9	POLH (CLL, BCL)	0.012	0	0.004	0.052	0.039	0.07	0.017	0	0	0	0	0.03
10	POLE (ultra-hypermutation)	0.002	0.003	0.009	0.01	0.015	0.006	0	0	0	0	0	0
11	Alkylating agents	0.048	0.02	0	0.038	0.033	0.029	0	0	0	0	0	0.022
12	Unknown (liver cancer)	0.095	0.066	0.047	0.018	0.01	0.003	0	0	0	0	0	0.02
13	APOBEC	0	0	0	0	0	0	0.009	0	0	0.003	0.012	0.01
14	Unknown (uterine cancer and glioma / hypermutation)	0	0	0.032	0.023	0	0	0.001	0	0	0	0	0
15	Defective DNA MMR (small INDELs)	0.004	0.022	0.01	0.02	0.007	0.022	0	0	0	0	0	0
16	Unknown (liver cancer)	0	0	0	0.053	0	0	0	0	0	0	0	0.046
17	Unknown (different cancers)	0	0	0	0	0	0	0	0	0	0	0	0
18	Unknown (different cancers)	0.133	0.163	0.178	0.129	0.17	0.147	0.263	0.192	0.085	0.2	0.227	0.177
19	Unknown (pilocytic astrocytoma)	0	0	0.009	0	0	0.005	0	0	0	0	0	0
20	Defective DNA MMR (small INDELs)	0	0	0	0	0	0	0.154	0.251	0.299	0.144	0.051	0.006
21	Unknown (stomach cancer / MSI)	0	0	0	0	0.001	0	0	0	0	0	0	0
22	Aristolochic acid	0	0	0	0	0	0	0	0	0	0	0	0
23	Unknown (liver cancer)	0	0	0	0	0	0	0	0	0	0	0	0
24	Aflatoxin	0	0	0	0	0	0	0	0	0	0	0	0
25	Unknown (Hodgkin lymphoma)	0	0	0.029	0.032	0.045	0.026	0	0	0	0	0	0
26	Defective DNA MMR (small INDELs)	0	0	0	0	0	0	0	0	0	0	0	0
27	Unknown (kidney clear cell carcinomas / small INDELs)	0	0	0	0	0	0	0	0	0	0	0.05	0.082
28	Unknown (stomach cancer)	0.027	0.019	0.012	0.015	0.018	0.011	0.004	0	0	0.003	0	0
29	Tobacco chewing	0	0	0	0.034	0.055	0.035	0	0	0	0	0	0
30	Unknown (breast cancer)	0.014	0	0.053	0.025	0.049	0.046	0	0	0	0	0	0.011

Table 1.1: Mouse tumor signature etiology

Table showing contribution of each proposed tumor etiology for each of the 12 mouse tumors.

Numbers represent a proportion of the whole.

Translocation #	Position1	Position2	Supporting Reads (approximate)	Total Reads (exact)	% Support
1	3 _ 65552053	2 _ 20941004	10	61	16.39
2	2 _ 161669843	18 _ 78292047	10	119	8.40
3	15 _ 43944218	1 _ 112318855	6	89	6.74
4	13 _ 23307876	11 _ 88303305	11	104	10.58
5	5 _ 5609182	18 _ 56166475	14	77	18.18
6	5 _ 7058436	2 _ 8931319	4	49	8.16
7	16 _ 83532604	14 _ 96841358	15	83	18.07
8	3 _ 153288050	17 _ 10818207	15	88	17.05
9	16 _ 83532819	14 _ 96841377	19	94	20.21
10	16 _ 18368004	12 _ 80665928	10	85	11.76
11	8 _ 102861241	17 _ 67932443	0	57	0.00
12	14 _ 21312516	11 _ 9863013	16	234	6.84
13	9 _ 55224433	8 _ 85188141	13	82	15.85
14	X _ 38480149	9 _ 55983052	12	70	17.14
15	6 _ 73162349	16 _ 96121815	3	34	8.82
16	4 _ 43262573	3 _ 113857214	4	68	5.88
17	5 _ 62573934	13 _ 86796353	2	77	2.60
18	3 _ 135929183	1 _ 139635092	9	95	9.47
19	6 _ 67680744	4 _ 147419479	0	158	0.00
20	7 _ 79199005	19 _ 40536086	14	81	17.28

Table 1.2: Supporting reads for 20 randomly selected translocations from the tumor in figure 6

Random translocations were selected by inputting all translocations from the tumor into Excel, and using the RAND() function to assign a random number. The 20 highest translocations were then selected. Positions 1 and 2 represent the translocation breakpoint. Genome Ribbon was used to analyze read evidence.

Tumor	Translocations with Extensive* Read Support	Translocations with Low* Read Support	Translocations with no Read Support	% Extensive Support	% with at Least Some Support	Average Read Support (%)
Neu_E2F1KO_1	18	1	1	90	95	14.5
Neu_E2F1KO_2	17	0	3	85	85	14.78
Neu_E2F1KO_3	13	2	5	65	75	8.82
Neu_WT_1	17	1	2	85	90	15.83
Neu_WT_2	13	3	4	65	80	10.22
Neu_WT_3	18	1	1	90	95	12.42
PyMT_E2F1KO_1	16	2	2	80	90	13.21
PyMT_E2F1KO_2	18	1	1	90	95	13.48
PyMT_E2F1KO_3	17	1	2	85	90	10.8
PyMT_WT_1	14	4	2	70	90	13.62
PyMT_WT_2	13	3	5	65	80	14.94
PyMT_WT_3	15	4	1	75	95	14.06
*Extensive read support is deemed greater than 5% of reads supporting the translocation						
*Low read support is deemed greater than 0, but less than 5% of reads supporting the translocation						

Table 1.3: Table showing read support for 20 randomly drawn translocations within each of the 12 mouse tumors

To pick 20 random translocations, for each tumor, all translocation events were imported into excel and a random number was assigned using RAND() function. These were then sorted highest to lowest, and the 20 highest translocations were taken. Translocation read support was analyzed using GenomeRibbon.

Cosmic Cancer Genes Exclusive to E2F1 KO Tumors	Mutation Type	Cosmic Cancer Gene Mutations Exclusive to E2F1 WT Tumors	Mutation Type
ABL1	SNV	AFF4	SNV
AFF1	SNV	ATP1A1	SNV
AKT2	SNV	BAP1	SNV
ALK	SNV	BCL2	SNV
ANK1	SNV	BCL7A	SNV
AR	SNV	CARD11	SNV
ARHGEF10	SNV	CASP3	SNV
ARID1A	SNV	CHEK2	SNV
ATM	SNV	CPEB3	SNV
ATRX	SNV	CTNNB1	SNV
AXIN1	SNV	ETV4	SNV
BAZ1A	Translocation	FLI1	SNV
BCL11A	SNV	FOXO4	SNV
BCL9	SNV	LHFP	Translocation
BCL9L	SNV	LMNA	SNV
BRD4	SNV	MSH2	SNV
CAMTA1	SNV	NRG1	SNV
CASP9	SNV	PIK3R1	SNV
CBLB	SNV	PLAG1	SNV
CCDC6	SNV	POLD1	SNV
CD274	SNV	PREX2	SNV
CD79A	SNV	RANBP2	SNV
CDKN1A	SNV	ROBO2	SNV
CNTRL	SNV	RSPO3	SNV
CREB1	SNV	SF3B1	SNV
DNMT3A	SNV	SMAD4	SNV
ELF4	SNV	SUZ12	SNV
ELK4	SNV	TGFBR2	SNV
ELN	SNV	ZBTB16	SNV
EPS15	SNV	ZEB1	SNV
ERCC2	SNV		
ERCC3	SNV		
ERCC4	SNV		
ETV5	SNV		
EZR	SNV		
FAM47C	SNV		
FAT3	SNV		
FGFR2	SNV		
FLNA	SNV		
FLT3	SNV		
FOXP1	SNV		

Table 1.4: Cosmic associated genes

Table 1.4 (cont'd)

Table shows Cosmic cancer associated genes that are mutated exclusively within E2F1 KO or E2F WT mouse tumors.

GAS7	SNV		
GPC5	SNV		
GRM3	SNV		
H3F3A	SNV		
HOXD11	SNV		
IL6ST	SNV		
JAK2	SNV		
KAT7	SNV		
KCNJ5	SNV		
KDM6A	SNV		
KDSR	SNV		
KEAP1	SNV		
KMT2A	SNV		
KMT2C	SNV		
KMT2D	SNV		
LZTR1	SNV		
MAF	SNV		
MALT1	SNV		
MAP2K4	SNV		
MAP3K13	SNV		
MITF	Translocation		
MLLT1	SNV		
MLLT10	SNV		
MSN	SNV		
MUTYH	SNV		
NACA	SNV		
NBEA	Translocation		
NF1	SNV		
NFKB2	SNV		
NIN	SNV		
NTRK3	SNV		
NUP98	SNV		
NUTM1	SNV		
PAX8	SNV		
PDGFRA	SNV		
PDGFRB	SNV		
PHOX2B	SNV		
PICALM	SNV		
POU2AF1	SNV		
PTCH1	SNV		

Table 1.4 (cont'd)

PTK6	SNV		
PTPN6	SNV		
PTPRT	SNV		
PWWP2A	SNV		
RARA	SNV		
REL	SNV		
RET	SNV		
RMI2	SNV		
RNF213	SNV		
ROS1	SNV		
SDHAF2	SNV		
SETD2	SNV		
SFPQ	SNV		
SIRPA	SNV		
SIX1	SNV		
SKI	SNV		
SMARCE1	SNV		
SOCS1	SNV		
SPEN	SNV		
SRC	SNV		
SRGAP3	SNV		
STAG1	SNV		
STK11	SNV		
STRN	SNV		
TAF15	SNV		
TBX3	SNV		
TCF3	SNV		
TEC	SNV		
TET1	SNV		
TET2	SNV		
TFEB	SNV		
THRAP3	SNV		
TMPRSS2	SNV		
TRAF7	SNV		
TRIM24	SNV		
TRIM27	SNV		
TRIP11	SNV		
TSC1	SNV		
TSHR	SNV		
VAV1	SNV		
VHL	SNV		
WT1	SNV		
ZFHX3	SNV		

CHAPTER 2

P*TPRH* MUTATIONS IN PYMT MOUSE TUMORS

ABSTRACT

Genetically engineered mouse models are an important means for investigating a variety of cancers. While their relevancy to human cancer has been well documented on a histological and molecular level, sequencing of mouse tumors has not been as common. Through whole genome sequencing of PyMT mouse mammary tumors, we have uncovered a mutation in the protein tyrosine phosphatase receptor type H gene (*Ptprh)*. This conserved mutation is present in 80% of PyMT tumors, and correlates with increased phosphorylation of EGFR, a known target of PTPRH. Interestingly, *Ptprh* mutations also correlated with increased p-AKT, an important signaling molecule downstream of EGFR.

INTRODUCTION

PHOSPHATE SIGNALING WITHIN THE CELL

The human body is a highly organized, functional system. To achieve this high degree of functionality, cells need to communicate effectively with their neighbors and within themselves. Communicating with neighboring cells is usually accomplished through a variety of extra-cellular ligands that act as messages travelling from cell to cell, and throughout the body. Within each cell, signaling is achieved through a complicated network of specialized proteins that are often activated through a series of reactions catalyzing ATP to phosphorylate amino acid residues on target substrates. Many of these proteins are classified as kinases and broken down into two large groups based on which amino acid residues are phosphorylated, including serine/threonine kinases and tyrosine kinases. Within these cascades are a number of other proteins including guanine nucleotide exchange factors (GEF) that act as intermediaries, and become active upon exchange of their bound GDP for GTP. Some downstream targets of these cascades are transcription factors. Activation or repression of transcription factors by signaling cascades eventually leads to transcription of various genes. While these signaling pathways are complicated, they are often initiated through various receptor tyrosine kinases (RTK)s.

RECEPTOR TYROSINE KINASES

Receptor tyrosine kinases are perhaps some of the most important signaling molecules in cellular communication and cancer. Prior to the name 'receptor tyrosine kinase' being coined in the late 1970's, important work involving the elucidation of this class of proteins had been done. Experiments in the early 1960s were responsible for the discovery of epidermal growth factor (EGF), the ligand eventually found to be responsible for stimulation of the epidermal growth factor receptor (EGFR) and other RTKs. EGF was found to prompt early tooth eruption and eyelid formation in 8 day old mice [208, 209]. Work in the 70s also demonstrated the ability of the protein SRC and the growth factor EGF to stimulate serine and threonine phosphorylation, with the eventual seminal paper showing phosphorylation of tyrosine

residues by the SRC kinase [210–214]. It was a short time later when the phosphor-tyrosine activity of EGFR was also found [215].

Decades later, we have a much clearer understanding of RTKs and how they operate within the cell. While there are numerous classes of RTKs capable of being activated in a number of fashions, canonical RTK activation typically relies on dimerization of RTK monomers residing within the cell's cytoplasmic membrane ([216]. These RTKs consist of an extracellular binding domain, transmembrane domain, and intracellular catalytic domain. The basic RKT activation process consisting of ligand binding, dimerization, and phosphorylation of tyrosine residues on the C-terminal tail is conserved across varying RTK families, however the details of the process can differ significantly between individual receptors. While extracellular ligand binding appears necessary for RTK activation, and is usually associated with dimerization of RTK monomers, it isn't always required for dimerization [217]. Early RTK paradigm, applicable to a number of RTKs, shows dimerization is driven by ligands that are themselves dimerized, as is the case with VEGF and Axl [218–220]. Dimerization of other RTKs is driven by monomeric ligands, such as FGF [221]. Certain RTKs also require accessory molecules to aid in dimerization [222, 223]. With certain RTKs, the ligands directly facilitate dimerization by binding to each other. However, some RTKs dimerize by binding directly to themselves, with ligand binding facilitating activation by inducing a conformation shift. In some cases, RTKs are capable of dimerizing with other members in their family, which is common within the ERBB family of RTKs [224].

After ligand binding and dimerization, activation of the RTK dimer occurs, usually through a conformational shift that releases cis-auto inhibition in the intracellular domain. In most cases, the conformational shift opens up the active site, allowing ATP binding to occur. Interestingly, while most RTKs have vastly different crystal structures in an inactive state, structures of active RTK catalytic domains are strikingly similar [225]. Other modes of activation are seen, including by-passing allosteric inhibition as well as inhibition by c-terminal sequences [226]. After the active site is made accessible, tyrosine

residues near the c-terminal end of the intracellular domain become phosphorylated. Many RTKs have numerous tyrosine residues capable of being phosphorylated, and some evidence shows the residues are phosphorylated in a specific order [227].

Once tyrosine residues on the C-terminal tail become phosphorylated, a number of signaling molecules are recruited to propagate downstream signaling. Many of these molecules have SRC homology 2 (SH2) or phosphotyrosine binding (PTB) domains [228, 229]. These signaling cascades can achieve deregulated cell growth through numerous mechanisms, including activation or repression of numerous transcription factors capable of altering cellular programing, as well as differential control of the cell cycle. Numerous mechanisms act as a negative feedback loop to keep RTK signaling in check, including RTK degradation through ubiquitination and phosphate removal by protein tyrosine phosphatases (PTPs) [230, 231]. Overall, the complexity of RTK signaling is vast, and disruptions to all facets of these processes can induce tumor formation. Disruptions to RTKs themselves include chromosomal rearrangements, amplification events, and gene mutations resulting in a gain of function [226], something commonly seen within the epidermal growth factor receptor.

EPIDERMAL GROWTH FACTOR RECEPTOR

EGFR plays a role in numerous cancers including glioma and lung cancer. EGFR is a member of the ERBB family of RTKs, and is involved in numerous signaling pathways responsible for increasing cellular growth, proliferation, and an evasion of apoptotic signals. Pathways stimulated by EGFR activation include Pi3k/Akt and Ras/Raf/Mek/Erk. While EGFR follows the basic RTK activation process, there are notable differences compared to more 'canonical' receptor tyrosine kinases. For example, some evidence has shown EGF is capable of activating pre-existing EGFR dimers [232, 233], and further evidence has shown increased expression of EGFR can stimulate ligand independent dimerization [234]. Even though ligand binding is capable of stimulating dimerization through conformational shifts, EGFR dimerization is entirely mediated by the extracellular domains [235, 236]. Furthermore, EGFR seems to differ in that the

receptor doesn't require trans-autophosphorylation to phosphorylate and open the active domain in the C-terminal tail. Instead, the intracellular region of EGFR contains a C-lobe and an N-lobe. Once dimerized, the C-lobe is capable of swinging around to connect with the N-lobe allowing a disruption to the auto-inhibited state, and an active conformation to be taken [237].

After the active conformation is taken, phosphorylation can occur on the many tyrosine residues in EGFR's c-terminal tail [238–240]. Interestingly, various ligands seem capable of inducing differential tyrosine phosphorylation and various downstream signaling pathways [241, 242]. Gene mutations are also capable of inducing EGFR's active state, and these mutations are common in multiple cancers. Common mutations leading to constitutively active EGFR in non-small cell lung cancer (NSCLC) include a deletion in exon 19, and the L858R point mutation [243, 244]. EGFR stimulation can lead to eventual transcription of numerous genes, from immediate early genes such as the transcription factors FOS and JUN within minutes, to secondary late response genes over 120 minutes after stimulation [245]. After signaling, EGFR is internalized and returned to the cell surface or marked for degradation [246, 247]. Some research has indicated the cell's 'decision' process involving EGFR internalization is pH dependent [248].

EGFR has also been seen in the nucleus of regenerating liver tissue [249], and various cancers including ovarian and bladder [250, 251]. Furthermore, EGFR has been found to act as a transcriptional activator via direct binding to A/T-rich sequences (ATRS) in the promoters of certain genes, such as cyclin D1 [252]. Nuclear EGFR is also capable of acting as a co-activator through interactions with transcription factors, such as STAT3, which recruits nuclear EGFR to the *iNOS* gene [253]. This has led to nuclear EGFR having prognostic value for a variety of cancers, including breast and non-small cell lung cancer [254, 255]. Overall, EGFR is extensively involved in cancer progression through a variety of mechanisms. Its importance is illustrated by the successful treatment of EGFR mutant cancers with tyrosine kinase inhibitors, which will be discussed in more detail further below.

PHOSPHATASES

Just as RTKs are responsible for propagating phosphate signaling within the cell, phosphatases are responsible for regulating these signaling pathways through the removal of phosphate groups from target residues. While conventional wisdom suggested kinases were the most important aspect regarding intercellular signaling, phosphatases are just as important in that regard. Since some of the earliest work on tyrosine phosphatases [256–260], the field has expanded rapidly as a sign of appreciation for how important these proteins are in the regulation of cellular pathways.

Typically, phosphatases are broadly classified into two groups, including serine/threonine phosphatases and tyrosine phosphatases. These groups are further delineated into a number of classifications dependent on the subcellular location and substrate specificity of the phosphatase. Here I will focus more on protein tyrosine phosphatases (PTPs), and more specifically receptor like PTPs (RPTPs). RPTPs largely consist of a variable extracellular region, transmembrane domain, and largely conserved intra-cellular phosphatase domain [261]. Often, the extracellular regions of RPTPs are comprised of a number of immunoglobulin-like or fibronectin type III domains, which are thought to mediate substrate binding and cell-cell contacts [261]. The intracellular phosphatase domains of RPTPs consist of the highly conserved HC-(X_5)-R motif responsible for catalytic activity, as well as nine other conserved motifs that play a role in selectivity and catalysis of target substrates [262]. Many PTPs contain a cleft within the conserved catalytic motif that is responsible for recognition of phosphorylated tyrosine [263]. This cleft is too deep for phosphorylated serine and threonine residues, which is thought to mediate the selectivity of PTPs for pTYR. Catalysis of phosphorylated tyrosine occurs during a two step chemical process involving a conformational shift of the PTP active site, which makes the PTP catalytically competent [261].

While PTPs are generally thought to shut down pathway signaling, their impact is entirely context dependent. In fact, the regulation of certain pathways by PTPs can result in activation or repression in an entirely context dependent manner. In the case of the RPTP CD45, dephosphoylation of SRC results in

activation of signaling downstream of SRC, rather than an inhibition of the pathways [264, 265]. These context dependent processes complicate the narrative of PTPs, allowing them to be viewed as having tumor suppressor or oncogenic properties depending on their cellular location and target pathways.

PROTEIN TYROSINE PHOSPHATASE RECEPTOR TYPE H

PTPRH, otherwise known as Stomach Cancer-Associated Phosphatase 1 (SAP-1) is a member of the receptor like protein phosphatases. Like many other RPTPs, PTPRH has an extracellular region consisting of fibronectin domains, a transmembrane domain, and an intracellular phosphatase domain. The structure of PTPRH is largely conserved between humans and mice, with humans having eight fibronectin domains and mice having six [266]. PTPRH was first cloned in the early 1990's (as SAP-1) from human gastrointestinal cancers, and much of its characterization has been in that context [267]. Like CD45, PTPRH is capable of activating pathways downstream of SRC in a context dependent manner [266]. Interestingly, PTPRH becomes inactive upon dimerization, which is regulated by the extracellular fibronectin domains [266].

While the literature base for PTPRH is small, it was found to be a regulator of EGFR through a screening approach in 2017 [268]. Within that study, Yao et. al found PTPRH to be a negative regulator of EGFR, specifically at EGFR tyrosine residue 1197. PTPRH deficient ovarian cell lines were also found to have ERK activation downstream of EGFR [269]. Overall, this indicates *PTPRH* mutations may contribute to deregulated cellular pathways within tumors, through multiple mechanisms.

RESULTS

DISCOVERY OF *PTPRH* MUTATIONS IN MOUSE PYMT TUMORS

Previous research in the lab discovered a *Ptprh* mutation within the mammary tumors of PyMT FVB mice [193]. Upon targeted resequencing, 81% of PyMT mice (n = 45) were found to have a conserved V483M mutation within *Ptprh*. Further addition of 22 samples to this dataset found this ratio held, with *Ptprh* mutations occurring in 82% of PyMT tumors (Figure 2.1A). Interestingly, mammary tumors that

arose within the same mouse had the same pattern of *Ptprh* mutations, so if one tumor from mouse A had a heterozygous *Ptprh* mutation, other tumors from that same mouse had heterozygous mutations as well (Table 2.1). While a mechanism for this has yet to be explored, previous work has ruled out these mutations being germline [193]. Previous analysis of whole exome sequencing (WES) data acquired from a collaborator showed *Ptprh* mutations occurred throughout the *Ptprh* exome in PyMT mice of various backgrounds other than FVB. This is in contrast to FVB mice, where the mutation in *Ptprh* always results in a valine to methionine shift at amino acid 483 (Figure 2.1B). Interestingly, analysis of the WES data also found *Ptprh* mutation status to be conserved between primary tumors and their metastasis (Figure 2.1C). A student's T-test found no statistical difference ($p = .39$) when comparing the number of exonic mutations in primary tumors, to exonic mutations in metastatic tumors. These data suggest that when *Ptprh* mutations occur, they occur early within the primary tumor progression.

PTPRH **MUTANT TUMORS CORRELATE WITH HIGH EGFR ACTIVITY**

As mentioned above, PTPRH has known interactions with the epidermal growth factor receptor. Therefore, we hypothesized PyMT tumors with a mutation in *Ptprh* would have increased phosphorylation of EGFR, specifically at EGFR residue 1197. To correlate mouse *Ptprh* mutations with increased p-EGFR, western blots were run using an antibody specific for 1197-EGFR [193] (Figure 2.2A). These blots show a clear correlation between mutated *Ptprh* and increased phosphorylation of EGFR. In fact, homozygous mutant tumors have an even further increase in p-EGFR than heterozygous mutant tumors. Suggesting a dominant negative mechanism may be occurring. To further explore the relationship of mutated *Ptprh* with signaling pathways downstream of EGFR, western blots for phosphorylated AKT, ERK, and the transcription factor STAT3 were completed using mouse tumor lysates that were wildtype for *Ptprh*, or had a homozygous mutation (Figure 2.3A/B). AKT, ERK, and STAT3 are both important regulators of pathways downstream of EGFR. [270–274]. As you can see, *Ptprh* mutant PyMT tumors have a clear increase in phosphorylated AKT, but not of ERK or STAT3. This suggests mutated *Ptprh* is only responsible

for regulating some of the tyrosine residues on the c-terminal tail of EGFR. Based on our data we believe PTPRH may specifically be targeting EGFR residue 1197, however, previous characterization of tyrosine residues on the c-terminal tail of EGFR has illustrated the complicated nature of these signaling pathways, and it may not be as simple as PTPRH targeting a single residue.

DISCUSSION

Through whole genome sequencing of PyMT mammary tumors from FVB mice, we have uncovered a conserved V483M mutation within the *Ptprh* gene. This gene was found mutated in 82% of tumors (n = 67) and was determined not to be germline. Further analysis of WES data found a conservation of *Ptprh* mutations within primary mammary tumors and their matched metastasis, suggesting *Ptprh* mutations occur within the early stages of tumor progression. Correlative western blot analysis found increased phosphorylation of EGFR at residue 1197, as well as increased phosphorylation of AKT further downstream, but not of ERK or STAT3.

It may be prudent in the future to determine if in fact *Ptprh* mutations are occurring early within tumor formation. This may give insight as to whether PTPRH can be a driving force of tumor progression. 82% of tumors harboring mutations in *Ptprh* also begs the question of whether this mutation is selected for in this particular oncogenic model. While the above questions were not addressed within this work, the answers could provide beneficial insight into the role of PTPRH in PyMT carcinogenesis.

We have yet to uncover a mechanism behind the failure of mutant *Ptprh* dephosphorylating EFGR, however heterozygous mutants resulting in increased p-EGFR suggest the mechanism may be dominant negative. Furthermore, dimerization of PTPRH is known to cause loss of activity, suggesting a mutation in the fibronectin domain could lead to increased ability of PTPRH to bind to itself. Uncovering this mechanism in future work, potentially through a series of co-immunoprecipitations and other biochemical assays, could lead to valuable insight as to how the V483M mutation impacts PTPRH's ability to dephosphorylate EGFR.

With EGFR being a well know regulator of numerous cellular signaling pathways, *Ptprh* mutant tumors could have deregulated cellular growth dynamics through some of these pathways. In fact, we saw increased phosphorylated AKT within *Ptprh* mutant tumors that also had increased p-EGFR. AKT is an important regulator of pathways leading to increased cellular proliferation and evasion of pro apoptotic signals, so PTPRH mutations resulting in increased AKT activation could result in increased cellular proliferation and enhanced tumor growth. In fact, this has been noted in PyMT tumors [193], which is a striking phenotype given the already fast growth of PyMT tumors. A further exploration into how increased p-AKT is linked to increased –EGFR would be interesting. That mechanism could occur through canonical signaling mechanisms, such as through GRB2 and PI3K and intermediaries, or perhaps through another mechanism. Overall, further exploration V483M mutant *Ptprh*'s contribution to tumorigenesis in PyMT mice would provide valuable insights into phosphatase biology.

MATERIALS AND METHODS

TARGETED RESEQUENCING OF PYMT TUMORS

DNA was extracted from flash frozen tumors using lysis buffer (50 mL Tris HCl, 5 mL 500 mM EDTA, 10 mL 10% SDS, 20 mL 5M NaCl, H20 up to 500 mL), or FFPE tissue using Qiagen FFPE extraction kit. The region flanking V483M was PCR amplified using the following primers, F = GGCCTTAGGTTCAATTGTGAATAC, R = CCTTAGCTTCCCGAGTATTGGTT. Amplified DNA was sent to GeneWiz for Sanger sequencing with the following primer TCATCCAAACTACATCTATGATCCA. Geneious software was used for alignment to reference DNA.

ANALYSIS OF *PTPRH* MUTATIONS IN WES DATA

Pre-annotated VCF files were downloaded for 64 tumors from GEO ascension number GSE142387. Data was processed within R by reading in VCF files, then filtering to only keep mutations within the Chr 7 bp 4548992 – 4604041 range (location of *Ptprh* in mouse genome). These files were then converted to Annovar format, exported, and annotated using Annovar. Statistical analysis was completed using a

student's t test (unequal variance, 2 tailed) between the metastasis group (mutations per met sample), and the primary group (mutations per primary tumor).

WESTERN BLOTTING

Tumor lysates were harvested from flash frozen tumors by crushing with a mortar and pestle, then dissolving in TNE lysis buffer (5 mL 1 M Tris HCl pH 8, 3 mL 5M NaCl, 1 mL NP40, 400 uL .5M EDTA, 2.0 mL .5M NaF, H2O to 100 mL). Roche mini protease tablets and sodium orthovanadate were used and protease and phosphatase inhibitors respectively. Sample concentrations were read using BCA assay, and were diluted to same concentration using extra lysis buffer. SDS was added and samples were heated to 95C for 10 min. Samples were loaded onto an 8% gel and run for ~2 hours, then transferred onto .45 uM PVDF at 70 volts for 2 hours. Blocking occurred for 1 hour at room temp in 5% BSA. Primary antibodies were incubated overnight in blocking buffer. Blots were rinsed with TBST and incubated at room temp with secondary for 1 hour before being rinsed and imaged again. Antibodies were as follows; total EGFR (Cell sig. D38B1), p-EGFR (Invitrogen PA5-37553), AKT (Cell sig. 11E7), p-AKT (Cell sig. D9E), STAT3 (cell sig. 79D7), p-STAT3 (D3A7), B-Tubulin (Proteintech 10094-1-AP), Vinculin (E1E9V).

APPENDIX

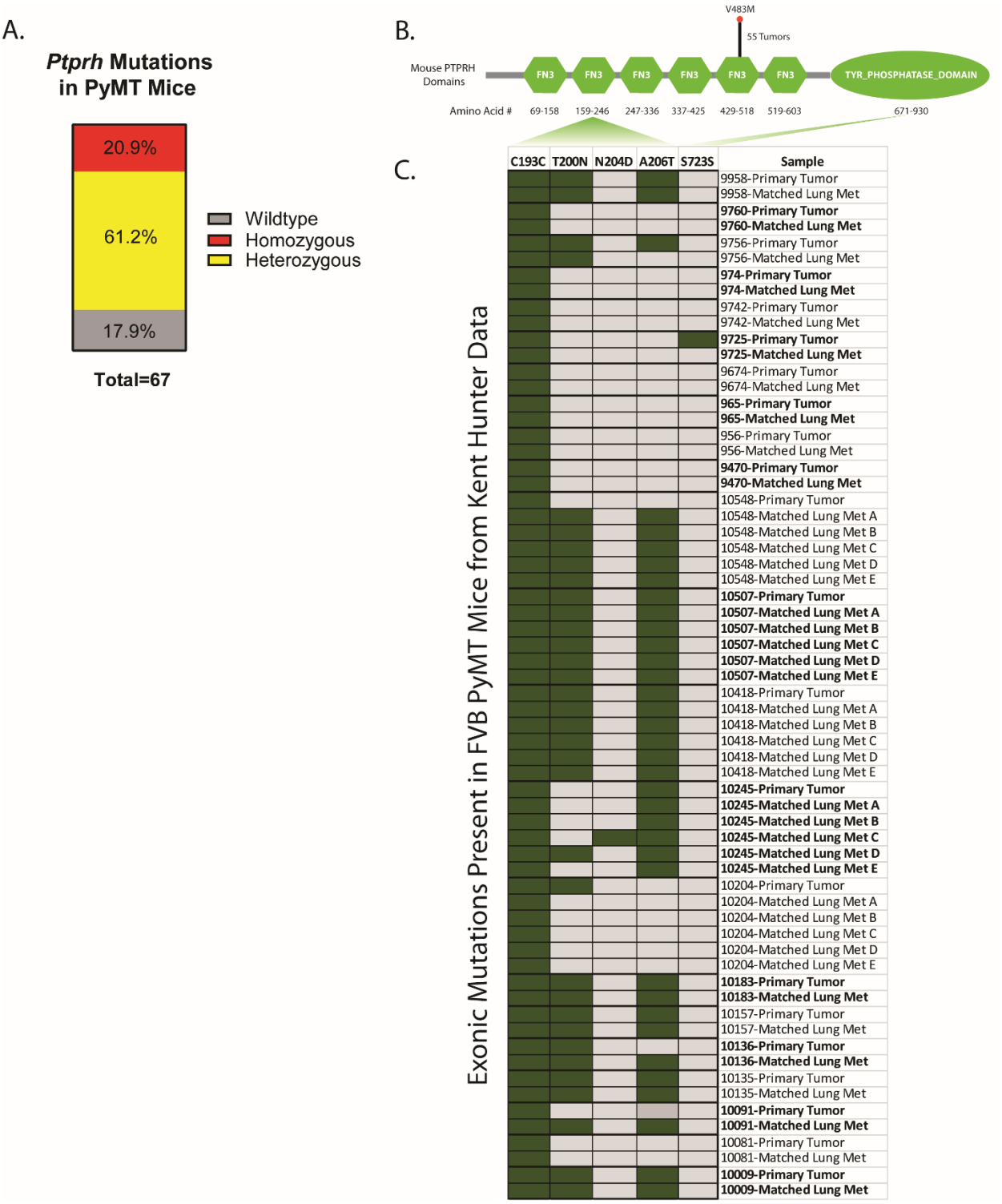


Figure 2.1: *Ptprh* mutations in PyMT mouse tumors

A) *Ptprh* V483M mutation frequency seen in PyMT mammary tumors of FVB background mice. B) Lollipop plot of PTPRH exome showing location of V483M mutation within the predicted PTPRH fibronectin

Figure 2.1 (cont'd)

domain. C) Table of *Ptprh* exonic mutations seen in primary PyMT FVB tumors and their matched metastasis. WES data obtained from a collaborator.

A.

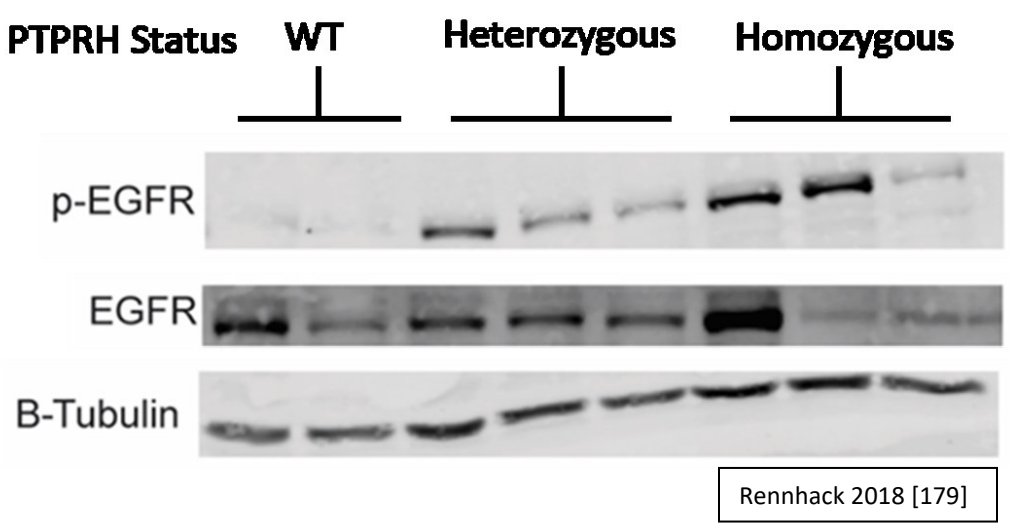


Figure 2.2: Increased p-EGFR in *Ptprh* mutant mouse tumors

A) Increase in phosphorylated 1197 EGFR seen in heterozygous and homozygous *Ptprh* mutant PyMT mouse tumors.

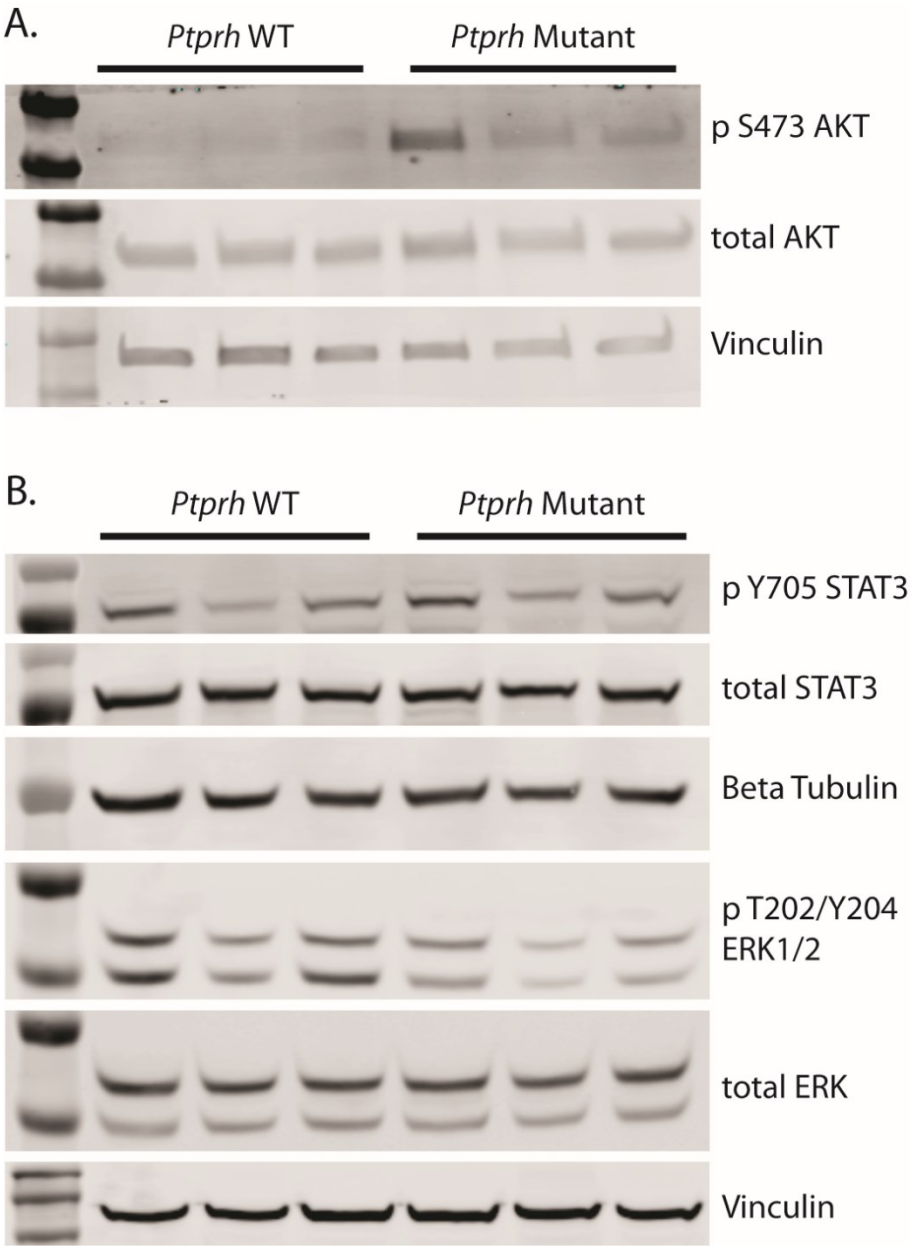


Figure 2.3: Downstream pathway activity in *Ptprh* mutant mouse tumors

A) Increase in phosphorylated S473 AKT seen in homozygous *Ptprh* mutant PyMT mouse tumors. B) No differences seen in phosphorylated ERK1/2 or Y705 STAT3.

Mouse #	Tumor # (mammary gland #)	*Ptprh* Status
2139	2	Heterozygous
2139	6	Heterozygous
273	1	WT
273	5	WT
273	6	WT
273	8	WT
274	3	WT
274	7	WT
274	8	WT
2831	4	Homozygous
2831	5	Homozygous
2831	6	homozygous
300	3	Heterozygous
300	7	Heterozygous
300	8	Heterozygous
300	9	Heterozygous
3146	6	Homozygous
3146	7	Homozygous
3304	1	Heterozygous
3304	8	Heterozygous
3720	2	Homozygous
3720	8	Homozygous
379	2	WT
379	5	WT
455	1	Heterozygous
455	2	Heterozygous
455	6	Heterozygous
456	1	Heterozygous
456	2	Heterozygous
456	3	Heterozygous
456	4	Heterozygous
547	2	Homozygous
547	5	Homozygous
547	6	Homozygous
563	1	Heterozygous
563	2	Heterozygous
563	4	Heterozygous
563	6	Heterozygous
592	1	Heterozygous
592	3	Heterozygous
616	1	Homozygous
616	2	Homozygous
618	1	WT
618	4	WT
618	6	WT
628	5	Heterozygous
628	6	Heterozygous
693	3	Heterozygous
693	4	Heterozygous

Table 2.1: Mammary gland *Ptprh* mutation status in PyMT mice

Table 2.1 (cont'd)

Ptprh mutation status is conserved amongst different mammary gland tumors from the same mouse.

CHAPTER 3

RELATIONSHIP OF PTPRH AND EGFR IN HUMAN CANCER

ABSTRACT

While mouse models of cancer can be beneficial tools for studying the disease, not all genomic mutations found within mouse tumors are relevant to human tumor development. Here we investigate the importance of *PTPRH* mutations in human cancer, finding that 5% of NSCLC cases have mutations within *PTPRH*. Many of these mutations are predicted to have increased EGFR activity, and activation of the PI3K/AKT pathway downstream of EGFR. We show PTPRH ablation through CRISPR leads to increased phosphorylation of EGFR, as well as AKT. A phosphorylated receptor tyrosine kinase array also discovered other RTKs potentially targeted by PTPRH, including a confirmed increase in phosphorylated FGFR1 upon loss of PTPRH. Interestingly, *Ptprh* mutant mouse tumors and PTPRH KO lung cancer cells also display increased EGFR localization to the nucleus of cells, which has been noted in other cancers and regenerating liver tissue.

INTRODUCTION

Previous data found a *Ptprh* mutation within PyMT mammary tumors, with these tumors exhibiting increased p-EGFR and p-AKT as compared to *Ptprh* WT tumors. While this data was striking, it does not show whether *PTPRH* mutations are relevant within human cancers. Genetic aberrations found within mouse models of cancer are not always applicable to human forms of the disease [275, 276]. This is especially the case for certain oncogenic drivers in mice, such as the PyMT oncogene used to drive carcinogenesis within the PyMT model [277]. While tumor induction within the PyMT model relies on the activation of certain pathways known to be important for carcinogenesis, such as Pi3K/AKT, the main oncogenic driver (PyMT) is not found within human cancers.

Determining whether genetic mutations found in mouse tumors are applicable to human cancers can also be complicated by a large number of passenger mutations, whose effect on tumor progression can be ambiguous [278]. Sorting through mutations found via whole genome sequencing is often completed by applying numerous filtering steps, including but not limited to the following;

1. Annotating variants to determine their coding classification (nonsynonymous, etc.)
2. Correlating particular mutations to survival data or another phenotype across multiple samples
3. Analyzing human datasets to determine whether the mutation is present in human tumors
4. Cross referencing mutation lists with known oncogenes and tumor suppressor genes
5. Using pathway or gene set databases (such as Gather) to find relationships between lists of mutated genes
6. Pairing mutations with transcriptomic data to check for a corresponding alteration in gene expression
7. Determining whether the mutation results in a potential protein conformational shift

Overall, the resources available to aid in determining whether a particular mouse model gene mutation is relevant are vast, and numerous resources should be combined to shift through the noise occurring within the mutational landscape of mouse model tumors. In this chapter, some of the listed resources are applied to show the relevancy of PTPRH to human cancers. The relationship between PTPRH and EGFR is also flushed out.

RESULTS

PTPRH MUTATIONS IN HUMAN CANCER

To determine whether *PTPRH* mutations were present within human tumors, data from The Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC) were analyzed taking a pan-cancer data mining approach. Initial analysis of these two data collections showed high rates of *PTPRH* mutations in skin, uterus, and lung cancers (Figure 3.1A). Interestingly, when analyzing data from ICGC, the percentage of patients with *PTPRH* mutations within the same cancer type was noted to be variable across datasets from different countries. For instance, a higher percentage of melanoma patients in Australia were noted to have *PTPRH* mutations than in the United States. A closer look at the individual datasets however, revealed differences in data processing and reporting that accounted for most of the mutation percentage discrepancies. When focusing on lung cancer however, it was noted that a higher percentage of patients in South Korea had *PTPRH* mutations as compared to the United States (Figure 3.1B). This analysis only considered exonic PTPRH mutations.

Because of the known relationship between PTPRH and EGFR, we decided to focus more closely on *PTPRH* mutations within non-small cell lung cancer (NSCLC) patients, since *EGFR* activating mutations occur in a large subset of those patients. This would give us a patient group that has already been characterized in the context of increased EGFR signaling and treatment with EGFR inhibitors. With PTPRH known to target EGFR, we hypothesized a mutation in *PTPRH* could lead to increased EGFR signaling in patient tumors that have no canonical activating mutations in EGFR. Importantly, we see that NSCLC

patients with mutations in *PTPRH* are mutually exclusive from NSCLC patients with activating mutations in *EGFR* (Figure 3.1C). This means the subset of patients with *PTPRH* mutations could have increased activation of EGFR, but are not classified as such and are therefore missing out on potentially efficacious EGFR therapies. Analyzing the TCGA NSCLC dataset for potential discrepancies in age, overall survival, sex, or race found no statistical differences (Figure 3.1D). Interestingly, while *EGFR* mutant lung cancers are not typically associated with smoking, *PTPRH* mutant tumors have previously been associated smoking [279].

BIOINFORMATICS PREDICTS ACTIVATION OF EGFR AND DOWNSTREAM PATHWAYS

To further explore whether *PTPRH* mutations in NSCLC tumors lead to increased EGFR activity, a number of bioinformatics predictions were used. First, we predicted EGFR activity in *PTPRH* mutant NSCLC tumors using single sample gene set enrichment analysis (ssGSEA) on RNA-sequencing data from these tumors (Figure 3.2A). This analysis showed certain *PTPRH* mutant tumors had predicted high EGFR activity. In fact, there three 'hotspot' regions within the PTPRH exome where *PTPRH* mutations were predicted to have increased EGFR activity. Two of these hotspot regions occur within PTPRH fibronectin domains where we also discovered our mouse *Ptprh* mutation, and the third region occurs within the phosphatase domain. Interestingly, phosphatase domain *PTPRH* mutations with predicted high EGFR activity are located just downstream of the conserved $HC(X_5)R$ activity motif, but not within the motif.

With correlative predictions showing increased p-EGFR in *PTPRH* mutant lung cancer tumors, we wanted to determine whether pathways downstream of EGFR were also being impacted within those tumors. To begin, ssGSEA was completed on 12 tumors in each of the three groups; *PTPRH* mutants with predicted high EGFR from the previous GSEA analysis, *EGFR* L858R mutants, and tumors WT for both *PTPRH* and *EGFR*. The pathway predictions from ssGSEA were then clustered into a heatmap using hierarchical K-means clustering (Figure 3.2B). This analysis showed certain *PTPRH* mutant tumors to cluster with *EGFR* mutant tumors, suggesting they have a similar pathway activation profile. To further

investigate pathways downstream of EGFR, GSEA was completed on tumors the same 12 tumors used for the ssGSEA and pathway clustering. GSEA showed predicted activation of the PI3K/AKT pathway, which matches the increased p-AKT seen in *Ptprh* mutant mouse tumors (Figure 3.2C).

PTPRH TARGETS EGFR IN HUMAN LUNG CANCER LINE

With bioinformatics analysis showing *PTPRH* mutations occurring in 5% of NSCLC tumors and predicting activation of EGFR and EGFR pathways within those tumors, we wanted to determine whether non-functional PTPRH could indeed lead to activated EGFR. CRISPR knockouts were created in the H23 NSCLC cell line, targeting exon four of PTPRH. Sanger sequencing of some CRISPR clones confirmed a disruption to *PTPRH* sequence a few base pairs upstream of the PAM sequence, where an adenosine insertion occurred (Figure 3.3A). Adenosine insertion leads to truncation of the PTPRH mRNA through multiple early stop codons.

While we struggled to find a working antibody for PTPRH, we used Y1197 phosphorylated EGFR as a screen for determining the effectiveness of PTPRH knockout. Increased Y1197 phosphorylation was seen in PTPRH KO clones harboring a disruption at the cut site, but not in clones without the disruption (Figure 3.3B). To determine whether expressing PTPRH within the PTPRH KO clones could rescue the increased p-y-1197 EGFR phenotype, wild type PTPRH was transiently expressed within one of the PTPRH KO clones (Figure 3.3C). This resulted in a decrease of phosphorylated tyrosine at EGFR site 1197. Overexpressing a catalytically dead version of PTPRH did not rescue the increased phosphorylation of EGFR tyrosine 1197 (Figure 3.3D). Overall, these analysis show PTPRH is responsible for dephosphorylating EGFR at tyrosine residue 1197.

To determine whether there were increases in p-AKT, p-STAT3, and p-ERK within PTPRH KO cells, western blots were completed using lysates from the same PTPRH KO CRISPR clones that showed increased p-EGFR. Interestingly, the same pattern of phosphorylation seen in mouse tumor lysates occurred within human cell line lysates (Figure 3.4A). PTPRH KO clones showed increased p-AKT, but no

increases in p-STAT3 or p-ERK. It was noted however, that one CRISPR clone did not have the same increase in p-AKT, suggesting the possibility of clonal effects. To determine whether clonal effects were indeed occurring, a PTPRH KO CRISPR clone with high p-AKT was subjected to CRISPR homologous recombination repair to yield a Y1197F mutation in EGFR. Y1197F mutants were confirmed through the addition of an ECORI cut site, as well as Sanger sequencing. CRISPR repair yielded two mutant clones, one with a heterozygous mutation at Y1197, and one with a homozygous mutation. Western blots completed on lysates from Y1197F mutant clones show decreases in p-EGFR and p-AKT as compared to the parent cell, confirming the increase in p-AKT was indeed due to the loss of PTPRH (Figure 3.4B/C).

TARGETING OF OTHER KINASES BY PTPRH

Certain phosphatases are known to have multiple targets. To determine whether loss of PTPRH may impact other kinases within H23 cells, a human receptor tyrosine kinase array was completed. A membrane arrayed with RTK antibodies for specific phosphorylation sites was incubated with either H23 WT lysate, or H23 PTPRH KO lysate. The membrane was then incubated with biotinylated antibody followed by labelled streptavidin. Numerous RTK's were found to have different phosphorylation profiles between the membranes incubated with WT lysate as compared to PTPRH KO lysate (Figure 3.5A). After quantifying the signals, two RTKs in particular had increased phosphorylation on the PTPRK KO blot as compared to the WT blot. These were fibroblast growth factor receptor 1 (FGFR1), with an approximate 3.5 fold increase, and insulin like growth factor 1 receptor (IGF1R) with an approximate 2.4 fold increase. Western blots were completed to confirm increased phosphorylation of FGFR1 in H23 PTPRH KO cells (Figure 3.5B). Indeed, a substantial increase in phosphorylated FGFR1 was seen in PTPRH KO cell lysate as compared to WT lysate, when looking at the 145 KD band.

To predict whether FGFR1 and IGF1R may have increased signaling within *PTPRH* mutant human tumors, we completed the same analysis as above for prediction of EGFR activation in human tumors. ssGSEA was completed to predict pathway activation status of FGFR1 and IGF1R within *PTPRH* mutant

tumors, and pathway activation status was correlated to each sample via a lollipop plot (Figure 3.5C). Interestingly, the same predicted hotspots for EGFR activation seem conserved for FGFR1 and IGF1R activation. In other words, if one of the three kinases are predicted to be active, the other two kinases are most likely predicted to be active as well.

NUCLEAR EGFR WITHIN *PTPRH* MUTANT TUMORS

As mentioned in the introduction of chapter two, EGFR has been noted within the nucleus of cells in times of cellular stress and deregulation. To determine the subcellular location of EGFR within *Ptprh* mutant mouse tumors, immunohistochemistry was completed using an antibody specific for p-y-1197 EGFR (Figure 3.6A). IHC showed vast increases in EGFR staining within *Ptprh* mutant mouse tumors, with EGFR localized to the nucleus. With the above mouse analysis being correlative, we wanted to determine whether loss of PTPRH resulted in increased EGFR translocation to the nucleus. To accomplish this, H23 PTPRH KO cells were injected into the left flank of nude mice. After reaching 8-10mm in size, mice were necropsied and both flash frozen and formalin fixed tumor tissue was harvested. No metastasis were noted in these mice. IHC was completed using a p-y-1197 specific EGFR antibody, and an increase in nuclear EGFR was noted in mouse tumors derived from PTPRH KO cells (Figure 3.6B). These data suggest a failure of PTPRH to dephosphorylate EGFR at tyrosine residue 1197 leads to increased localization of EGFR to the nucleus. Future analysis to determine whether full or partial length EGFR is located within the nucleus, as well as putative targets of EGFR within the nucleus would be highly beneficial.

DISCUSSION

A pan cancer analysis of human *PTPRH* mutations found numerous cancers harboring mutations in at least 5% of patients, suggesting mutated PTPRH may play a role in tumor development for various other cancers. With PTPRH affecting cell signaling pathways in a context dependent manner, it is possible *PTPRH* mutations could have an oncogenic or tumor suppressive effect depending on the cancer site and cell type. *PTPRH* mutations were found in approximately 5% of NSCLC patients, with these mutations

spread across the PTPRH exome. This is an interesting contrast to the conserved V645M mutation found within our PyMT tumors, and has implications for which mutations may be impactful on tumor growth. While a mechanism has yet to be explored for these various mutations, it is possible the mutations are acting in different fashion from each other. Mutations within phosphatase domain may abrogate catalytic activity, while mutations in the fibronectin domains may prevent dimerization and binding of target substrates. Since some of the phosphatase domain mutations with predicted high EGFR activity lie outside the conserved activity $HC(X_5)R$ motif, it is also possible these mutations are occurring within other conserved PTP motfis, and preventing recognition of substrate binding sites. More biochemical analysis will be needed to explore these hypothesis.

In the previous chapter, mouse tumors showed a correlation between mutant *Ptprh* and high phosphorylation of EGFR. This was confirmed in a human NSCLC cell line through CRISPR ablation of PTPRH. Overexpressing WT PTPRH within PTPRH KO cells rescued the increased p-EGFR phenotype, confirming PTPRH does indeed regulate EGFR within this context. Bioinformatics predictions showed predicted activation of the PI3K/AKT pathway, and this was confirmed through western blotting of PTPRH KO clones. Interestingly, phosphorylation of STAT3 (a transcription factor known to be regulated by EGFR) or ERK were not affected by PTPRH ablation. This suggests PTPRH is only regulating certain tyrosine residues on the c-terminal tail of EGFR. A more robust analysis of how other pathways downstream of EGFR may be affected by PTPRH loss would be a prudent next step. In generating Y1197F EGFR mutants within the PTPRH KO clone with higher p-AKT levels, we noted a decrease in phosphorylation of AKT. However that phosphorylation did not reduce completely to wild type levels. It is possible this failure to reduce p-AKT levels to those seen in WT is due to other activated pathways within the PTPRH KO cells, especially since we see increased phosphorylation of other kinases within PTPRH KO cells.

A kinase array showed increased phosphorylation of numerous RTKs within PTPRH KO cells, including FGFR1 and IGFR1. Interestingly, increased phosphorylation of EGFR was not shown on the array.

However, when checking the phosphorylated antibodies used on the blot, tyrosine 1197 site was not included. This is further confirmation that PTPRH is targeting tyrosine 1197 on EGFR, and not other tyrosine sites. As the array was only designed for RTK interactions, other intracellular signaling molecules may have been impacted by loss of PTPRH, but would have been missed. A mass-spec approach may be beneficial in the future, to determine what other signaling molecules may be impacted by loss of PTPRH. Increased phosphorylation of FGFR1 was confirmed through western blotting. This has interesting implications for both cellular pathways that may be affected, as well as potential treatment options for those with non-functional PTPRH. Perhaps a dual drug inhibition approach of targeting FGFR1 and EGFR would be prudent.

Finally, *Ptprh* mutant mouse tumors, and PTPRH KO human tumors implanted in mice have increased staining of nuclear EGFR. Nuclear EGFR has been noted in times of cellular stress, as well as regenerating liver tissue. While in the nucleus, EGFR can act as a cofactor, or direct transcriptional activator by binding to the promoters of certain genes, such as cyclin D1. Increased nuclear EGFR upon loss of PTPRH activity could have profound impacts on cellular signaling pathways. The mechanism behind increased nuclear localization of EGFR has not been explored, but warrants further exploration. It is possible that loss of PTPRH activity leading to increased activation of EGFR could result in increased internalization of EGFR, although this hypothesis would need to be further explored.

One potential caveat to this work is the lack of other cell lines with PTPRH ablation. While the addition of another PTPRH KO cell line would have added robustness to these data, we feel the current data sufficiently demonstrates PTPRH is responsible for regulating EGFR signaling due to two key experiments. First, overexpression of WT PTPRH within leads to reduced phosphorylation of EGFR within PTPRH KO cells, while overexpression of a catalytically dead version of PTPRH does not result in this reduction. Second, heterozygous and homozygous Y1197F EGFR mutants having a step-wise reduction in p-Y 1197 EGFR within PTPRH knockout cells, meaning the heterozygous mutant had some reduction of p-

Y 1197, and the homozygous mutant had a larger reduction in p-Y 1197. Overall, these data suggest PTPRH is indeed responsible for regulating EGFR signaling.

MATERIALS AND METHODS

DETERMINING PTPRH MUTATIONS IN HUMAN CANCERS

Pan-Cancer datasets from numerous sources, including TCGA and ICGC, were analyzed through CBioPortal and the ICGC portal. Lung cancer mutation percentage were analyzed specifically using TCGA 2016 dataset accessed through CBioPortal. The South Korean and U.S datasets showing discrepancy in percentage of *PTPRH* mutations were analyzed on the ICGC portal. Both datasets were filtered to include only patients with exonic mutations.

MUTUAL EXLCUSIVITY

All NSCLC datasets available on CBioPortal were used for this analysis, and are listed below. *PTPRH* and *EGFR* SNV mutation data were downloaded and combined. Duplicate samples were removed, and any sample with a *PTPRH* or *EGFR* mutation was considered. A 2x2 contingency table was run to determine mutual exclusivity. Datasets include; MSK - cancer cell 2018, MSKCC - J clin oncol 2018, TRACERx - NEJM 2017, University of Turnin, 2017, MSK - Science 2015, TCGA - Nat Genet 2016 (Pan), Broad - cell 2012, MSKCC - Science 2015, TCGA - Firehose Legacy, TCGA - Nature 2014, TCGA - Pan-cancer Atlas, TSP - Nature 2008, MSKCC - Cancer Discov 2017, TCGA - Nature 2012

DEMOGRAPHICS OF *PTPRH* MUTATIONS

Age, overall survival, and race demographics were analyzed using the Lung Adenocarcinoma TCGA Pan-Cancer Atlas data set downloaded from CBioPortal. This was one of the few datasets with race data. Two-tailed Student's T-Tests assuming unequal variance were completed for *PTPRH* mutant VS. *EGFR* mutant samples, as well as *PTPRH* mutant VS. WT (non-EGFR mutant) samples for age of diagnosis and overall survival. Samples without age or OS data were excluded. Only samples with missense or

truncating mutations were included, and overexpression samples were excluded. Race was analyzed using a 2x2 contingency table.

EGFR ACTIVITY AND PATHWAY ACTIVITY PREDICTION

TCGA pan-cancer RNA-seq dataset (downloaded from UCSC Xena) was analyzed for PTPRH, EGFR, FGFR1, and IGF1R mutations. This mutation list was downloaded and filtered to keep samples that had a mutation in PTPRH, EGFR, or that were WT for PTPRH, EGFR, FGFR1, and IGF1R. Any sample with a mutation in PTPRH was kept, resulting in 53 samples. 10 samples of each of the two categories were kept; WT for PTPRH and the above three RTKs, and L858R mutant EGFR that were WT for PTPRH, FGFR1, or IGF1R. To decide which WT and EGFR samples to keep, the samples from those subsequent groups were assigned a random number using the RAND() function in excel. These numbers were then sorted from highest to lowest, keeping the top 10 samples. RSEM(log2 X+1) normalization was applied to the filtered sample list, resulting in 47 *PTPRH* mutant samples (WT for the kinases), 9 samples that WT for *PTPRH* and the three kinases, and 8 samples with EGFR mutations (WT for *PTPRH*, *FGFR1*, and *IGF1R*). ssGSEA was run on the samples to predict pathway activation status. Pathways for each kinase were filtered down, selecting the most relevant and robust pathway. In Microsoft Excel, a ranking sum score was applied to the pathway prediction data for each sample using the following formula;

=(B4-MIN(B$4:B$475))/(MAX(B$4:B$475)-MIN(B$4:B$475))

For GSEA analysis of *PTPRH* mutant tumors, the pan-cancer RNA-seq dataset was again downloaded from UCSC Xena. Twelve tumors for each of the three categories were kept; *PTPRH* mutant tumors predicted to have high EGFR activity, *EGFR* L858R mutants, and tumors that were WT for both *PTPRH* and *EGFR*. GSEA was completed using the GenePattern server.

CRISPR KNOCKOUT

Benchling [280] was used to design the guide RNA (AGCACACACTAACATCACCG) targeting the fourth exon of PTPRH. The guide was cloned into px458 using AgeI and EcoRI, and transformed into DH5a.

Transient transfection of px458 into H23 cells was completed using Promega's Viafect. GFP positive cells were sorted into single cell clones into 96 well plates using FACS. Once clones had grown into a colony, they were subsequently moved to 24-well plates, then 6-well plates. DNA was harvested and sent to ACTG for sanger sequencing.

CRISPR KNOCK-IN MUTATION

Guide RNA was designed in Benchling with the PAM (NGG) sequence 5 bp downstream of the desired EGFR a.a. 1197 mutation site. The single stranded region of homology was designed in Benchling by choosing desired length for homology arms as well as the desired mutation, then taking the reverse complement of that strand. The oligo was designed with 36 bp upstream of the desired mutation site and 90 bp downstream. The desired mutation resulting in a Y1197F amino acid substitution was added. Luckily, this mutation also resulted in the addition of an EcoRI cut site, which was used for downstream screening. The mutation also altered the guide RNA enough to prevent re-annealing once HR mediated repair occurred. Guide RNA was cloned into px458 in a manner similar to the CRISPR knockout protocol. For transfection, H23 PTPRH KO cells were seeded at ~85% confluency, then transfected using Viafect in a 6:1 ratio. 1 ug of px458 with guide, and 4 ug of ss repair template were transfected. Sorting was completed using FACS for GFP. Clones were screened using a digest for EcoRI, and confirmed with sequencing.

WESTERN BLOTTING

Blocking was completed at room temperature for one hour, using manufactures recommended buffer. Primary antibody was incubated overnight at four degrees C. Blots were imaged using LiCOR system. Antibodies used were as follows; total EGFR (Cell Signaling D38B1), 1197 EGFR (Invitrogen PA5-37553), total AKT (Cell Signaling 11E7), p-s473 AKT (Cell Signaling D9E), total STAT3 Cell Signaling 79D7 (), p-Y705 STAT3 (Cell Signaling D3A7), total FGFR1 (Cell Signaling D8E4), p-Y653/654 FGFR1 (Cell Signaling 3471s), beta tubulin (Proteintech 10094-1), vinculin (Cell Signaling E1E9V).

OVEREXPRESSION EXPERIMENTS

PTPRH c-DNA within plasmid PRc-CMV was kindly provided by Dr. Takashi Matozaki at Kobe University. Site directed mutagenesis was used to achieve a D986A mutant. 5% DMSO and a 2 minute/kb extension time were used during SDM due to the high GC content of PTPRH. Both WT and D986A mutant *PTPRH* plasmid constructs were transiently expressed in PTPRH KO cells using Viafect. G418 Gentacin was used as a selection marker. Once all control cells were dead, protein lysate was harvested using TNE lysis buffer with protease and phosphatase inhibitors.

RECEPTOR TYROSINE KINASE ARRAY

Protocol for RayBiotech Human RTK Phosphorylation Array C1 kit was followed. Membranes were incubated with lysate from H23 WT cells or H23 PTPRH KO cells. Lysate concentration was read using a Bradford assay, then diluted and read again to ensure accuracy.

IHC NUCLEAR EGFR

Human cell lines H23 PTPRH WT or H23 PTPRH KO were injected into the left flank of nude mice. H23 cell line tumors were grown to approximately 10 mm in the largest direction prior to necropsy. Mouse PyMT tumors, and tumors grown from human H23 cells were necropsied with portions of tumor tissue preserved in formalin, and portions of tumor flash frozen for further downstream analysis. Formalin fixed paraffin embedded tumors were subjected to staining using an antibody specific for 1197 EGFR (Thermo PA5-37553).

APPENDIX

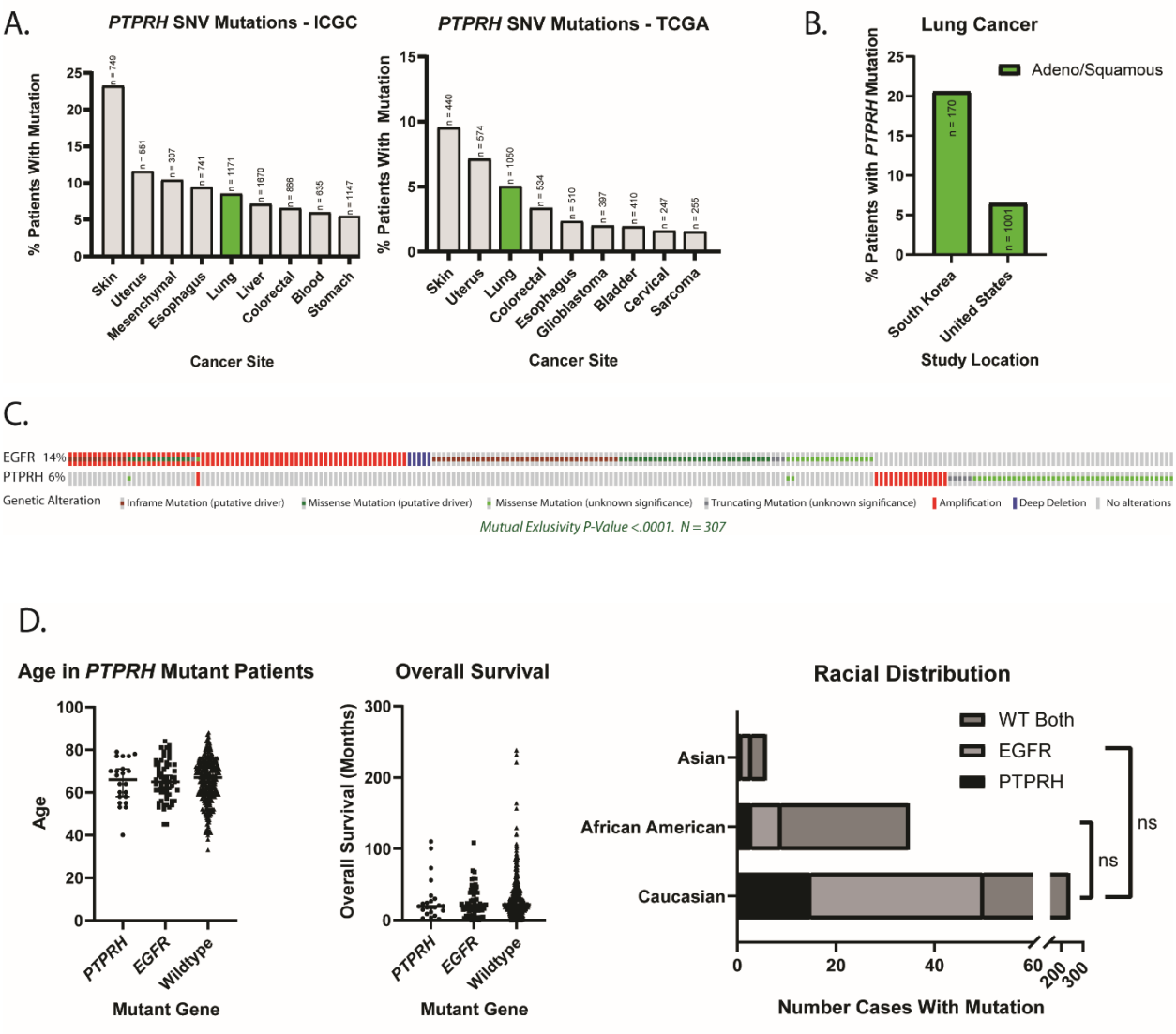


Figure 3.1: ***PTPRH* mutations within human cancers**

A) Pan-cancer analysis using data from ICGC and TCGA shows *PTPRH* mutations present within numerous cancers. Lung cancer is highlighted. B) *PTPRH* mutation rates can vary within NSCLC, depending on study site. C) Oncoplot of TCGA data showing *EGFR* and *PTPRH* mutation rates with NSCLC. Each rectangle represents a patient tumor. *PTPRH* mutations are mutually exclusive from *EGFR* mutations. D) Analysis completed on TCGA data shows no relationship seen between *PTPRH* mutations and age, overall survival, or race.

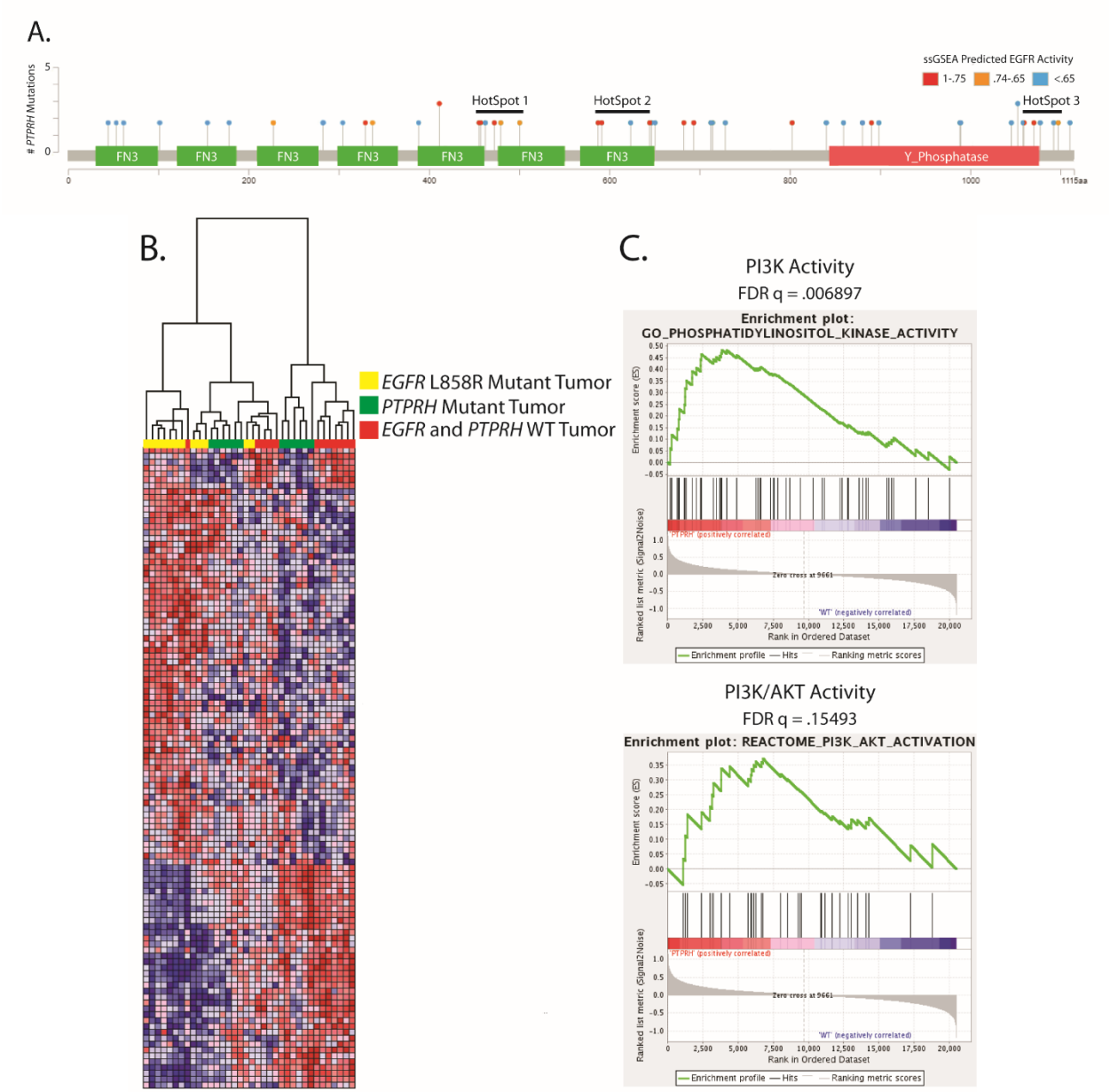


Figure 3.2: Pathway activation predictions in *PTPRH* mutant tumors

A) Lollipop plot correlates predicted EGFR activity with human *PTPRH* mutations. Each dot represents a human tumor with its *PTPRH* mutation corresponding to that location on the PTPRH exome. B) ssGSEA was used to predict gene set enrichment in *EGFR* or *PTPRH* mutant NSCLC tumors. Enriched gene sets were subjected to hierarchical clustering and visualized with a heatmap. C) GSEA predicts activation of the PI3K/AKT pathway downstream of EGFR.

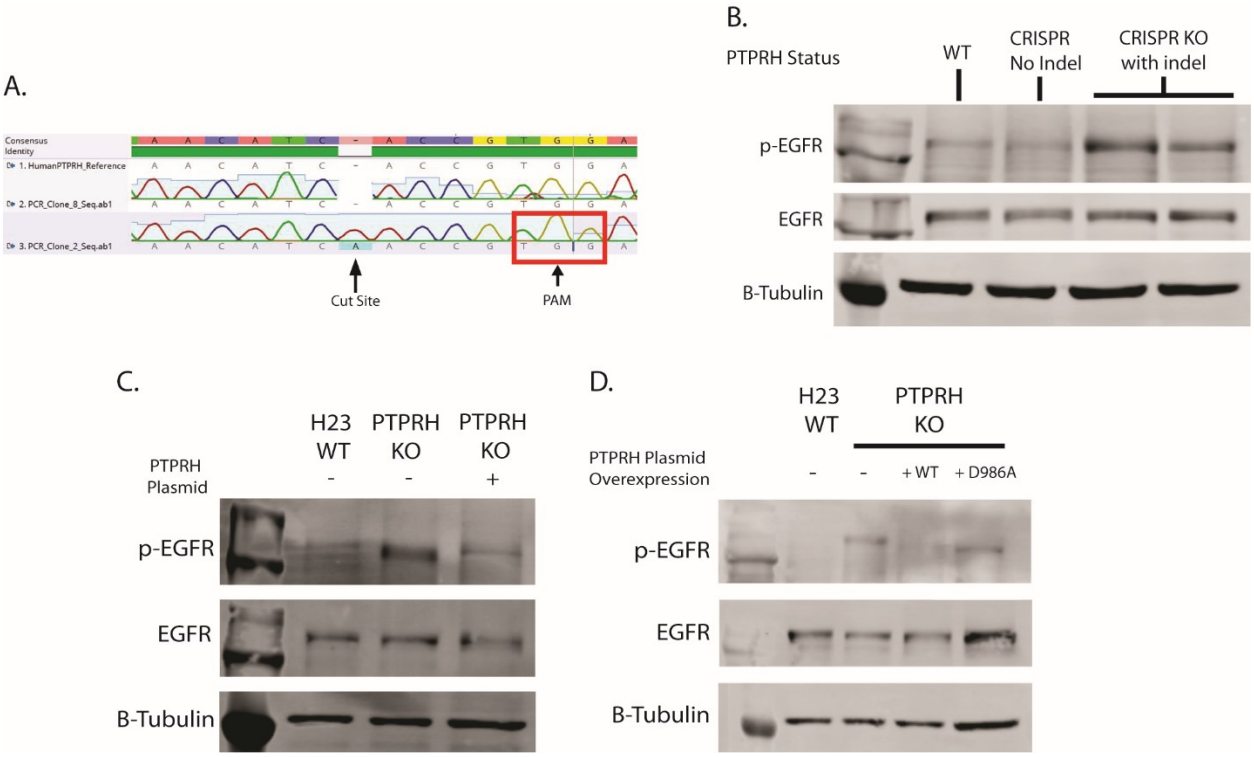


Figure 3.3: PTPRH knockout cells have increased p-EGFR

A) Electropherogram shows indel of A insertion a few base pairs upstream of the PAM sequence within H23 NSCLC cells. B) Western blotting for 1197 p-EGFR shows increased p-EGFR in PTPRH KO cells with indel. Both KO clones had same A insertion seen in electropherogram shown in 3.3A. C) Overexpression of a WT PTPRH plasmid in PTPRH KO cells reduces p-EGFR to WT H23 levels. D) Overexpression of a D986A catalytically dead version of PTPRH does not rescue increased p-EGFR phenotype.

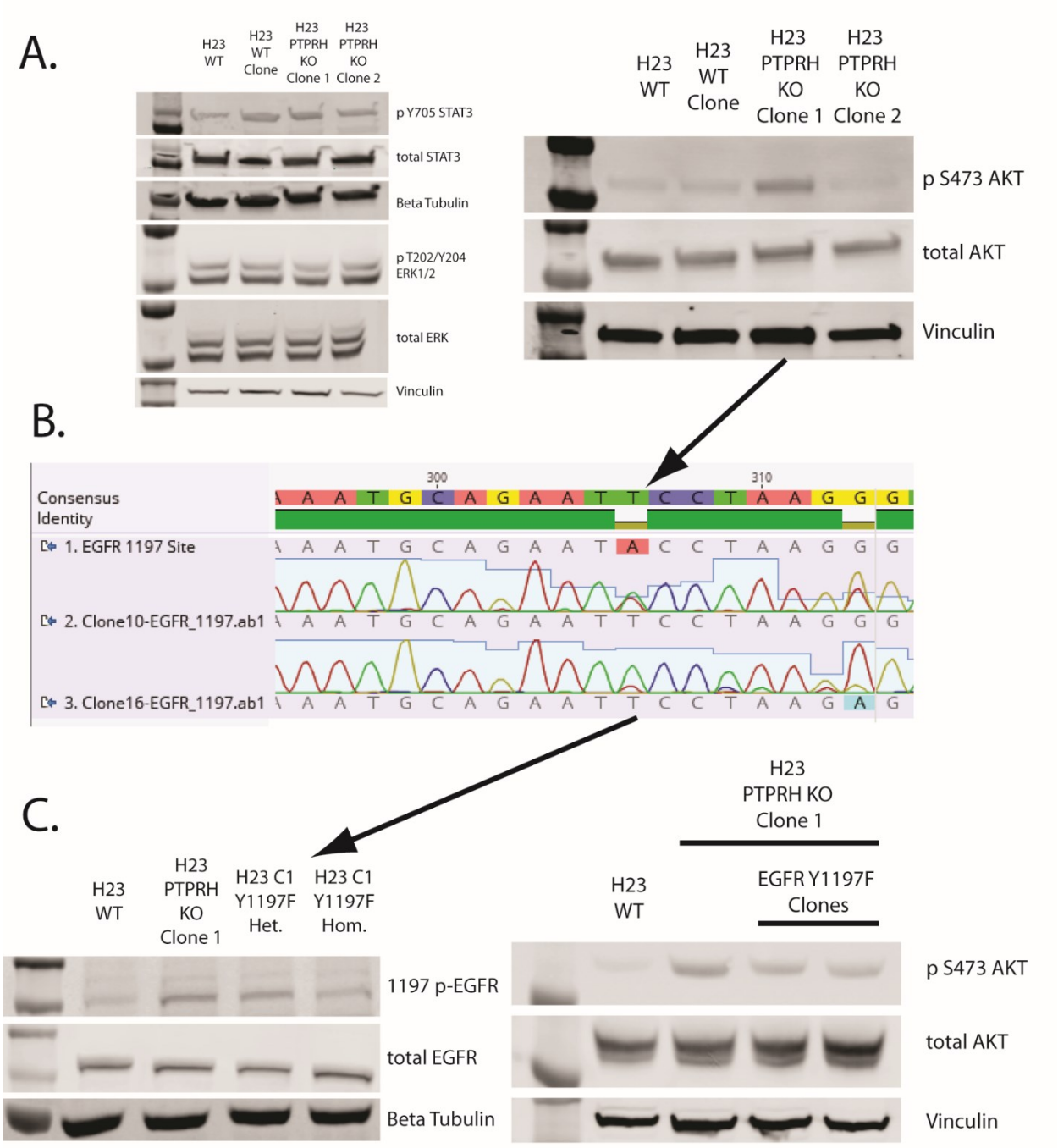
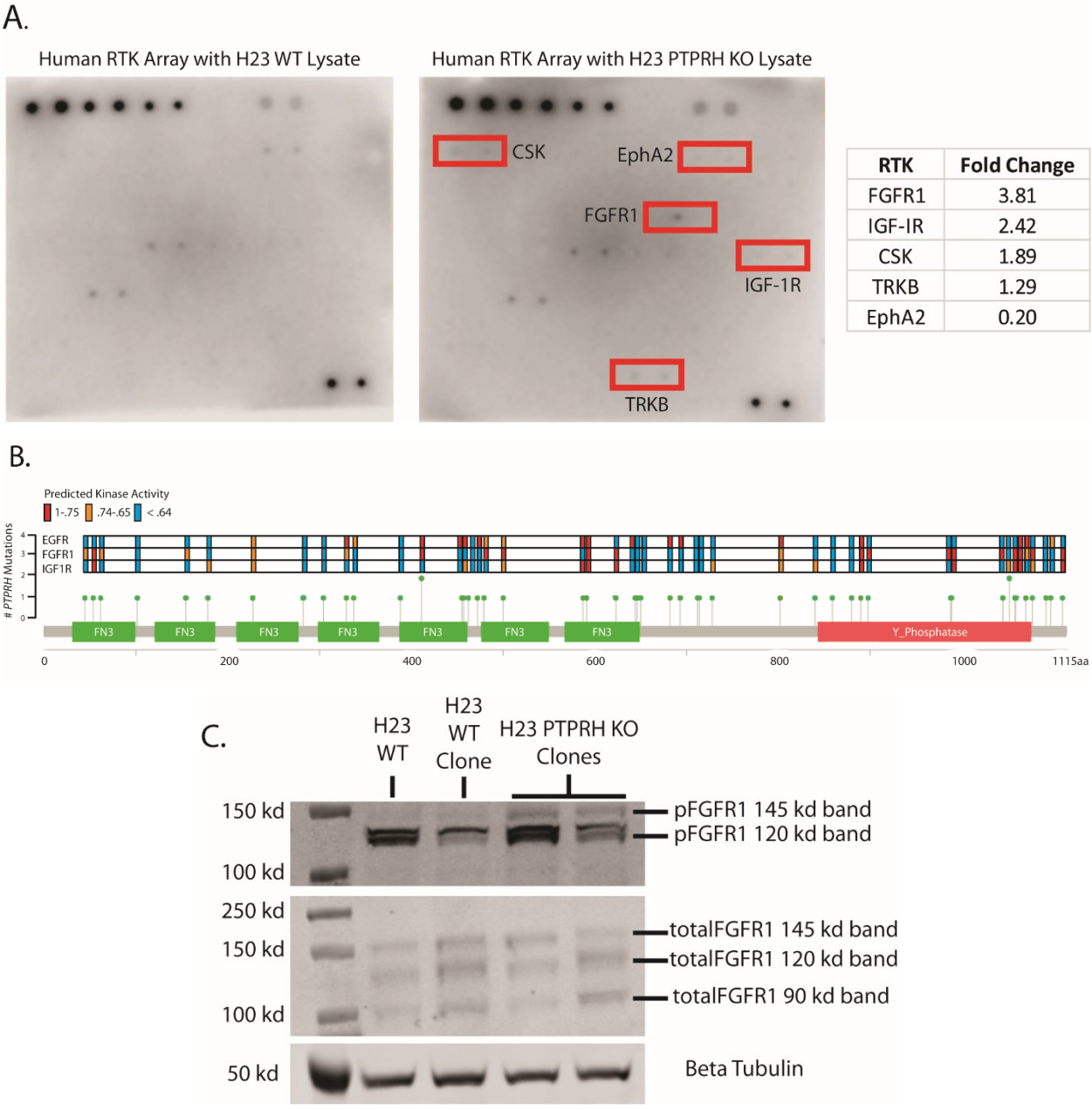


Figure 3.4: Downstream signaling of H23 PTPRH KO cells

Western analysis shows activation of AKT pathway, but not STAT3 in PTPRH KO cells. A) Western blotting shows no increase in p-ERK or p-Y705 STAT3 in PTPRH KO cells, but increased p-S473 AKT in one clonal population. B) Electropherogram of H23 PTPRH KO Clone 1 cells subjected to CRISPR for mutation

Figure 3.4 (cont'd)

of EGFR tyrosine 1197. Clone 16 had heterozygous mutation, and clone 10 had homozygous mutation to achieve tyrosine to phenylalanine amino acid substitution. C) Western blot of H23 PTPRH KO/EGFR Y1197F mutants shows decreased phosphorylation of EGFR at residue 1197, and decreased p S473 when mutating the tyrosine at 1197 to phenylalanine.



RTK	Fold Change
FGFR1	3.81
IGF-IR	2.42
CSK	1.89
TRKB	1.29
EphA2	0.20

Figure 3.5: PTPRH regulates other kinases

A) Human phosphorylated RTK array shows variable phosphorylation of FGFR1, IGF1R, and other kinases between H23 PTPRH KO lysate and H23 WT lysate. B) Lollipop plot showing predicted activity of RTKs in *PTPRH* mutant NSCLC tumors. Hotspot regions are similar to those of the EGFR lollipop plot in figure 3.2A. C) Western showing increased p-FGFR1 in H23 PTPRH KO clones, as compared to PTPRH WT cells.

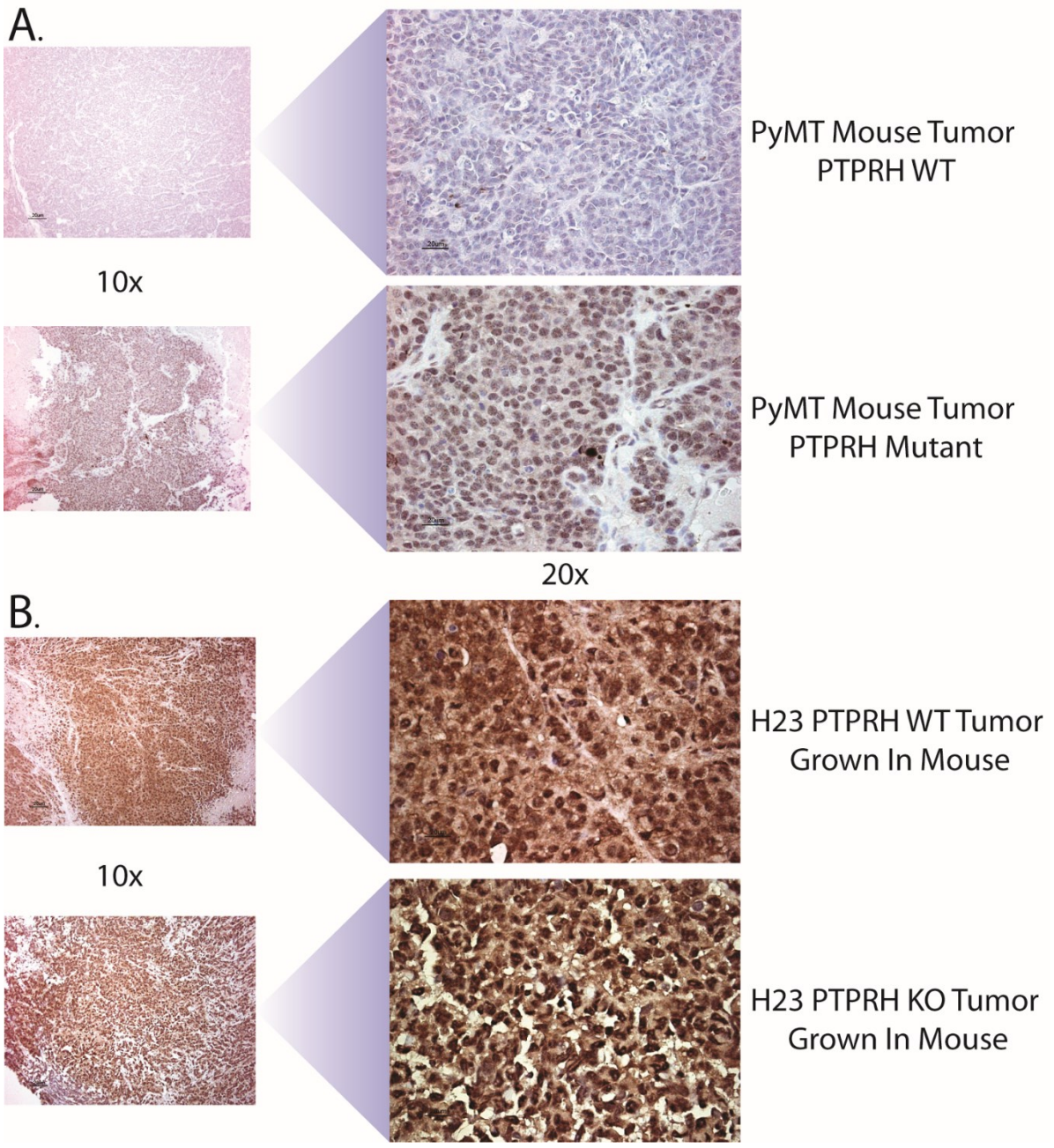


Figure 3.6: Localization of EGFR to the nucleus in PTPRH ablated tumors

Immunohistochemistry using an antibody specific for 1197 p-Y-EGFR shows increased nuclear localization of PTPRH in mouse and human tumors with PTPRH activity loss. A) PyMT tumors with V486M mutation correlate with increased localization of 1197 EGFR to the nucleus. B) H23 PTPRH WT or KO cells were injected into the left flank of nude mice. Tumors grown from H23 PTPRH KO cells have increased EGFR staining within the nucleus.

CHAPTER 4

TREATMENT OPPORTUNITIES FOR *PTPRH* MUTATIONS IN NON-SMALL CELL LUNG CANCER

ABSTRACT

Previous data has shown increased phosphorylation of EGFR upon loss of PTPRH in H23 NSCLC cells. Pooled knockout of PTPRH within H23 cells leads to increased proliferation and cellular growth, suggesting PTPRH loss contributes to tumor growth through EGFR pathway activation. We show *PTPRH* mutant non-small cell lung cancer lines respond to osimertinib treatment *in vitro*, and the H2228 PTPRH mutant cell line responds to osimertinib treatment *in vivo*. Furthermore, treatment of H2228 tumors with osimertinib reduces cellular proliferation as seen through KI67 staining on formalin fixed tumors. Overall, these data suggest *PTPRH* mutant NSCLC patients may benefit from tyrosine kinase inhibitor treatment of EGFR.

INTRODUCTION

PTPRH DEREGULATION IN HUMAN CANCERS

While some phosphatases, such as PTEN [281, 282], have well defined tumor suppressive capabilities, many phosphatases are undefined in the context of cancer. Overall, the importance of cell signaling changes through phosphatase regulation is becoming more appreciated. Even with the literature on PTPRH being sparse, there have been investigations into the roles of PTPRH within some cancers. Expression levels of PTPRH are thought to be low within normal colon epithelial tissue, however increased expression has been seen within severe dysplasia of the colon, and colon cancer [283]. An inverse of this expression profile deregulation is seen within cancers of the liver, where lower PTPRH expression is seen within poorly differentiated hepatocellular carcinomas (HCC) while normal liver tissue has high expression of PTPRH. Furthermore, expression of PTPRH within two HCC cell lines having low PTPRH expression drastically reduced cellular motility and growth rate *in vitro*, suggesting PTPRH has a tumor suppressive role within hepatocellular carcinoma.

While the differing nature of PTPRH expression between colon cancers and hepatocellular carcinomas seems contradictory, it is important to remember PTPRH can affect signaling pathways in a context dependent manner. Loss of PTPRH expression within hepatocellular carcinomas aligns with canonical thinking that phosphatases abrogate downstream signaling of RTKS through removal of phosphate groups, while overexpression of PTPRH in colon cancers highlights the ability of PTPRH to act as an oncogene due to activation of SRC.

Overexpression of PTPRH has been noted in NSCLC, with correlative hypomethylation of PTPRH being suggested as the cause [279]. Furthermore, PTPRH overexpression has been noted as a prognostic indicator for poor survival. This seems to be in contrast to our data, which suggests loss of PTPRH leads to increased oncogenic signaling. On the surface it may seem logical that if loss of PTPRH function leads to increased oncogenic signaling through EGFR, then high expression of PTPRH should abrogate this

signaling. However, there are two important pieces of information that could explain this discrepancy. First, PTPRH function has been shown to decrease upon homodimerization. It is entirely possible that overexpression of PTPRH leads to increased homodimerization through increased contact of PTPRH with itself, although this would need to be further explored. Second, overexpression of PTPRH could lead to increased targeting of other signaling molecules such as SRC. As dephosphorylation of tyrosine residues on SRC activates downstream signaling, this mechanism could also explain the potential discrepancy. Overall, PTPRH deregulation has been noted in numerous cancers.

NON-SMALL CELL LUNG CANCER

Lung cancer accounts for the greatest amount of U.S. cancer deaths in both men and women, and 5 year survival rates remain poor [152]. Broadly, lung cancer is classified into two major histologies, including small-cell (SC) and non-small cell lung cancer (NSCLC). SC lung cancer typically has a poorer prognosis than NSCLC, and is typically associated with smoking. The mutation profile between the two histologies also varies, with SC lung cancer patients typically having mutations in the tumor suppressor genes *Rb* and *Tp53*, and NSCLC patients having mutations in oncogenes *EGFR* and *KRAS*.

Overall, NSCLC accounts for approximately 85% of all lung cancer cases, and is further delineated into three histologies including Adenocarcinoma, Squamous cell carcinoma, and Large cell carcinoma [284]. The prognosis for NSCLC patients is markedly improved compared to that of patients with small cell lung cancer, however prognosis varies widely depending on whether the tumor has metastasized. 5 year survival rates for localized NSCLC approach 63%, but with distant metastasis 5 year survival rates drop to 7% (American Cancer Society). Prognosis is complicated by a number of factors however, including smoking status, *EGFR* mutation status, and initial response to treatment [285].

Approximately 15% of NSCLC patients have tumors presenting with EGFR activating mutations, or amplification of EGFR, however this percentage is substantially higher in Asian patients [286]. 80% of these *EGFR* mutations are putative oncogenic drives, with the vast majority of these mutations being

missense L858R mutations, or a small deletion around amino acids 750. Activating *EGFR* mutations are indicators of responsiveness to tyrosine kinase inhibitors, however this is not the case for tumors with *EGFR* amplification [287]. Overall, patients with activating mutations in *EGFR* have better 5-year survival outcomes, as TKIs are capable of increasing survival time.

TYROSINE KINASE INHIBITORS

Development of drugs targeting PTPs has proved difficult in many cases. This is potentially due to the context dependent nature of many PTPs, as well as their tumor suppressive roles. Targeting a PTP with potential tumor suppressive qualities would result in the opposite of the intended effect. With these difficulties, drugs have been developed to target certain PTPs. Typically, these drugs target PTPs with oncogenic properties that increase cellular pathway signaling. Shp099 is an inhibitor of the protein tyrosine phosphatase SHP2, a PTP with SRC homology like domains known to activate MAPK signaling [288, 289].

With the apparent difficulty of targeting PTPs for drug treatment, targeting PTP substrates may be another viable option. Since we have shown non-functional PTPRH to enhance EGFR signaling, targeting PTPRH mutant tumors with tyrosine kinase inhibitors directed at EGFR may be a viable option. Tyrosine kinase inhibitors are often used to treat NSCLC patients who have tumors presenting with canonical EGFR activating mutations. First generation TKIs, such as erlotinib and gefitinib, were designed to target the ATP binding domain of EGFR. These TKIs successfully enhance progression free survival, however resistance mechanisms eventually develop, usually in the form of a T790M *EGFR* mutation which causes a structural shift and prevents binding of TKIs to the ATP binding domain [290]. Third generation TKIs, such as osimertinib, have been developed to get around this structural inhibition by binding to a nearby cysteine residue. While third generation TKIs are capable of overcoming T790M resistance, new resistance mechanisms eventually develop. Currently, 4^{th} generation TKIs are being developed based on allosteric inhibition of EGFR.

RESULTS

POOLED PTPRH KNOCKOUTS HAVE INCREASED GROWTH

Previous data had shown *Ptprh* mutant PyMT tumors to have decreased tumor latency [193]. With loss of PTPRH function in the human H23 cell line resulting in increased PI3K/AKT pathway activation downstream of EGFR, we hypothesized this may lead to increased cellular proliferation within PTPRH KO cells. To address this, growth curves were completed using H23 PTPRH WT cells, as well as two PTPRH KO clones (Figure 4.1). Growth curves involving clones were ambiguous, with one clone showing clear increased growth, but a second clone growing at a similar rate as the wild type cells. To determine whether clonal effects were responsible for the discrepancy in phenotype, PTPRH pooled knockouts were created in the H23 cell line. Sanger sequencing of pooled knockout cells and subsequent TIDE (Tracking of Indels by Decomposition) analysis showed a knockout efficiency of approximately 45%. Even with low efficiency of knockout, MTT assays and growth curves using PTPRH pooled knockout cells showed increased proliferation and growth of KO cells over wild type cells (Figure 4.2). Overall, these data show loss of PTPRH leads to increased cellular growth.

PTPRH MUTANT CELL LINES ARE SENSITIVE TO TYROSINE KINASE INHIBITION THROUGH OSIMERTINIB TREATMENT

Non-small cell lung cancer patients whose tumors present with EGFR mutations often benefit from tyrosine kinase inhibitor therapy. With loss of PTPRH function leading to increased activation of EGFR and pathways downstream of EGFR, we hypothesized that *PTPRH* mutant tumors would benefit from treatment with tyrosine kinase inhibitors. Previous work showed *Ptprh* mutant PyMT mouse tumors to be sensitive to the TKI erlotinib [193]. To explore whether human *PTPRH* mutations sensitize tumors to TKI therapy, we obtained two NSCLC cells lines with *PTPRH* mutations. Cell line H1155 has an M188I *PTPRH* mutation within one of the fibronectin domains (similar to the mutation we found within our mouse tumors), and cell line H2228 has a Q887P mutation within the phosphatase domain. Subjecting

these cell lines to a dose response curve with the TKI erlotinib showed no response (Figure 4.3A). However, when completing a dose response curve using the TKI osimertinib, a third generation TKI, *PTPRH* mutant cell lines showed a response (Figure 4.3B). However, subjecting H23 PTPRH KO cell lines to the same osimertinib dose regime showed no response as compared to H23 WT cells (data not shown). This may be due to the high mutational burden of the H23 cell line, which includes a mutation in the TP53 and KRAS genes, well characterized tumor suppressor and oncogenes respectively. To explore whether H23 PTPRH KO cells would show enhanced response to KRAS and EGFR inhibition, a dual drug dose response curve was completed. However, no enhanced response was seen (Figure 4.3C). With PTPRH KO cells also showing increased phosphorylation of FGFR1, it was hypothesized these cells may respond to dual inhibition of FGFR1 and EGFR. A dose response curve was completed using osimertinib and the FGFR1 inhibitior PD166866. PD166866 was chosen due to its high selectivity for FGFR1 over other members of the FGFR family. Even with increased FGFR1 noted within PTPRH KO cells, no increased sensitivity was seen upon inhibition with FGFR1 (Figure 4.3D).

TREATING MICE WITH HUMAN *PTPRH* MUTANT TUMORS

With *PTPRH* mutant NSCLC lines responding to osimertinib *in vitro*, we wanted to determine whether tumors grown from these cell lines would respond *in vivo*. To explore this, *PTPRH* mutant H2228 cells or *EGFR* mutant H1975 cells serving as a positive control were injected into the left flank of nude mice. After tumors reached approximately 6 mm in the largest direction, mice were randomized into vehicle control or drug treatment groups. H1975 injected mice were subjected to an osimertinib dose of 25 mg/kg, and H2228 injected mice were subjected to either 25 mg/kg or 50 mg/kg as seen in the literature. As expected, H1975 injected mice serving as the positive control responded extremely well to osimertinib treatment (Figure 4.4A). While H2228 mice receiving 25 mg/kg of osimertinib failed to respond to treatment, mice treated with 50 mg/kg responded favorably (Figure 4.4B). However, 50 mg/kg treatment had to be stopped after 14 days due to weight loss.

With *PTPRH* mutant tumors showing response to osimertinib *in vivo*, we wanted to determine whether tumors experienced reduced proliferation and increased apoptosis. After completion of drug course, H2228 injected mice were necropsied with portions of the tumor preserved in formalin, as well as flash frozen for future analysis. To assess proliferation and apoptosis within H2228 tumors, immunohistochemistry was completed for KI67 and TUNEL staining. As seen via KI67 staining, tumors from mice treated with 50 mg/kg of osimertinib had vastly reduced proliferation when compared to tumors from mice given vehicle control (Figure 4.5A). Interestingly, mice given vehicle control actually had slightly increased apoptosis as compared to osimertinib treated mice (Figure 4.5B), which was unexpected.

DISCUSSION

Initial findings show increased phosphorylation of EGFR upon loss of PTPRH in the NSCLC line H23. Furthermore, 5% of NSCLC patients are shown to have mutations in *PTPRH*, with certain mutations having predicted high EGFR and PI3K/AKT activity. With an estimated 235,000 cases of lung cancer occurring yearly within the United States (cancer.gov), over 10,000 patients (85% of all lung cancer cases are NSCLC, and 5% of those have *PTPRH* mutations) who present with *PTPRH* mutations could potentially benefit from EGFR targeted TKI therapy. Two NSCLC lines with *PTPRH* mutations were found to respond to the TKI osimertinib *in vitro*, with the H2228 cell line also responding *in vivo*.

Interestingly, *PTPRH* mutant cell lines responded to osimertinib, but not the first line TKI erlotinib, even with erlotinib having more affinity for wild type EGFR and osimertinib having more affinity for T790M mutant EGFR. A possible explanation for this may lie in the conformational state of EGFR, which may remain in an activated state upon PTPRH failing to dephosphorylate tyrosine residues on the c-terminal tail of EGFR. However, this hypothesis would need to be further explored. Osimertinib treatment of H2228 *PTPRH* mutant tumors in mice resulted in tumor shrinkage, showing proof of principal that *PTPRH* mutant tumors may benefit from treatment with TKIs. Other potential options for treatment of PTPRH

targets include dual inhibition of kinases whose signaling pathways are altered by PTPRH loss, or targeting RTKs with proteolysis targeting chimera (PROTAC) molecules, which target them for degradation.

Overall, treatment of downstream targets regulated by phosphatases, rather than the phosphatases themselves, may be a viable solution, although this will would require considerable characterization of the pathways affected by deregulated phosphatases. This is especially important to consider with the context dependent nature of PTP regulation, such as PTPRH deactivating EGFR, but activating SRC.

MATERIALS AND METHODS

POOLED CRISPR KNOCKOUT

Guide RNA (AGCACACACTAACATCACCG) for PTPRH was designed using Benchling. Guide was cloned in lentiviral Cas9 plasmid Addgene # 52961. Viral generation was completed through transfection of 293T cells with packaging plasmid psPAX2 and envelop plasmid pMD2.G in a ratio of 3.7:1.2:5 with the Cas9 plasmid respectively. Viafect was used for transfection. Viral supernatant was collected from 293T cells 3 days after transfection, and filtered through a .22 uM syringe filter. 1 mL of filtered viral supernatant was applied to H23 WT cells at ~30% confluency. Puromycin at 2.5 ug/mL was used as a selectable marker. Sanger sequencing was used to confirm knockout, and for TIDE analysis [291].

MTT ASSAY

H23 WT and H23 pooled PTPRH KO cells were subjected to an MTT assay. Assay kit (Roche 11465007001) instructions were followed. Assay was completed in triplicate. Graphpad was used to plot and statistically analyze results. A Welch's two-tailed t-test yielded a p-value of .0137.

GROWTH CURVES

On day 0, 1.0 x 10^5 cells were plated in triplicate within 6-well plates. On days 1-5, cells were trypsinized and cell number was read using an automated cell counter. Graphpad was used to plot results.

DOSE RESPONSE CURVES

Cells were trypsinized and cell concentration was read using an automated cell counter. Cells were then diluted to 5.0 x 10^4 cells per mL, and 20 uL of cell suspension was added to wells of an opaque 384 well plate using an electronic multichannel pipette. After overnight recovery, cells were subjected to a dose response curve of increasing drug concentration in half log steps. For single drug curves, osimertinib (Cayman AZD9291) range was .00003 to 30 uM. For dual drug curves, osimertinib range was .03 to 10 uM, and either KRAS inhibitor (ARS853, Cayman) or FGFR1 inhibitor (PD166866, Cayman) range was .00003 to 30 uM. 10 mM stocks of drugs were made by diluting with DMSO, and half-log drug series were diluted fresh with complete media. Cell viability was read after 48 hours using Promega's Cell Titer Glo. Luminescence values were normalized to non-drug treated controls, and plotted using Graphpad.

***IN VIVO* MOUSE TREATMENT**

H2228 or H1975 cell lines were injected into the left flank of 6-12 week old nude mice. Cells were trypsinized and suspended in PBS at a concentration of 10,000 cells/uL. Mice were briefly anesthetized using isofluorane, and injected using a 25 gauge needle. After tumors reached 6mm in the largest dimension, mice were randomized into one of three treatment groups; vehicle control, 25 mg/kg osimertinib, or 50 mg/kg osimertinib. The 50 mg/kg dose was only used for mice with H2228 tumors. Osimertinib (AZD9291 Cayman) was diluted using the following in order to achieve a final ratio: 5% DMSO, 40% polyethylene glycol, 5% tween-80, 50% H2O. Max volume of treatment was 10 uL for 1 gram of body weight. Mice were weighed on first day of treatment, and volume of drug was adjusted to achieve proper dose. After endpoint (28 days or tumors reaching 20 mm in largest direction), mice were euthanized using CO2, and necropsied. Portions of tumors were preserved in formalin for histology as well as flash frozen for future experiments. Mice were also checked for metastasis.

APPENDIX

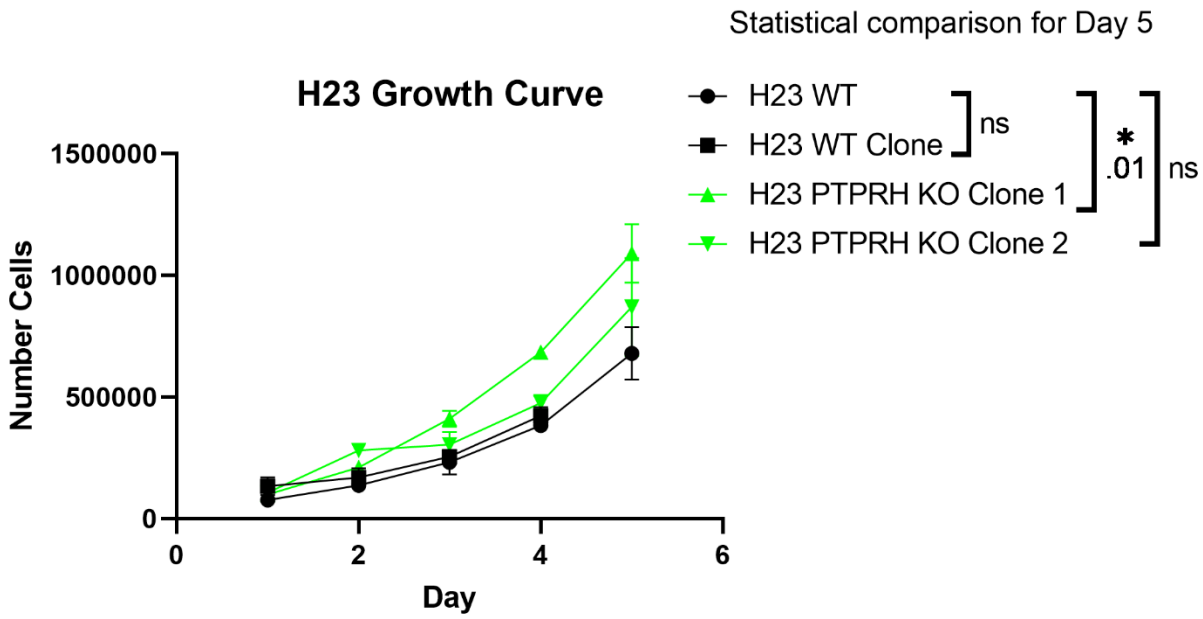


Figure 4.1: Variable growth of PTPRH KO clones

Growth curves of H23 WT cells and two H23 PTPRH KO clones show variable growth of knockout clones.

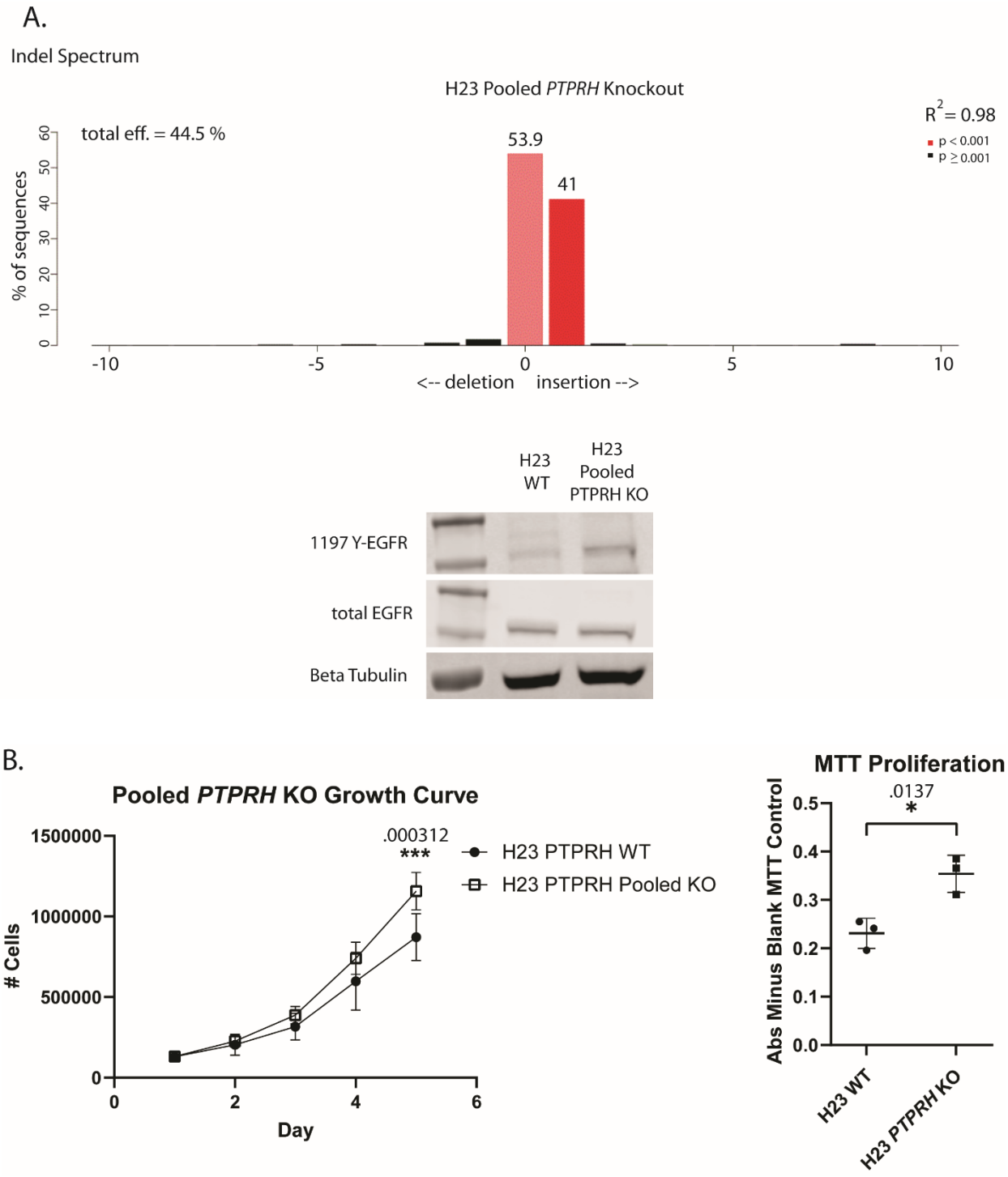


Figure 4.2: Increased cellular growth and proliferation upon pooled PTPRH knockdown

H23 cells were subjected to pooled knockout of PTPRH using CRISPR. A) TIDE analysis estimates 45% knockout efficiency from sequencing data. Western blotting of pooled PTPRH KO cells showed increased

Figure 4.2 (cont'd)

phosphorylation of EGFR at Y1197. B) Growth curves and MTT assays using PTPRH pooled KO cells show increased growth and proliferation of PTPRH KO cells compare to PTPRH WT cells.

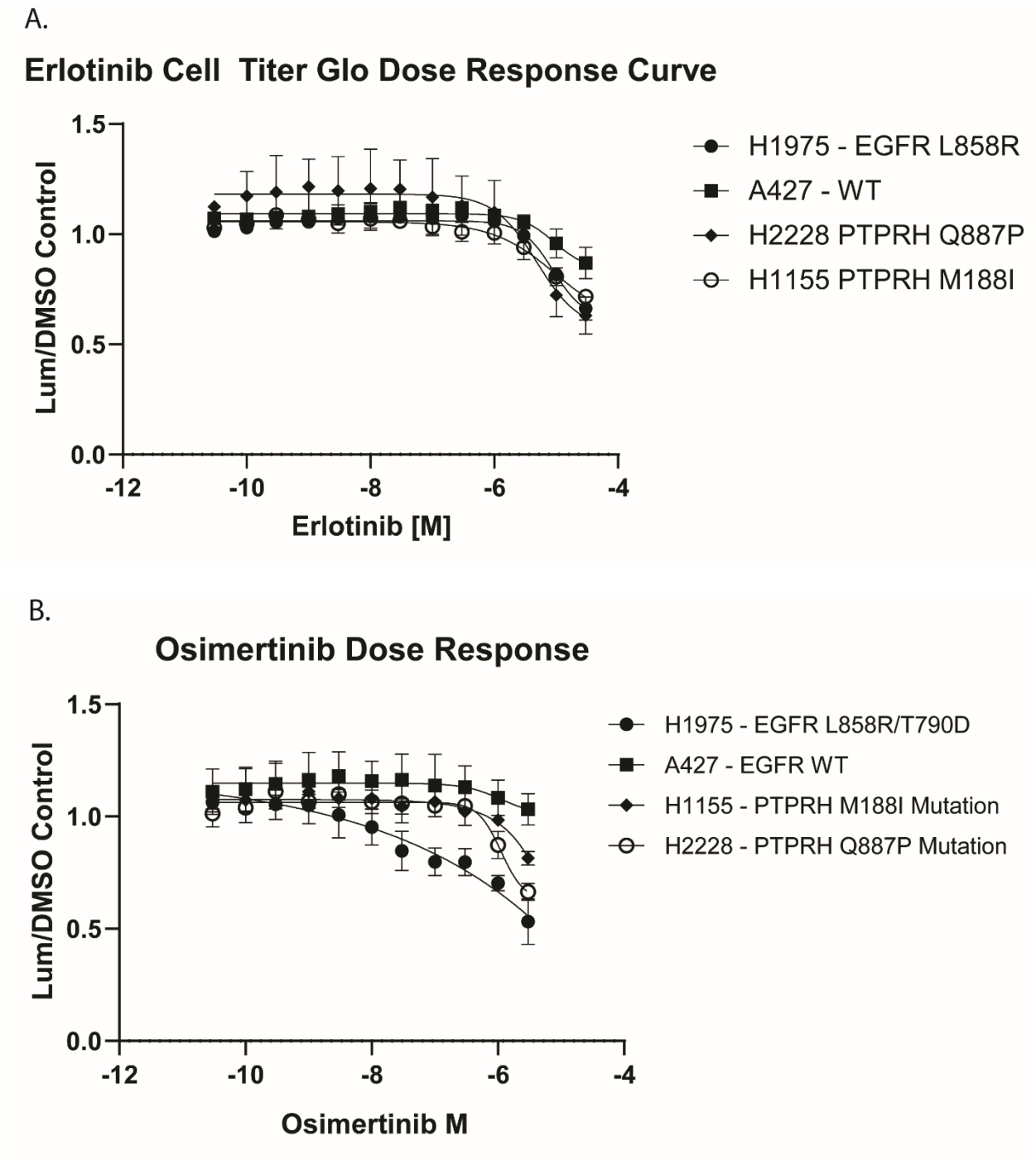


Figure 4.3: Tyrosine kinase inhibitor treatment of *PTPRH* mutant cell lines

Figure 4.3 (cont'd)

PTPRH mutations found within cell lines derived from human NSCLC tumors were subjected to dose response curves. H1975 has canonical L858R activating *EGFR* mutation and T790M resistance mutation. H1975 serves as positive control, but is not inhibited by erlotinib due to T790M resistance mutation. A427 serves as negative control, and has no mutations in *EGFR* or *PTPRH*. H1155 line has M188I *PTPRH* mutation residing within a fibronectin domain. H2228 line has Q887P *PTPRH* mutation residing in phosphatase domain. A) Cell lines treated with erlotinib, a 1st generation tyrosine kinase inhibitor. B) Cell lines treated with osimertinib, a 3rd generation tyrosine kinase inhibitor used to overcome EGFR T790M resistance mutation. C) H23 PTPRH KO cells don't respond to TKI inhibition. Since H23 has a *KRAS* G12C mutation, H23 PTPRH KO cells were subjected to dual inhibition of EGFR (osimertinib) and KRAS (ARS853). No response was seen upon the addition of KRAS inhibitor. D) PTPRH KO cells have increased activation of FGFR1. H23 PTPRH KO cells were subjected to a dual inhibition curve of EGFR inhibitor osimertinib and FGFR1 inhibitor PD166866, however no increased sensitivity to FGFR1 was noted.

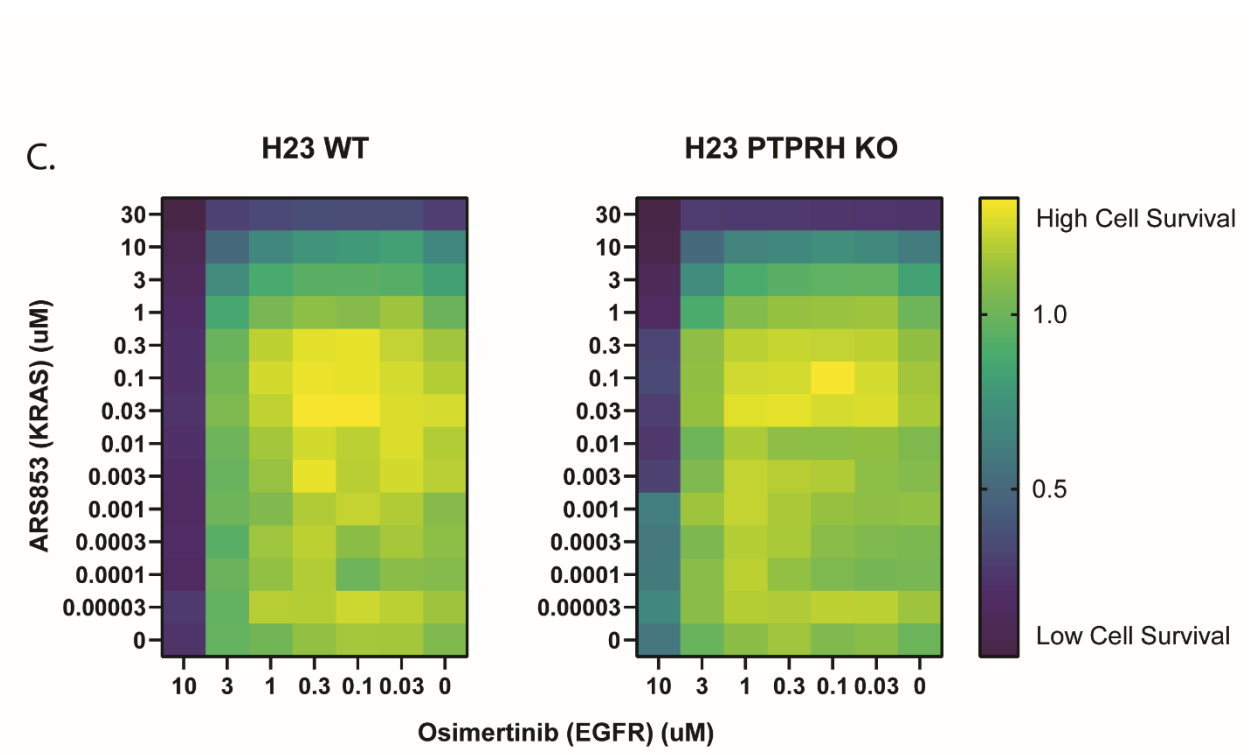
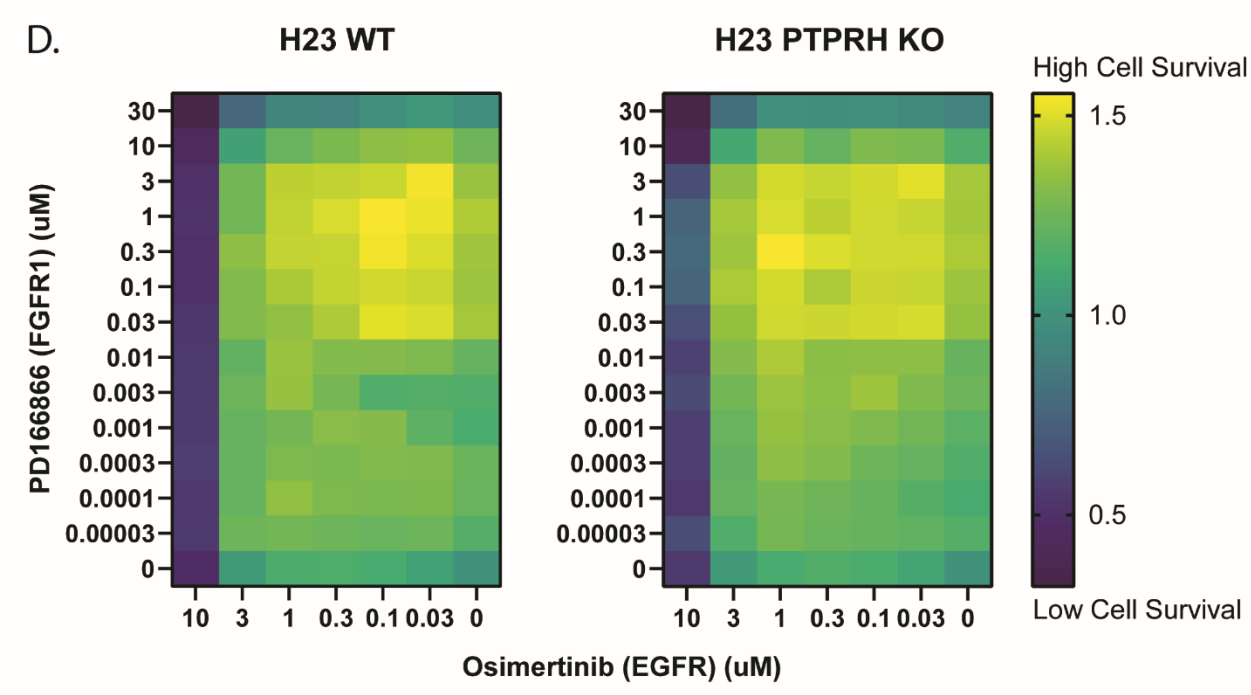


Figure 4.3 (cont'd)



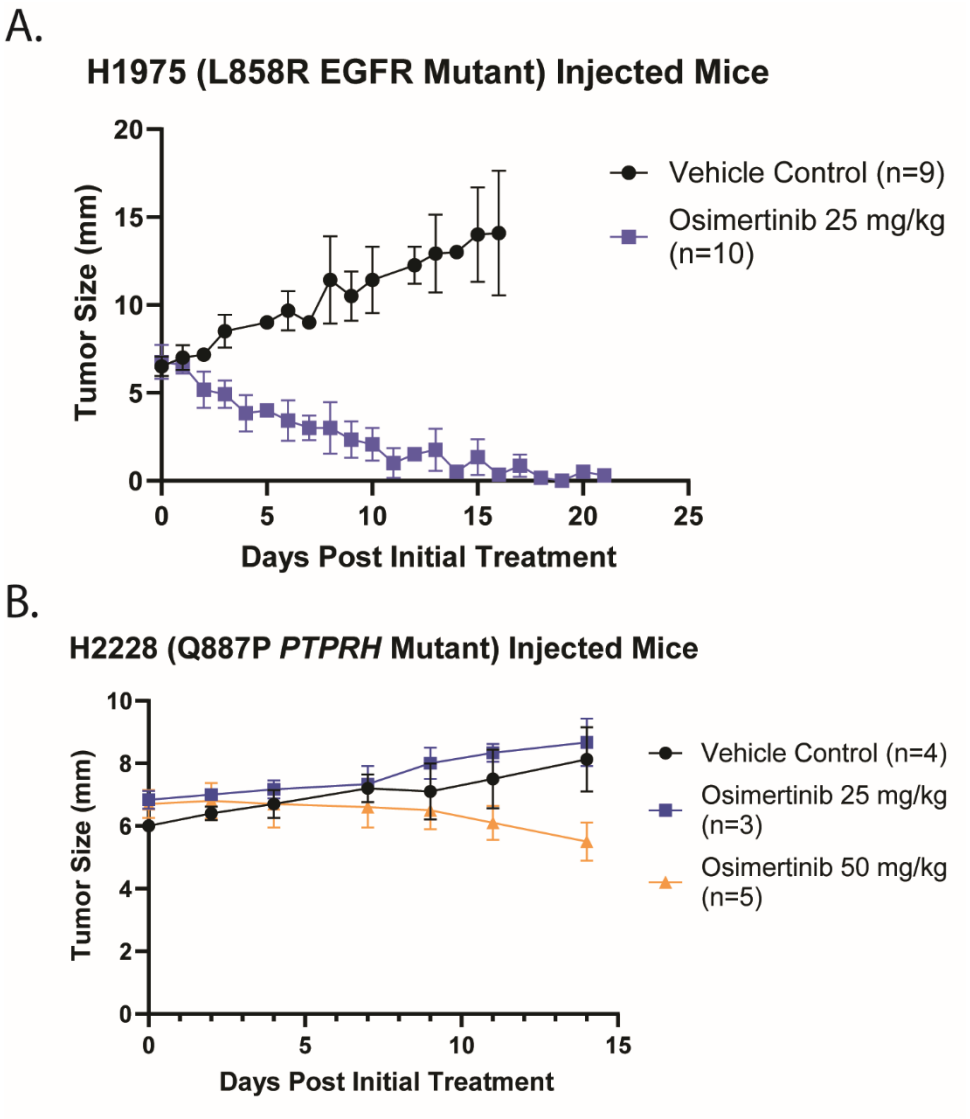


Figure 4.4: *In vivo* treatment of H2228 *PTPRH* mutant tumors

Graphs showing tumor size (measured in largest dimension) of nude mice injected with human NSCLC lines, and treated with osimertinib via oral gavage. Treatment began when tumors reached ~6.0 mm. X-axis indicates measurements taken post initiation of treatment. A) H1975 injected mice served as positive control arm for drug treatment. B) Experimental arm H2228 injected mice was divided into two treatment arms, 25 mg/kg and 50 mg/kg. Treatment was stopped after 14 days in 50 mg/kg treated mice due to weight loss.

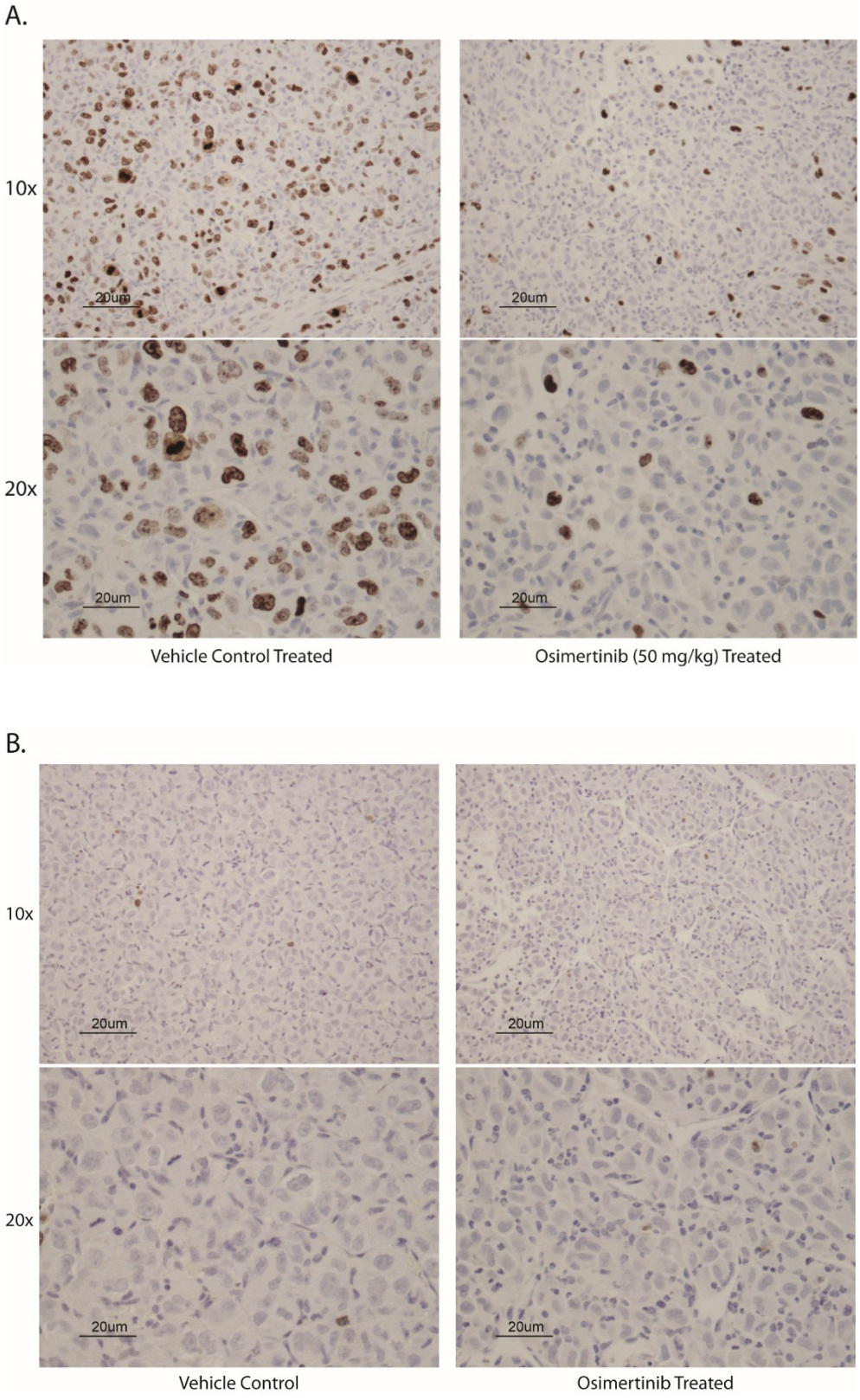


Figure 4.5: TUNEL and KI67 staining in *PTPRH* mutant tumors treated with osimertinib

Figure 4.5 (cont'd)

Representative pictures of KI67 or TUNEL stained slides, from FFPE preserved mouse tumors. Mouse tumor slides are from osimertinib treatment (50 mg/kg) or vehicle control groups of H2228 (*PTPRH* mutant) injected mice. A) KI67 staining shows decreased proliferation in mouse tumors treated with 50 mg/kg osimertinib. B) TUNEL staining shows mild increase in vehicle control treated tumors.

CHAPTER 5

FUTURE DIRECTIONS

METASTASIS IN E2F1 KNOCKOUT MOUSE MODELS

While this work characterized the genomes of E2F1 KO mice using extensive bioinformatics analysis, bench work validation is needed to further explore potentially mutated pathways. One of the most fruitful follow ups would be investigation into whether cell adhesion pathways are indeed disrupted within E2F1 KO mice. Previous research from the lab found a decrease in circulating tumor cells within PyMT E2F1 KO mice, supporting the hypothesis that mutated cell adhesion genes allow potentially metastatic cells to leave the primary tumor in greater numbers.

A possible exploration into this could involve immunohistochemistry staining for cadherin and other adhesion molecules, in PyMT WT and PyMT E2F1 KO tumors. This experiment however wouldn't differentiate between potentially disrupted collagen and cadherin fibers, so if these proteins were still present in PyMT E2F1 KO tumors, but were non-functional, staining wouldn't be able to determine that. Another potential experiment would be the use of cell adhesion assays such as Vybrant. These assays however, are just measures of whether cells are able to bind. They are not capable of measuring the strength of that binding. If disruptions occurred to cell adhesion molecules that allowed them to bind, but the strength of that binding was diminished, these assays would not make that distinction. Advanced microscopy techniques may be prudent to investigate the binding forces of these cells, and could be used on cell lines derived from PyMT tumors. An extremely interesting experiment would be to determine if E2F1 loss is indeed driving an increased mutational burden in cell adhesion genes, and how this may be occurring if it is the case.

We also discovered a variation in the mutation profiles of E2F1 KO tumors, with PyMT E2F1 KO tumors having increased association with defective miss-match repair. This may be tied to an increase in mutational burden within cell adhesion genes, as mentioned above. All of this evidence points towards E2F1 loss leading to a shift in the mutation profile of these tumors. As mentioned in chapter two, E2F1 is involved in numerous DNA repair mechanisms, including recruitment of double stranded break and

nucleotide excision repair processing factors. Loss of recruitment of these factors, and a disruption to the S phase of the cell cycle could explain why we see a shift in mutation profile in E2F1 KO tumors, although this needs to be confirmed.

In our investigation, we also discovered potential disruptions to the WNT pathway, a pathway with known involvement in the epithelial to mesenchymal transition (EMT). Further investigations into WNT and Beta Catenin may prove fruitful. A good place to start may be determining whether these pathways are actually disrupted through a series of western blots for active beta catenin or other downstream signaling molecules. It may also be beneficial to investigate whether PyMT E2F1 KO cells have a reduced ability to undergo EMT. If that is indeed the case, a deeper dive into how E2F1 loss is preventing the epithelial to mesenchymal transition would be a potentially interesting paper.

PTPRH MUTATIONS IN HUMAN CANCERS

We have uncovered a *Ptprh* mutation within PyMT mouse tumors. *Ptprh* mutant tumors correlated with increased phosphorylation of EGFR, a known oncogene. Throughout this work we have shown *PTPRH* mutations are present in 5% of human NSCLC patients, and many of these patient tumors have predicted high EGFR activity. CRISPR knockout of PTPRH in the H23 NSCLC cell line results in increased phosphorylation of EGFR and AKT, and *PTPRH* mutant cell lines respond to the TKI osimertinib *in vitro* and *in vivo*. This work suggests patients with *PTPRH* mutant tumors may respond to FDA approved TKI therapy, however there are a lot of avenues left to explore. Below we will discuss the following three research areas that we believe will be the most fruitful going forward.

First, it would be prudent to explore the impact of various human *PTPRH* mutations of the mechanism of interaction between PTPRH and EGFR. Our bioinformatics prediction data suggests certain mutations are more likely to result in the increase of phosphorylated EGFR, with some of these mutations occurring within the fibronectin domains of PTPRH, and other occurring within the phosphatase domain. Mutations within the fibronectin domains could result in a failure of PTPRH to bind target substrates, or

it could result in increased homodimerization of PTPRH. Increased dimerization of PTPRH has been shown to decrease activity of the phosphatase. A series of co-immunoprecipitation experiments of various overexpressed *PTPRH* mutants could potentially answer these questions.

It would seem obvious that mutations within the phosphatase domain of PTPRH would abolish catalytic activity, however the majority of phosphatase domain mutations within NSCLC tumors appear outside of the conserved HC-(X_5)-R activity motif. This suggests other mechanisms may be at play for the disruption of PTPRH de-phosphorylating EGFR. Other conserved motifs within the phosphatase domain are involved with recognition of phosphorylated tyrosines, so mutations within these motifs might result in a failure to recognize target substrates. Further characterization of the *PTPRH* mutations occurring in NSCLC, as well as other cancers including melanoma, may prove fruitful for future genetic screening to determine whether patients may benefit from TKI therapy.

The second area involves a deeper investigation into potential treatment methods for patients whose tumors harbor *PTPRH* mutations. While we have shown the *PTPRH* mutant cell line H2228 to respond *in vitro* and *in vivo* to the TKI osimertinib, the response was not as robust as the *EGFR* mutant line H1975. Interestingly, the H2228 cell line did not respond to the TKI erlotinib, a first generation TKI that has a higher affinity for WT EGFR that osimertinib. If mutations in *PTPRH* led to increased EGFR signaling, one would expect an inhibitor with higher affinity for WT EGFR would have a greater impact, due to *EGFR* being WT in this scenario. One potential explanation for this could be EGFR undergoing a conformational shift due to PTPRH failing to remove phosphate residues on the c-terminal tail, however this would need to be further explored. Another potential explanation is the response of H2228 to osimertinib is simply due to other factors outside of *PTPRH* mutation, however we would point out that another cell line (H1155) with mutant *PTPRH* also responded to osimertinib inhibition. To rule out this possibility, addback of WT PTPRH through an overexpression experiment would be prudent. However, this experiment would need to be carefully managed, and perhaps done under the direction of the endogenous promoter. This

is due to dimerization of PTPRH reducing PTPRH activity. Strong overexpression of WT PTPRH could conceivably result in increased homodimerization due to saturation of the protein in the membrane leading to increased proximity.

With this data in mind, there are a number of other avenues to explore. The first is using combination therapies to determine if *PTPRH* mutant tumors are more responsive to multiple TKIs. We have shown PTPRH KO cells to have increased FGFR1, therefore it is feasible *PTPRH* mutant tumors may have increased signaling of other RTKs. Profiling the activation of these RTKs, and subsequent dual TKI inhibition may prove a fruitful endeavor for treating *PTPRH* mutant tumors. While our data is not promising for dual EGFR and FGFR1 inhibition, it is only the result of one FGF1 inhibitor test. With dozens of other FGFR1 inhibitors on the market, it may be prudent to test some of these as well. Further characterization of RTK activation upon *PTPRH* mutation may lead to other RTKs being discovered as potential targets.

Since PTPRH targets are often WT and uninterrupted themselves, PROTACs targeting RTKs and other signaling molecules downstream may be another area to explore. Targeting EGFR or AKT with a PROTAC may be a beneficial treatment, however further characterization of what happens to EGFR molecules after failed interactions with PTPRH needs to be further explored. If mutations in *PTPRH* simply cause higher turnover of EGFR, and EGFR is already being internalized and marked for ubiquitination at a high rate, PROTAC treatment may prove unbeneficial. With *in vivo* CRISPR experiments for the treatment of mouse tumors beginning to be explored, as well as viral overexpression of genes *in vivo,* another avenue may be targeting expression of WT PTPRH to the tumor, although this is most likely a long way off if feasible at all.

A third area of interest would be the impact of increased nuclear EGFR within *PTPRH* mutant tumors. Many questions remain here including determining a mechanism behind increased EGFR within the nucleus, what EGFR is doing within the nucleus, and whether this provides further treatment

opportunities. To determine the mechanism behind increased EGFR localization to the nucleus, a prudent first step would be assessing whether nuclear EGFR was full length or truncated in this case. This could be further explored through determining whether canonical mechanisms are responsible for EGFR internalization at the membrane through clathrin-mediated pits, and trafficking through the cell. Recycling of EGFR back to the membrane can be affected by cellular pH. With tumors being known to have increased cellular acidity, it is also possible that mutant *PTPRH* is leading to increased internalization of EGFR due to a failure to remove phosphate groups, and then altered pH within tumor cells is resulting in increased trafficking to the nucleus.

Determining the impact of increased nuclear EGFR could be prudent for investigating tumor biology as well as other potential treatments. EGFR is known to act as a transcriptional coactivator by binding AT rich regions directly on the promotor of certain genes such as cyclin-D1. To determine what molecules and signaling pathways may be affected by nuclear EGFR, a Mass-spec experiment may prove fruitful. Other pathways found deregulated through increased nuclear EGFR, may provide other targets for treatment. Overall, this project has many potential areas for future exploration.

WORKS CITED

WORKS CITED

1. Stehelin D, Varmus HE, Bishop JM, Vogt PK (1976) DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. Nature 260:170–173

2. Tabin CJ, Bradley SM, Bargmann CI, Weinberg RA, Papageorge AG, Scolnick EM, Dhar R, Lowy DR, Chang EH (1982) Mechanism of activation of a human oncogene. Nature 300:143–149

3. Stratton MR, Campbell PJ, Futreal PA (2009) The cancer genome. Nature 458:719–724

4. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW (2013) Cancer genome landscapes. Science (80-) 340:1546–1558

5. Nagase H, Nakamura Y (1993) Mutations of the APC (adenomatous polyposis coli) gene. Hum Mutat 2:425–434

6. Powell SM, Zilz N, Beazer-Barclay Y, Bryan TM, Hamilton SR, Thibodeau SN, Vogelstein B, Kinzler KW (1992) APC mutations occur early during colorectal tumorigenesis. Nature 359:235–237

7. Forrester K, Allmoguera C, Perucho M, Han K, Grizzle WE (1987) Detection of high incidence of K-ras oncogenes during human colon tumorigenesis. Nature 327:298–303

8. Hayakumo T, Nakajima M, Yasuda K, et al (1991) Prevalence of K-ras gene mutations in human colorectal cancers. Nippon Shokakibyo Gakkai Zasshi 88:1539–1544

9. Fearon ER, Vogelstein B (1990) A genetic model for colorectal tumorigenesis. Cell 61:759–767

10. Pao W, Miller V, Zakowski M, et al (2004) EGF receptor gene mutations are common in lung cancers from “never smokers” and are associated with sensitivity of tumors to gefitinib and erlotinib. Proc Natl Acad Sci 101:13306–13311

11. Prior IA, Lewis PD, Mattos C (2012) A comprehensive survey of ras mutations in cancer. Cancer Res 72:2457–2467

12. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, Mcguire WL (1987) Human Breast Cancer: Correlation of Relapse and Survival with Amplification of the HER-2lneu Oncogene. Science (80-) 235:177–182

13. Nowell P (1960) A minute chromosome in human chronic granulocytic leukemia.

14. Rowley JD (1973) A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. Nature 243:290–293

15. Collins SJ, Groudine MT (1983) Rearrangement and amplification of c-abl sequences in the human chronic myelogenous leukemia cell line K-562. Proc Natl Acad Sci U S A 80:4813–4817

16. Taberlay PC, Statham AL, Kelly TK, Clark SJ, Jones PA (2014) Reconfiguration of nucleosome-depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer. Genome Res 24:1421–1432

17. Aran D, Sabato S, Hellman A (2013) DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. Genome Biol 14:R21

18. Yegnasubramanian S, Wu Z, Haffner MC, et al (2011) Chromosome-wide mapping of DNA methylation patterns in normal and malignant prostate cells reveals pervasive methylation of gene-associated and conserved intergenic sequences. BMC Genomics 12:313

19. Nielsen FC, Van Overeem Hansen T, Sørensen CS (2016) Hereditary breast and ovarian cancer: New genes in confined pathways. Nat Rev Cancer 16:599–612

20. Mendoza PR, Grossniklaus HE (2015) The Biology of Retinoblastoma. In: Prog. Mol. Biol. Transl. Sci. Elsevier B.V., pp 503–516

21. Epstein SS (1978) The politics of cancer. Sierra Club Books

22. Haenszel W (1966) Epidemiological Approaches to the Study of Cancer and Other Chronic Diseases. In: Natl. Cancer Inst. Monogr. 19. https://books.google.com/books?hl=en&lr=&id=VoxrAAAAMAAJ&oi=fnd&pg=PR7&dq=haenszel+epidemiological+study+of+cancer+and+other+chronic+diseases&ots=ZEbNQrYnCc&sig=bRfwbC72wz-_hHfQtFtFwX6GrOU#v=onepage&q&f=false. Accessed 14 Apr 2020

23. Mason T, McKay F (1974) US cancer mortality by county, 1950-1969. Dhew Publ 74–615

24. Hoover R, Fraumeni JF (1975) Cancer mortality in U.S. counties with chemical industries. Environ Res 9:196–207

25. Blot WJ (1977) Geography of Cancer. Sciences (New York) 17:12–15

26. Jaehn P, Kaucher S, Pikalova L V., Mazeina S, Kajüter H, Becher H, Valkov M, Winkler V (2019) A cross-national perspective of migration and cancer: incidence of five major cancer types among resettlers from the former Soviet Union in Germany and ethnic Germans in Russia. BMC Cancer 19:869

27. Maskarinec G, Noh JJ (2004) THE EFFECT OF MIGRATION ON CANCER INCIDENCE AMONG JAPANESE IN HAWAII.

28. (1957) SMOKING and health; joint report of the Study Group on Smoking and Health. Science 125:1129–33

29. London TRC of P of (1962) Smoking and Health.

30. Cutler SJ (1955) A Review of the Statistical Evidence on the Association Between Smoking and Lung Cancer. J Am Stat Assoc 50:267–282

31. Gou LY, Niu FY, Wu YL, Zhong WZ (2015) Differences in driver genes between smoking-related and non-smoking-related lung cancer in the Chinese population. Cancer 121:3069–3079

32. Alexandrov LB, Ju YS, Haase K, et al (2016) Mutational signatures associated with tobacco smoking in human cancer. Science (80-) 354:618–622

33. Tomasetti C, Vogelstein B (2015) Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science (80-) 347:78–81

34. Wu S, Powers S, Zhu W, Hannun yusuf (2016) Substantial contribution of extrinsic risk factors to cancer development. Nature. doi: 10.1038/nature16166

35. Wild C, Brennan P, Plummer M, Bray F, Straif K, Zavadil J (2015) Cancer risk: Role of chance overstated. Science (80-) 347:728

36. Song M, Giovannucci EL (2015) Cancer risk: many factors contribute. Science 347:728–729

37. Ashford NA, Bauman P, Brown HS, Clapp RW, Finkel AM, Gee D, Hattis DB, Martuzzi M, Sasco AJ, Sass JB (2015) Cancer risk: Role of environment. Science (80-) 347:727

38. Tomasetti C, Li L, Vogelstein B (2017) Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. Science (80-) 355:1330–1334

39. Zhu L, Finkelstein D, Gao C, et al (2016) Multi-organ Mapping of Cancer Risk. Cell 166:1132-1146.e7

40. McFarland CD, Korolev KS, Kryukov G V, Sunyaev SR, Mirny LA (2013) Impact of deleterious passenger mutations on cancer progression. Proc Natl Acad Sci U S A 110:2910–2915

41. Castro-Giner F, Ratcliffe P, Tomlinson I (2015) The mini-driver model of polygenic cancer evolution. Nat Rev Cancer 15:680–685

42. Kumar S, Warrell J, Li S, et al (2020) Passenger Mutations in More Than 2,500 Cancer Genomes: Overall Molecular Functional Impact and Consequences. Cell 180:915-927.e16

43. Melnikov A, Rogov P, Wang L, Gnirke A, Mikkelsen TS (2014) Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. Nucleic Acids Res 42:112

44. Giacomelli AO, Yang X, Lintner RE, et al (2018) Mutational processes shape the landscape of TP53 mutations in human cancer HHS Public Access Author manuscript. Nat Genet 50:1381–1387

45. Yamagiwa K, Ichikawa K (1918) Experimental study of the pathogenesis of carcinoma. J Cancer Res 27:123–81

46. Pazos P, Lanari C, Meiss R, Charreau EH, Dosne Pasqualini C (1991) Mammary carcinogenesis induced by N-methyl-N-nitrosourea (MNU) and medroxyprogesterone acetate (MPA) in BALB/c mice. Breast Cancer Res Treat 20:133–138

47. Bonser M (1954) The Evolution of mammary cancer induced in female IF mice with minimal doses of locally acting methylcholanthrene.

48. Abba MC, Zhong Y, Lee J, Kil H, Lu Y, Takata Y, Simper MS, Gaddis S, Shen J, Marcelo Aldaz C (2016) DMBA induced mouse mammary tumors display high incidence of activating Pik3ca and loss of function Pten mutations. Oncotarget 7:64289–64299

49. Currier N, Solomon SE, Demicco EG, et al (2005) Oncogenic Signaling Pathways Activated in DMBA-Induced Mouse Mammary Tumors. Toxicol Pathol 33:726–737

50. Rehm S (1990) Chemically induced mammary gland adenomyoepitheliomas and myoepithelial carcinomas of mice. Immunohistochemical and ultrastructural features. Am J Pathol 136:575–84

51. Behbod F, Kittrell FS, LaMarca H, et al (2009) An intraductal human-in-mouse transplantation model mimics the subtypes of ductal carcinoma in situ. Breast Cancer Res 11:R66

52. Valdez KE, Fan F, Smith W, Allred DC, Medina D, Behbod F (2011) Human primary ductal carcinoma in situ (DCIS) subtype-specific pathology is preserved in a mouse intraductal (MIND) xenograft model. J Pathol 225:565–573

53. D'Cruz CM, Gunther EJ, Boxer RB, et al (2001) c-MYC induces mammary tumorigenesis by means of a preferred pathway involving spontaneous Kras2 mutations. Nat Med 7:235–239

54. Andrechek ER, Cardiff RD, Chang JT, Gatza ML, Acharya CR, Potti A, Nevins JR (2009) Genetic heterogeneity of Myc-induced mammary tumors reflecting diverse phenotypes including metastatic potential. Proc Natl Acad Sci U S A 106:16387–16392

55. Ivics ZN, Hackett PB, Plasterk RH, Izsvá Z (1997) Molecular Reconstruction of Sleeping Beauty, a Tc1-like Transposon from Fish, and Its Transposition in Human Cells its original location and promotes its reintegration else- where in the genome (Plasterk, 1996). Autonomous mem- bers of a transposon family can express an active trans- posase, the trans-acting factor for transposition, and thus are capable of transposing on their own. Nonauton. Cell 91:501–510

56. Drabek D, Zagoraiou L, DeWit T, Langeveld A, Roumpaki C, Mamalaki C, Savakis C, Grosveld F (2003) Transposition of the Drosophila hydei Minos transposon in the mouse germ line. Genomics 81:108–111

57. Ding S, Wu X, Li G, Han M, Zhuang Y, Xu T (2005) Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. Cell 122:473–483

58. Collier LS, Carlson CM, Ravimohan S, Dupuy AJ, Largaespada DA (2005) Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. Nature 436:272–276

59. Kas SM, De Ruiter JR, Schipper K, et al (2017) Insertional mutagenesis identifies drivers of a novel oncogenic pathway in invasive lobular breast carcinoma. Nat Genet 49:1219–1230

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

61. Andres A-C, Schonenberger C-A, Groner B, Hennighausent L, Lemeur M, Gerlinger P (1987) Ha-ras oncogene expression directed by a milk protein gene promoter: Tissue specificity, hormonal regulation, and tumor induction in transgenic mice (whey acidic protein gene/whey acidic protein-ras transgene/Y chromosome integration/mammary gland tumors/salivary gland tumors). Dev Biol 84:1299–1303

62. Vassar R, Rosenberg M, Rosst S, Tyner A, Fuchs E (1989) Tissue-specific and differentiation-specific expression of a human K14 keratin gene in transgenic mice (stratified squamous epithelia).

63. Jackson EL, Willis N, Mercer K, Bronson RT, Crowley D, Montoya R, Jacks T, Tuveson DA (2001) Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. Genes Dev 15:3243–3248

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

65. Gunther EJ, Belka GK, Wertheim GBW, Wang J, Hartman JL, Boxer RB, Chodosh LA (2002) A novel doxycycline-inducible system for the transgenic analysis of mammary gland biology. FASEB J 16:283–92

66. Debies MT, Gestl SA, Mathers JL, Mikse OR, Leonard TL, Moody SE, Chodosh LA, Cardiff RD, Gunther EJ (2008) Tumor escape in a Wnt1-dependent mouse breast cancer model is enabled by p19Arf/p53 pathway lesions but not p16Ink4a loss. J Clin Invest 118:51–63

67. Moody SE, Sarkisian CJ, Hahn KT, Gunther EJ, Pickup S, Dugan KD, Innocent N, Cardiff RD, Schnall MD, Chodosh LA (2002) Conditional activation of Neu in the mammary epithelium of transgenic mice results in reversible pulmonary metastasis. Cancer Cell 2:451–461

68. Podsypanina K, Politi K, Beverly LJ, 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 105:5242–5247

69. Demarest RM, Dahmane N, Capobianco AJ (2011) Notch is oncogenic dominant in T-cell acute lymphoblastic leukemia. Blood 117:2901–2909

70. Wang X, Cunningham M, Zhang X, Tokarz S, Laraway B, Troxell M, Sears RC (2011) Phosphorylation regulates c-Myc's oncogenic activity in the mammary gland. Cancer Res 71:925–936

71. Andrechek ER (2000) Amplification of the neu/erbB-2 oncogene in a mouse model of mammary tumorigenesis. Proc Natl Acad Sci 97:3444–3449

72. Andrechek ER, Hardy WR, Laing MA, Muller WJ (2004) Germ-line expression of an oncogenic erbB2 allele confers resistance to erbB2-induced mammary tumorigenesis.

73. Liu DP, Song H, Xu Y (2010) A common gain of function of p53 cancer mutants in inducing genetic instability. Oncogene 29:949–956

74. Yuan W, Stawiski E, Janakiraman V, et al (2013) Conditional activation of Pik3ca H1047R in a knock-in mouse model promotes mammary tumorigenesis and emergence of mutations. Oncogene 3253:318–326

75. Cressman VL, Backlund DC, Hicks EM, Gowen LC, Godfrey V, Koller BH (1999) Mammary tumor formation in p53- and BRCA1-deficient mice. Cell Growth Differ 10:1–10

76. Rao T, Ranger JJ, Smith HW, Lam SH, Chodosh L, Muller WJ (2014) Inducible and coupled expression of the polyomavirus middle T antigen and Cre recombinase in transgenic mice: an in vivo model for synthetic viability in mammary tumour progression. doi: 10.1186/bcr3603

77. Ranger JJ, Levy DE, Shahalizadeh S, Hallett M, Muller WJ (2009) Identification of a Stat3-dependent transcription regulatory network involved in metastatic progression. Cancer Res 69:6823–6830

78. Masuda T, Xu X, Dimitriadis EK, Lahusen T, Deng CX (2016) “DNA binding region“ of BRCA1 affects genetic stability through modulating the intra-S-phase checkpoint. Int J Biol Sci 12:133–143

79. Annunziato S, Kas SM, Nethe M, et al (2016) Modeling invasive lobular breast carcinoma by CRISPR / Cas9-mediated somatic genome editing of the mammary gland. Genes Dev 30:1470–1480

80. Xue W, Chen S, Yin H, et al (2014) CRISPR-mediated direct mutation of cancer genes in the mouse liver HHS Public Access. Nature 514:380–384

81. Platt RJ, Chen S, Zhou Y, et al (2014) CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. Cell 159:440–455

82. Gaj T, Gersbach CA, Barbas CF (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol 31:397–405

83. Ahronian LG, Lewis BC (2014) Using the RCAS-TVA System to Model Human Cancer in Mice. Cold Spring Harb Protoc 2014:pdb.top069831

84. Guy CT, Cardiff RD, 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

85. Golovkina T V, Prakash O, Ross SR (1996) Endogenous Mouse Mammary Tumor Virus Mtv-17 Is Involved in Mtv-2-Induced Tumorigenesis in GR Mice. Virology 218:14–22

86. Andrechek ER (2015) HER2/Neu tumorigenesis and metastasis is regulated by E2F activator transcription factors. Oncogene. doi: 10.1038/onc.2013.540

87. Jhan J-R, Andrechek ER (2016) Stat3 accelerates Myc induced tumor formation while reducing growth rate in a mouse model of breast cancer. Oncotarget 7:

88. Hollern DP, Honeysett J, Cardiff RD, Andrechek ER (2014) The E2F Transcription Factors Regulate

Tumor Development and Metastasis in a Mouse Model of Metastatic Breast Cancer. Mol Cell Biol 34:3229–3243

89. Cardiff RD, Anver MR, Gusterson B a, et al (2000) The mammary pathology of genetically engineered mice: the consensus report and recommendations from the Annapolis meeting. Oncogene 19:968–988

90. Cardiff RD, Wellings SR (1999) The Comparative Pathology of Human and Mouse Mammary Glands. J. Mammary Gland Biol. Neoplasia 4:

91. Ponzo MG, Lesurf R, Petkiewicz S, et al (2009) Met induces mammary tumors with diverse histologies and is associated with poor outcome and human basal breast cancer. Proc Natl Acad Sci U S A 106:12903–8

92. Hollern DP, Swiatnicki MR, Andrechek ER (2018) Histological subtypes of mouse mammary tumors reveal conserved relationships to human cancers. PLoS Genet. doi: 10.1371/journal.pgen.1007135

93. Lifsted T, Le Voyer T, Williams M, Muller W, Klein-Szanto A, Buetow KH, Hunter KW (1998) Identification of inbred mouse strains harboring genetic modifiers of mammary tumor age of onset and metastatic progression. Int J Cancer 77:640–644

94. Jeffers M, Fiscella M, Webb CP, Anver M, Koochekpour S, Vande Woude GF (1998) The mutationally activated Met receptor mediates motility and metastasis. Med Sci 95:14417–14422

95. Seth P, Porter D, Lahti-Domenici J, Geng Y, Richardson A, Polyak K, Kang K-W, Frank SA, Lee W-H, Lee EY-HP (2002) Cellular and molecular targets of estrogen in normal human breast tissue. Cancer Res 62:4540–4

96. Perou CMCM, Sørile T, Eisen MBMB, et al (2000) Molecular portraits of human breast tumours. Nature 406:747–752

97. Lukes L, Crawford NPS, Walker R, Hunter KW (2009) The Origins of Breast Cancer Prognostic Gene Expression Profiles. Cancer Res 69:310–318

98. Flowers M, Schroeder JA, Borowsky AD, Besselsen DG, Thomson CA, Pandey R, Thompson PA (2010) Pilot study on the effects of dietary conjugated linoleic acid on tumorigenesis and gene expression in PyMT transgenic mice. Carcinogenesis 31:1642–1649

99. Eilon T, Barash I (2011) Forced activation of Stat5 subjects mammary epithelial cells to DNA damage and preferential induction of the cellular response mechanism during proliferation. J Cell Physiol 226:616–626

100. Lou Y, Preobrazhenska O, Auf Dem Keller U, Sutcliffe M, Barclay L, McDonald PC, Roskelley C, Overall CM, Dedhar S (2008) Epithelial-Mesenchymal Transition (EMT) is not sufficient for spontaneous murine breast cancer metastasis. Dev Dyn 237:2755–2768

101. Kretschmer C, Sterner-Kock A, Siedentopf F, Schoenegg W, Schlag PM, Kemmner W (2011)

Identification of early molecular markers for breast cancer. Mol Cancer 10:15

102. Zhang M, Tsimelzon A, Chang C-H, Fan C, Wolff A, Perou CM, Hilsenbeck SG, Rosen JM (2015) Intratumoral Heterogeneity in a Trp53-Null Mouse Model of Human Breast Cancer. Cancer Discov 5:520–533

103. McBryan J, Howlin J, Kenny PA, Shioda T, Martin F (2007) ERalpha-CITED1 co-regulated genes expressed during pubertal mammary gland development: implications for breast cancer prognosis. Oncogene 26:6406–6419

104. Herschkowitz JI, Simin K, Weigman VJ, et al (2007) Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biol 8:R76

105. Hollern DP, Andrechek ER (2014) A genomic analysis of mouse models of breast cancer reveals molecular features of mouse models and relationships to human breast cancer. Breast Cancer Res. doi: 10.1186/bcr3672

106. Kirouac DC, Du J, Lahdenranta J, Onsum MD, Nielsen UB, Schoeberl B, McDonagh CF (2016) HER2+ Cancer Cell Dependence on PI3K vs. MAPK Signaling Axes Is Determined by Expression of EGFR, ERBB3 and CDKN1B. PLoS Comput Biol 12:1004827

107. Rennhack J, To B, Wermuth H, Andrechek ER (2017) Mouse models of breast cancer share amplification and deletion events with human breast cancer. J Mammary Gland Biol Neoplasia. doi: 10.1007/s10911-017-9374-y

108. Santarpia L, Lippman SL, El-Naggar AK Targeting the Mitogen-Activated Protein Kinase RAS-RAF Signaling Pathway in Cancer Therapy. doi: 10.1517/14728222.2011.645805

109. Martini M, Chiara M, Santis D, Braccini L, Gulluni F, Hirsch E (2014) PI3K/AKT signaling pathway and cancer: an updated review. doi: 10.3109/07853890.2014.912836org/10.3109/07853890.2014.912836

110. Bild AH, Yao G, Chang JT, et al (2006) Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature 439:353–357

111. Pfefferle AD, Herschkowitz JI, Usary J, et al (2013) Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts. Genome Biol. doi: 10.1186/gb-2013-14-11-r125

112. Nik-Zainal S, Davies H, Staaf J, et al (2016) Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 534:47–54

113. Bose R, Kavuri SM, Searleman AC, et al (2013) Activating HER2 mutations in HER2 gene amplification negative breast cancer. Cancer Discov 3:224–237

114. Samuels Y, Wang Z, Bardelli A, et al (2004) High Frequency of Mutations of the PIK3CA Gene in Human Cancers. Science (80-) 304:554

115. McFadden DG, Politi K, Bhutkar A, et al (2016) Mutational landscape of EGFR- , MYC- , and Kras-driven genetically engineered mouse models of lung adenocarcinoma. Proc Natl Acad Sci 113:E6409–E6417

116. Govindan R, Ding L, Griffith M, et al (2012) Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell 150:1121–34

117. Rennhack J, Swiatnicki M, Zhang Y, Li C, Bylett E, Ross C, Szczepanek K, Hanrahan W, Jayatissa M, Hunter K (2018) Integrated sequence and gene expression analysis of mouse models of breast cancer reveals critical events with human parallels. bioRxiv 375154

118. Pfefferle AD, Agrawal YN, Koboldt DC, Kanchi KL, Herschkowitz JI, Mardis ER, Rosen JM, Perou CM (2016) Genomic profiling of murine mammary tumors identifies potential personalized drug targets for p53-deficient mammary cancers. Dis Model Mech 9:749–757

119. Gerlinger M, Rowan AJ, Horswell S, et al (2012) Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. N Engl J Med 366:883–892

120. Navin N, Kendall J, Troge J, et al (2011) Tumor evolution inferred by single-cell sequencing. Nature 472:90–95

121. Pal B, Chen Y, Vaillant F, et al (2017) Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling. Nat Commun 8:1627

122. Dai C, Arceo J, Arnold J, Sreekumar A, Dovichi NJ, Li J, Littlepage LE Metabolomics of oncogene-specific metabolic reprogramming during breast cancer. doi: 10.1186/s40170-018-0175-6

123. Pitteri SJ, Faca VM, Kelly-Spratt KS, et al (2008) Plasma Proteome Profiling of a Mouse Model of Breast Cancer Identifies a Set of Up-Regulated Proteins in Common with Human Breast Cancer Cells. J Proteome Res 7:1481–1489

124. Schoenherr RM, Kelly-Spratt KS, Lin C, et al (2011) Proteome and Transcriptome Profiles of a Her2/Neu-driven Mouse Model of Breast Cancer. Proteomics Clin Appl 5:179–188

125. Andrechek ER, Cardiff RD, Chang JT, Gatza ML, Acharya CR, Potti A, Nevins JR (2009) Genetic heterogeneity of Myc-induced mammary tumors reflecting diverse phenotypes including metastatic potential.

126. Ponzo MG, Lesurf R, Petkiewicz S, et al (2009) Met induces mammary tumors with diverse histologies and is associated with poor outcome and human basal breast cancer. Proc Natl Acad Sci U S A 106:12903–8

127. Usary J, Zhao W, Darr D, et al (2013) Predicting drug responsiveness in human cancers using genetically engineered mice. Clin Cancer Res 19:4889–4899

128. Jhan J-R, Andrechek ER (2017) Effective personalized therapy for breast cancer based on predictions of cell signaling pathway activation from gene expression analysis. Oncogene 36:3553–3561

129. Portier WS (2020) Cancer Clinical Trials: Implications for Oncology Nurses. Semin Oncol Nurs 150998

130. Johnson JI, Decker S, Zaharevitz D, et al (2001) Relationships between drug activity in NCI preclinical in vitro and in vivo models and early clinical trials. Br J Cancer 84:1424–1431

131. Olive KP, Tuveson DA (2006) The use of targeted mouse models for preclinical testing of novel cancer therapeutics. Clin Cancer Res 12:5277–5287

132. Van Norman GA (2019) Phase II Trials in Drug Development and Adaptive Trial Design. JACC Basic to Transl Sci 4:428–437

133. Van Norman GA (2019) Limitations of Animal Studies for Predicting Toxicity in Clinical Trials: Is it Time to Rethink Our Current Approach? JACC Basic to Transl Sci 4:845–854

134. Bailey J, Thew M, Balls M (2014) An analysis of the use of animal models in predicting human toxicology and drug safety. ATLA Altern to Lab Anim 42:181–199

135. Weinstein BS, Ciszek D (2002) The reserve-capacity hypothesis: Evolutionary origins and modern implications of the trade-off between tumor-suppression and tissue-repair. Exp Gerontol 37:615–627

136. Hemann MT, Greider CW (2000) Wild-derived inbred mouse strains have short telomeres.

137. Uhl EW, Warner NJ (2015) Mouse Models as Predictors of Human Responses: Evolutionary Medicine. Curr Pathobiol Rep 3:219–223

138. Day CP, Merlino G, Van Dyke T (2015) Preclinical Mouse Cancer Models: A Maze of Opportunities and Challenges. Cell 163:39–53

139. Institute. NHGR (2016) The Cost of Sequencing a Human Genome | NHGRI. In: Cost Seq. a Hum. Genome. https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost. Accessed 24 Apr 2020

140. Hwang B, Lee JH, Bang D (2018) Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp Mol Med 50:1–14

141. Thomas A, Rajan A, Giaccone G (2012) Tyrosine Kinase Inhibitors in Lung Cancer. Hematol Oncol Clin North Am 26:589–605

142. Zhao S, Fung-Leung W-P, Bittner A, Ngo K, Liu X (2014) Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. PLoS One 9:e78644

143. Huang CT, Hsieh CH, Chung YH, Oyang YJ, Huang HC, Juan HF (2019) Perturbational Gene-Expression Signatures for Combinatorial Drug Discovery. iScience 15:291–306

144. Li L, Zhao GD, Shi Z, Qi LL, Zhou LY, Fu ZX (2016) The Ras/Raf/MEK/ERK signaling pathway and its role in the occurrence and development of HCC (Review). Oncol Lett 12:3045–3050

145. Gatza ML, Lucas JE, Barry WT, et al (2010) A pathway-based classification of human breast cancer. Proc Natl Acad Sci U S A 107:6994–9

146. Subramanian A, Tamayo P, Mootha VK, et al (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102:15545–15550

147. Veronesi U, Banfi A, del Vecchio M, et al (1986) Comparison of Halsted mastectomy with quadrantectomy, axillary dissection, and radiotherapy in early breast cancer: Long-term results. Eur J Cancer Clin Oncol 22:1085–1089

148. Goodman LS, Wintrobe MM, Dameshek W, Goodman MJ, Gilman A, McLennan MT (1984) Nitrogen Mustard Therapy: Use of Methyl-Bis(Beta-Chloroethyl)amine Hydrochloride and Tris(Beta-Chloroethyl)amine Hydrochloride for Hodgkin's Disease, Lymphosarcoma, Leukemia and Certain Allied and Miscellaneous Disorders. JAMA J Am Med Assoc 251:2255–2261

149. Pegram MD, Lipton A, Hayes DF, et al (1998) Phase II study of receptor-enhanced chemosensitivity using recombinant humanized anti-p185(HER2/neu) monoclonal antibody plus cisplatin in patients with HER2/neu-overexpressing metastatic breast cancer refractory to chemotherapy treatment. J Clin Oncol 16:2659–2671

150. Slamon DJ, Leyland-Jones B, Shak S, et al (2001) Use of chemotherapy plus a monoclonal antibody against her2 for metastatic breast cancer that overexpresses HER2. N Engl J Med 344:783–792

151. Masoud V, Pagès G (2017) Targeted therapies in breast cancer: New challenges to fight against resistance. World J Clin Oncol 8:120–134

152. (2019) SEER Stat Database: Mortality.

153. Woodard GA, Jones KD, Jablons DM (2016) Lung cancer staging and prognosis. In: Cancer Treat. Res. Kluwer Academic Publishers, pp 47–75

154. Ilic M, Ilic I (2016) Epidemiology of pancreatic cancer. World J Gastroenterol 22:9694–9705

155. Dickson P V, Gershenwald JE (2011) Staging and prognosis of cutaneous melanoma. Surg Oncol Clin N Am 20:1–17

156. Mitri ZI, Parmar S, Johnson B, et al (2018) Implementing a comprehensive translational oncology platform: From molecular testing to actionability. J Transl Med 16:358

157. Vagner T, Spinelli C, Minciacchi VR, et al (2018) Large extracellular vesicles carry most of the tumour DNA circulating in prostate cancer patient plasma. J Extracell Vesicles. doi: 10.1080/20013078.2018.1505403

158. Guy CT, Webster MA, Schaller M, Parsons TJ, Cardiff RD, Muller WJ (1992) Expression of the neu protooncogene in the mammary epithelium of transgenic mice induces metastatic disease. Proc

Natl Acad Sci 89:10578–10582

159. Nevins JR (1992) E2F: A link between the Rb tumor suppressor protein and viral oncoproteins. Science (80-) 258:424–429

160. Gorgoulis VG, Zacharatos P, Mariatos G, Kotsinas A, Bouda M, Kletsas D, Asimacopoulos PJ, Agnantis N, Kittas C, Papavassiliou AG (2002) Transcription factor E2F-1 acts as a growth-promoting factor and is associated with adverse prognosis in non-small cell lung carcinomas. J Pathol 198:142–156

161. Qin G, Kishore R, Dolan CM, et al (2006) Cell cycle regulator E2F1 modulates angiogenesis via p53-dependent transcriptional control of VEGF.

162. Rouaud F, Hamouda-Tekaya N, Cerezo M, et al (2018) E2F1 inhibition mediates cell death of metastatic melanoma. Cell Death Dis. doi: 10.1038/s41419-018-0566-1

163. Field SJ, Tsai FY, Kuo F, Zubiaga AM, Kaelin WG, Livingston DM, Orkin SH, Greenberg ME (1996) E2F-1 Functions in mice to promote apoptosis and suppress proliferation. Cell 85:549–561

164. Cancer Genome Atlas Research Network JN, Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM (2013) The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet 45:1113–20

165. Chaffer CL, Weinberg RA (2011) A perspective on cancer cell metastasis. Science (80-) 331:1559–1564

166. Fidler IJ (2003) The pathogenesis of cancer metastasis: The “seed and soil” hypothesis revisited. Nat Rev Cancer 3:453–458

167. Welch DR, Hurst DR (2019) Defining the Hallmarks of Metastasis. Cancer Res 79:3011–3027

168. Burnier J V., Wang N, Michel RP, et al (2011) Type IV collagen-initiated signals provide survival and growth cues required for liver metastasis. Oncogene 30:3766–3783

169. Chang TT, Thakar D, Weaver VM (2017) Force-dependent breaching of the basement membrane. Matrix Biol 57–58:178–189

170. Walker C, Mojares E, del Río Hernández A (2018) Role of Extracellular Matrix in Development and Cancer Progression. Int J Mol Sci 19:3028

171. Cox TR, Rumney RMH, Schoof EM, et al (2015) The hypoxic cancer secretome induces pre-metastatic bone lesions through lysyl oxidase. Nature 522:106–110

172. Daves MH, Hilsenbeck SG, Lau CC, Man TK (2011) Meta-analysis of multiple microarray datasets reveals a common gene signature of metastasis in solid tumors. BMC Med Genomics. doi: 10.1186/1755-8794-4-56

173. Cosphiadi I, Atmakusumah TD, Siregar NC, Muthalib A, Harahap A, Mansyur M (2018) Bone

Metastasis in Advanced Breast Cancer: Analysis of Gene Expression Microarray. Clin Breast Cancer 18:e1117–e1122

174. Lawrence MS, Stojanov P, Polak P, et al (2013) Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 499:214–218

175. Ding L, Ellis MJ, Li S, et al (2010) Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature 464:999–1005

176. Yachida S, Jones S, Bozic I, et al (2010) Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature 467:1114–1117

177. Wu Y, Ginther C, Kim J, Mosher N, Chung S, Slamon D, Vadgama J V (2012) Expression of Wnt3 Activates Wnt/β-Catenin Pathway and Promotes EMT-like Phenotype in Trastuzumab-Resistant HER2-Overexpressing Breast Cancer Cells. Mol Cancer Res 10:1597–1606

178. Zhan T, Rindtorff N, Boutros M (2017) Wnt signaling in cancer. Oncogene 36:1461–1473

179. Rennhack JP, To B, Swiatnicki M, et al (2019) Integrated analyses of murine breast cancer models reveal critical parallels with human disease. Nat Commun 10:3261

180. Alexandrov LB, Nik-Zainal S, Wedge DC, et al (2013) Signatures of mutational processes in human cancer. Nature 500:415–421

181. Chen J, Zhu F, Weaks RL, Biswas AK, Guo R, Li Y, Johnson DG (2011) E2F1 promotes the recruitment of DNA repair factors to sites of DNA double-strand breaks. Cell Cycle 10:1287–1294

182. Geyer FC, Weigelt B, Natrajan R, Lambros MB, De Biase D, Vatcheva R, Savage K, Mackay A, Ashworth A, Reis-Filho JS (2010) Molecular analysis reveals a genetic basis for the phenotypic diversity of metaplastic breast carcinomas. J Pathol 220:562–573

183. Dietz S, Harms A, Endris V, Eichhorn F, Kriegsmann M, Longuespée R, Stenzinger A, Sültmann H, Warth A, Kazdal D (2017) Spatial distribution of EGFR and KRAS mutation frequencies correlates with histological growth patterns of lung adenocarcinomas. Int J Cancer 141:1841–1848

184. Greaves M, Maley CC (2012) Clonal evolution in cancer. Nature 481:306–313

185. Johnson BE, Mazor T, Hong C, et al (2014) Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. Science (80-) 343:189–193

186. Hao JJ, Lin DC, Dinh HQ, et al (2016) Spatial intratumoral heterogeneity and temporal clonal evolution in esophageal squamous cell carcinoma. Nat Genet 48:1500–1507

187. Nattestad M, Chin C-S, Schatz MC (2016) Ribbon: Visualizing complex genome alignments and structural variation. bioRxiv 0344:82123

188. Chang JT, Nevins JR (2006) GATHER: a systems approach to interpreting genomic signatures. Bioinformatics 22:2926–2933

189. Choi EH, Kim KP (2019) E2F1 facilitates DNA break repair by localizing to break sites and enhancing the expression of homologous recombination factors. Exp Mol Med. doi: 10.1038/s12276-019-0307-2

190. Guo R, Chen J, Zhu F, Biswas AK, Berton TR, Mitchell DL, Johnson DG (2010) E2F1 localizes to sites of UV-induced DNA damage to enhance nucleotide excision repair. J Biol Chem 285:19308–19315

191. Francis JC, Melchor L, Campbell J, et al (2015) Whole-exome DNA sequence analysis of Brca2 - And Trp53 -deficient mouse mammary gland tumours. J Pathol 236:186–200

192. Campbell KM, O'Leary KA, Rugowski DE, Mulligan WA, Barnell EK, Skidmore ZL, Krysiak K, Griffith M, Schuler LA, Griffith OL (2019) A Spontaneous Aggressive ERα+ Mammary Tumor Model Is Driven by Kras Activation. Cell Rep 28:1526-1537.e4

193. Rennhack JP, To B, Swiatnicki M, et al (2019) Integrated analyses of murine breast cancer models reveal critical parallels with human disease. Nat Commun 10:3261

194. Shah SP, Roth A, Goya R, et al (2012) The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature 486:395–399

195. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP (2006) GenePattern 2.0. Nat Genet 38:500–501

196. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120

197. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

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

199. Larson DE, Harris CC, Chen K, Koboldt DC, Abbott TE, Dooling DJ, Ley TJ, Mardis ER, Wilson RK, Ding L (2012) Somaticsniper: Identification of somatic point mutations in whole genome sequencing data. Bioinformatics 28:311–317

200. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. doi: 10.1038/nbt.2514

201. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK (2012) VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22:568–576

202. Wang K, Li M, Hakonarson H (2010) ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. doi: 10.1093/nar/gkq603

203. Layer RM, Chiang C, Quinlan AR, Hall IM (2014) LUMPY: A probabilistic framework for structural

variant discovery. Genome Biol 15:R84

204. Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO (2012) DELLY: Structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 28:333–339

205. Díaz-Gay M, Vila-Casadesús M, Franch-Expósito S, Hernández-Illán E, Lozano JJ, Castellví-Bel S (2018) Mutational Signatures in Cancer (MuSiCa): A web application to implement mutational signatures analysis in cancer samples. BMC Bioinformatics 19:224

206. Ahmadinejad N, Troftgruben S, Maley C, Wang J, Liu L (2019) MAGOS: Discovering Subclones in Tumors Sequenced at Standard Depths. bioRxiv 790386

207. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: An information aesthetic for comparative genomics. Genome Res 19:1639–1645

208. Levi-Montalcini R, Booker B (1960) EXCESSIVE GROWTH OF THE SYMPATHETIC GANGLIA EVOKED BY A PROTEIN ISOLATED FROM MOUSE SALIVARY GLANDS. Proc Natl Acad Sci 46:373–384

209. COHEN S (1962) Isolation of a mouse submaxillary gland protein accelerating incisor eruption and eyelid opening in the new-born animal. J Biol Chem 237:1555–1562

210. Purchio AF, Erikson E, Brugge JS, Erikson RL (1978) Identification of a polypeptide encoded by the avian sarcoma virus src gene. Proc Natl Acad Sci U S A 75:1567–1571

211. Collett MS, Erikson RL (1978) Protein kinase activity associated with the avian sarcoma virus src gene product. Proc Natl Acad Sci U S A 75:2021–2024

212. Carpenter G, King L, Cohen S (1978) Epidermal growth factor stimulates phosphorylation in membrane preparations in vitro. Nature 276:409–410

213. Sefton BM, Hunter T, Beemon K (1979) Product of in vitro translation of the Rous sarcoma virus src gene has protein kinase activity. J Virol 30:311–8

214. Eckhart W, Hutchinson MA, Hunter T (1979) An activity phosphorylating tyrosine in polyoma T antigen immunoprecipitates. Cell 18:925–933

215. Ushiro H, Cohen S (1980) Identification of phosphotyrosine as a product of epidermal growth factor-activated protein kinase in A-431 cell membranes. J Biol Chem 255:8363–8365

216. Weinberg RA (2014) The Biology of Cancer, 2nd Edition.

217. Lawrence MC, McKern NM, Ward CW (2007) Insulin receptor structure and its implications for the IGF-1 receptor. Curr Opin Struct Biol 17:699–705

218. Leppänen VM, Prota AE, Jeltsch M, Anisimov A, Kalkkinen N, Strandin T, Lankinen H, Goldman A, Ballmer-Hofer K, Alitalo K (2010) Structural determinants of growth factor binding and specificity by VEGF receptor 2. Proc Natl Acad Sci U S A 107:2425–2430

219. Wiesmann C, Fuh G, Christinger HW, Eigenbrot C, Wells JA, De Vos AM (1997) Crystal structure at 1.7 Å resolution f VEGF in complex with domain 2 of the Fit-1 receptor. Cell 91:695–704

220. Sasaki T, Knyazev PG, Clout NJ, Cheburkin Y, Göhring W, Ullrich A, Timpl R, Hohenester E (2006) Structural basis for Gas6-Axl signalling. EMBO J 25:80–87

221. Plotnikov AN, Schlessinger J, Hubbard SR, Mohammadi M (1999) Structural basis for FGF receptor dimerization and activation. Cell 98:641–650

222. Stauber DJ, DiGabriele AD, Hendrickson WA (2000) Structural interactions of fibroblast growth factor receptor with its ligands. Proc Natl Acad Sci U S A 97:49–54

223. Schlessinger J, Plotnikov AN, Ibrahimi OA, Eliseenkova A V., Yeh BK, Yayon A, Linhardt RJ, Mohammadi M (2000) Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. Mol Cell 6:743–750

224. Graus-Porta D, Beerli RR, Daly JM, Hynes NE (1997) ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling. EMBO J 16:1647–1655

225. Huse M, Kuriyan J (2002) The conformational plasticity of protein kinases. Cell 109:275–282

226. Lemmon MA, Schlessinger J (2010) Cell signaling by receptor tyrosine kinases. Cell 141:1117–1134

227. Favelyukis S, Till JH, Hubbard SR, Miller WT (2001) Structure and autoregulation of the insulin-like growth factor 1 receptor kinase. Nat Struct Biol 8:1058–1063

228. Pawson T (2004) Specificity in Signal Transduction: From Phosphotyrosine-SH2 Domain Interactions to Complex Cellular Systems. Cell 116:191–203

229. Schlessinger J, Lemmon MA (2003) SH2 and PTB domains in tyrosine kinase signaling. Sci STKE. doi: 10.1126/stke.2003.191.re12

230. Kirkin V, Dikic I (2007) Role of ubiquitin- and Ubl-binding proteins in cell signaling. Curr Opin Cell Biol 19:199–205

231. Avraham R, Yarden Y (2011) Feedback regulation of EGFR signalling: Decision making by early and delayed loops. Nat Rev Mol Cell Biol 12:104–117

232. Clayton AHA, Walker F, Orchard SG, Henderson C, Fuchs D, Rothacker J, Nice EC, Burgess AW (2005) Ligand-induced dimer-tetramer transition during the activation of the cell surface epidermal growth factor receptor-A multidimensional microscopy analysis. J Biol Chem 280:30392–30399

233. Gadella TWJ, Jovin TM (1995) Oligomerization of epidermal growth factor receptors on A431 cells studied by time-resolved fluorescence imaging microscopy. A stereochemical model for tyrosine kinase receptor activation. J Cell Biol 129:1543–1558

234. Chung I, Akita R, Vandlen R, Toomre D, Schlessinger J, Mellman I (2010) Spatial control of EGF receptor activation by reversible dimerization on living cells. Nature 464:783–787

235. Ogiso H, Ishitani R, Nureki O, et al (2002) Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. Cell 110:775–787

236. Garrett TPJ, McKern NM, Lou M, et al (2002) Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor α. Cell 110:763–773

237. Zhang X, Gureasko J, Shen K, Cole PA, Kuriyan J (2006) An Allosteric Mechanism for Activation of the Kinase Domain of Epidermal Growth Factor Receptor. Cell 125:1137–1149

238. Schulze WX, Deng L, Mann M (2005) Phosphotyrosine interactome of the ErbB-receptor kinase family. Mol Syst Biol 1:2005.0008

239. Erba EB, Bergatto E, Cabodi S, Silengo L, Tarone G, Defilippi P, Jensen ON (2005) Systematic analysis of the epidermal growth factor receptor by mass spectrometry reveals stimulation-dependent multisite phosphorylation. Mol Cell Proteomics 4:1107–1121

240. Honegger A, Dull TJ, Szapary D, Komoriya A, Kris R, Ullrich A, Schlessinger J (1988) Kinetic parameters of the protein tyrosine kinase activity of EGF-receptor mutants with individually altered autophosphorylation sites. EMBO J 7:3053–60

241. Guo L, Kozlosky CJ, Ericsson LH, Daniel TO, Cerretti DP, Johnson RS (2003) Studies of ligand-induced site-specific phosphorylation of epidermal growth factor receptor. J Am Soc Mass Spectrom 14:1022–1031

242. Knudsen SLJ, Wai Mac AS, Henriksen L, Van Deurs B, Grøvdal LM (2014) EGFR signaling patterns are regulated by its different ligands. Growth Factors 32:155–163

243. Lynch TJ, Bell DW, Sordella R, et al (2004) Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non–Small-Cell Lung Cancer to Gefitinib. N Engl J Med 350:2129–2139

244. Paez JG, Jänne PA, Lee JC, et al (2004) EGFR mutations in lung, cancer: Correlation with clinical response to gefitinib therapy. Science (80-) 304:1497–1500

245. Conte A, Sigismund S (2016) Chapter Six - The Ubiquitin Network in the Control of EGFR Endocytosis and Signaling. In: Prog. Mol. Biol. Transl. Sci. Elsevier B.V., pp 225–276

246. Stern KA, Place TL, Lill NL (2008) EGF and amphiregulin differentially regulate Cbl recruitment to endosomes and EGF receptor fate. Biochem J 410:585–594

247. Raymond JR, Baldys A, Göoz M, Morinelli TA, Lee MH, Luttrell LM, Raymand JR (2009) Essential role of c-Cbl in amphiregulin-induced recycling and signaling of the endogenous epidermal growth factor receptor. Biochemistry 48:1462–1473

248. French AR, Tadaki DK, Niyogi SK, Lauffenburger DA (1995) Intracellular trafficking of epidermal

growth factor family ligands is directly influenced by the pH sensitivity of the receptor/ligand interaction. J Biol Chem 270:4334–4340

249. Marti U, Burwen SJ, Wells A, Barker ME, Huling S, Feren AM, Jones AL (1991) Localization of epidermal growth factor receptor in hepatocyte nuclei. Hepatology 13:15–20

250. Psyrri A (2005) Effect of Epidermal Growth Factor Receptor Expression Level on Survival in Patients with Epithelial Ovarian Cancer. Clin Cancer Res 11:8637–8643

251. Lipponen P, Eskelinen M (1994) Expression of epidermal growth factor receptor in bladder cancer as related to established prognostic factors, oncoprotein (c-erbB-2, p53) expression and long-term prognosis. Br J Cancer 69:1120–1125

252. Lin SY, Makino K, Xia W, Matin A, Wen Y, Kwong KY, Bourguignon L, Hung MC (2001) Nuclear localization of EGF receptor and its potential new role as a transcription factor. Nat Cell Biol 3:802–808

253. Lo HW, Hsu SC, Ali-Seyed M, Gunduz M, Xia W, Wei Y, Bartholomeusz G, Shih JY, Hung MC (2005) Nuclear interaction of EGFR and STAT3 in the activation of the iNOS/NO pathway. Cancer Cell 7:575–589

254. Lo H-W, Xia W, Wei Y, Ali-Seyed M, Huang S-F, Hung M-C (2005) Novel prognostic value of nuclear epidermal growth factor receptor in breast cancer. Cancer Res 65:338–48

255. Traynor AM, Weigel TL, Oettel KR, et al (2013) Nuclear EGFR protein expression predicts poor survival in early stage non-small cell lung cancer. Lung Cancer 81:138–141

256. Sefton BM, Hunter T, Beemon K, Eckhart W (1980) Evidence that the phosphorylation of tyrosine is essential for cellular transformation by Rous sarcoma virus. Cell 20:807–16

257. Chernoff J, Li HC, Cheng YS, Chen LB (1983) Characterization of a phosphotyrosyl protein phosphatase activity associated with a phosphoseryl protein phosphatase of Mr = 95,000 from bovine heart. J Biol Chem 258:7852–7857

258. Pallen CJ, Valentine KA, Wang JH, Hollenberg MD (1985) Calcineurin-mediated dephosphorylation of the human placental membrane receptor for epidermal growth factor urogastrone. Biochemistry 24:4727–30

259. Chan CP, Gallis B, Blumenthal DK, Pallen CJ, Wang JH, Krebs EG (1986) Characterization of the phosphotyrosyl protein phosphatase activity of calmodulin-dependent protein phosphatase. J Biol Chem 261:9890–9895

260. Lin MF, Clinton GM (1988) The epidermal growth factor receptor from prostate cells is dephosphorylated by a prostate-specific phosphotyrosyl phosphatase. Mol Cell Biol 8:5477–5485

261. Tonks NK (2013) Protein tyrosine phosphatases - from housekeeping enzymes to master regulators of signal transduction. FEBS J 280:346–378

262. Tonks NK, Diltz CD, Fischer EH (1988) Characterization of the major protein-tyrosine-phosphatases of human placenta. J Biol Chem 263:6731–6737

263. Jia Z, Barford D, Flint AJ, Tonks NK (1995) Structural basis for phosphotyrosine peptide recognition by protein tyrosine phosphatase 1B. Science (80-) 268:1754–1758

264. Pingel JT, Thomas ML (1989) Evidence that the leukocyte-common antigen is required for antigen-induced T lymphocyte proliferation. Cell 58:1055–1065

265. Koretzky GA, Picus J, Thomas ML, Weiss A (1990) Tyrosine phosphatase CD45 is essential for coupling T-cell antigen receptor to the phosphatidyl inositol pathway. Nature 346:66–68

266. Wälchli S, Espanel X, Van Huijsduijnen RH (2005) Sap-1/PTPRH activity is regulated by reversible dimerization. Biochem Biophys Res Commun 331:497–502

267. Matozaki T, Suzuki T, Uchida T, Inazawa J, Ariyama T, Matsuda K, Horita K, Noguchi H, Mizuno H, Sakamoto C (1994) Molecular cloning of a human transmembrane-type protein tyrosine phosphatase and its expression in gastrointestinal cancers. J Biol Chem 269:2075–2081

268. Yao Z, Darowski K, St-Denis N, et al (2017) A Global Analysis of the Receptor Tyrosine Kinase-Protein Phosphatase Interactome. Mol Cell 65:347–360

269. Yao Z, Darowski K, St-Denis N, Babu M, Gingras A-C, Correspondence IS (2017) A Global Analysis of the Receptor Tyrosine Kinase-Protein Phosphatase Interactome. Mol Cell 65:347–360

270. Olayioye MA, Beuvink I, Horsch K, Daly JM, Hynes NE (1999) ErbB receptor-induced activation of Stat transcription factors is mediated by Src tyrosine kinases. J Biol Chem 274:17209–17218

271. Leaman DW, Pisharody S, Flickinger TW, Commane MA, Schlessinger J, Kerr IM, Levy DE, Stark GR (1996) Roles of JAKs in activation of STATs and stimulation of c-fos gene expression by epidermal growth factor. Mol Cell Biol 16:369–375

272. Levy DE, Darnell JE (2002) STATs: Transcriptional control and biological impact. Nat Rev Mol Cell Biol 3:651–662

273. Bjorge JD, Chan TO, Antczak M, Kung HJ, Fujita DJ (1990) Activated type I phosphatidylinositol kinase is associated with the epidermal growth factor (EGF) receptor following EGF stimulation. Proc Natl Acad Sci U S A 87:3816–3820

274. Vivanco I, Sawyers CL (2002) The phosphatidylinositol 3-kinase-AKT pathway in humancancer. Nat Rev Cancer 2:489–501

275. Zhong W, Myers JS, Wang F, et al (2020) Comparison of the molecular and cellular phenotypes of common mouse syngeneic models with human tumors. BMC Genomics 21:2

276. De Ruiter JR, Wessels LFA, Jonkers J (2018) Mouse models in the era of large human tumour sequencing studies. Open Biol. doi: 10.1098/rsob.180080

277. Guy CT, Cardiff RD, 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

278. Kumar S, Warrell J, Li S, et al (2020) Passenger Mutations in More Than 2,500 Cancer Genomes: Overall Molecular Functional Impact and Consequences. Cell 180:915-927.e16

279. Sato T, Soejima K, Arai E, et al (2015) Prognostic implication of PTPRH hypomethylation in non-small cell lung cancer. Oncol Rep 34:1137–1145

280. (2021) Benchling [Biology Software]. https://help.benchling.com/en/articles/1413662-cite-benchling-in-your-research. Accessed 26 Mar 2021

281. Li J, Yen C, Liaw D, et al (1997) PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. Science (80-) 275:1943–1947

282. Chen CY, Chen J, He L, Stiles BL (2018) PTEN: Tumor suppressor and metabolic regulator. Front Endocrinol (Lausanne) 9:338

283. Seo Y, Matozaki T, Tsuda M, Hayashi Y, Itoh H, Kasuga M (1997) Overexpression of SAP-1, a transmembrane-type protein tyrosine phosphatase, in human colorectal cancers. Biochem Biophys Res Commun 231:705–711

284. Travis WD, Brambilla E, Nicholson AG, et al (2015) The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances since the 2004 Classification. J Thorac Oncol 10:1243–1260

285. Crvenkova S (2015) Prognostic factors and survival in non-small cell lung cancer patients treated with chemoradiotherapy. Maced J Med Sci 3:75–79

286. Shi Y, Au JSK, Thongprasert S, Srinivasan S, Tsai CM, Khoa MT, Heeroma K, Itoh Y, Cornelio G, Yang PC (2014) A prospective, molecular epidemiology study of EGFR mutations in Asian patients with advanced non-small-cell lung cancer of adenocarcinoma histology (PIONEER). J Thorac Oncol 9:154–162

287. Fukuoka M, Wu YL, Thongprasert S, et al (2011) Biomarker analyses and final overall survival results from a phase III, randomized, open-label, first-line study of gefitinib versus carboplatin/paclitaxel in clinically selected patients with advanced non - small-cell lung cancer in Asia (IPASS). In: J. Clin. Oncol. J Clin Oncol, pp 2866–2874

288. Bennett AM, Tang TL, Sugimoto S, Walsh CT, Neel BG (1994) Protein-tyrosine-phosphatase SHPTP2 couples platelet-derived growth factor receptor β to Ras. Proc Natl Acad Sci U S A 91:7335–7339

289. Vogel W, Ullrich A (1996) Multiple in vivo phosphorylated tyrosine phosphatase SHP-2 engages binding to Grb2 via tyrosine 584. Cell Growth Differ 7:1589–1597

290. Nagano T, Tachihara M, Nishimura Y (2018) Mechanism of Resistance to Epidermal Growth

Factor Receptor-Tyrosine Kinase Inhibitors and a Potential Treatment Strategy. Cells 7:212
291. Brinkman EK, Kousholt AN, Harmsen T, Leemans C, Chen T, Jonkers J, van Steensel B (2018) Easy quantification of template-directed CRISPR/Cas9 editing. Nucleic Acids Res 46:e58