MODELING CANCER PREDISPOSITION: PROFILING LI-FRAUMENI

SYNDROME PATIENT-DERIVED CELL LINES USING BIOINFORMATICS AND

THREE-DIMENSIONAL CULTURE MODELS

Amruta Rajendra Phatak

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements

for the degree

Doctor of Philosophy

in the Department of Medical and Molecular Genetics,

Indiana University

December 2015

Accepted by the Graduate Faculty, Indiana University, in partial

fullfillment of the requirements for the degree of Doctor of Philosohphy.

_____

Brittney-Shea Herbert, Ph.D., Chair

Doctoral Committee

_____

Yunlong Liu, Ph.D.

_____

Marc Mendonca, Ph.D.

October 07, 2015

_____

Clark Wells, Ph.D.

ii

DEDICATION

I dedicate this dissertation to my beloved mentors Drs. Brittney-Shea Herbert and Madhuri Sharon and *Aai* (my mother), without whom this work would not have been possible.

ACKNOWLEDGEMENTS

Amruta Rajendra Phatak

MODELING CANCER PREDISPOSITION: PROFILING LI-FRAUMENI

SYNDROME PATIENT-DERIVED CELL LINES USING BIOINFORMATICS AND

THREE-DIMENSIONAL CULTURE MODELS

Although rare, classification of over 200 hereditary cancer susceptibility

syndromes accounting for ~5-10% of cancer incidence has enabled the discovery and

understanding of cancer predisposition genes that are also frequently mutated in sporadic

cancers. The need to prevent or delay invasive cancer can partly be addressed by

characterization of cells derived from healthy individuals predisposed to cancer due to

inherited "single-hits" in genes in order to develop patient-derived samples as preclinical

models for mechanistic *in vitro* studies. Here, we present microarray-based transcriptome

profiling of Li-Fraumeni syndrome (LFS) patient-derived unaffected breast epithelial

cells and their phenotypic characterization as *in vitro* three-dimensional (3D) models to

test pharmacological agents. In this study, the epithelial cells derived from the unaffected

breast tissue of a LFS patient were cultured and progressed from non-neoplastic to a

malignant stage by successive immortalization and transformation steps followed by

growth in athymic mice. These cell lines exhibited distinct transcriptomic profiles and

were readily distinguishable based upon their gene expression patterns, growth

characteristics in monolayer and *in vitro* 3D cultures. Transcriptional changes in the

epithelial-to-mesenchymal transition gene signature contributed to the unique phenotypes

observed in 3D culture for each cell line of the progression series; the fully transformed

LFS cells exhibited invasive processes in 3D culture with disorganized morphologies due to cell-cell miscommunication, as seen in breast cancer. Bioinformatics analysis of the deregulated genes and pathways showed inherent differences between these cell lines and targets for pharmacological agents. After treatment with small molecule APR-246 that restores normal function to mutant p53, we observed that the neoplastic LFS cells had reduced malignant invasive structure formation from 73% to 9%, as well as an observance of an increase in formation of well-organized structures in 3D culture (from 27% to 91%) by stereomicroscopy and confocal microscopy. Therefore, the use of well-characterized and physiologically relevant preclinical models in conjunction with transcriptomic profiling of high-risk patient derived samples as a renewable laboratory resource can potentially guide the development of safer and more effective chemopreventive approaches.

Brittney-Shea Herbert, Ph.D., Chair

TABLE OF CONTENTS

APPENDICES

CURRICULUM VITAE

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

.cls	class file or template file
.txt file format	tab-delimited file
2D	Two-dimensional
3D	Three-dimensional
ACC	Adrenal Cortical Carcinomas
AF594	AlexaFluor 594
ANOVA	Analysis of Variance
APR-246	PRIMA-1^{MET} (APREA)
BSA	Bovine Serum Albumin
Cas9	CRISPR associated protein 9
CCLE	Cancer Cell Line Encyclopedia
CRISPR	clustered regularly interspaced short palindromic repeat
CV	Coefficient of variation
DAPI	4',6-diamidino-2-phenylindole
DMSO	Dimethyl Sulfoxide
ECM	Extracellular matrix
EGCG	Epigallocatechin gallate
EHS	Engelbreth-Holm-Swarm
EMT	Epithelial Mesenchymal Transition
ES	Enrichment score
EU	Endotoxin Units
F-actin	Filamentous actin
FBS	Fetal Bovine Serum
FDR	false discovery rate
GEO	Gene Expression Omnibus
GFR	Growth Factor Reduced
GSEA	Gene Set Enrichment Analysis
HBOC	Hereditary breast and ovarian cancer
HBSS	Hank's Balanced Salt Solution
HC	Hierarchical clustering
HER2	Human Epidermal Growth Factor Receptor 2
HMEC	Human Mammary Epithelial Cell
hTERT	human Telomerase Reverse Transcriptase
Hu EGF	Recombinant Human Epidermal Growth Factor
HUGO	Human Genome Organization
IARC	International Agency for Research on Cancer
IF	Immunofluorescence
IPA®	Ingenuity® Pathway Analysis
iPSC	induced Pluripotent Stem Cell
LFS	Li-Fraumeni Syndrome
lrECM	laminin-rich extracellular matrix
M171	Medium 171

MET	Mesenchymal to Epithelial Transition
MSigDB	Molecular Signature Database
mTOR	Mechanistic Target of Rapamycin
NES	Normalized Enrichment Score
NF-κB	Nuclear Factor Kappa-light-chain-enhancer of activated B cells
PBS	Phosphate buffered saline
PCA	Principal Components Analysis
PD	Population Doubling
PDL	Population Doubling Level
PDX	Patient-derived xenografts
PI3K	Phosphatidylinositol 3-Kinase
PRIMA-1	p53 re-activation and induction of massive apoptosis
qRT-PCR	quantitative real-time polymerase chain reaction
RMA	Robust Multi-chip Average
TCGA	The Cancer Cell Genome Atlas
TNBC	Triple-Negative Breast cancer
WT	Wild-type

CHAPTER ONE: Introduction

1.1. Cancer predisposition syndromes: insights to mechanisms of cancer and genotypic/phenotypic aspects

Cancer predisposition can be defined as an increased susceptibility to the development of cancer due to rare germline mutations that either occur *de novo* or are inherited; the genes in which the cancer risk-conferring mutations occur are called cancer predisposition genes. The inherited cancers develop due to highly penetrant germline mutation whereas familial cancer may arise due to low-penetrance genes constituting for ~5-10% and 10-15% (**Figure 1**) of global cancer burden respectively [1].

Over the past three decades, more than 100 cancer predisposition genes and the associated cancer predisposition has been identified which has enabled characterization of important genes such as *RB1*, *PTEN, BRCA2, MLH1* and *TP53* and revolutionized the field of cancer genetics at an unprecedented rate. An overlap of mutations in these genes is also found in the general population that contributes to sporadic cancer development. In women, a number of autosomal-dominant, highly penetrant cancer predisposition genes such as *PTEN*, *BRCA1* and *BRCA2*, and *TP53* contribute to breast cancer development in Cowden syndrome (CS), hereditary breast and ovarian cancer syndrome (HBOC) and Li-Fraumeni Syndrome (LFS), respectively.
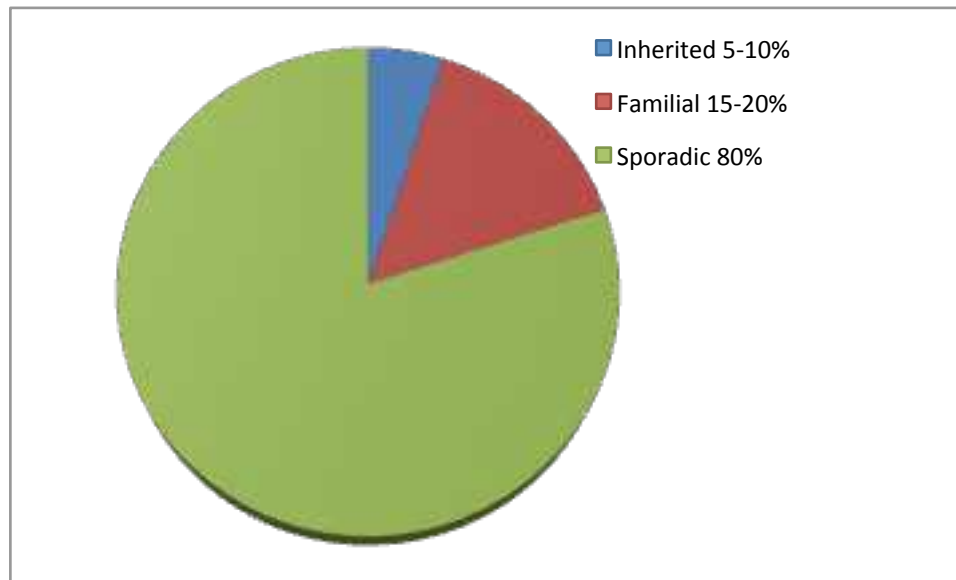
Figure 1. Inherited cancer predisposition syndromes are rare. Most cancer are sporadic; however, germline mutations in approximately 114 cancer predisposition genes that confer high or moderately high risks of cancer called have been identified. About 5-10% cancers are inherited due to highly penetrant germline mutations in cancer predisposition genes. Familial cancers account to 15-20% of cancer burden which may result from by interaction(s) between one or multiple low-penetrance alleles and/or the environment (Figure adapted from Nagy, 2004).

Li-Fraumeni Syndrome

In 1969, Drs. Frederick Li and Joseph Fraumeni, based on retrospective data analysis of over 600 medical records, described a cancer syndrome (later termed Li-Fraumeni syndrome) in four families that comprised of multiple members who developed pediatric and diverse early onset cancers such as rhabdomyosarcomas, soft tissue sarcomas [2, 3]. Further epidemiological research confirmed Li-Fraumeni syndrome (LFS) as a rare autosomal-dominant cancer predisposition syndrome characterized by a spectrum of tumors - soft-tissue and bone sarcomas, premenopausal breast cancers, central nervous system tumors, acute leukemias and adrenal cortical carcinomas (ACC) being most frequent and known as the "core" component class of LFS tumors [4, 5]. Germline heterozygous mutations in *TP53* gene that encodes tumor suppressor protein p53 were identified in individuals in LFS families [6]. The Li-Fraumeni syndrome (MIM 151623) is defined as clinically and genetically heterogeneous characterized by autosomal dominant inheritance, early onset of tumors with multiple tumors occurring within an individual and multiple family members affected. Sporadic mutations in *TP53* gene are also common in most human cancers [7] and have been observed in every region of *TP53* gene [8], however, some mutations are more frequent than others (i.e., hotspot mutations). The tumor suppressor plays critical roles in multiple cellular processes including maintenance of genome stability, cell-cycle arrest and apoptosis in response to oncogenic insults, and thus heterozygous mutations in p53 may be an early step necessary in cancer progression [9, 10] as observed in LFS [11]. A conservative estimate for Li-Fraumeni syndrome causing *TP53* germline mutation carrier rate of 1/5000 has been calculated [12]. These LFS-associated *TP53* mutations can be grouped based on the

mutation type; the dominant-negative missense *TP53* alterations that may also display gain-of-function properties are more common than the second group comprising of truncating and frame shift mutations, partial or whole gene deletions that exhibit a loss of function. After over two decades of establishing the association of *TP53* mutations with LFS, the genotype-phenotype correlation between the underlying *TP53* mutation and the clinical phenotype including the age-specific cancer risks, tumor type, pathological features, penetrance and expressivity, host and/or environmental factors is still not well understood [5, 13]. This demonstrates the need of patient-derived tissues as an important resource to understand the genotype-phenotype correlation. To model LFS, mouse models representing the frequently mutated *TP53* alleles have been utilized, however, they do not represent broad spectrum of mutations observed [14, 15]. Only few of the most frequent mutation are studied for their contribution in cancer progression, and since the missense *TP53* mutations that give rise to mutant p53 proteins with unique features (neomorphic gain of function) [16, 17], mechanisms of p53 driven cancer progression are not been completely understood [18]. Recently, the use of Li-Fraumeni Syndrome patient iPSC-derived osteoblasts as a model to study the role of mutant p53 in development of osteosarcoma was reported [19]. Early-onset breast cancers are the most common cancer type in women (>70% cases) with LFS and this underscores the need for preventive measures such as bilateral mastectomy given limited management options [20, 21]. Since preventative surgical procedures such as oophorectomies and bilateral mastectomies are frequently recommended to predisposed ovarian and breast cancer individuals [22, 23], the use of primary cells derived from the excised tissues to generate cells lines therefore offers an avenue to model cancer predisposition.

1.2. Cancer Transcriptomics: burgeoning applications of the computational toolbox

Translational importance of gene expression analysis

Pathologists have characterized 18 distinct histopathological subtypes breast cancer (IARC 2012) and comprehensive molecular profiling has identified six clinically different intrinsic breast cancer subtypes (luminal A, luminal B, HER2- enriched, basal-like, claudin-low, and a normal-like group) characterized by distinct risk factors, incidence, and baseline prognosis, and treatment response [24-29]. Also, six unique triple-negative breast cancer (TNBC) subtypes (basal-like 1 and 2, immunomodulatory, mesenchymal, mesenchymal stem cell-like and luminal androgen receptor) with unique molecular profiles and ontologies have been identified [30]. The numerous published breast cancer gene-expression signatures aim to improve upon the prognostication provided by traditional clinical and pathological information [31]. The molecular subtyping of breast cancer has been used to accurately predict clinical response [27, 32-36], facilitate identification of novel treatment regimens and appropriate patient selection for clinical trials. Thus, gene expression profiling based classifications of breast cancer reveal complexity and molecular heterogeneity and presents an opportunity to tailor individualized therapies to a patient's tumor subtype.

Gene expression profiling classifies breast cancer into intrinsic subtypes based on the biology of the underlying disease pathways [26, 31]. Gene expression profiling is one of the robust approaches to discover novel predictive biomarkers and molecular targets and is the basis for rational pharmacological targeting, biomarker identification and

chemoprevention strategies. In the past decade, the important contribution of gene expression profiling to prognostic value in clinical settings has been widely acknowledged [27, 33, 37-43]. Recently, integrated molecular analysis of breast carcinomas has yielded a comprehensive catalogue of likely drivers in the four main breast cancer subtypes namely Luminal A, Luminal B, HER2-enriched and Basal-like has been published [44]. Furthermore pathway analyses of gene expression data have the potential to aid in specific therapeutic targeting. To overcome the limitations of cytotoxic chemotherapy, current cancer bioinformatics and systems biology approaches aim at responder identification and selective targeting of specific cell signaling pathways crucial for tumor growth and survival. This strategy requires mapping pathway deregulation patterns to therapeutically target and predict drug response in tumors using the gene expression signatures. The prospective use of MammaPrint® breast cancer recurrence signature to accurately determine beneficial versus ineffective chemotherapeutic options for breast cancer patients [45], thereby reducing the time, unnecessary cost and drug toxicities arising from a trial and error approach is an excellent example of how gene expression signatures can be translated in the clinic. Although some challenges exist in translating gene expression signature in the clinic, there have been promising advances in the past decade toward stratified personalized approach to patient treatment [46].

The use of genomics to understand genotype-phenotype correlations, distinct pathways, and biological principles of inherited cancer and the extrapolation of these finding to sporadic cancers and cancer prevention has been attempted [47]. Previously, Herbert et al. have performed gene expression profiling of breast epithelial and stromal

6

cells derived from two different LFS patient-derived and identified plausible targets for chemoprevention based on the differences in mutation type [48]. In summary, pathway deregulation analyses of samples derived from individuals with familial or inherited cancer syndromes can lead to prediction of therapy responders, molecular targeting of specific signaling pathways and development of rational combination therapies.

1.3. Relevance of cell lines as preclinical models for cancer research

A comprehensive drug screening project using a panel of 60 human cancer cell lines representing 9 different cancer types was undertaken in 1980s to perform primary high-throughput screening before progressing to xenograft models [49]. Since this endeavor did not meet expectations of identifying prominent drug candidates, a system utilizing hollow fiber assay using 12 cancer cell lines was used as an *in vitro* model as a preceeding step to novel drug testing in xenograft models [50]. The denunciation of cell lines as preclinical tumor models was owed to changes in genetic and transcriptional profiles, failure to represent tumor heterogeneity and lack of tumor microenvironmental components. Most of these limitations were due to improper tissue culture practice and in the recent years it has been shown that cell lines do in fact represent the features of tumor source. Also, with the advent of novel 3D culture models and capabilities of co-culture methods, some components of the microenvironemnt can be simulated *in vitro* without compromising on the economic and user-friendly applications of cell lines as preclinical models. The two-dimensional (2D) *in vitro* monolayer cultures largely retain genomic features of parent tissue [51, 52] and are ideal high-throughput screening platforms in

preclinical setting. During the past decade, molecular heterogeneity and need for classification of breast cancer has been acknowledged and alternative targeted therapies are being investigated [53]. Advances have been made in understanding major signaling pathways and developing new drugs yet unresolved questions, biomarker and therapeutic targets identification still remain. We expect that molecular classification on the basis of gene expression profiles for of non-malignant and tumorigenic breast cell lines from high-risk individuals will guide rationale targeting and biomarker prediction for use in preventative or therapeutic strategies to reduce the incidence of breast cancer. Recently, through the Cancer Cell Line Encyclopedia (CCLE) project, a comprehensive genetic and molecular characterization of about 1000 human cancer cell lines was performed [54]. The CCLE enables public access to mRNA expression, chromosomal copy number and mutation data for analysis and visualization with the goal of developing gene-expression based predictions of preclinical drug sensitivity and response. Use of human cancer cell line panels and genetic lesions dependent transcriptional signatures have been fruitful in studying heterogeneity and various subtypes in cancer [52, 55, 56]. The Cancer Cell Genome Atlas project (TCGA) project is another gargantuan effort ot characterize different form of cancer and their subtypes from patient-derived biospecimens to facilitate personalized medicine [57, 58]. Human clinical trials are indispensable and expensive, so the design of an ideal trial that reflects heterogeneity at genetic level of a population for promising preclinical drugs is restricted based on ethical, regulatory and econmic level [59]. This is especially true for cancer predisposition syndromes wherein the exisitng knowledge for chemopreventive or clinical management is inadequate.

High-risk patient-derived cell line models can serve as renewable resource in preclinical evaluation of novel cancer therapies and drug response prediction. To model breast cancer progression, a few breast epithelium derived series exist (for example, MCF10AT [60, 61], HMT-3522 [62], and the human 21T series [63]) that model different stages of cancer *in vitro*, however the LFS patient-derived HME50 cell lines are innovative in that they are the only breast epithelial cell line series from a high-risk patient; in addition, stromal cells from this patient have been cryopreserved. High-risk patient-derived cell line models serve as a renewable resource for preclinical evaluation of novel cancer therapies and drug response prediction.

1.4. Relevance of preclinical three-dimensional *in vitro* models

The weak correlation between preclinical screening and clinical efficacy of anticancer agents reflects the limitations of preclinical screening models to accurately predict clinical response. Ideally, preclinical studies should aid early identification of unpredictable toxicities and lack of efficacy [64]. Also, information on pharmacokinetic/pharmacodynamic properties of a drug, tissue concentrations and target modulation by the agent (or a surrogate biomarker predictive of drug action) are important [64]. The discovery of novel cancer therapies relies on selection preclinical models that recapitulate the heterogeneity and acknowledgment of limitations of each model [65]. The two-dimensional (2D) *in vitro* monolayer cultures largely retain genomic features of parent tissue [51, 52] and are ideal high-throughput screening platforms. However, monolayers cannot accurately recapitulate physiological environment and

complexity of cancer. Xenograft models are vital tools to investigate new

pharmacological agents but unreliable for drug eficacy studies due to inherent

interspecies differences. Xenograft models poorly predict response to therapy in humans

and are not expedient for proof-of-principle experiments for molecular targeted therapies.

Also, due to space, time and cost considerations, mouse xenograft models are unsuitable

for high-throughput screening [66].

The three-dimensional (3D) *in vitro* models although relatively new, are

physiologically relevant models that mimic morphology and signaling, are amenable to

rapid experimentation and can complement 2D cell cultures and xenograft models [67].

Three-dimensional (3D) Matrigel® assays have been used to study human breast-tumour

cell lines at different stages of progression in laminin-rich extracellular matrix (lrECM)

[68]. When grown in monolayers cultures, the nonmalignant cells are similar in

appearance to the malignant cells; however, the phenotypic differences become obvious

in 3D cultures. In 3D cultures, the non-malignant cells undergo growth arrest and form a

polarized, acini-like structure (**Figure 2**), whereas the malignant cells proliferate and

form amorphous structures [69, 70].

Bissell and others have shown that reversal of tumorigenic phenotype of the

malignant cells can be achieved by inhibiting different signaling pathways (inhibitory

antibodies and pathway inhibitors) and normal phenotype can be restored [71, 72].

However, the signal transduction pathways in 3D culture of non-malignant cells are not

always preserved in cells grown as monolayers [72]. Since tumor cells growing in 3D

cultures more closely mimic their counterparts *in vivo*, aspects such as invasive potential, changes in polarity, and drug sensitivity can be better studied in 3D cultures. Furthermore, a comparison between drug responses in 2D and 3D cell culture systems and *in vivo* drug responses must be determined [66, 73] to demonstrate that drug sensitivity data derived from 3D cultures captures clinically relevant response more faithfully than traditional 2D cultures.

During the past decade, molecular heterogeneity and need for classification of breast cancer has been acknowledged and alternative targeted therapies are being investigated. Advances have been made in understanding major signaling pathways and developing new drugs yet unresolved questions, biomarker and therapeutic targets identification still remain. We expect gene expression based molecular classification of non-malignant and tumorigenic breast cell lines from high- risk individuals will guide rationale targeting and biomarker prediction for use in preventative or therapeutic strategies for invasive breast cancer. We will use *ex vivo* 3D Matrigel® cultures to test investigational agents and we anticipate amenability of this example to other high-risk patient samples where gene expression analysis will guide chemoprevention and treatment regimens on the basis of a match between cell/sample type and underlying pathways [31].

3D culture offers an avenue to use both biologically derived and well-defined synthetic matrices and to incorporate cells types from the tumor microenvironment in 3D co-cultures. Some of the many differences observed in growth of tumor cells in 3D cultures in contrast to 2D monolayers are: (i) morphology similar to tumors *in vivo* [74]; (ii) slower growth rates [75] that reflect mathematical models of tumors *in vivo* [76]; (iii) increased glycolysis in 3D [77]; (iv) differential expression [78]. Some of the important reasons for using 3D matrices for anti-cancer drug development and need for advancement in 3D cell culture field are [79]: (i) applications in clinical setting that emphasizes personalized medicine; (ii) to overcome issues of misleading drug sensitivity data that arise due to limitations of 2D monolayer culture and animal models [67] and (iii) cells grown in 3D culture may better represent native behavior and expression profiles of cells in their *in vivo* environment [80]; both the ECM cues and the mechanical properties provided by 3D *in vitro* systems affect the behavior and gene expression [81].

Anatomy of breast

H&E staining of breast lobule

Fat pad

Lobule

Ducts

Connective tissue

BM
hemidesmosone
α6/β4 integrins
tight junction
ZO-1
myoepithelial cell
luminal cell

in vivo

3D culture
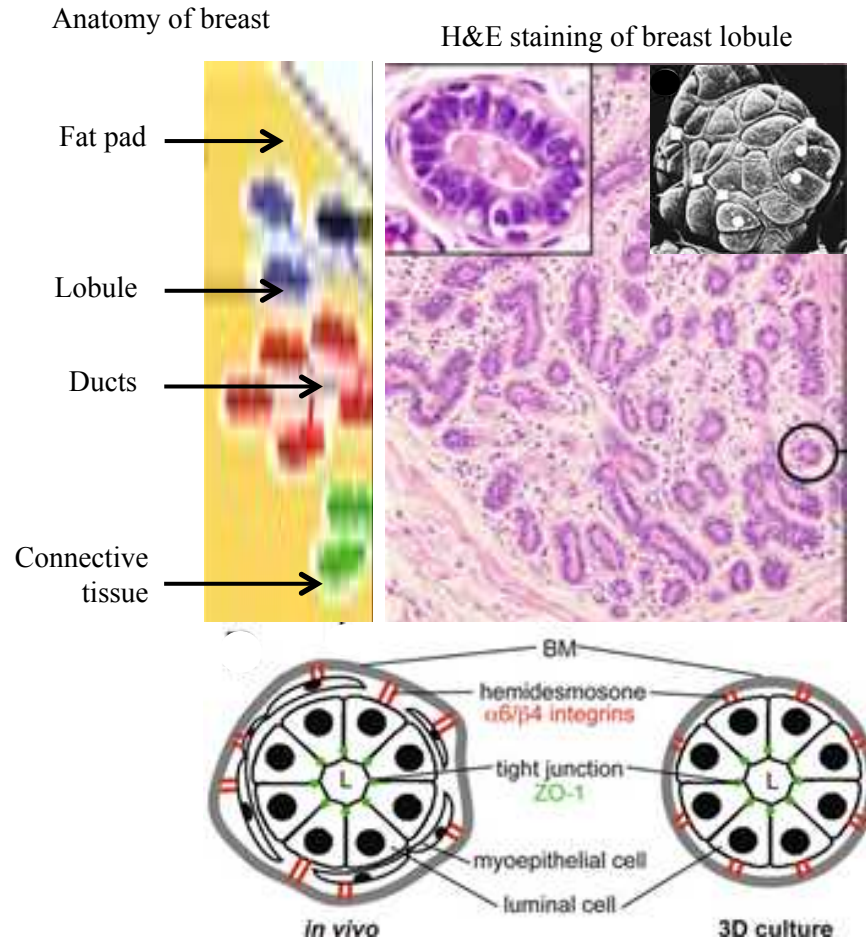
Figure 2. Schematic representation of human breast anatomy and *in vitro* epithelial acini. The human breast gland comprises of both epithelial and stromal components. The functional epithelial component is the ductolobular system comprising of about 15-20 lobes radially arranged and connected by a ductal network. Each lobe contains multiple lobules and each lobule is composed of acini. Each acinus rests on basement membrane and comprises of epithelial cells that surround a lumen. Breast epithelial cells when grown in 3D culture in presence of basement membrane components can adopt acinar morphology as observed *in vivo* (adapted from Vidi, P. et. al., 2013).

1.5. Epithelial-to-mesenchymal transition in cancer development

In cancer progression, epithelial-to-mesenchymal (EMT) plasticity is characterized by loss of polarity and loss of cell-cell adhesion that alters cytoskeletal properties such that epithelial characteristics are lost with concurrent acquisition of mesenchymal phenotype. Epithelial cells such as those derived from glandular breast form polarized, continuous layer bound by cell-adhesion molecules (such as claudin, occludin and E-cadherin) which rests on basement membrane that facilitates connection with extracellular matrix such that the apical-basal polarity is maintained. Loss of epithelial characteristics and acquisition of mesenchymal migratory and invasive program during EMT are initiated by loss of cell adhesion junctional proteins, reduced cell-to-cell contact, loss of cell polarity and adoption of mesenchymal transcriptional pattern. This process embodies loss of expression and function of epithelial markers such as E-cadherin and tight junction proteins (*ZO-1*, *OCLN*, *CLDN*), cytokeratins (*KRT19*, *KRT7*), Type IV collagen along with up-regulation of mesenchymal hallmarks such as N-cadherin, fibroblast specific protein 1 (*FSP1*), vimentin (*VIM*), type I and type III collagens, fibronectin (*FN1*) that together confers the transdifferentiated cells to invade and migrate through the extracellular matrix. Several transcriptional and signaling pathways that respond to external stimuli can operate the EMT switch and this reflects the complexity of multi-tiered crosstalks between regulatory factors and environment [82]. Some well characterized EMT mediating signaling pathways such as bone morphogenetic protein (BMP), epidermal growth factor (EGF), fibroblast growth factor (FGF), platelet derived growth factor (PDGF), transforming growth factor β (TGF-β),

integrin signaling etc. can induce EMT driving transcriptional programs via intracellular kinase signaling cascades [83, 84]. Several transcriptional factors that can induce and drive EMT have been characterized including basic helix-loop-helix (bHLH) factors (*ZEB1*, *ZEB2*, *TWIST*), zinc-finger binding transcription factors (*SNAI1* and *SNAI2/SLUG*), T cell factor (TCF)/Lymphoid enhancer binding factor-1 (*TCF/LEF-1*). The EMT transcriptional regulators SNAIL and *bLHLH* can repress the expression of epithelial markers claudins, occludins, E-cadherin, desmoplakin and concurrently activate fibronectin, vitronectin and N-cadherin associated with mesenchymal phenotype. The E-cadherin down-regulation destabilizes the adherens junction, repression of claudins and occludins causes disbanding of apical tight junctions, and the loss of desmoplakin and plakophilin results in loss of desomosomes. Either of the three major EMT transcriptional regulators namely *SNAIL*, *bHLH* and *ZEB* can repress E-cadherin expression and activate expression of N-cadherin, irrespective of the upstream signaling modules adopted by the cell. This repressed E-cadherin expression accompanied by cadherin switching by mesenchymal neural N-cadherin up-regulation is a critical event during EMT that alters cell adhesion that facilitates migration and invasion [83, 85]. Since EMT in cancer progression imparts metastatic potential of transformed cells with tumorigenic potential [86], EMT promoting signals serve as therapeutic target for inhibiting cancer cell dissemination resulting in localized disease that is amenable to surgical or radiotherapeutic intervention [87]. The function of tumor suppressor TP53 is compromised by either mutation or loss on over half of human cancers and is associated with poor prognosis [88]. The transcriptional factor p53 is a master regulator of genes that regulate pathways responsible for cell adhesion, ECM interaction, senescence,

15

migratory potential, anoikis, invasion, cell plasticity/stemness [9, 89] and so the loss of p53 function primes cells to sustain EMT, invasion and metastatic potential [90-93]. Studies have shown that cancer cells preferentially express mutant p53 relative to functional WT p53 thereby suggesting dependence on the gain-of-function properties of mutant p53 [10, 17, 90, 94]. Since the loss of wild-type p53 and mutant p53 contributes to EMT and metastatic processes, therapeutic targeting of mutant p53 and restoration of wild type p53 function may therefore be a promising anticancer strategy [18, 95-97].

1.6. Rationale and overall objective

For over 50 years, patient-derived breast cancer cell lines have proved indispensable as preclinical models for drug discovery and development. Although well characterized, these frequently used continuous cell lines are limited in number and fail to recapitulate clinically observed complexity and heterogeneity of breast cancer. In addition to the dearth of normal and pre-malignant cell lines that can model early disease stages, there is a limitation on the number of well-characterized cell lines that represent progressive stages of breast cancer derived from individuals predisposed to cancer.

By utilizing the tissues from individuals predisposed to cancer who harbor a germline mutation and opt for preventative surgeries to establish cell lines, an important step of *in vitro* (e.g. viral mediated) manipulation of tumor suppressors or cancer drivers can be bypassed. These individuals with unique mutations can thus be represented in preclinical studies by using patient-derived cells to model disease biology. The breast epithelial HMECs and stromal cells derived from high-risk individuals and women with no history of breast cancer that represent different ethnic or biological backgrounds can partly address the complexity and heterogeneity of breast biology and cancer to facilitate cancer prevention and drug development. In recent years, the preeminence of 3D cultures as physiologically relevant models to study biology and drug discovery has been appreciated. Since 3D cultures can mimic several properties of original tumor and normal patient-derived tissue, molecular subtypes of breast cancer can be represented in high-throughput assays.

The goal of this study was to develop Li-Fraumeni patient-derived breast tissue as a renewable resource by characterizing established pre-malignant and malignant cell lines that can model different stages of breast cancer progression and then use gene expression profiling to identify targets for pharmacological interventions. Toward this goal, the objectives were:

1. Characterization of cell lines from the unaffected breast tissue of Li-Fraumeni syndrome patient by genetically manipulated to progress from immortalization, transformation and malignant stages *in vitro* as monolayer and 3D cultures.

2. Use of gene expression profiling to identify changes accompanied by sequential genetic manipulations, predict drug targets and observe effects of targeted pharmacological agents on the Li-Fraumeni syndrome cell lines *in vitro* 3D culture.

Significance

Although well-defined breast cancer cell lines are widely used for mechanistic and therapeutic studies, their predictive value in the clinical setting is limited by failure to adequately reflect breast cancer heterogeneity or *in vivo* morphology [64, 98, 99]. Traditional 2D cell culture does not accurately mimic the three-dimensional (3D) environment in which cells function; resulting in inaccurate data and prediction of response of cancer cells to chemotherapeutics. On the other hand, xenografts from patient-derived tumors have proven to be invaluable, effective models for chemotherapeutic screening and translating efficacy to clinical trials [99-101]. However, use of xenograft models is limited by high cost, time and species-specific differences that

affect pharmacologic and mechanistic studies. These factors have led to the development

and interest in standardization of 3D *in vitro* systems for efficacy testing of anti-cancer

drugs [102-106]. The full potential of the cells derived from breast tissue of an individual

with cancer predisposition or one with high risk for cancer as a preclinical model for

chemoprevention and therapeutic targeting studies will be realized upon characterization

to ensure expression of molecular targets of interest and its use to confirm drug-target

interaction and efficacy. It is therefore necessary to characterize preclinical models to

ensure molecular targets are being expressed in studies in order that novel agents that

offer high anticancer activity and decreased side effects can be integrated in treatment

regimens. Also, this information can be used to determine and evaluate the most

promising drug combinations for therapy. The development of target-orientated agents

with defined mechanisms of action calls for molecular characterization and physiological

relevance of models, in addition to well-defined pharmacodyanamic and pharmacokinetic

properties of therapeutic agents [107]. A thorough outlook of how this proposal fits in

with the overall goals of cancer therapeutic and preventative intervention research is

shown in the schematic (**Figure 3**) shows that gene expression and molecular profiling of

tumors to identify responders to specific therapies can guide patient selection for

chemopreventive studies and clinical trials. Advances made in breast cancer subtyping

and pathway analysis; preclinical 3D culture models and generation of patient-derived

xenografts (PDX) can provide subtype specific outcome data, help reduce costs and tailor

effective treatment regimens for patients. A multipronged approach using various

preclinical models will be necessary to overcome trade-offs in cancer research and

treatment.

Central hypothesis of the project

```
┌─────────────────────────────┐      ┌───────────────────────────────────────┐
│     Healthy Individual      │      │    Individual predisposed to cancer     │
└─────────────────────────────┘      └───────────────────────────────────────┘
     │              │                        │                    │
     ▼              ▼                        ▼                    ▼
┌──────────┐  ┌──────────────┐  ┌───────────────────────────────┐  ┌──────────────┐
│  Tumor   │  │   Normal     │  │  Predisposed but normal tissue │  │ Tumor tissue │
│  tissue  │  │   tissue     │  │                               │  │              │
└──────────┘  └──────────────┘  └───────────────────────────────┘  └──────────────┘
     │              │                        │                    │
     ▼              ▼                        ▼                    ▼
┌────────────────────────────────────────────────────────────────────────────────┐
│         Tissue Biobanking – preliminary exploratory and discovery               │
└────────────────────────────────────────────────────────────────────────────────┘
```

3D 1⁰ Organoid Culture

$3D\ 1^0\ Organoid\ Culture$

- Predictive and prognostic assays
- Preclinical therapeutic testing
Tool box: Genome editing methods

PDX

Knowledge based rational therapeutic design

3D culture of established cell line

Renewable resource for
- Exploratory and mechanistic studies
- Drug discover, development
- Validation Studies

Figure 3. Schematic of preclinical model development using high-risk patient-derived tissues. Both normal and tumor tissue samples derived from sporadic and high-risk cases are an invaluable resource for preclinical exploratory studies. Advances in biobanking, organoid culture and generation of patient-derived xenografts (PDX) in conjugation with genome editing methods have enabled unprecedented accessibility to patient samples for *in vitro* mechanistic studies.

CHAPTER TWO: Materials and Methods

2.1. Reagents

Cell culture reagents: To propagate HME50 cells for microarray experiments and characterization in 3D culture, HME50 cell lines were grown in M171 modified basal medium 171 (Gibco®, Life Technologies™, U.S.A, catalog #M171500) supplemented with 10 ng/ml recombinant Human Epidermal Growth Factor (Hu EGF) (Catalog #PHG0311L, Gibco® Life Technologies™), 0.4% bovine pituitary extract (Hammond Cell Technologies), 5 μg/ml insulin (#I1882, Sigma Aldrich), 0.5 μg/ml hydrocortisone (#H4001, Sigma- Aldrich), isoproterenol hydrochloride (#I5627, Sigma-Aldrich), and 5 μg/ml human apo-transferrin (#T1147, Sigma-Aldrich); malignant HMET were additionally supplemented with 5% Fetal bovine serum (#S11150, Atlanta Biologicals, GA) was added to HMET growth media only. 10X Hank's balanced salt solution (HBSS buffer), pH 7.50 was filter sterilized and used for maintenance of HME50 cells.

For pharmacological targeting and experiments to study phenotypic reversion of HME50 cells, serum-free and chemically defined H14 medium was prepared by supplementing DMEM/F12 basal media (#MT10090CV, ThermoFisher) with 2.6ng/ml selenous acid-sodium salt (#354201,BD Biosciences), 0.15IU/ml prolactin (Sigma-Aldrich, #L6520-1000IU), 0.1nM β-estradiol (#E2758, Sigma-Aldrich), 500ng/ml insulin (#I1882, Sigma-Aldrich), 1.4μM hydrocortisone, 10μg/ml human apo-transferrin and 10ng/ml recombinant Human Epidermal Growth Factor (Hu EGF).

3D Matrigel® culture reagents: Growth Factor Reduced Corning® Matrigel® Matrix (Catalog #354230; Protein concentration within range of 9.0mg/ml to 10.0mg/ml and Endotoxin <1.5EU/ml), MatTek glass bottom dishes (#P35G-1.5-20-C & #P35G-1.5-14-C MatTek Corporation), were used for experiments involving confocal imaging of 3D embed culture.

Immunofluorescence reagents: Phalloidin CruzFluor^{TM} AF594 conjugate was dissolved in DMSO as 1000X stock (catalog #sc363795, Santa Cruz Biotech). AffiniPure F(ab')$_2$ Fragment Goat Anti-Mouse IgG, F(ab')$_2$ Fragment specific (Jackson ImmunoResearch Laboratories, Inc. Catalog#115-006-006). The primary antibodies used were mouse monoclonal ZO-1 (#339100, Molecular Probes®), rabbit monoclonal E-cadherin (#3195, Cell Signaling Technology®), rabbit Anti-Connexin-43 (#C6219, Sigma-Aldrich). The secondary antibodies Alexa Fluor® 488 F(ab')2 fragment goat anti-rabbit (#A11070) and rhodamine red-X goat anti-mouse (#R-6393) were used at 1:1000 dilution in immunofluorescence buffer (IF buffer, see method below).

The 3D Matrigel® cultures were mounted using ProLong® Gold antifade (P3693) or Vectashield® (H-1200) mounting medium with 4',6-Diamidino-2-Phenylindole, Dihydrochloride (DAPI). Invitrogen normal goat (#10000C), mouse (#10410), and rabbit (#10510) serum were aliquoted and stored at -20 °C for use.

Buffer solutions used for immunostaining: Immunofluorescence (IF) buffer for phalloidin AF594 Conjugate actin staining consisted of 2% BSA, 1X PBS, 0.2% Triton-X, 4% Goat serum. Quenching after PFA fixation was carried out using 0.1M Glycine PBS. For antibody staining, the IF wash buffer (10X, 500ml) stock was made using 10X PBS, sodium azide (2.5g,NaN3), bovine serum albumin (5g, BSA), Triton X-100 (10ml), Tween-20 (2.5ml), volume adjusted to 500 ml and pH to 7.4.

The IF Blocking buffer used during primary blocking step prior to addition of primary antibody, was made by adding 10% goat serum to 1X IF buffer. To mask the immunoreactive mouse IgG antibodies present in Matrigel® or EHS, 1% F(ab')$_2$ was added to 1X IF buffer in addition to 10% goat serum and incubated for 3 hours in secondary blocking step before addition of primary antibody as recommended by [108, 109].

Pharmacological agents: The selective mutant p53 binding small molecule called p53 reactivation and induction of massive apoptosis (PRIMA-1) was purchased from Cayman chemical and a 10mM stock solution was made by using DMSO as solvent, aliquoted and stored at -20 °C. The methylated small molecule and structural analog of PRIMA-1 that reactivates mutant p53 is currently in clinical Phase Ib/II trial called APR-246 (PRIMA-1^{MET}) and was purchased from Tocris Bioscience (catalog #3710). Resveratrol (catalog #R5010) and epigallocatechin gallate (EGCG, catalog #E4143) were purchased from Sigma-Aldrich. Rapamycin was purchased from LC Laboratories (catalog #R-5000).

Sterile 10mM PRIMA-1 stock in DMSO and 100mM APR-246 stock in water were aliquoted and stored at -20 °C; sterile EGCG stocks were made immediately before use by dissolving in media; 1mM rapamycin and 5mM resveratrol were dissolved in DMSO, aliquoted, stored at -80 °C and filter sterilized before use.

2.2. The Li-Fraumeni syndrome HME50 cell progression series

The HME50 cells were derived from a 31-year-old Li-Fraumeni syndrome (LFS) patient's noncancerous breast tissue [110]. The patient harbored a germline missense mutation at codon 133 in exon 5 in one of the two alleles of the p53 gene (Met to Thr substitution [M133T]) that affects wild-type p53 protein conformation. These cells undergo crisis around population doubling (PD) level 50-60 and spontaneously immortalize with a 0.5-1 x10^{-6} frequency [110]. For immortalization and transformation experiments of the HME50 cells, Dr. Brittney-Shea Herbert previously performed retroviral vector mediated transduction of hTERT, HRasV12, or controls (pBabe or pLXSN empty vectors) and selection with 150 ng/ml puromycin, or 200 μg/ml hygromycin, for hTERT, HRasV12, respectively, as previously described [111-113]. The medium was changed every 2-3 days and cells were monitored routinely for mycoplasma. The cell lines have been tested for *TP53* mutations by conventional sequencing and characterization of cell surface markers performed to authenticate cell lines retained mutational and cell surface marker characteristics; mycoplasma testing by immunocytology or thermocycler was routinely performed to ensure cell lines were mycoplasma-free. The establishment of HME50 cell series is schematically represented in **Figure 4**.
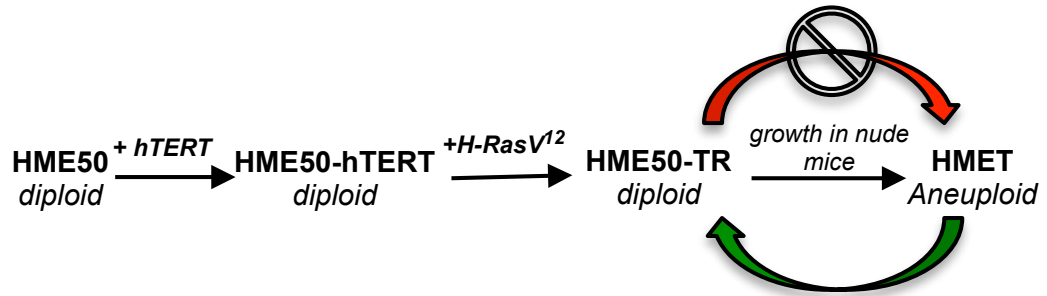
Figure 4. Establishment of HME50 cell progression series. Li-Fraumeni syndrome patient-derived HME50 cells harbor a heterozygous *TP53* [M133T] missense mutation and can spontaneously immortalize at a frequency of 5×10^{-7} (Shay JW et al., 1995) in culture (not shown). After successive immortalization and transformation of HME50 cells by addition of hTERT and HRasV12 to generate HME50hT and HME50TR cell lines, the transformed HME50TR cells capable of growth in soft agar and tumorigenesis in nude mice were collected and propagated as the tumor cell line HMET. The progression series can be useful for mechanistic studies and testing pharmacological agents that can inhibit (red arrow) or revert (green) the malignant progression.

2.3. Cell culture

HME50 cell lines were propagated and routinely maintained in aforementioned modified basal H14 medium (see cell culture reagents, 2.1) at 37 °C, 5% CO_2 and 95% humidity, media was changed every other day and cells passaged at 75% confluency.

2.4. Gene expression profiling of HME50 series

2.4.1. Microarray data import and normalization

RNA Extraction: To ensure that biological differences in each cell strain are much greater than differences resulting from technical manipulations of cell culture passage, quadruplicates of each cell strain were independently collected and isolated for RNA, according to the Center for Medical Genomics guidelines [42]. Total RNA was prepared from cultured cells using the Qiagen RNeasy kit by Dr. Brittney-Shea Herbert. All RNA samples were confirmed to have an A260/280 ratio of >1.8 by spectrophotometer and gel electrophoresis. Total RNA was diluted to a concentration of 1 µg/µl in RNAse-free water and 10 µg was given the Center for Medical Genomics for microarray processing. RNA integrity was further validated on an Agilent Bioanalyzer. All the samples showed distinct peaks corresponding to intact 28S and 18S ribosomal RNAs and therefore were included in the analysis.

Microarray Processing: Microarray processing was performed at the Center for Medical Genomics (CMG) at the Indiana University School of Medicine (IUSM). Gene expressions of the LFS HME50 series were profiled using the HG-U133_Plus_2 Affymetrix chip. Preparation of cDNA, cRNA, and labeling were carried out according to the protocols recommended by Affymetrix in the GeneChip® Expression Analysis Technical Manual (Affymetrix, Santa Clara, CA). Arrays were hybridized for 17h at 42 °C. The arrays were washed and stained protocol by fluidics stations controlled by GCOS software using the standard Affymetrix protocol and scanned using a dedicated Model 3000 scanner controlled by GCOS software. Following gene expression profiling of a total of 16 samples of HME50 series, the microarray data .cel files were imported followed by Robust Multi-array Average (RMA) normalization and quality control analysis as recommended for Affymetrix microarrays using the Partek® Genomics Suite®, version 6.6 ©; 2015 (Partek Inc., St Louis, MO, USA). The categorical attributes were then assigned to the microarray files to the random effects namely cell line and scan dates (**Appendix 1**).

Principal Components Analysis

Principal Component Analysis (PCA) was performed to visualize RMA normalized microarray data for global gene expression profiles of HME50, HME50-5E, HME50-hTERT and HMET cell lines. Each sample is represented by a single dot (*n=16*), each sample is indicated by distinct color for each cell line (*n=4*) and ellipse. The principal component #1 (PC1) depicted the highest variance between the cell lines.

Hierarchical clustering

The HME50 microarray data was preprocessed by filtering based on expression values for analytical hierarchical clustering to remove genes that have low variance across HME50 datasets. This effectively improved both the processing time by focusing on the most interesting genes that vary across the samples and by excluding genes that do not vary significantly from affecting the clustering pattern. The RMA normalized microarray data was filtered based on coefficient of variance (CV) parameter to exclude genes with CV less than 0.3 (i.e. "exclude genes CV <0.3"). This preprocessing resulted in a spreadsheet with 391 probesets with CV >0.3, which was subjected to hierarchical clustering using Euclidean distance for row/column dissimilarity measure that determined distance between samples, and average linkage as row/column method to determine the distance between two clusters. For gene expression normalization, the gene intensities were standardized such that mean equals zero and scaling the standard deviation equal one.

Applying the exclude CV<0.3 filter circumvented the use of gene list derived from Analysis of Variance (ANOVA) results that identified a large list of genes that vary between HME50 cell lines due to inherent differences in gene expression between the cell lines. This unsupervised hierarchical clustering was further used to discover groups based on the expression pattern of hallmarks of epithelial-mesenchymal transition.

2.4.2. Detection of differentially regulated genes

To identify differentially expressed genes amongst LFS HME50 cell lines, the microarray data was subjected to ANOVA analysis using Partek$^®$ Genomics Suite$^{™}$ 6.6 for characterization of LFS HME50 cell line series as breast cancer progression model. The categorical attributes were included as experimental factors in ANOVA with methods of moment to estimate variance components in this balanced study. To generate gene lists for individual contrasts included in the ANOVA design, threshold for significance of change of p-value with FDR <0.05 and size of fold change >1.3 or <-1.3 were used (**Appendix 2**).

2.4.3. IPA® pathway analyses

IPA® Core Analysis: For each list, Affymetrix ID was used as identifier, the GeneChip® HG U133 Plus 2.0 Affymetrix was used as reference platform, and to each identifier corresponding p-value and fold-change expression values were assigned per observation column. The statistical cut-off (p-value with step-up FDR <0.05; fold change >1.3 or <-1.3) for each ANOVA list was applied before uploading dataset to IPA®. Since these lists consisted of >3000 genes for IPA® analysis, the original ANOVA gene lists with statistical thresholds were further subjected to fold change cut-off of 2.0 to reduce the number of statistically significant genes for analysis (**Table 5**). Since the selection of IPA® eligible molecules depends on information present in Ingenuity® Knowledge base, and IPA® also resolved duplicate gene symbols and identifiers, the analysis ready molecule list may display fewer number of features than were originally uploaded.

Pathway analyses to identify deregulated functional networks, canonical pathways and to understand the implications of gene expression changes in each comparison (**Table 5**) was performed using Ingenuity® Pathways Analysis (IPA®; Build version 346717M; www.ingenuity.com). For HME50 progression series analysis, the Human Genome U133 Plus 2.0 Array was used as a reference set. Both direct and indirect relationships with high confidence prediction or experimentally observed were included for Core Analysis.

For network (and mechanistic) analysis, the differentially regulated genes between each comparison (p-value with FDR (step-up) <0.05 calculated using Benjamini and Hochberg method [114, 115] and fold change >1.3 or <-1.3 as thresholds) were formatted and uploaded into IPA® software (QIAGEN's, Ingenuity® Pathway Analysis, IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity). The analysis settings were set to flexible file format, Affymetrix as identifier type and Human Genome U133 Plus 2.0 array as reference set was used and both direct- and indirect relationships were considered. To determine the top biological functions associated with the gene expression profiles, we performed downstream effects analysis from which we focused on the top affected functions (overlap p-value <0.05, z-score >2.0 for activated regulator or increased function and <2.0 for inhibited regulator or decreased functions). The inference of IPA® results was based on overlap P-value that measures enrichment of network-regulated genes in dataset based on one-sided Fisher's Exact Test and the Z-score that assessed both significance of measure and predicts activation state of the regulator by measuring the match of observed and predicted up/down regulation patterns. We further

generated gene networks to identify molecular relationships and perturbed functions in an interactive mode and then overlaid the pharmacologic agents used in 3D phenotypic assay to identify the potential molecular targets. To explore the signaling pathways and genes differentially expressed in HME50 cell lines, we performed IPA® Core Analysis on gene lists for comparison guided by Venn diagram as well as for each contrast.

IPA® Canonical Pathways and Upstream Regulators: To gain biological insight into these differences, we focused on Analysis of Variance (ANOVA) gene lists with threshold of p-value with FDR <0.05 and size of fold change >2.0 or <-2.0 and performed IPA® analysis to study differences in molecular functions and canonical signaling. The IPA® functional analysis and canonical pathways measured the probability of association of focus genes in the study and given processes or pathway due to random chance; the smaller p-value denotes stronger association identified by statistically significant over-representation of focus genes in given process or pathways. For IPA® canonical pathways, the ratio for canonical pathways indicates which pathways overlap most with the genes in uploaded dataset and it is calculated by dividing the number of genes in the uploaded gene list that participate in a canonical pathway, by the total number of genes in that canonical pathway. The IPA® canonical pathway p-value measured the possibility of observing an association between a specific pathway and the uploaded dataset due to random chance, by also considering the total number of analysis eligible genes in dataset and reference set of genes. The calculated p-value is dependent on the size of the pathway and the p-value is more significant when the relative proportion of analysis eligible molecules in pathway in the pathway is greater. Therefore,

both the ratio that indicates the strength of association between a specific pathway and uploaded gene list, as well as the p-value that measures the statistical significance are considered for IPA® canonical pathways. The pathways with small p-values indicated significant association with uploaded dataset and ratios determine the pathways with most overlap with genes in uploaded dataset.

2.4.4. Gene Set Enrichment Analysis

Gene Set Enrichment Analysis (GSEA) [116] computationally determines whether an *a priori* defined set of genes has statistically significant and consistent differences between two phenotypes by examining the differences in expression values on comprehensive dataset without preprocessing to filter genes with low variance across HME50 datasets being compared. The (GSEA) v2.0.14 algorithm was used to analyze and review enrichment plots significantly associated with non-malignant phenotype (group comprising of HME50, HME50-5E, HME50-hTERT) relative to the malignant (HMET cell line) phenotype. Since GSEA algorithm does not filter the expression dataset, weighted statistic employed by GSEA algorithm ensures that genes with low expression values or variance do not contribute to the positive enrichment score and that they occupy the center of the ranked gene list, thus increasing the power of the statistic analysis. The GSEA algorithm was used as per instructions provided in Subramanian et al., 2005 (http://www.broadinstitute.org/gsea/index.jsp) as follows:

The expression dataset files: The HME50 progression series microarray data features (annotated genes) and expression values for each feature in each sample was saved as .txt file format. The microarray data was formatted to remove redundant identifiers, probeset IDs annotated to Human Genome Organization (HUGO) gene symbols as identifiers (feature) and consistently used as feature identifiers across all expression and gene datasets and this data was saved as .txt file. The expression dataset file contains features (genes/probes) and corresponding expression value for each sample in tab-delimited text file with .txt extension. The CEL files containing raw expression

values for each sample were merged, RMA normalized and each quadruplicate sample

was assigned cell line name followed by 1, 2, 3 and 4 rather than averaging the

expression values for the probesets to avoid redundancy. For example, the technical

quadruplicates for malignant HMET were labeled as HMET1, HMET2, HMET3 and

HMET4 and grouped together under malignant label. This resulted in dataset comprising

of 16 data column (quadruplicates for each cell line in HME50 progression series). After

uploading and specifying expression datasets, the parameter 'collapse dataset to gene

symbols' was specified to 'True' to indicate use of HUGO gene symbols in GSEA to

avoid inflated enrichment scores and facilitate the biological interpretation of GSEA and

leading edge results.

Phenotype label file: The phenotype labels define a discrete phenotype (e.g.

malignant and non-malignant) and associate each sample in the dataset with the

categorically labeled phenotype. The phenotype labels were assigned to dataset files 'on-

the-fly' in GSEA application dialog box to assign each dataset columns of HME50,

HME50-5E and HME50 to 'NMAL' for non-malignant and HMET datasets to 'MAL' for

malignant while setting analysis parameters in GSEA and the data was uploaded to

GSEA.

Gene set files: The MSigDb Gene sets files contain one or more curated gene sets

and the files can be used from Broad ftp site or exported from the publically accessible

Molecular Signature Database. Next, analysis parameters were ascribed and gene sets

obtained from the MSigDB [117] (C2, C3, C5 and C6 gene set collections) were selected

based on gene sets in each collection to compute overlaps between uploaded data for non-malignant and malignant phenotypes and gene sets available in MSigDB. Briefly, collection C2 (4725 gene sets), C3 (836 gene sets), C5 (1454 GO gene sets) and C6 (189 oncogenic signature gene set) comprise of curated gene sets from online pathway databases, motif genes based on conserved cis-regulatory motifs comprising of microRNA targets and transcription factor targets, GO term annotated gene set and oncogenic signatures defined from microarray gene expression data respectively. The GSEA analysis was performed after uploading the expression dataset, gene sets and assigning phenotype labels to each sample.

GSEA statistics and report: The Enrichment Score (ES) representing the primary result of GSEA is calculated in reference to a ranked list of genes, the running-sum statistic increases when a genes in present the gene set and decreases when it is absent and it reflects the degree to which a particular gene set is over-represented at the top or bottom of the ranked gene list. The magnitude of running-sum statistic increment depends on the correlation of the gene with the phenotype. The ES is maximum deviation from 0 encountered in walking the list (peak in the plot); and the most interesting genes are with positive ES indicates enrichment of gene set at the top of ranked list (peak in the beginning) and the negative ES reflecting gene set enrichment at the bottom of the ranked list (end of enrichment profile). The leading edge subset of genes contributes the most to the positive ES and appears in ranked list just before the peak while the negative ES genes appear after the peak score. The ranking metric (signal to noise ratio) measures genes correlation with phenotype and its value progresses from positive to negative while

descending down the ranked list. The positive ranking metric value indicates correlation with the first phenotype and a negative value indicates correlation with the second phenotype. The GSEA results are compared across datasets by primary statistic Normalized Enrichment Score (NES) that is based on GSEA ES for all dataset permutations and it accounts for differences in gene set size and in correlation between user-selected MSigDB gene sets and expression dataset. The false discovery rate (FDR) estimates the probability of a gene set with a given NES represents a positive result finding at random. The GSEA results with FDR <25% are most interesting and were considered for this study. The nominal p-value was used to estimate statistical significance of the ES; in GSEA report, a p-value of zero (0.0) indicates an actual p-value of <(1/# permutations) i.e. an actual p-value is <0.001 is reported if the analysis performed 1000 permutations. In this study, the statistical significance was ascertained by comparing the enrichment results generated from 1,000 random permutations of the gene set to obtain P values (nominal P value) and including GSEA results ranked according to the nominal P-value (<0.05) and false discovery rate (≤0.25).

2.5. Modeling stages of breast cancer and phenotypic reversion in 3D culture

2.5.1. Three-dimensional (3D) Matrigel® overlay culture and stereomicroscope imaging of HME50 cell lines

The Matrigel® (Engelbreth-Holm-Swarm (EHS) mouse sarcoma derived extra cellular matrix (ECM) protein-rich reconstituted basement membrane) aliquots were thawed at 4 °C overnight to coat the chilled culture surfaces (for example, dish or well) with an appropriate volume of "Matrigel®/EHS coat" for a thin layer of Matrigel® (**Table 1**) directly onto surface and spread evenly with a pipette tip/1-ml syringe plunger for smaller areas, or cell scraper for large surface areas. The coated surfaces were incubated for 20 minutes at 37 °C to allow the Matrigel® layer to gel without it over-drying in a humidified chamber/large petri dish. HME50 cell lines and cell lines used as controls (non-malignant MCF10A, ME16C and malignant MDA-MB-231 controls) cultured at 37 °C with 5% CO_2 were trypsinized from a monolayer to a single-cell suspension at 70% confluency. The cell number required for seeding was calculated (**Table 1**) and appropriate volume of cell suspension was aliquoted into a 1.5 ml microcentrifuge tube. The cells were pelleted by centrifugation at ~115 g, re-suspended in half the "medium volume" (**Table 1**) and plated onto surface coated with Matrigel®. The cells were allowed to settle and attach to the EHS for 20 min at 37 °C without disturbance with an intermittent agitation in the x-y plane at interval of 10 minutes during incubation. Long incubation (>25-30 minutes) and over-agitation of cells were avoided as they lead to the drying of Matrigel® layer and deposition of cells in clusters.

38

To perform drug response assays in 3D overlay culture, the addition of pharmacological agent depended on its mode of action. Depending on concentration and the cytotoxic nature of the pharmacological agent, the addition of the drug was carried out either during seeding by adding drug to cell suspension medium during seeding or after short incubation of 4-24 hours after seeding. Addition of small molecule inhibitor APR-246 and the antioxidant polyphenol flavonoid EGCG was carried out 3 hours post incubation by adding the agent directly to the medium instead of mixing along with cell suspension with Matrigel® and was repeated for all subsequent media changes for the duration of the culture. HME50 cells were treated with resveratrol, rapamycin and DMSO vehicle control post 12 hour of incubation and during subsequent media changes every other day. The remaining medium was chilled on ice and Matrigel® was added as per 10% total media volume (10% Matrigel® during initial plating) and this Matrigel®-medium mixture was gently added to the plated culture down the side of the well to avoid disturbance of the cells or Matrigel®. Care was taken that medium remains thoroughly chilled before addition of EHS to ensure homogenous mixing and even deposition of EHS onto cells in culture. This overlay 3D culture can be maintained for 4 days or more depending on endpoint of assay and completion of morphogenesis while the Matrigel®-medium (5% Matrigel® total media volume) mixture is replaced every 2 d. For better resolution during imaging and to avoid excess Matrigel® deposition, only 5% of Matrigel® per well was sufficient during subsequent media changes.

The 3D overlay culture method is an economic alternative to the 3D embedded assay for performing drug-response assay, as it reduces duration of morphogenesis (7-8 days versus 10-12 days) and Matrigel® use, and facilitates stereomicroscope imaging for acinar size determination as cells seeded on the coated surface form acini in a single plane. Therefore, the on-top assay was ideally used for time-lapse imaging and also for in situ immunostaining of cell lines that form invasive stellate structures in 3D. The number of cells to be plated per square centimeter of culture surface area has been optimization based on the growth properties of each of the HME50 cell lines (refer to **Table 1**). If incubated too long after plating on Matrigel® coat, HMET cells tend migrate across the surface leading to crowding in certain areas. Secondly, if HMET cells are seeded in higher concentration, toward the completion of morphogenesis they tend aggregate with one another forming gross multi-acinar "nodular" structures (>1mm in size) due to their invasive properties. The recommended cell numbers for plating (refer to **Table 1**) ensures the completion of deregulated morphogenesis of HMET cells resulting in distinct "stellate" like cells which can be easily visualized and characterized.

Table 1. Standardized protocol for Matrigel® 3D Overlay culture of HME50 cell lines.

Dish	# of wells	Diameter (mm)	Area (cm^2) Per well/ chamber	3D on-top			HME50 cells/cm^2	
				EHS coat (μl)	Half Medium volume (μl)	10% EHS (in medium) (μl)	Non-malignant 25,000 cells/cm^2	Malignant 17,500 cells/cm^2
Petri	------	60	28.3	850	2,500	500	500,000	350,000
Plates	6	35	9.6	500	1,000	200	180,000	120,000
	24	16	2.0	120	250	50	37,500	25,000
	48	10	0.75	80	100	20	14,000	95,00
	96	6	0.26	15	30	6	5,000	3,500
Slides	4	-------	1.8	120	250	50	33,750	24,000
	8	-------	0.8	90	100	20	15,000	10,500

The table shows the cell numbers used for 3D overlay culture of HME50 cell lines; for seeding, the non-malignant group comprised of HME50, HME50-hTERT, HME50-5E and HME50-TR* whereas HMET indicate malignant cell line (modified from Lee et al., 2007). (Based on the growth properties of HME50-TR, cell densities used for non-malignant cells were also used for transformed HME50-TR which were considered as "non-malignant" for calculation of seeding densities)

2.5.2. Modified 3D Matrigel® embed culture and confocal imaging of HME50 cell lines

The Matrigel® (EHS/ECM) aliquots were thawed at 4 °C overnight to coat the chilled 20mm glass-bottom culture (**Figure 5**) surface with 50µl of "EHS coat" for a thin layer of Matrigel® (**Table 2**) with a pre-chilled cell scraper/lifter. These coated glass-bottom dishes were incubated for 3 minutes at 37 °C to allow the EHS to gel without over-drying. For small volumes and thin coats of Matrigel® as used in this protocol, it was ensured that the plates do not over incubate (< 3 minutes) which results in drying of Matrigel®. For the entire duration of experiment starting from this step, the glass-bottom dishes were placed along with 35mm sterile petri-dish containing sterile PBS/water without lid to maintain humidity in the large 100mm sterile carrier dish. This is crucial to prevent media evaporation and drying of Matrigel® surface. HME50 cells were trypsinized from a monolayer to a single-cell suspension as previously described and strictly healthy cells at less than 75% confluency were used. The appropriate cell suspension volume based on number of cells to be plated was aliquoted into a 1.5 ml microcentrifuge tube, gently centrifuged at ~115g and then re-suspended in 200µl chilled Matrigel® (on ice) without introducing air bubbles. This cell-Matrigel® suspension was pipetted evenly onto the precoated surface and incubated for 6 minutes at 37 °C ensuring Matrigel® to gel in humidified 100mm carrier petri dish. For embed culture, 2ml of chilled H14 media (supplement with 5% FBS for HMET) was added.

This modified 3D embed method allowed high-resolution confocal imaging and economic use of Matrigel® in a complete 3D format which is physiologically more relevant than the "overlay" (2.5D culture) format. All pharmacological agents were added

in appropriate concentration and cultures re-fed after 12 hours of seeding. The 3D embedded Matrigel® cultures were maintain for 10 days, with media changes and appropriate concentration of drug treatment every other day.

Table 2. Standardized protocol for Matrigel® 3D Embed culture of HME50 cell lines.

Area (cm^2) of glass coverslip	3D Embed		Recommended cell numbers	
	EHS coat (µl)	EHS plate (µl)	Non-malignant: 0.6 x 10^6/ml	Malignant: 0.45 x 10^6/ml
28.3	50	200	120,000 cells ∕ 200 µl EHS	90,000 cells ∕ 200 µl EHS

The table shows the cell numbers used for 3D embed culture of HME50 cell lines for 35mm glass bottom dish format; for seeding, the non-malignant group comprised of HME50, HME50-hTERT, HME50-5E and HME50-TR* whereas HMET indicates malignant cell line (modified from Lee et al., 2007). (Based on the growth properties of HME50-TR, cell densities used for non-malignant cells were also used for transformed HME50-TR that was considered as "non-malignant" for calculation of seeding densities).

Uncoated 35mm dish MatTek™ glass bottom dish (No. 1.5) coverglass

50μl GFR-Matrigel™ coat,
200μl Matrigel™ embed

Laser excitation
Zeiss LSM-700
confocal microscope
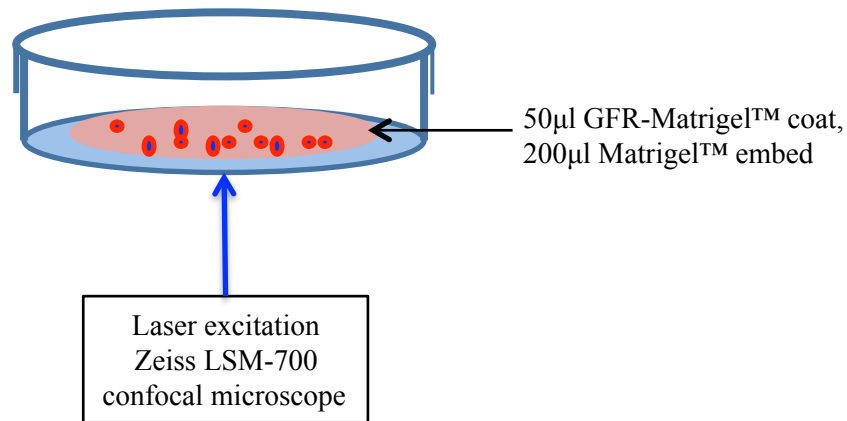
Figure 5. Schematic representation of the modified 3D Matrigel® embedded culture method using glass bottom dishes. Use of glass bottom dishes allows laser excitation and imaging of acini embedded close to glass coverslip without disturbing the acinar organization. The use of embedded culture better preserves the complex mechano-signaling and growth characteristics as compared to the 3D overlay culture format.

2.5.3. Treatment of HME50 cell lines with pharmacological agents

Working solutions from stock solutions of APR-246, rapamycin and resveratrol stored at -20 °C were prepared immediately before use. To make EGCG working stock, EGCG was weighed, dissolved in media and filter sterilized immediately before use. The methylene blue cell proliferation assay was used to determine IC50 values for PRIMA-1, APR-246, EGCG and resveratrol for each of the HME50 cell lines in 2D culture. These 2D culture based IC50 values were used to empirically treat the HME50 cells in 3D culture with following drugs at various concentrations: PRIMA-1 (20 μM), APR-246 (5μM, 10μM, 20 μM and 40 μM); with Rapamycin (25nM, 50nM, 75nM and 100nM; DMSO 100nM was used as vehicle control); resveratrol (2.5μM, 5, 7.5μM and 10μM; DMSO 10μM used as vehicle control) and EGCG (3.75μM, 7.5μM, 15μM and 30μM). All the drug treatments were repeated during media changes every other day. For APR-246 and EGCG treatment started on day 0, at 3 hours post seeding whereas rapamycin and resveratrol was added 12 hours post seeding in 3D overlay culture. All drug treatments for 3D embed cultures were initiated 12 hours post-seeding.

2.5.4. Immunofluorescence, image acquisition and statistical analyses of HME50 acini in 3D cultures

Actin staining of 3D Matrigel® (embed/overlay) cultures: After morphogenesis or treatment of 3D cultures, the media was aspirated and the Matrigel® surface was gently rinsed twice with 1X HBSS. The 35mm glass bottom dishes with 200µl Matrigel® on 20mm coverslip were fixed with 500µl of 2% PFA for 10 minutes with gentle intermittent rocking. The volumes were adjusted to 300µl of 2% paraformaldehyde (PFA) for fixation of 100µl Matrigel® on 14mm coverslip glass bottom dish. Care was taken during the fixation step to avoid over-fixation that depolymerizes Matrigel® and loss of acini as well as negatively affects antibody binding. Also, under-fixation of the culture that negatively affects the immunostaining of acini was avoided; generally, a visible conversion of gel to translucent gel was used as a good indicator for successful fixation and culture was closely monitored to further avoid loss of acini and depolymerization. Especially, HMET cells formed large, invasive, multi-acinar structures and needed to be closely monitored during fixation as Matrigel® depolymerization tends to occur relatively more due to the inherent ECM degrading property of HMET cells. Next, the cultures were quenched by three washed with 1ml of 0.1M Glycine PBS for 10 minutes at room temperature by gentle intermittent rocking. The quenching step in some cases was repeated as necessary until gel-like consistency of Matrigel® culture was regained satisfactorily. The cultures were then permeabilized by 500µl of 0.2% Triton-X PBS for 5 minutes at room temperature and then blocked for 2 hours with 1000µl of IF blocking buffer. Next, 1000µl Phalloidin AF594 conjugate at 1:9000 concentration was added with 2% BSA in 1X PBS to culture and incubated for 6 hours (sodium azide added if

incubating overnight). After incubation, three quick washes were followed by 4 washes for 15 minutes with 1X PBS. Excess moisture was dabbed off and the stained cultures briefly air-dried for 5 minutes before adding the Vectashield or Prolong Gold mountant with DAPI. These cultures were cured for at least 72 hours before confocal imaging.

Confocal Imaging: The fixed and stained 3D cultures mounted in ProLong Gold antifade reagent were imaged on Laser Confocal Scanning Microscope FV1000D (spectral type inverted Microscope) or Zeiss LSM-700 confocal microscope (inverted). To evaluate size and acini morphology, images and optical sections were acquired via an Apotome (Zeiss) X63 oil immersion objective (Zeiss). Images were processed with Zeiss Axiovision 4.8. Stereo-images were acquired using a Nikon SMZ1500 microscope.

CHAPTER THREE: Results

3.1. HME50 cell lines exhibit distinct morphologies and growth characteristics in monolayer and 3D Matrigel® culture

3.1.1. Characteristics of HME50 progression series in monolayer culture

The expression of hTERT in context of heterozygous mutant p53 (TP53 [M133T]) resulted in stable immortalization and extended life span of primary HME50 cells derived from the non-cancerous breast tissue of Li-Fraumeni patient. The patient-derived HME50 cells entered crisis at population doubling level (PDL) of 50-60. The spontaneous immortalization of these HME50 cells has been observed to occur at a frequency of 0.5-1 $x10^{-6}$ [110] in HME50 cells that escape the crisis stage and grow indefinitely beyond PDL 70. The successive immortalization and transformation strategy involved retroviral transfection of HME50 cells with hTERT and constitutively active HRasV12 as described previously [111-113]. The spontaneously immortalized HME50-5E cells exhibit a near tetraploid status but are incapable of anchorage independent growth or tumor growth in athymic mice. After retroviral transfection of HME50 cells with hTERT, diploid karyotype was maintained by the immortalized cells which lacked anchorage-independent growth potential and were non-tumorigenic in athymic mice. The non-malignant HME50, HME50-5E and HME50-hTERT cell lines in HME50 progression series adopted a cobblestone-like morphology at confluency, a characteristic of epithelial cells grown *in vitro* monolayer culture (**Figure 6**).

The successive HME50-TR cells that are transformed due to constitutively active H-Ras with active telomerase in p53 heterozygous background maintained diploid status while displaying anchorage independent growth and tumorigenic potential in nude mice when co-injected with Matrigel®. In contrast, the transformed pre-malignant HME50-TR cells showed a less than typical cobblestone-like appearance with pronounced elongated feature as compared to the non-malignant HME50 cells.

The HMET cells derived from tumors generated by injection of HME50-TR cells, exhibited near tetraploid status and tumorigenic potential independent of Matrigel co-injection. The mammary fat pad injection of HMET cells in athymic mice resulted in high-grade carcinoma evident in the sections of the orthotopic xenografts that exhibited high mitotic grade, inflammation and alveolar growth patterns (Herbert et al. unpublished observations). The HMET cell lines in culture showed a prominent spindle-shaped morphology typical of mesenchymal cell as well as microscopic protrusions and high nucleus/cytoplasmic ratio in 2D *in vitro* culture (**Figure 6**).

2-D culture   **3-D culture**

(rows labeled: HME50-5E, HME50-hT, HME50-TR, HMET)
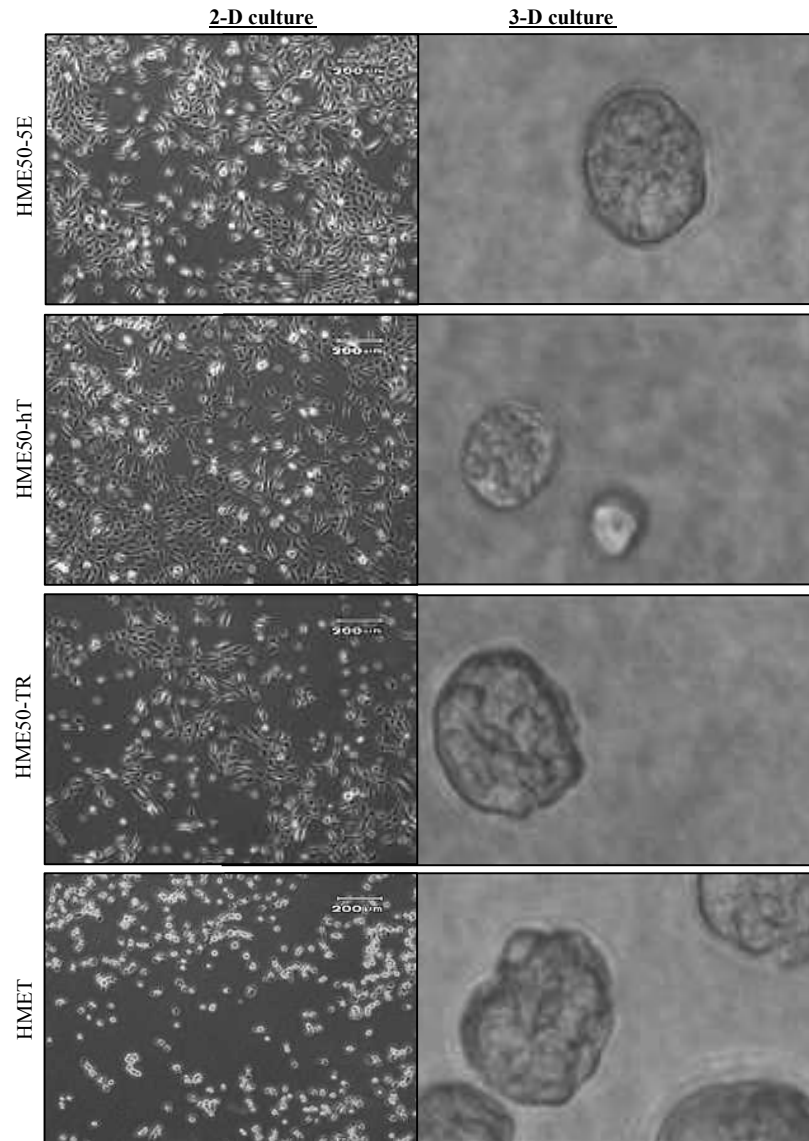
Figure 6. The morphologies of HME50 cell lines in monolayer versus 3D culture condition. Phase contrast images of non-malignant HME50-5E, HME50-hTERT, HME50-TR cells and malignant HMET cell grown as monolayer (2D; left column) or in 3D overlay culture format on day four after seeding (right column).

3.1.2. HMET cells adopt stellate-like acinar morphology in 3D culture

The morphogenesis and organization of glandular breast epithelial into polarized three-dimensional structures can be used to distinguish cells with or without malignant potential. In 3D Matrigel® culture, the diploid HME50-hTERT cells underwent a well-defined and orchestrated program of morphogenesis in 3D culture to form acinar structures of spherical morphology with nuclei radially organized around a hollow lumen and exhibited cell-cell adhesion typical of normal epithelial cells. The average HME50-hTERT acini size observed was 50 microns, and the range of between 31μm to 50μm after day 10 of morphogenesis; acini exceeding the size of ≥55μm are very unusual (**Figure 7-8**).

The spontaneously immortalized HME50-5E cell line also exhibited properties similar to HME50-hTERT cells upon completion of acinar morphogenesis in 3D culture. The average size of HME50-5E was measured to be 57μm, and the range observed was 50μm - 80μm (**Figure 9**). Also in agreement with the microarray analysis, the mRNA level for gap junction (connexin 43) protein was up-regulated compared to HMET suggesting presence of robust gap junctions in HME50-5E acini (**Figure 10**).

52
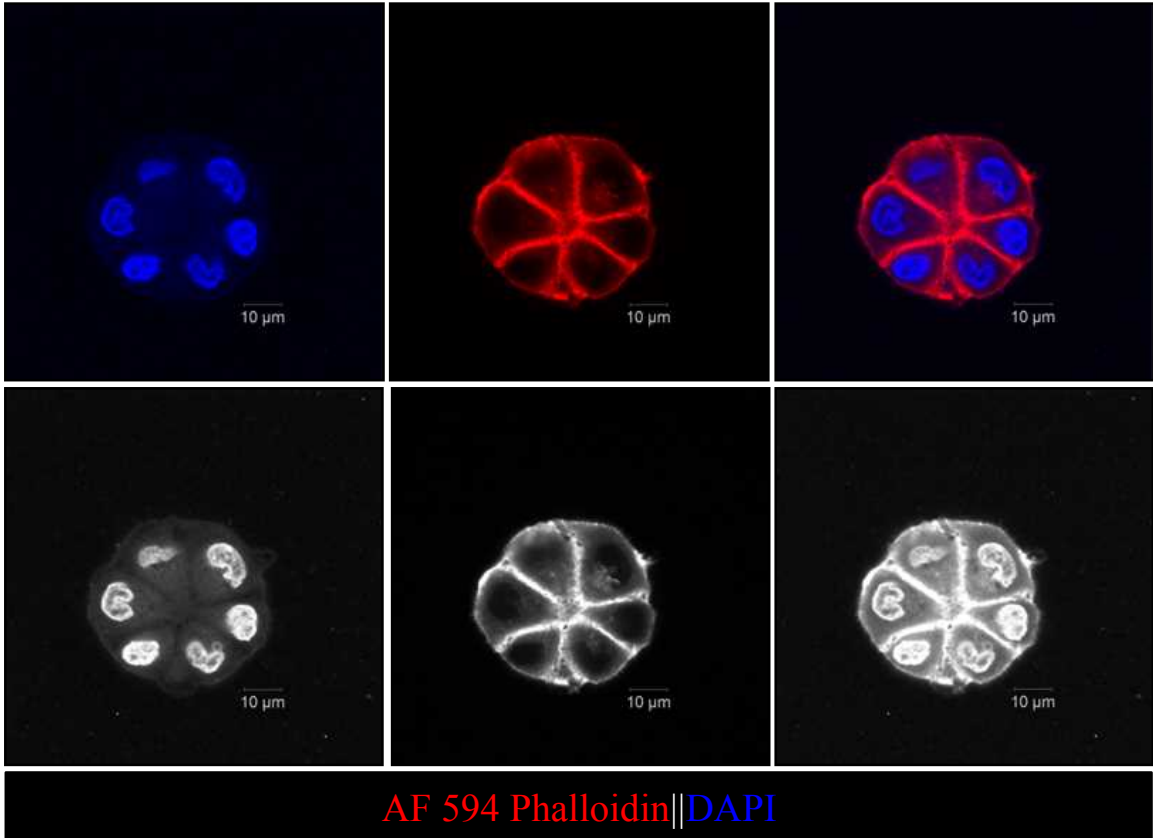
Figure 7. Non-malignant HME50-hTERT acini adopt spherical morphology in 3D embedded culture. Non-malignant HME50-hTERT cells embedded in 3D Matrigel culture grown for 10 days followed by F-actin and nuclear staining by AF 594 Phalloidin and DAPI respectively. The round morphology with radially organized nuclei and robust cell-cell adhesion is evident in the representative center z-stack of acini by confocal imaging, scale bar 10µm.
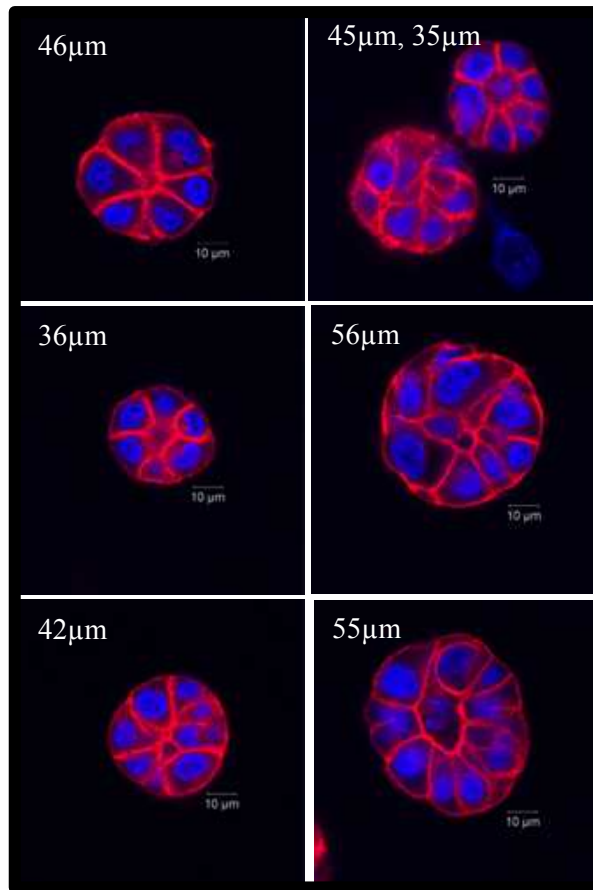
HME50-hTERT

Figure 8. Representative confocal images of center Z-stack of HME50-hTERT acini in 3D embedded culture. The sizes of HME50-hTERT were observed to vary between 31-50μm; the average size of HME50-hTERT acini observed to be 50μm. The feature of HME50-hTERT cells that phenotypically distinguished from its counterparts in HME50 series is relatively its smaller acini size and development of hollow lumen in the spherical acini. Representative confocal images show center z-stack stained with F-actin binding AF 594 Phalloidin (red) and nuclear stain DAPI (blue). Scale bar, 10μm
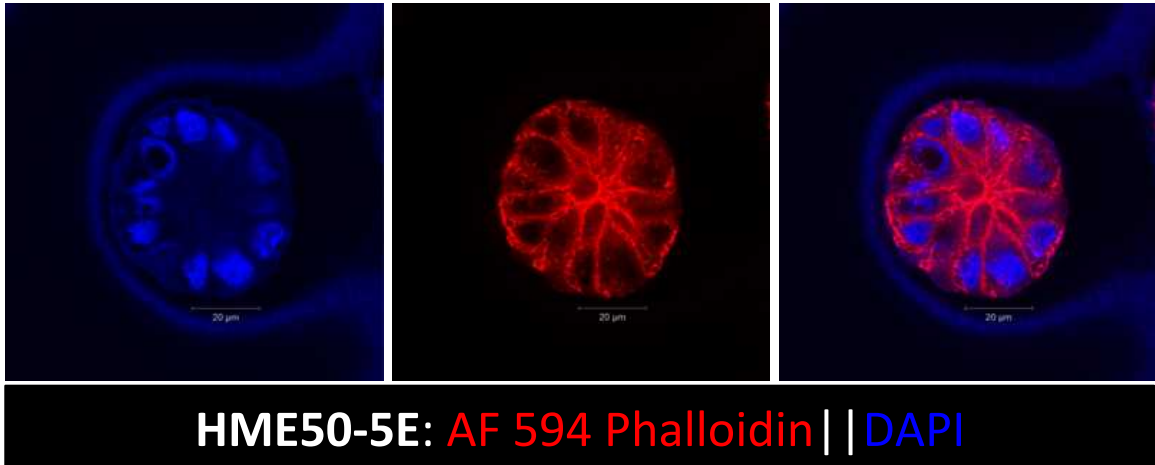
Figure 9. Representative center Z-stack confocal image of spherical HME50-5E acini. Spontaneously immortalized HME50-5E acini adopt spherical morphology in 3D embedded culture. HME50-5E cells embedded in 3D Matrigel culture grown for 10 days followed by F-actin and nuclear staining by AF 594 Phalloidin (red) and DAPI (blue) respectively. The spherical HME50-5E acini displayed organized nuclei, robust cell-cell adhesion and a hollow lumen evident in the representative center z-stack of the acini observed by confocal imaging. Scale bar, 10µm

HME50-5E : Cx43||HOECHST
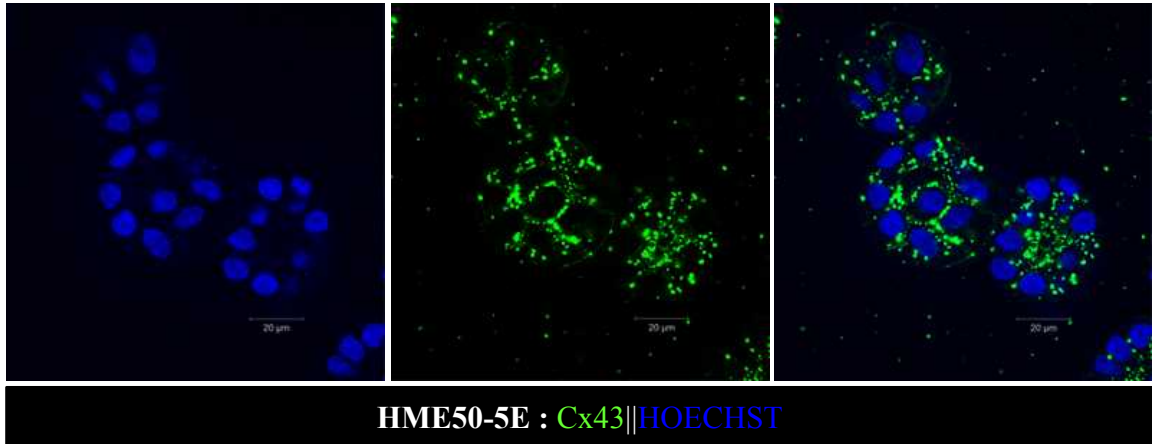
Figure 10. Spontaneously immortalized HME50-5E acini exhibit robust gap junction channels and hollow lumen. Immunofluorescence staining of spontaneously immortalized HME50-5E acini grown for 10 days embedded in 3D Matrigel culture. Connexin-43 (green) and Hoechst staining (blue) of HME50-5E acini showed well-organized nuclei and the centermost z-stack showed clear lumen showed HME50-5E cells while aneuploid, follow stages of normal epithelial morphogenesis. The immunofluorescence of HME50-5E for the gap junction marker Connexin-43 agrees with the HME50 microarray data described later.

The pre-malignant HME50-TR cells develop acini with mass-like morphology characterized by disorganized nuclei and lack of hollow, cleared lumen (**Figure 11-12**). The average size of HME50-TR acini was observed to be 78 microns, however, variation in HME50-TR size exists and structures ranging from 80-100 microns in size are not uncommon (**Figure 11**). This suggested lack of radial nuclear organization and hyperproliferation leading to cellular crowding resulting in multilayered acini. Also, HME50-TR acinar structures displayed cell-cell adhesion and expressed E-cadherin (**Figure 13**) but were considerably larger and relatively disorganized than HME50-hTERT and HME50-5E acini.
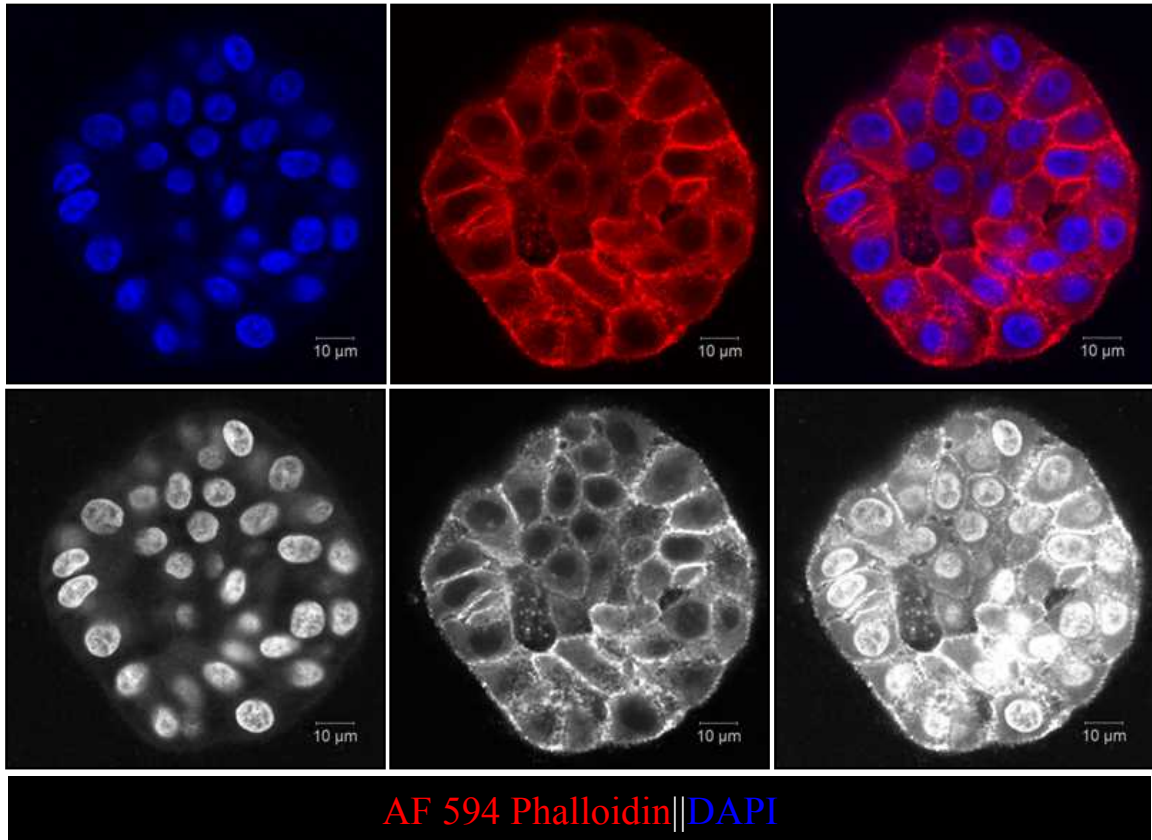
Figure 11. Pre-malignant HME50-TR acini adopt mass-like morphology in 3D embedded culture. Pre-malignant HME50-TR cells embedded in 3D Matrigel culture grown for 10 days followed by nuclear stain and F-actin staining by AF 594 Phalloidin and DAPI respectively. The mass-like morphology with disorganized nuclei albeit robust cell-cell adhesion was evident in the representative center z-stack of acini by confocal imaging, scale bar 10μm.
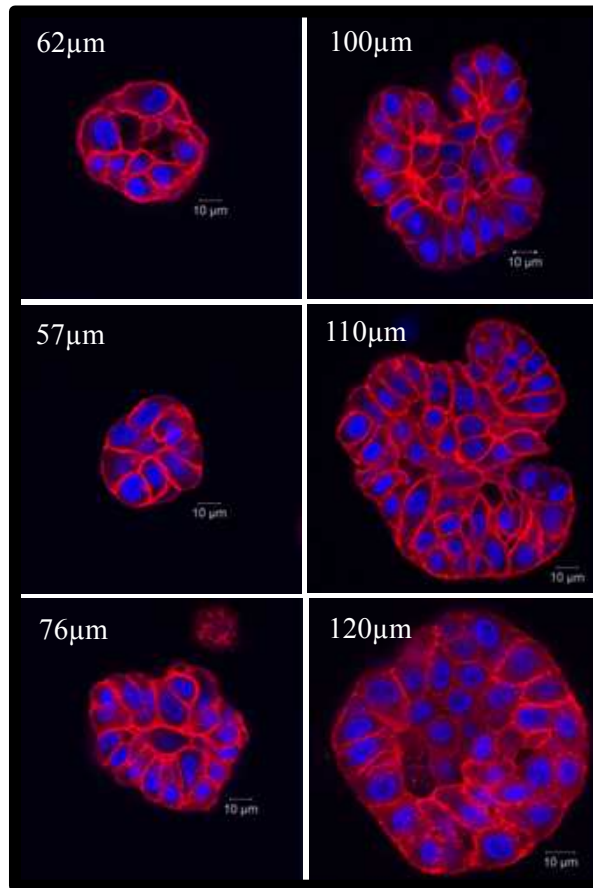
HME50-TR

Figure 12. Representative confocal images of center Z-stack of HME50-TR acini in 3D embedded culture. The sizes of HME50-TR acini were observed to vary between 80-100μm; the average size of HME50-TR acini observed to be 78μm. The feature of HME50-TR cells that phenotypically distinguished them from their counterparts in HME50 series was the formation of large acini by day 10 of morphogenesis and the lack of hollow lumen in the disorganized mass-like acini. Representative confocal images show center z-stack stained with F-actin binding AF 594 Phalloidin (red) and nuclear stain DAPI (blue). Scale bar, 10μm

HME50-TR : E-cadherin||HOECHST
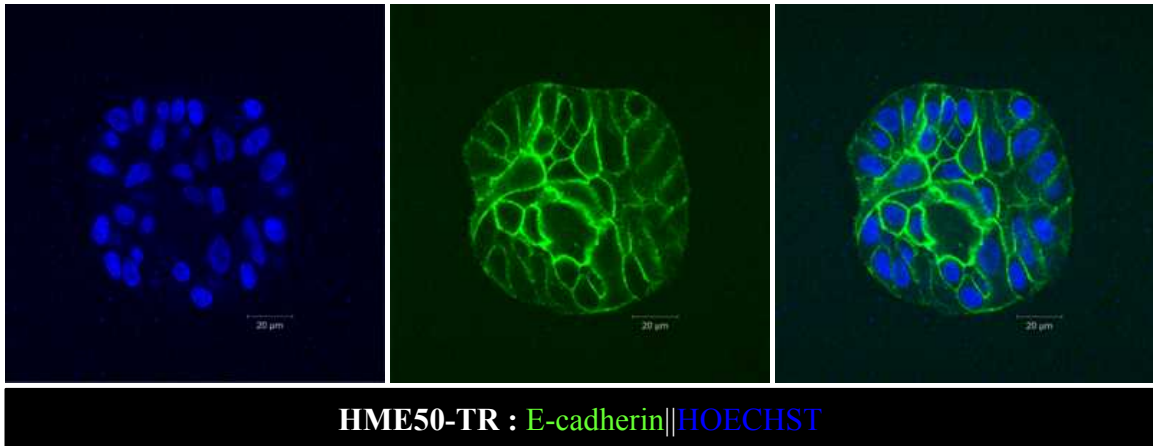
Figure 13. Basolateral E-cadherin staining of pre-invasive HME50-TR acini shows lack of hollow lumen. Immunofluorescence staining of pre-invasive HME50-TR acini grown for 10 days embedded in 3D Matrigel culture displayed basolateral E-cadherin (green) and nuclear Hoechst staining (blue) of HME50TR acini. Dis-organized nuclei and absence of a cleared lumen was evident in the centermost z-stack indicating HME50TR acini do not strictly respond the cues of normal epithelial morphogenesis.

On the other extreme of the progression series, the tumorigenic cell line HMET exhibit grossly disorganized nuclei, filled lumen and lack of cell-cell adhesion and evident by stellate or grape-like morphology classified by stellate invasive processes that bridge multiple colonies by day ten of morphogenesis (**Figure 14**). The HMET cells seeded at a density such that the cells do not crowd or are not in immediate proximity, initiate morphogenesis at a faster pace as evident by development of relatively large acini compared to its HME50 counterparts by day 4, eventually losing cellular organization, and invade surrounding matrix quickly developing as large multi-acinar stellate colonies by day 8 (**Figure 15**). HMET invasive stellate colonies generally exceed 100 microns in size and show protrusions. The gene expression pattern of HMET suggested an EMT signature characterized by loss of E-cadherin and acquisition of invasive potential by deregulated expression of genes that control extracellular matrix modeling and cell adhesion, which was also supported by morphology observed in 3D culture. In conclusion, retroviral hTERT expression extended the life span of HME50 cells without drastically altering the morphological properties of HME50 cells (**Figures 7-8**) whereas distinct changes in morphology and growth characteristics were observed after immortalization and transformation strategies used to develop the HME50 progression series (**Figures 11-15**). Thus, the acinar phenotypes observed for HME50 cell lines were distinct that prompted the analysis of microarray transcriptomic profiles to understand how HME50 cell lines may reflect features of breast cancer progression.

AF 594 Phalloidin‖DAPI
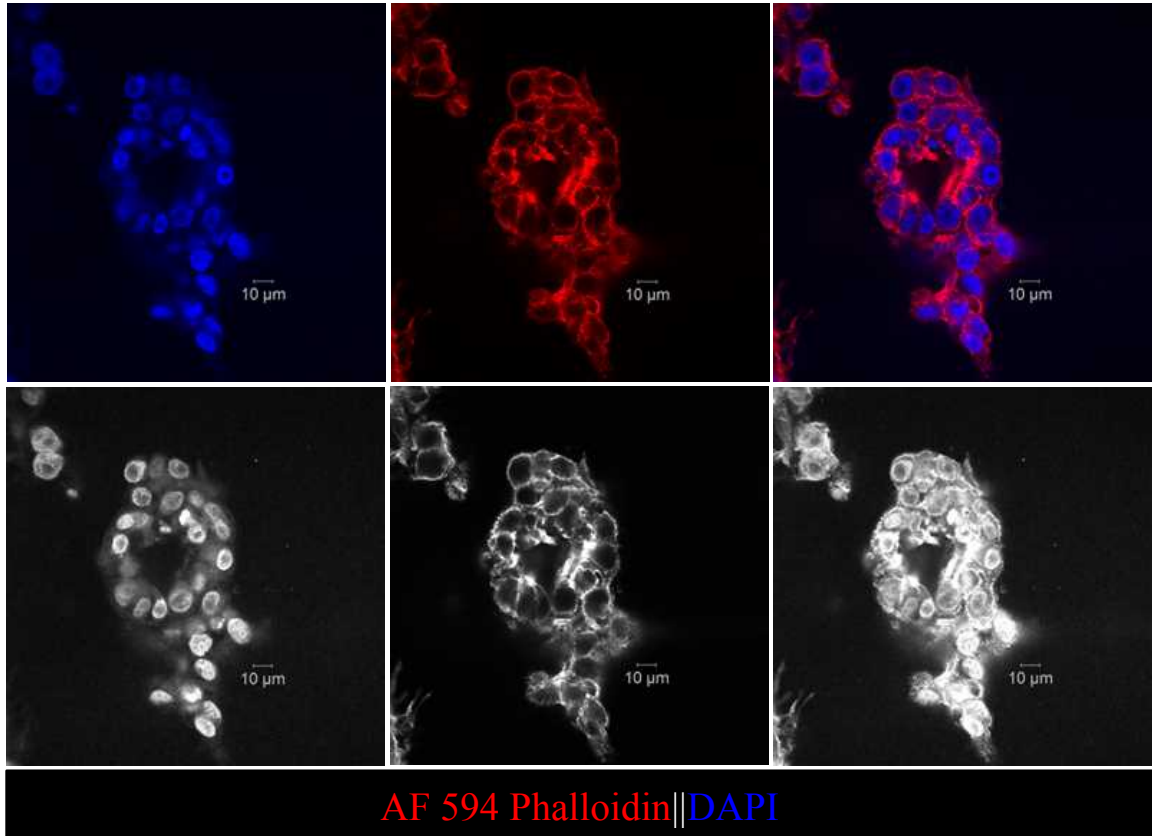
Figure 14. Malignant HMET acini adopt stellate-like morphology in 3D embedded culture. Malignant HMET cells embedded in 3D Matrigel culture grown for 10 days followed by F-actin and nuclear staining by AF 594 Phalloidin and DAPI respectively. The stellate-like morphology with disorganized nuclei, poor cell-cell adhesion, and invasive processes was evident in the representative center z-stack of acini by confocal imaging, scale bar 10µm.

HMET

127µm 151µm

135µm 129µm

151µm 100µm

10 µm 10 µm 10 µm 10 µm 10 µm 10 µm
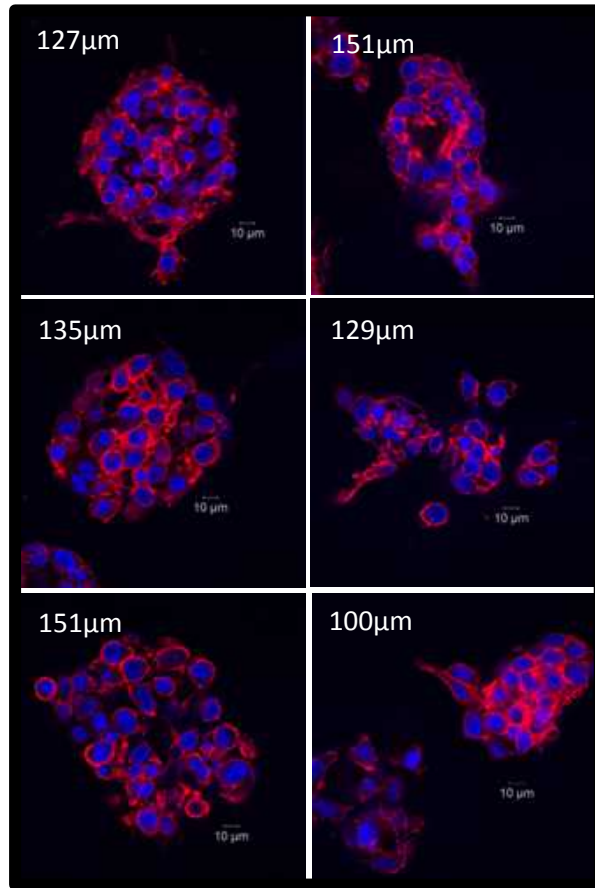
Figure 15. Representative confocal images of center Z-stack of HMET acinar structures in 3D embed culture. The size and morphology of HMET malignant acinar structures grown in 3D embed cultures varied drastically and often exceeded >100µm in size. The HMET acinar structures were readily distinguishable from their non-malignant counterparts in HME50 progression series due to the distinct stellate-like invasive colonies that are formed by day 8 of seeding. Most phenotypic heterogeneity was visually observed amongst the disorganized HMET structures; formation of invasive processes, migratory behavior and lack of organization being the redundant features. Representative confocal images show center z-stack stained with F-actin binding AF 594 Phalloidin and nuclear stain DAPI. Scale bar, 10µm.

3.2. HME50 cell lines can be delineated by their distinct gene expression patterns

3.2.1. Principal Components Analysis conclusively delineates HME50 cell lines based on differences in global gene expression profiles

Exploratory investigation of our microarray data using principal component analysis (PCA) enabled dimensionality reduction and 3D visualization of the variation, thus providing evidence that the HME50 cells cluster distinctly in accord to their inherent differences in global gene expression patterns with no outliers and major effects in the data (**Figure 16**). Each dot in PCA scatter plot represents a sample; samples with similar intensity values across the probesets on the whole chip are grouped close together whereas samples that are far apart in the plot are dissimilar. The technical quadruplicates are more closely related to one another ensuring that the observed differences are owing to inherent biological differences and not due to technical manipulations during sample processing.

Each cell line represented in the PCA plot is classified based on the amount of variability captured in each of the three principal components; with most variability captured in Principal component (PC) #1 (X-axis 22.9%), followed by PC #2 (Y-axis 15.9%) and PC 3# (Z-axis12.6%). This showed HMET samples (purple) are distinct and resolve farthest from HME50 (red), HME50-5E (blue) and HME50-hTERT (green) clusters while the quadruplicates samples for each cell line resembled each other as shown in **Figure 16**. Thus, the PCA plot graphically depicted the differences between HME50 cell lines by distinct clustering based on the gene expression changes after sequential genetic manipulation during malignant progression – the LFS patient-derived

HME50 cells cluster together, closer to hTERT-immortalized HME50-hTERT and spontaneously immortalized HME50-5E cell line cluster than with the malignant HMET cluster on the PCA plot. Therefore, all of the cell lines clustered very distinctly in the PCA plot indicating that these cell lines indeed have very different gene expression pattern owing to sequential genetic manipulations although they had HME50 as the common source.

Moreover, after filtering the microarray data such that genes with low coefficient of variance (CV <0.3) across dataset are excluded, malignant and non-malignant cluster distinctly indicating gene expression changes (**Figure 17**). In conclusion, the gene expression changes essentially acquired during each step of sequential immortalization, transformation and gain of metastatic potential dramatically altered the global gene signature of HME50 cells and resulted in separation of non-malignant samples from malignant samples.
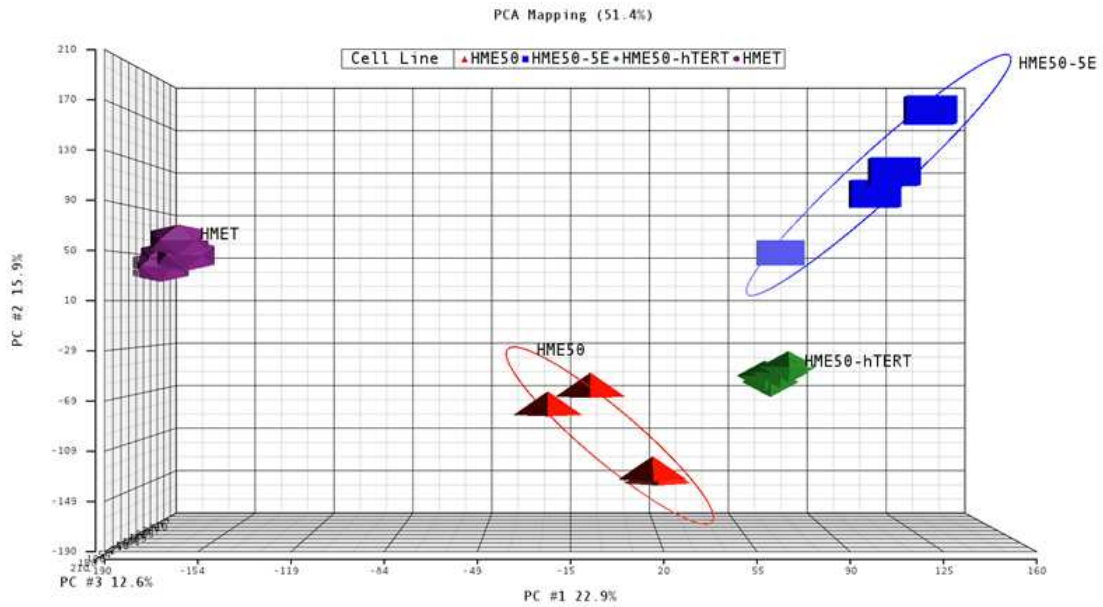
Figure 16. Principal Components Analysis (PCA) of HME50 progression series cell lines. To enable multidimensional data reduction and visualization of gene expression data of each sample profiled using HG_U133_Plus 2.0 Affymetrix Array as a dot; the color and shape of the dot denotes the HME50 cell line; HME50-hTERT (green), HME50-5E (blue), HME50 (red), and HMET (purple). The scatter plot depicted each cell line separated in space, and most separation was observed between malignant HMET and non-malignant HME50 cell lines; no distinct separation between the technical replicates; this cluster pattern is further highlighted by the ellipse that grouped samples according to the cell line.
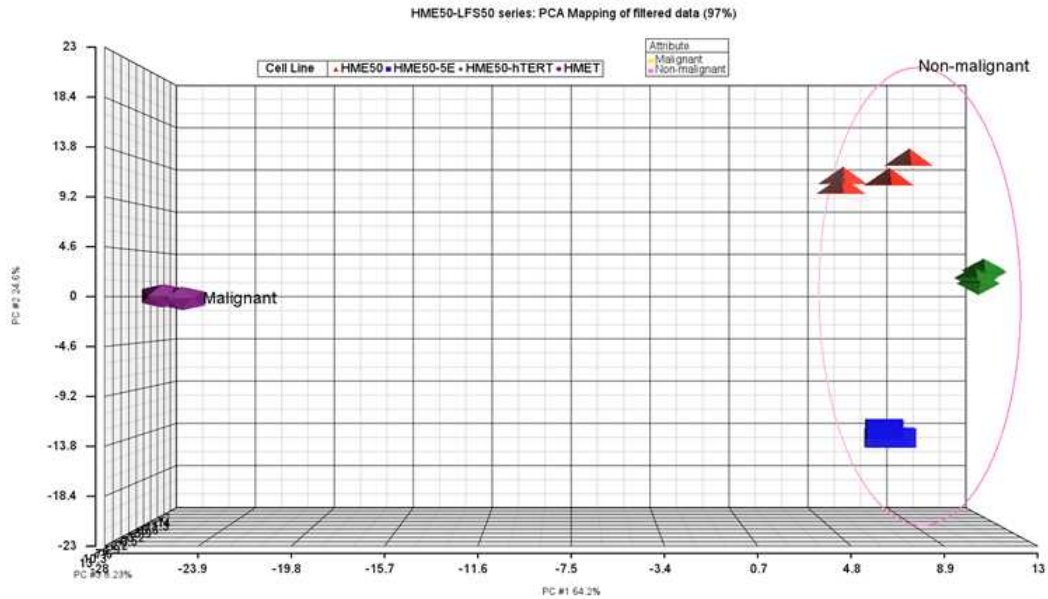
Figure 17. Principal Components Analysis scatter plot of filtered HME50 data. The

scatter plot enabled the visualization of expression of 391 Affymetrix features with

variable gene expression filtered based on coefficient of variation >0.3 across the 16

samples (see **Figure 16**). The malignant and non-malignant clusters are well separated in

space also as observed in unsupervised hierarchical dendrogram.

3.2.2. Hierarchical clustering shows distinct clustering between non-malignant HME50

cell lines

The PCA results were further supported by unsupervised hierarchical clustering of

genes that vary across the dataset by a coefficient of variation of >0.3 in the microarray

dataset (**Figure 18**). The unsupervised hierarchical clustering analyses revealed that the

non-malignant HME50 cell lines, namely HME50, HME50-5E, and HME50-hTERT,

have a relatively similar gene expression pattern and segregate together in contrast to the

malignant HMET cell line, which has a visually evident difference in gene expression

profile (**Figure 18**). Since the hierarchical clustering dendrogram showed that HME50

and HME50-hTERT cell lines have the most similar gene expression, we noted that the

retroviral expression of hTERT for immortalization did not alter the gene expression

pattern drastically, therefore justifying the use of the HME50-hTERT cells as a baseline

for the progression series for therapeutic/prevention studies. The hierarchical clustering

dendrogram illustrated the standardized gene expression level of each of the 391

Affymetrix identifiers in each samples. Genes with unchanged expression levels were

displayed as a value of zero and were colored grey; up-regulated genes with positive

values were colored red; down-regulated genes with negative values were displayed in

blue. The clusters were distinguished using different tree colors (**Figure 19**), each cell

line sample was represented in rows (16 samples; malignant HMET separated from non-

malignant cell lines by branches colored red and blue based on tree distance) while genes

were represented as columns (5 distinct genes clusters of colored azure (21 genes), brown

(9 genes), grey (50 genes), pink (81 genes), and orange (148 genes) in column

dendrogram based on distance). This hierarchical clustering dendrogram suggests

HME50-hTERT has gene-expression signature that closely correlates with the parental

HME50 cell line and HMET has the most distinct gene expression profile as compared to

non-malignant cell lines in HME50 series. To further understand the biological

differences in context of molecular pathways between the cell lines, IPA® Core Analysis

was performed on each gene cluster discovered by the clustering algorithm.
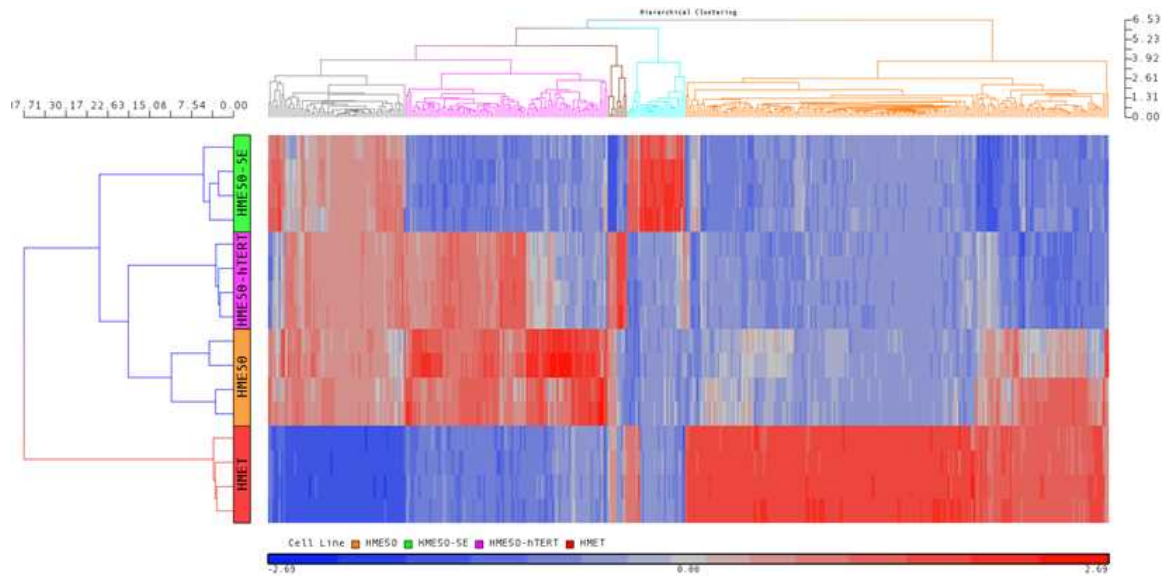
Figure 18. Unsupervised hierarchical clustering pattern of HME50 progression series expression data. The rows in the hierarchical cluster represent quadruplicates for the HME50 cell lines and columns represent features (individual genes). From top to bottom of the dendrogram on left hand of cluster, quadruplicate HME50-5E (label color green), HME50-hTERT (magenta), HME50 (orange), and HMET (red) samples. Genes with coefficient of variation >0.3 across all samples were selected for hierarchical clustering analysis (391 Affymetrix Identifiers); technical quadruplicates cluster together. The red and blue colors represent up-regulation and down-regulation of genes expression levels, respectively, whereas grey represents no change in gene expression as shown in the scale bar at bottom. The row dendrogram colored based on tree spacing by distance shows two major clusters: cluster red contains only malignant HMET cell line, whereas the tree colored blue consisted of non-malignant HME50, HME50-hTERT and HME50-5E cell lines. The HME50-hTERT samples aligned closest to parent HME50 cell line within the blue colored non-malignant cluster.

71

To identify the differentially expressed genes across the HME50 represented in dendrogram as 391 features (306 unique genes) genes, ANOVA analysis on the 391 identifiers of hierarchical cluster was performed using statistical threshold of p-value with FDR <0.05 and >1.3 or <-1.3 fold change. The ANOVA results revealed the number of genes differentially regulated across each comparison (**Table 3**) further broken into in 5 distinct clusters across each comparison (**Table 4**). Interestingly, the grey cluster consists of 50 genes that are all down-regulated whereas the orange cluster consists of 148 genes that are all up-regulated in malignant HMET cells relative non-malignant HME50 cell lines (**Figure 19**). The azure cluster shows unique set of 21 genes that are up-regulated in HME50-5E cells relative to HME50 as well as HME50-hTERT cell lines and this cluster includes *LIMCH1*, *TENM1*, *SLAMF7*, *SLCO1B3*, *FAM201A*, *WISP3*, *SPP1*, *CC2D2A*, *UPK1B*, *SLCO1B3*, *CUX2*, *BCHE* and *GHR* as most up-regulated. The pink cluster consists of 81 genes, all of which are down-regulated in HME50-5E comparison with the parental HME50 cell line. Contrastingly, a subset of 63 genes in the pink cluster, 62 genes are down-regulated in HME50-5E relative to HME50-hTERT with the exception of *PDPN* gene (podoplanin; FC -3.5). This revealed the pink cluster in HME50-hTERT shows an overlap of subset of genes with both parental HME50 and spontaneously immortalized HME50-5E.

The top up-regulated genes in HME50-hTERT relative to HME50-5E are *S100A7*, *KLK7*, *LOC728613*, *RARRES1*, *SLC6A14*, *TCEAL8*, *CDC42EP5*, *TENM2*, *KLK5*, and *KLK10*. A considerable similarity between HMET and HME50-5E is observed in the pink cluster, consistent with only 7 genes up-regulated and 14 down-

regulated in HMET relative to HME50-5E. The genes up-regulated in HMET are

S100A7, *GPM6B*, *ZNF677*, *MUCL1*, *TENM2*, *ZNF880* and *ZNF415*; whereas *CARD18*,

SLC1A6, *PDPN*, *SPRR3*, *PHACTR3*, *ZNF750*, *KLK5*, *SYT14*, *TMPRSS11E* and *BEX4* are

down-regulated relative to HME50-5E.

148 genes 21 genes 9 genes 81 genes 50 genes

Mesenchymal Markers:
ZEB1, ZEB2 SMAD6, FN1, CDH2

Epithelial Markers:
FBN2, TP63, DSG3, KRT6A, KRT6B
Adhesion proteins:
ITGB6, CDH1, DSC2

Cell Line HMESO HMESO-SE HMESO-hTERT HMET
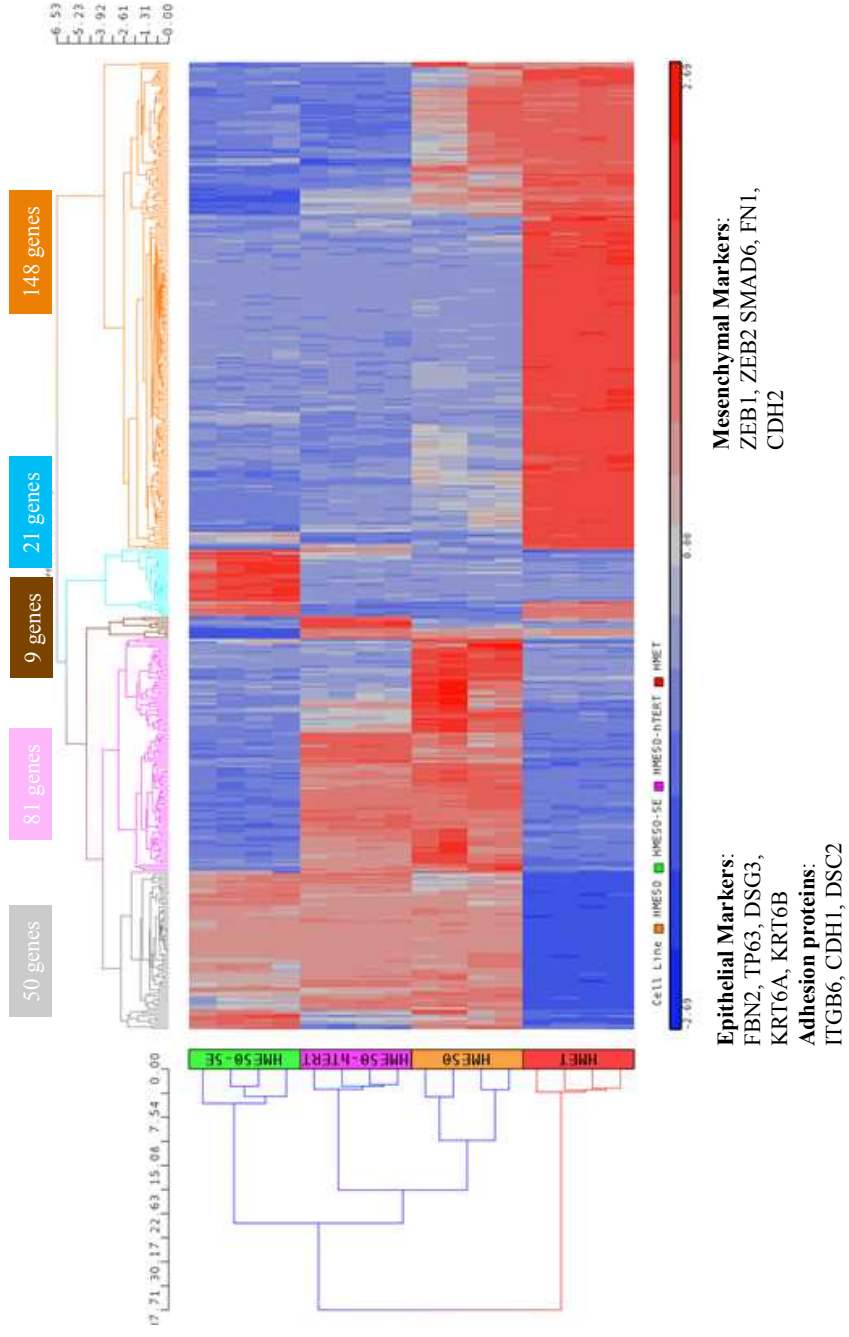
Figure 19. Five major clusters discovered by unsupervised hierarchical clustering of HME50 cell lines. Unsupervised hierarchical clustering shows dendrogram of HME50 cell lines (technical quadruplicates rows) using genes with coefficient of variation >0.3 (columns) for expression across all HME50 cell line microarray samples. The columns represent 391 features differentially expressed across the HME50 datasets which distinctly cluster into five groups shown on the column dendrogram colored grey, pink brown, azure and orange based on tree spacing by distance. The grey cluster (50 genes) comprises of epithelial markers which are down-regulated whereas the orange cluster represents mostly mesenchymal markers (148 genes) up-regulated in HMET cell line samples. The red and blue colors represent up-regulation and down-regulation of genes expression levels, respectively, whereas grey represents no change in gene expression as shown in the scale bar.

Table 3. Differentially regulated genes in hierarchical dendrogram across HME50 cell lines.

	Compared HME50 cell lines	Up-reg	Down-reg
1	HME50-5E vs HME50 (249)	49	200
2	HME50-hTERT vs HME50 (214)	46	168
3	HMET vs HME50 (282)	142	140
4	HME50-hTERT vs HME50-5E (189)	125	64
5	HMET vs HME50-5E (247)	164	83
6	HMET vs HME50-hTERT (277)	154	123
7	HMET vs HME50 & HME50-5E & HME50-hTERT (306)	157	149

The numbers in parenthesis indicate total number of differentially expressed genes using a statistical threshold of p-value with FDR <0.05 and >1.3 or <-1.3 fold change for each comparison; cells colored pink and green indicate number of up-regulated and down-regulated genes respectively.

Table 4. Differentially regulated genes in each cluster of dendrogram across HME50 cell lines.

	Contrasts	Dendrogram Clusters									
		GREY		PINK		BROWN		AZURE		ORANGE	
1	HME50-5E vs HME50 (249)	**36**		**81**		**6**		**21**		**105**	
		18	18	0	81	0	6	21	0	10	95
2	HME50-hTERT vs HME50 (214)	**27**		**67**		**9**		**6**		**105**	
		13	14	13	54	9	0	4	2	7	98
3	HMET vs HME50 (282)	**50**		**81**		**6**		**8**		**137**	
		0	50	0	81	2	4	4	4	136	1
4	HME50-hTERT vs HME50-5E (189)	**35**		**63**		**9**		**21**		**61**	
		18	17	62	1	9	0	0	21	36	25
5	HMET vs HME50-5E (247)	**50**		**21**		**8**		**20**		**148**	
		0	50	7	14	8	0	1	19	148	0
6	HMET vs HME50-hTERT (277)	**47**		**65**		**9**		**8**		**148**	
		0	47	2	63	0	9	4	4	148	0
7	HMET vs HME50 & HME50-5E & HME50-hTERT (306)	**50**		**79**		**8**		**21**		**148**	
		0	50	0	79	5	3	4	17	148	0

The numbers in parenthesis indicate total number of differentially expressed genes using a statistical threshold of p-value with FDR <0.05 and >1.3 or <-1.3 fold change for each comparison. The numbers in bold indicate total number of genes in each cluster colored grey, pink, brown, azure and orange for comparison between each cell line; cells colored pink and green indicate number of up-regulated and down-regulated genes respectively.

We analyzed the differentially expressed genes grouped together by hierarchical clustering. First, 2-way Analysis of Variance (ANOVA) on hierarchically clustered 391 identifiers was performed to identify statistically significant differential gene expression among the LFS HME50 cell lines and five distinct groups evidenced by clustering depicted using different color for branches. To find the signaling pathways and functions orchestrated by the differentially expressed genes in clusters discovered by unsupervised hierarchical clustering (**Figure 19**), we performed IPA® Core Analysis on each cluster comprising of differentially regulated genes in each comparison (**Table 4**). Based on the experimentally observed change and direction of gene expression in HME50 cell lines, IPA® upstream regulator analyses predicted upstream regulators that orchestrate the signaling cascade might prompt observed gene expression changes. Further, IPA® algorithm processed directional mechanistic networks using the information from upstream regulator predictions and gene expression changes observed in the dataset. The changes in biological state and processes across the HME50 cell lines were analyzed by IPA® Core Analysis of each gene group defined by hierarchical clustering (**Figure 19**) were as follows:

In the azure cluster, *RASSF1* (p-value 9.76E-05), *GHR* (p-value 5.82E-04) and *SST* (p-value 8.66E-04) were predicted to be upstream regulators in HME50-5E cells responsible for the observed up-regulation of the 21 genes. The cluster grouped by pink branches in **Figure 19** comprised of 81 genes that were all down-regulated relative to HME50. We performed IPA® comparison analysis on these 81 statistically significant genes after ANOVA analysis to visualize significant genes and regulatory pathways in

HME50E relative to HME50 cells. The IPA® comparison analysis showed TREM1 signaling was inhibited in both HME50-5E and HMET (p-value 2.85E-04, z-score -2.0).

The pink cluster being similar in both HME50-hTERT and HME50 cell lines did not show any significant differences with respect to signaling pathways. However, in the brown cluster (**Figure 19**), we observed 9 genes that are up-regulated in HME50-hTERT relative to HME50 but down-regulated relative to HMET. In HMET and parental HME50 relative to HME50-hTERT comparison, down-regulation of *LOXL1-AS1*, *VNN1*, *C15orf48*, *IL1RL1*, *SLC44A5*, *ZNF260*, *HTATIP2*, *KYNU* and *ZNF529* was observed. Also, in the pink cluster comparing the 62 up-regulated genes in HME50hTERT relative to HME50-5E, IPA® predicted *ROCK2* (z-score 2.0, p-value 2.77E-06) and *TNF* (z-score 2.319, p-value 8.04E-04) as activated upstream regulators. Skin formation and neutrophil activation were predicted as activated functional annotations (z-scores 2.2 and 2.0 respectively).

The grey cluster comprises of 50 genes down-regulated in HMET cell line compared to the non-malignant HME50 cell lines. The IPA® analysis predicted *ZEB1* (p-value 2.7E-08), *EHF* (p-value 1.85E-06), *TAF4* (p-value 3.29E-06) and *KRT14* (p-value 3.78E-06) as activated upstream regulators. *ZEB1* is known to down-regulate *MPZL2*, *TP63*, *CDH1* (and affect *PTPRZ1* and *ESRP1*) that were experimentally observed to be down-regulated in the grey cluster dataset and thus consistent with predicted activation of *ZEB1*. Also, inhibition of the focus molecule *NFKB1A* (consistency score 1.155) was predicted (Z-score -2.0, p-value 4.68E-03) based on observed down-regulation of

PTPRZ1, *IL1RN*, *EREG*, *CDH1* and *PCDH7*. The *KRT14* down-regulation was accompanied by observed down-regulation of *DSC2*, *DSC3*, *DSG3* and *KRT6B*. Cancer, dermatological and connective tissue disorders top disease bio-functions; malignant tumor being the top-most function predicted to increase due to down-regulation of 49 out of the 50 genes in grey cluster.

On the other hand, the orange cluster comprises of genes that are all up-regulated in malignant HMET cells relative to non-malignant cell lines. Based on these 148 genes, the activation of top upstream regulators *TGFB1* (p-value 2.93E-18, z-score 3.759), *TNF* (p value 3.41E-12, z-score 3.683), choriogonadotropin (p- value 1.40E-11, z score 2.29), *CCND1* (p-value 3.45E-11, z score 2.778), *FGF2* (p-value 4.76E-11, z-score 2.189), *IL6* (p-value 4.23E-10, z-score 2.139) was predicted in HMET relative to HME50, HME50-5E and HME50-hTERT. On the other hand, Alpha catenin (p-value 1.92E-09, z-score -3.273), let-7 (p-value 9.14E-09, z-score -3.215), *NUPR1* (p value 4.09E-07 z-score -3.638), *AHR* (p-value 4.19E-07, z-score -2.315) and *RB1* (p-value 6.82E-04, z-score -2.618) were predicted as inhibited.

Interestingly, preliminary analysis based on the direction of gene expression in the orange cluster (**Figure 19**), EGCG (chemical drug) was predicted to be inhibited (p-value 2.4E-05, z-score -2.376) based on the mechanistic network consisting of 60 genes. Specifically, the evidence for inhibition of EGCG is statistically drawn from the expression direction of *TLR4*, *MMP3*, *MMP1*, *ITGB3*, *FN1*, *CSF3*, *CCL2* and *ANGPT1* which are known to be down-regulated by EGCG were experimentally observed as up-

regulated in the HMET dataset, thus consistent with inhibition of EGCG. The top diseases/function predicted to be increased are metastasis (z-score 2.655), angiogenesis (z-score 2.771), vasculogenesis (z-score 2.655), invasiveness (z-score 2.039), proliferation (z-score 4.539), cell movement (z-score 3.4); whereas, functions such as apoptosis (z-score -3.307), cell death (z-score -2.949), organismal death (z-score -5.845) and necrosis (z-score -2.281) are predicted to be decreased.

In conclusion, the grey cluster comprised of 50 statistically significant genes that are down-regulated in HMET compared to all non-malignant HME50 cell lines; the top down-regulated genes in this cluster identified were *SPRR1B*, *CDH1*, *KRT6B*, *KRT6A*, *DSG3*, *PNLIPRP3*, *ESRP1*, *S100A14*, *ITGB6* and *TP63* (**Figure 19**). On the other extreme, the orange cluster comprising of 148 statistically significant genes up-regulated in HMET compared to all non-malignant HME50 cell lines; the top up-regulated genes in this cluster were *LAPTM5*, *SRGN*, *TFPI*, *MMP3*, *KLRC2*, *EDIL3*, *STC1*, *PHOBTB3*, *SLC16A6* and *MMP1* in addition to important EMT hallmarks such as *ZEB1* and *CDH2*. Based on unsupervised hierarchical clustering, HMET cells displayed an up-regulation of hallmark EMT genes and down-regulation of epithelial markers. Further analysis by breaking the clusters into group of genes showed evidence of down-regulation of epithelial genes (grey cluster) that included genes such as *TP63*, *CDH1*, *DSC2*, *KRT6A* and, *DSC2*. The mesenchymal markers were up-regulated as shown in the dendrogram as the orange cluster that consists of mesenchymal markers such as *ZEB1*, *ZEB2*, *SMAD6*, *FN1*, *CDH2* when compared to the non-malignant HME50 cell lines. Validation of the E-cadherin down-regulation in HMET was carried out by immunostaining of 3D culture

(**Figure 20**) as well as qRT-PCR (data not shown). Thus, dendrograms generated using the gene expression signature showed that HMET cells may represent a mesenchymal phenotype with up-regulated mesenchymal markers and down-regulated epithelial genes when compared to the non-malignant HME50 cell lines. Based on the hierarchical clustering and IPA® predictions, acquisition of EMT plasticity in HMET cells during malignant progression were further studied using publically available epithelial-to-mesenchymal (EMT) signatures.
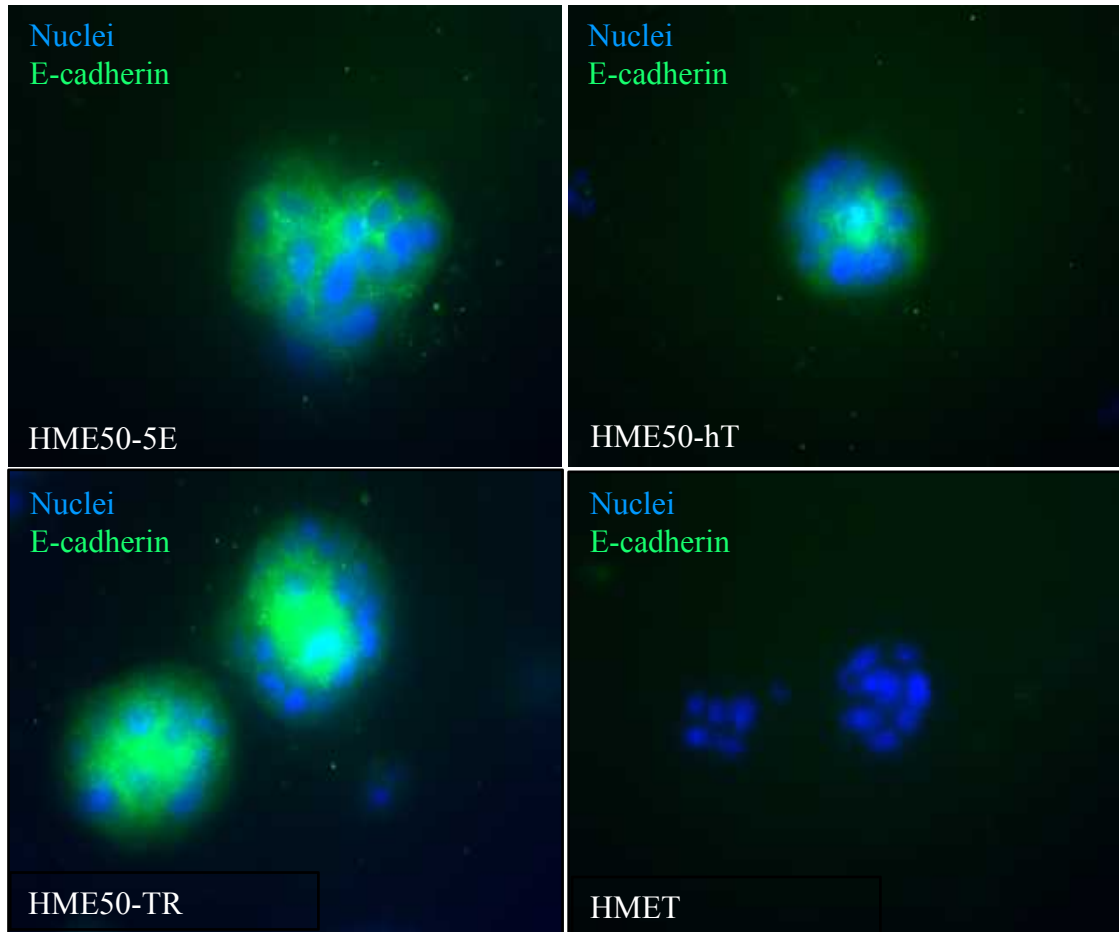
Figure 20. Malignant HMET acinar cells do not express E-cadherin. The
immunofluorescence staining of HME50 cell lines in 3D Matrigel overlay culture, stained
for E cadherin (green) and DAPI (blue). Note the non-malignant HME50-5E (top left)
and HME50-hTERT (top right) and pre-invasive HME50-TR (bottom row, left) stain for
E-cadherin; malignant HMET (bottom row, right) have lost E-cadherin expression, in
accordance to HME50 microarray data.

3.3. Characterization of distinct gene expression patterns

3.3.1. Overview of gene expression changes associated with step-wise genetic

manipulation of HME50 cells in the HME50 cell progression series

We next analyzed the RMA (Robust Multi-chip Average) normalized microarray

data by 2-way Analysis of Variance (ANOVA) model by using Method of Moments and

found differentially expressed genes between each contrast (**Table 5**) that passed the

specified statistical criteria (see Methods and **Appendix 2**). For any given comparison,

the gene list created consists of genes that have any change in group 1 relative to group 2

with FDR (step-up) <0.01 and fold-change >1.3 or <-1.3.

Not surprisingly, many features were observed to be differentially expressed each

comparison, specifically the number of differentially expressed genes between HME50-

5E relative to HME50, HME50-hTERT relative to HME50, and HMET relative to

HME50 was 11773, 7108 and 12051 respectively (**Table 5**). It is interesting to note that

close to twice as many IPA® analysis ready molecules were differentially regulated in

HME50-5E (7463 genes) than in HME50-hTERT (3146 genes) relative to parental

HME50 cells. Moreover, 9031 features were differentially regulated between

spontaneously immortalized HME50-5E and hTERT immortalized HME50-hTERT.

Since HME50-hTERT was used as non-malignant control in lieu of primary HME50

patient-derived HMECs, we also compared the gene expression changes between HME0-

hTERT and HMET. These gene lists with number of up-regulated and down-regulated

probesets, IPA® analysis ready molecules, and number of identifiers used in IPA®

84

analysis threshold of fold-change 2.0 are in (**Table 5**). The most significantly altered

genes driven by sequential genetic manipulation are shown in **Tables 6 –12** (fold change

and p-values with HUGO Gene Nomenclature Committee (HGNC) approved gene

symbols and gene name).

Table 5. Two-way ANOVA detects differentially regulated probesets for each contrast.

	2-way ANOVA Contrasts used	Total # features in each gene list		#IPA analysis ready molecules	IPA FC 2.0 cut-off
1	HME50-5E vs HME50	**11773**		7463	2499
		6409	5364		
2	HME50-hTERT vs HME50	**7108**		3146	1229
		3753	3355		
3	HMET vs HME50	**12051**		5286	2737
		6317	5734		
4	HME50-hTERT vs HME50-5E	**9031**		4052	1719
		4272	4759		
5	HMET vs HME50-5E	**14833**		6669	3822
		7199	7634		
6	HMET vs HME50-hTERT	**13599**		6013	3315
		6832	6767		
7	HMET vs HME50 & HME50-5E & HME50-hTERT	**13347**		5565	2903
		6712	6635		

The numbers in bold denote total number of features differentially regulated between each contrast; cells in pink and green indicate up-regulated and down-regulated features respectively. Column with IPA® analysis ready molecules indicates number of identifiers/probesets that meet the user-specified statistical cutoffs and after resolution of duplicates are mapped in IPA® Knowledge Base as eligible for network analysis, functions, canonical pathways and lists.

Table 6. Most significantly altered genes in HME50-5E corresponding to spontaneous immortalization of primary HME50 cells.

Gene Symbol	Gene title (HGNC approved)	Fold change	p-value
Genes up-regulated in HME50-5E relative to HME50			
LIMCH1	LIM and calponin homology domains 1	64.1754	4.94E-12
ODZ1	odz, odd Oz/ten-m homolog 1(Drosophila)	31.2445	5.29E-10
SLAMF7	SLAM family member 7	28.4145	3.26E-13
AGR2	anterior gradient homolog 2 (Xenopus laevis)	23.6229	7.84E-12
SLCO1B3	solute carrier organic anion transporter family, member 1B3	22.0318	7.81E-11
LCP1	lymphocyte cytosolic protein 1 (L-plastin)	21.9535	3.33E-11
LY6K	lymphocyte antigen 6 complex, locus K	20.4748	2.76E-10
FAM201A	family with sequence similarity 201, member A	16.6576	3.14E-08
WISP3	WNT1 inducible signaling pathway protein 3	16.3316	2.20E-11
SPP1	secreted phosphoprotein 1	15.6198	3.58E-12
Genes down-regulated in HME50-5E relative to HME50			
KLK7	kallikrein-related peptidase 7	-218.053	3.36E-12
S100A7	S100 calcium binding protein A7	-187.868	4.64E-14
MGP	matrix Gla protein	-152.09	3.94E-10
RARRES1	retinoic acid receptor responder (tazarotene induced) 1	-146.255	8.50E-12
KLK10	kallikrein-related peptidase 10	-116.345	3.07E-10
KRT23	keratin 23 (histone deacetylase inducible)	-100.704	1.04E-08
HTATIP2	HIV-1 Tat interactive protein 2, 30kDa	-96.0171	6.35E-16
MUCL1	mucin-like 1	-92.7008	4.15E-11
PI3	peptidase inhibitor 3, skin-derived	-83.2755	8.16E-13
CST6	cystatin E/M	-73.3769	2.54E-09

Table 7. Most altered genes in response to hTERT mediated immortalization of primary HME50 cells.

Gene Symbol	Gene title (HGNC approved)	Fold change	p-value
Genes up-regulated in HME50-hTERT vs HME50			
IL1RL1	interleukin 1 receptor-like 1	18.3505	6.96E-09
LOC100132891	hypothetical LOC100132891	13.7522	1.89E-12
SHOX2	short stature homeobox 2	10.7751	2.07E-09
VNN1	vanin 1	10.5412	1.24E-10
KYNU	kynureninase (L-kynurenine hydrolase)	9.88821	2.55E-08
CADPS2	Ca++-dependent secretion activator 2	8.71469	3.01E-12
LOC100287616	Hypothetical protein LOC100287616	8.68678	2.69E-09
SPRR3	small proline-rich protein 3	8.15087	4.43E-07
SCG5	secretogranin V (7B2 protein)	7.40939	3.45E-07
CTH	cystathionase (cystathionine gamma-lyase)	7.11168	2.07E-09
Genes down-regulated in HME50-hTERT vs HME50			
MGP	matrix Gla protein	-89.1987	1.31E-09
GPX3	glutathione peroxidase 3 (plasma)	-68.3997	5.15E-15
DLGAP5	discs, large (Drosophila) homolog-associated protein 5	-39.9116	2.46E-08
KIF20A	kinesin family member 20A	-38.99	3.42E-08
MUCL1	mucin-like 1	-36.7317	4.89E-10
CEP55	centrosomal protein 55kDa	-27.618	7.15E-08
NEK2	NIMA (never in mitosis gene a)-related kinase 2	-26.906	4.77E-09
PDPN	podoplanin	-26.0751	3.84E-11
DEPDC1	DEP domain containing 1	-23.6916	1.65E-08
MGP	matrix Gla protein	-89.1987	1.31E-09

Table 8. Gene expression changes in HMET cell line driven by the successive immortalization, transformation and gain of metastatic potential during malignant progression of HME50 cells.

Gene Symbol	Gene title (HGNC approved)	Fold change	p-value
Genes up-regulated in HMET vs HME50			
LAPTM5	lysosomal protein transmembrane 5	124.405	7.37E-12
TFPI	tissue factor pathway inhibitor	103.833	6.91E-13
KLRC1/C2	killer cell lectin-like receptor subfamily C, member 1	99.3924	1.89E-15
MMP3	matrix metallopeptidase 3 (stromelysin 1, progelatinase)	90.5187	1.11E-11
SRGN	serglycin	80.5829	2.47E-11
GNG11	guanine nucleotide binding protein (G protein), gamma 11	72.1448	2.29E-10
EPB41L3	erythrocyte membrane protein band 4.1-like 3	54.3098	2.74E-15
ADAM12	ADAM metallopeptidase domain 12	53.3512	2.84E-13
SLC16A6	solute carrier family 16, member 6	52.2974	8.13E-13
EDIL3	EGF-like repeats and discoidin I-like domains 3	43.8861	1.36E-10
Genes down-regulated in HMET vs HME50			
SPRR1B	small proline-rich protein 1B	-672.113	2.04E-15
KRT6B	keratin 6B	-585.22	8.08E-17
CDH1	cadherin 1, type 1, E-cadherin (epithelial)	-489.567	7.28E-19
KRT6A	keratin 6A	-480.842	3.13E-17
DSG3	desmoglein 3	-411.451	2.60E-16
SPRR1A	small proline-rich protein 1A	-230.315	1.10E-09
KLK7	kallikrein-related peptidase 7	-195.331	1.75E-11
PNLIPRP3	pancreatic lipase-related protein 3	-194.591	8.65E-13
S100A14	S100 calcium binding protein A14	-192.977	1.92E-14
KRT5	keratin 5	-191.37	5.98E-19

Table 9. Genes differentially regulated by hTERT mediated immortalization of HME50 cells relative to spontaneously immortalized HME50-5E cell line.

Gene Symbol	Gene title (HGNC approved)	Fold change	p-value
Genes up-regulated in HME50-hTERT vs HME50-5E			
S100A7	S100 calcium binding protein A7	327.973	1.53E-14
HTATIP2	HIV-1 Tat interactive protein 2, 30kDa	120.931	3.70E-16
C15orf48	chromosome 15 open reading frame 48	103.402	3.97E-11
KLK7	kallikrein-related peptidase 7	75.7801	3.60E-11
LOC728613	programmed cell death 6 pseudogene	54.5309	1.64E-13
RARRES1	retinoic acid receptor responder (tazarotene induced) 1	51.4613	1.09E-10
SLC6A14	solute carrier family 6 (amino acid transporter), member 14	49.5139	2.11E-08
TCEAL8	transcription elongation factor A (SII)-like 8	45.2155	7.42E-12
CDC42EP5	CDC42 effector protein (Rho GTPase binding) 5	44.1833	5.76E-12
SERPINB2	serpin peptidase inhibitor, clade B (ovalbumin), member 2	42.695	3.72E-10
Genes down-regulated in HME50-hTERT vs HME50-5E			
LIMCH1	LIM and calponin homology domains 1	-70.2882	3.91E-12
LCP1	lymphocyte cytosolic protein 1 (L-plastin)	-31.7598	9.79E-12
SLCO1B3	solute carrier organic anion transporter family, member 1B3	-26.1057	4.38E-11
ODZ1	odz, odd Oz/ten-m homolog 1(Drosophila)	-25.737	9.85E-10
VCAN	versican	-21.5239	3.95E-08
CYB5A	cytochrome b5 type A (microsomal)	-20.1536	5.07E-12
GHR	growth hormone receptor	-19.8971	4.64E-09
LY6K	lymphocyte antigen 6 complex, locus K	-18.3478	4.11E-10
SPP1	secreted phosphoprotein 1	-17.2749	2.42E-12
WISP3	WNT1 inducible signaling pathway protein 3	-17.2175	1.80E-11

Table 10. Top differentially expressed genes in malignant HMET relative to spontaneously immortalized aneuploid HME50-5E cell line.

Gene Symbol	Gene title (HGNC approved)	Fold change	p-value
Genes up-regulated in HMET vs HME50-5E			
LOC100288985	Hypothetical protein LOC100288985	507.555	8.41E-12
SRGN	serglycin	392.187	1.35E-16
MGP	matrix Gla protein	330.049	8.41E-11
LAPTM5	lysosomal protein transmembrane 5	271.22	1.44E-12
MMP1	matrix metallopeptidase 1 (interstitial collagenase)	234.896	1.19E-12
TFPI	tissue factor pathway inhibitor	166.309	1.66E-13
MMP3	matrix metallopeptidase 3	156.913	3.18E-12
HS3ST3B1	heparan sulfate (glucosamine) 3-O-sulfotransferase 3B1	135.82	2.30E-12
KLRC1	killer cell lectin-like receptor subfamily C, member 1	129.147	1.03E-15
RRM2	ribonucleotide reductase M2	127.019	1.86E-08
Genes down-regulated in HMET vs HME50-5E			
SPRR1B	small proline-rich protein 1B	-706.859	1.88E-15
CDH1	cadherin 1, type 1, E-cadherin (epithelial)	-512.318	6.72E-19
KRT6A	keratin 6A	-469.098	3.27E-17
KRT6B	keratin 6B	-400.566	1.58E-16
ESRP1	epithelial splicing regulatory protein 1	-255.965	1.84E-17
DSG3	desmoglein 3	-252.779	6.57E-16
PNLIPRP3	pancreatic lipase-related protein 3	-215.928	6.98E-13
S100A14	S100 calcium binding protein A14	-204.327	1.70E-14
FAM129A	family with sequence similarity 129, member A	-196.371	7.62E-14
KRT5	keratin 5	-148.712	1.03E-18

Table 11. Top differentially expressed genes in malignant HMET relative to

HME50-hTERT used as experimental non-malignant control.

Gene Symbol	Gene title (HGNC approved)	Fold change	p-value
Genes up-regulated in HMET vs HME50-hTERT			
LAPTM5	lysosomal protein transmembrane 5	305.882	1.14E-12
EDIL3	EGF-like repeats and discoidin I-like domains 3	242.188	2.38E-12
MGP	matrix Gla protein	193.569	2.38E-10
RHOBTB3	Rho-related BTB domain containing 3	170.227	2.07E-13
TFPI	tissue factor pathway inhibitor	163.69	2.49E-13
SRGN	serglycin	161.801	4.98E-12
MMP3	matrix metallopeptidase 3 (stromelysin 1, progelatinase)	145.294	3.76E-12
KLRC1	killer cell lectin-like receptor subfamily C, member 1	134.724	9.36E-16
GNG11	guanine nucleotide binding protein (G protein), gamma 11	107.517	8.77E-11
KIF20A	kinesin family member 20A	83.5605	4.69E-09
Genes down-regulated in HMET vs HME50-hTERT			
SPRR1B	small proline-rich protein 1B	-578.952	2.64E-15
KRT6B	keratin 6B	-545.759	9.12E-17
CDH1	cadherin 1, type 1, E-cadherin (epithelial)	-536.98	6.19E-19
KRT6A	keratin 6A	-492.614	3.00E-17
DSG3	desmoglein 3	-393.093	2.83E-16
PNLIPRP3	pancreatic lipase-related protein 3	-273.684	4.36E-13
ESRP1	epithelial splicing regulatory protein 1	-223.472	2.41E-17
FAM129A	family with sequence similarity 129, member A	-217.212	6.19E-14
LCN2	lipocalin 2	-189.544	1.06E-12
SPRR1A	small proline-rich protein 1A	-181.722	1.77E-09

Table 12. Top differentially expressed genes in malignant HMET relative to non-malignant group of HME50 cell lines.

Gene Symbol	Gene title (HGNC approved)	Fold change	p-value
Genes up-regulated in HMET vs HME50 & HME50-5E & HME50-hTERT			
LAPTM5	lysosomal protein transmembrane 5	217.724	2.44E-13
SRGN	serglycin	166.482	5.13E-13
TFPI	tissue factor pathway inhibitor	141.183	3.73E-14
MMP3	matrix metallopeptidase 3	127.316	5.53E-13
KLRC1	killer cell lectin-like receptor subfamily C, member 1	120.032	1.31E-16
EDIL3	EGF-like repeats and discoidin I-like domains 3	100.173	1.76E-12
LOC100288985	Hypothetical protein LOC100288985	88.2517	3.33E-11
RHOBTB3	Rho-related BTB domain containing 3	82.5846	3.11E-14
SLC16A6	solute carrier family 16, member 6	54.3836	7.96E-14
MMP1	matrix metallopeptidase 1 (interstitial collagenase)	53.252	4.14E-12
Genes down-regulated in HMET vs HME50 & HME50-5E & HME50-hTERT			
SPRR1B	small proline-rich protein 1B	-650.332	2.33E-16
CDH1	cadherin 1, type 1, E-cadherin (epithelial)	-512.585	7.23E-20
KRT6B	keratin 6B	-503.88	1.13E-17
KRT6A	keratin 6A	-480.751	3.37E-18
DSG3	desmoglein 3	-344.494	3.90E-17
PNLIPRP3	pancreatic lipase-related protein 3	-225.713	6.96E-14
ESRP1	epithelial splicing regulatory protein 1	-216.978	2.76E-18
S100A14	S100 calcium binding protein A14	-185.895	2.24E-15
KRT5	keratin 5	-170.8	8.19E-20
ITGB6	integrin, beta 6	-136.23	2.04E-14

93

3.3.2. Venn diagram summarizes the gene expression changes during HME50 progression relative to parent HME50 cell line

The number of differentially expressed unique features relative to HME50 was highest in HMET, 42% (5056/12051) followed by HME50-5E, 39.5% (3997/11773) and the least in HME50-hTERT, 18.5% (1316/7109). All three cell lines in HME50 progression cell lines share a set of 2674 features that are differentially regulated as compared to parental HME50 cell line (**Figure 21**). The HME50-5E cell line (3152) shares more identifiers than HME50-hTERT (1950) with malignant HMET. Since these original ANOVA results comprised of comprehensive lists with Affymetrix probeset ID as identifiers/features, duplicates genes were resolved and appropriate statistical thresholds applied to understand the molecular mechanisms involved in the HME50 progression series.
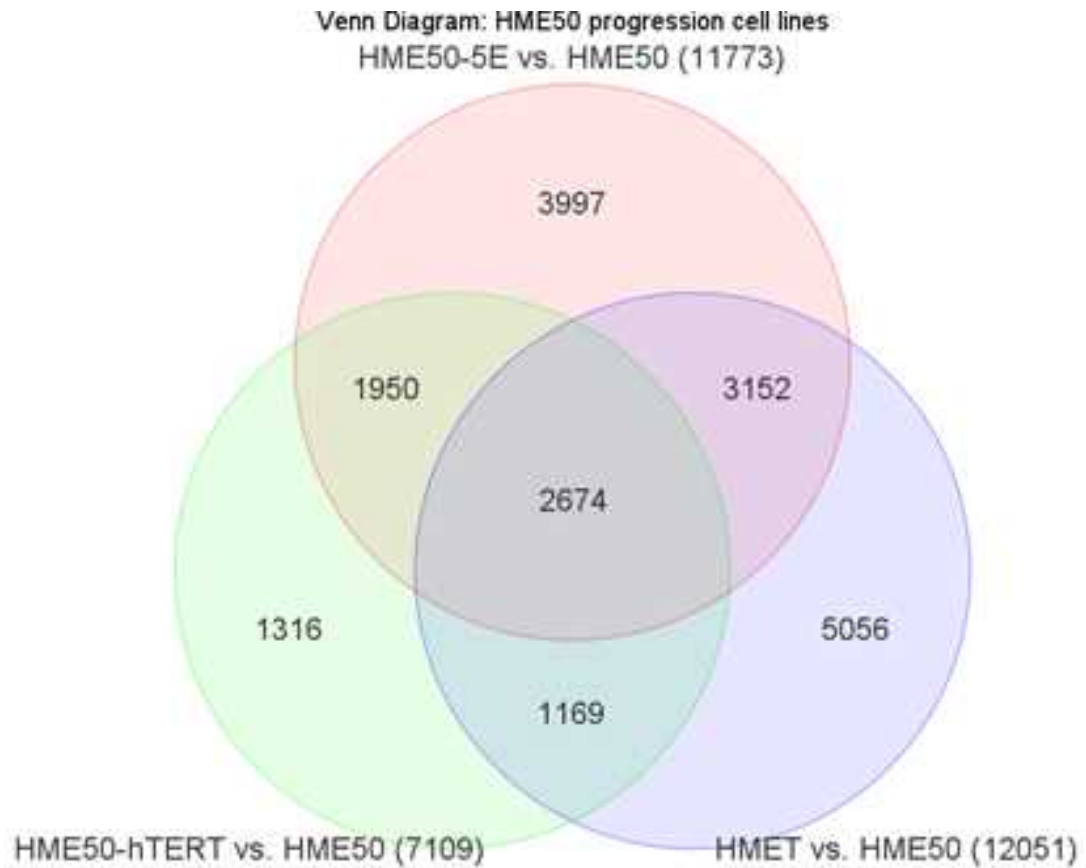
Figure 21. Venn diagram shows overlapping identifiers among the HME50 progression cell lines relative to parental HME50 cells. Venn diagram showing the overlapping common and unique features expressed in HME50-5E, HME50-hTERT and HMET lines relative to the patient-derived HME50 cells (P-value (step up FDR) <0.05; Fold Change >1.3 or <-1.3). The numbers in parenthesis denote the total number of features expressed in each comparison.

3.3.3. IPA® analyses of pathways enriched based on unique gene expression changes acquired during sequential genetic manipulations

Based on the Venn diagram (**Figure 21**), the list of 3997 probesets that were uniquely altered in HME50-5E expression dataset, 2858 genes were mapped by IPA® as analysis ready molecules that were enriched in oxidative phosphorylation (p-value 2.47E-05), HIPPO signaling (p-value 2.00E-04, 31.6% overlap), OX40 signaling pathway (p-value 6.96E-04), estrogen receptor signaling (p-value 7.45E-04) and PPAR signaling (p-value 1.28E-03) pathways. The toxicity related molecules enriched in HME50-5E cells were implicated in mitochondrial dysfunction, gene regulation by peroxisome proliferators via *PPARα*, *NF-κB* signaling, *PPARα/RXRα* activation and *TGF-β* signaling. Unique to HME50-5E, the inhibition of many microRNA molecules was predicted along with *PARP9* (z-score -2.157, 1.06E-03), *DOCK8* (z-score -2.828, 1.91E-03), *RICTOR* (z-score -3.959, p-value 4.02E-03), *SASH1* (z-score -3.300, p-value 4.22E-03), *NLRC5* (z-score -2.372, p-value 1.15E-02) and *SAMSN1* (z-score -2.982, p-value 1.75E-02).

Similarly, analysis of 1316 unique HME50-hTERT altered probesets (**Figure 21**) led to enrichment of protein ubiquitination (p-value 1.48E-04), caveolar-mediated endocytosis signaling (p-value 4.87E-04), Phospholipase C signaling (p-value 1.20E-03) canonical pathways based on 1016 analysis ready molecules. Given the observed gene expression changes in this list, IPA® predicted activation of *XBP1* (z-score 4.826, p-value 3.95E-11), *FSH* activated (z-score 2.6, p-value 1.93E-04), *TMBIM6* (z-score 2.0, p-value 4.41E-03) and *THBS4* activated (z-score 2.2, p-value 8.95E-03). The biological functions such as cell survival (z-score 2.3, p-value 7.26E-08), proliferation (z-score 2.2,

96

p-value 2.94E-05) and differentiation of epithelial cells and tissue (z-score 2.398, p-value 9.8E-03) were predicted to be increased; whereas organismal death (z-score -4.721, p-value 8.55E-03) and cellular senescence (z-score -2.33, p-value 1.25E-02) were predicted to decrease.

IPA® mapped 2858 genes from the given list of 5056 unique probesets altered in HMET cell line (**Figure 21**) which showed enriched in HIPPO signaling (p-value 3.58E-06, 40% overlap), Ephrin B signaling (p-value 6.2E-05), CCR3 signaling (p-value1.53E-04) and Wnt/β-catenin signaling (p-value 1.87E-04) canonical pathways. In the 2858 genes in the HMET, genes that mediate toxicity associated with p53 signaling, RAR activation and cell cycle: G1/S checkpoint regulation. In contrast to the inhibited microRNAs predicted for HME50-5E expression data, IPA® predicted activation of various mature microRNAs. Moreover, activation of *TGFB3* (z-score 3.5, p-value 1.69E-04), *TGFB1* (z-score 3.881, p-value 6.05E-04), *MYC* (z-score 2.95, p-value 1.46E-03) with simultaneous inhibition of *NUPR1* (z-score -5.484, p-value 4.52E-07), *MGEA5* (z-score -2.321, p-value 1.78E-04), *SPDEF* (z-score -3.14, p-value 6.69E-04), *IRS1* (z-score -2.7, p-value 6.75E-04) was predicted. Interestingly, the annotated disease/function based on HMET gene expression predicted increased proliferation (z-score 2.76, p-value 7.73E-07), cell movement (z-score 2.45, p-value 1.83E-03), G2/M phase of tumor cell lines (z-score 2.5, p-value 2.55E-03) and cell migration (z-score 2.148, p-value 3.49E-03).

97

The IPA® comparison analysis of overlapping genes in all three HME50-5E, HME50-hTERT and HMET progressed from HME50 exhibit commonly shared 2674 probesets altered compared to HME50 (**Figure 21**). For an overview of unique and overlapping pathways affected between the three observations, IPA® comparison analysis was performed on the core results. In summary, interferon and NF-κB signaling was inhibited in all three cell lines (p-value >0.01; z-score >-2.0, not very significant) based on observed gene expression changes in the 2015 analysis ready molecules shared by all three cell lines. However, the cell cycle G2/M DNA damage checkpoint was activated in HME50-5E and HME50-hTERT but inhibited in HMET dataset. Similarly, opposing directional change was observed for the *ATM* signaling and the cyclin and cell cycle regulation which was activated in HMET but inhibited in HME50-5e and HME50-hTERT cell lines. Furthermore, *NUPR1*, let-7, and *RBL2* were activated in HME50-5E and HME50-hTERT while inhibited in HMET. Simultaneous up-regulation of *TBX2*, *CSF2*, *HGF* and *FOXM1* in HMET and down-regulation in HME50-hERT and HME50-5E was predicted. Interestingly, *TP53* was predicted as inhibited upstream regulator in HMET but not in the non-malignant HME50-5E and HME50-hTERT dataset. Also, *CDKN1A* showed activated state in HME50-5E but inhibited state in HMET cells; whereas *RABL6* and *MYC* was inhibited in both HME50-hTERT and HME50-5E whereas up-regulated in HMET. This analysis suggested that HME50-5E cells maintain a "normal like" phenotypic state by robust cell cycle checkpoints and inhibition of migratory programs.

3.3.4. GSEA reveals significant enrichment of basal-like breast cancer phenotype in non-malignant HME50 cell lines relative to malignant HMET cell line

GSEA was performed, and the rank order of genes from the HME50 progression series in two phenotypes non-malignant versus malignant was compared within the molecular signature database (MSigDB) and significant enrichment scores for group each reported. GSEA revealed that the gene expression profiles of non-malignant HME50, HME50-5E, and HME50-hTERT relative to malignant HMET are significantly enriched and positively correlate with the *a priori* established "Charafe breast cancer basal vs. mesenchymal up signature" indicating that a group of genes expressed in basal-like breast cancer cell lines shared a well-defined directional change with non-malignant HME50 cell lines as reflected by the high normalized enrichment score (NES) and the leading edge genes in the enrichment plot (**Figure 22**). The *a priori* "basal vs mesenchymal up" curated gene signature in MSigDB comprises of genes up-regulated in basal-like breast cancer cell lines as compared to the mesenchymal-like breast cancer cell lines discovered by gene and protein expression profiling of breast cancer cell lines [118]. This gene set comprised of 121 genes, comprising of transcription factors, tumor suppressors, cell differentiation markers, oncogenes, as described in **Table 13**. This correlation strongly suggested HME50 non-malignant cell lines have basal like breast cancer features and the malignant HMET cell line has a mesenchymal-like gene expression signature. The enrichment plot for malignant HMET group showed significant enrichment for negatively correlated genes in the a priori "Kobayashi EGFR signaling" gene set (**Figure 23**) consisting of 251 genes that were down-regulated in gefitinib resistant H1975 NSCLC cell line after treatment with EGFR inhibitor [119]. This suggested that cyclin D1

activation with subsequent E2F-responsive gene suppression and proliferation arrest is lacking in HMET cell lines. The functional overview and categorization of genes contributing to leading edge are in **Table 14**. The GSEA algorithm used family wise error rate (FWER) for multiple testing correction and false discovery rate (FDR) to determine statistical significance of the results; results according to the nominal P-value <0.05 and FDR≤0.25; for both GSEA plots shown, FDR q-value of zero was calculated b the software indicating an actual p-value of ≤0.01 for 1000 permutations for the phenotype performed.
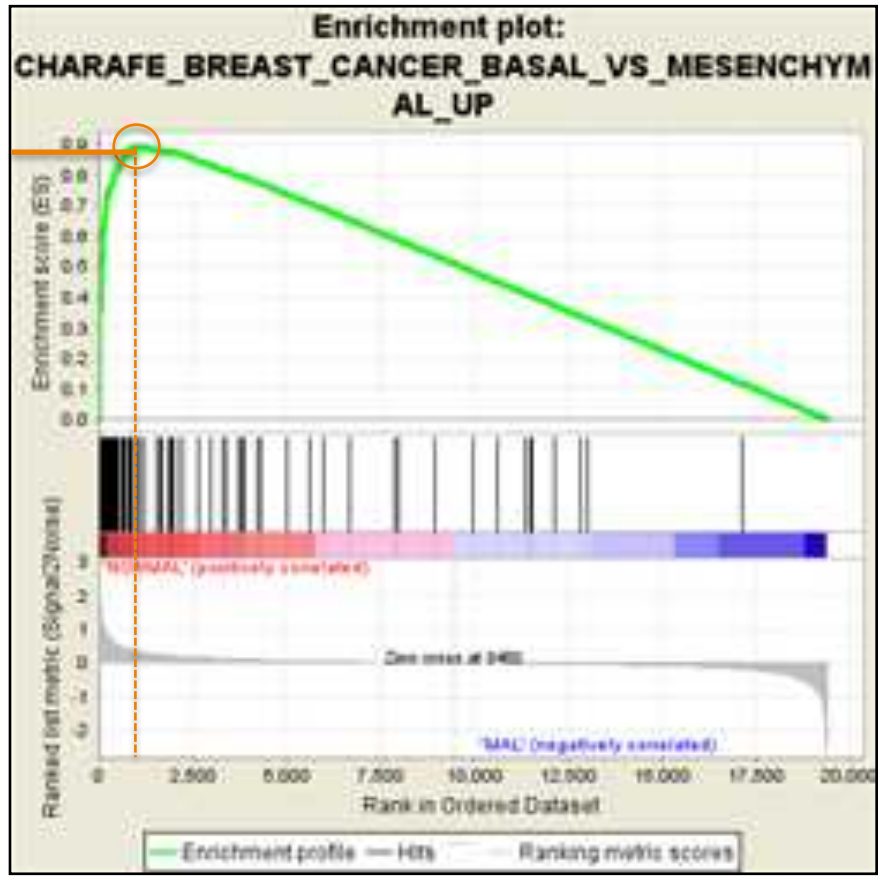
Enrichment Score

Rank at Maximum

Figure 22. Gene Set Enrichment Analysis plot for non-malignant HME50 cell lines.

Representative enrichment plot shown for non-malignant group (depicted as NONMAL; comprising of HME50, HME505E, HME50hTERT). Black bars represent the position of members of the category in the ranked list, together with the running enrichment score (plotted in green). For the non-malignant HME50 cell lines, GSEA resulted in enrichment of gene signature specific to 'basal compared to mesenchymal breast cancer cell lines' available in Molecular Signature Database (MSigDB). Statistical significance of GSEA results was based on 1000 permutations; the green curve represents enrichment score, orange dotted line indicates the maximum enrichment score (FDR q-value = 0.000) and the leading edge represents the core gene set that accounted for the enrichment signal. The non-malignant cell lines positively correlate with the CHARAFE BREAST CANCER BASAL VS MESENCHYMAL UP (M12795) in MSigDB.

Table 13. Functional overview and categorization of Charafe MSigDB gene set.

	Transcritpion factors	Cell differentiation markers	Protein kinases	Translocated cancer genes	Oncogenes	Tumor Suppresors
Tumor Suppresors		CDH1				CBLC, CDH1
Oncogenes	MAF	MUC1		MAF, MUC1	MAF, MUC1	
Translocated cancer genes	MAF	MUC1		MAF, MUC1		
Protein kinases			LIMK2, EPHA1			
Cell differentiation markers		MUC1, F11R, CDH1				
Transcritpion factors	MAF, ZFP36, ELF1, ELF3, RBL2, LITAF, EHF, FOSB, ZNF165, TSC22D3, KIAA0040, IRF6, OVOL1					

The table represents the functional categorization of genes in the MSigDB CHARAFE

BREAST CANCER BASAL VS MESENCHYMAL UP (M12795) that also contribute to

enrichment plot.

Enrichment plot:
KOBAYASHI_EGFR_SIGNALING_24HR_DN
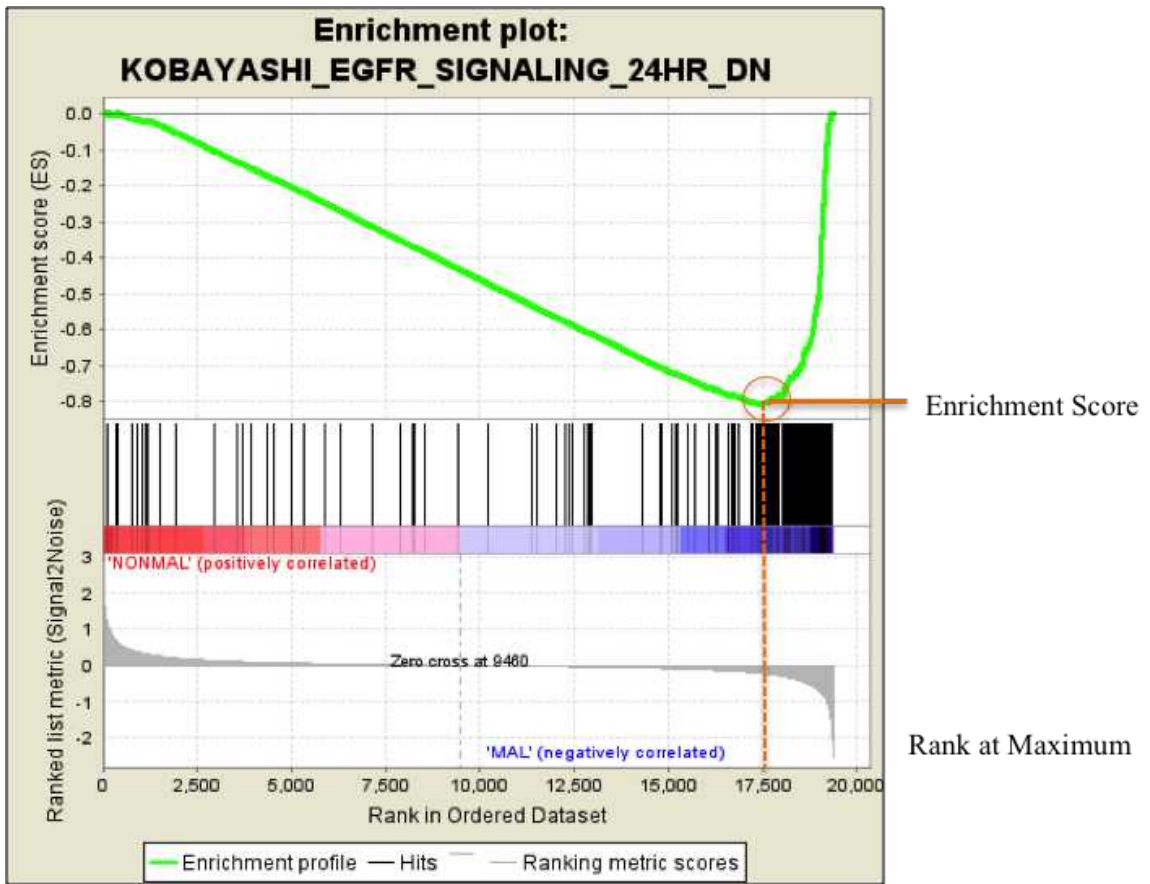
Figure 23. Gene Set Enrichment Analysis plot for malignant HMET cell line.

Representative enrichment plot shown for malignant (MAL; HMET) group. Black bars represent the position of members of the category in the ranked list, together with the running enrichment score (plotted in green). For the malignant HMET cell line, GSEA resulted in Enrichment plot of gene signature specific to 'EGFR signaling' available in MSigDB. Statistical significance of GSEA results was based on 1000 permutations; the green curve represents enrichment score, orange dotted line indicates the maximum enrichment score (FDR q-value = 0.000) and the leading edge subsequent to the peak score represents the core gene set that accounted for the enrichment signal. The malignant cell line positively correlate (ranking metric measured negative enrichment score for second phenotype) with the KOBAYASHI EGFR SIGNALING 24HR DN (M16010) FDR q-value = 0.000 in MSigDB.

Table 14. Functional overview and categorization of Kobayashi MSigDB gene set.

	Cytokines and growth factors	Transcritpion factors	Cell differentiation markers	Protein kinases	Translocated cancer genes	Oncogenes	Tumor Suppresors
Tumor Suppresors				BUB1B			MSH6, BUB1B, BLM, MSH2, BRCA1
Oncogenes		HMGA2, ETV1, ETV5		MET	ETV1, STIL, CCND1, HMGA2, CCND3, ETV5	ETV1, STIL, CCND1, HMGA2, CCND3, ETV5, MET	
Translocated cancer genes		HMGA2, ETV1, ETV6			ETV1, STIL, CCND1, HMGA2, CCND3, ETV6		
Protein kinases				CDK1, NEK2, MET, AXL, TTK, PKMYT1, AURKA, AURKB, PBK, CDK2, PLK4, VRK1, PLK1, BUB1, BUB1B, MELK			
Cell differentiation markers			ITGA6, NT5E, TNFRSF12A, HMMR				
Transcritpion factors		E2F1, MAFF, HMGB2, EZH2, MYBL1, MYBL2, SOX9, HMGA2, TIMELESS, ETV1, ETV5, FOSL1, TFDP1, TRIP13					
Cytokines and growth factors	DKK1, EREG, CKLF, TGFA, STC1, AREG, CX3CL1, NRG1, IL11						

The table represents the functional categorization of genes in the MSigDB KOBAYASHI

EGFR SIGNALING 24HR DN (M16010) that also contribute to enrichment plot.

3.3.5. IPA® Core Analysis of HME50 progression cell lines relative to parental HME50 cells

       IPA® Core Analysis of altered gene expression in HME50-5E relative to HME50 cells: The top canonical pathways in **Figure 24** displayed the enrichment of differentially regulated genes observed between HME50-5E and HME50 cells in the signaling and metabolic pathways. *TP53* is predicted as a top upstream regulator in HME50-5E and IPA® predicted *CDKN1A*, *IRGM1*, *NKX2-3*, and *MAPK1* as activated upstream regulators (**Table 15**). The upstream regulators predicted to be inhibited in HME50-5E relative to HME50 include *TGFB1*, *ERBB2*, *TNF*, *CSF2*, *OSM*, *RABL6*, *VEGF*, *HGF*, *ESR1*, *CCND1* and *IL1B*. The biological functions related to observed HME50-5E altered genes are involved in cell movement (z-score -5.384, p-value 1.04E-10), migration (z-score -5.241, p-value 5.44E-11) and survival (z-score -4.6, p-value 1.25E-08) that were predicted to be decreased. Simultaneously, biological functions such as organismal death (z-score 6.809, p-value 2.42E-08), hyperkeratosis (z-score 3.196, p-value 3.56E-08) and inflammation (z-score 2.15, p-value 7.00E-05) are predicted as increased in HME50-5E cell line.
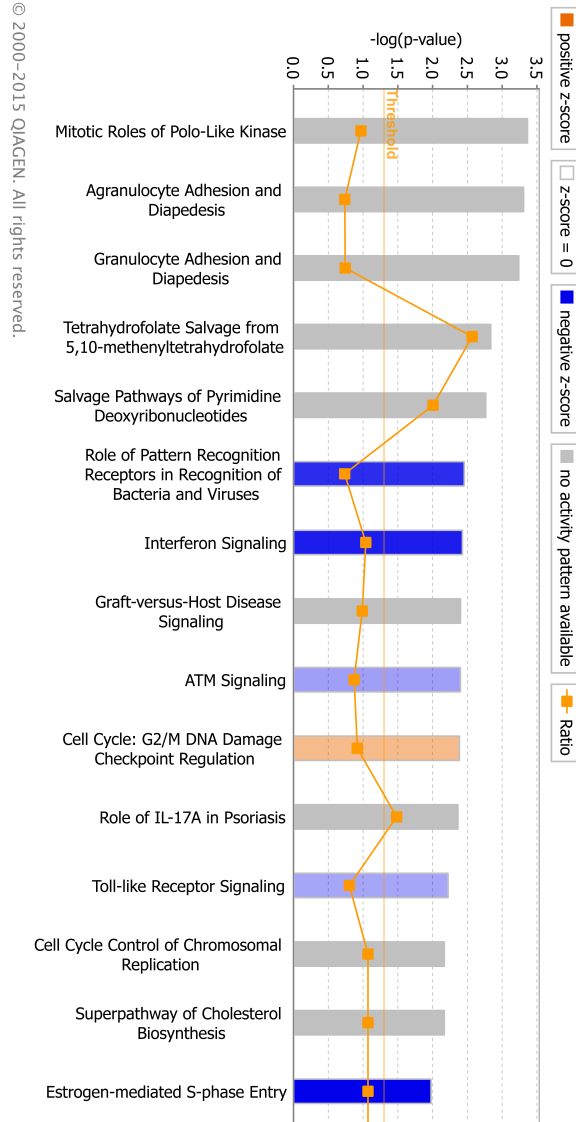
Figure 24. Canonical pathways enriched in HME50-5E relative to parental HME50.
The bar graph depicts the 15 most significant canonical pathways associated with gene expression differences observed in HME50-5E relative to HME50. The significant p-values based on right-tailed Fisher Exact Test are denoted as –log(p-value) on X-axis and ratios are displayed as line graph indicating number of molecules from dataset relative to total number of molecules in that pathway; orange and blue bars denote activated and inhibited state respectively.

Table 15. IPA® predicted upstream regulators responsible for observed gene expression changes in HME50-5E relative to HME50.

| GENE | HME-505E vs HME50 | | |
	z-score	p-value	FC
TP53	1.86	2.95E-32	
CDKNIA	3.745	7.82E-23	1.83
Irgm1	5.005	9.82E-20	
NKX2-3	2.34	1.93E-14	
KDM5B	4.019	1.51E-13	1.332
KRAS	2.509	1.83E-13	1.832
VEGF	-6.834	1.61E-22	
TGFB1	-4.342	2.77E-28	
ERBB2	-4.815	8.26E-27	
TNF	-6.249	-1.53E+00	-1.53
CSF2	-7	1.39E-23	-1.66
OSM	-2.268	7.42E-23	
HGF	-5.813	5.53E-21	
ESR1	-2.335	1.38E-19	-2.04
CCND1	-3.737	2.78E-19	-1.575
IL1B	-6.281	5.43E-10	-3.263
IFNL1	-5.466	2.39E-17	
MAPK1	-4.07	9.96E-14	
NUPR1	4.67	1.00E-13	1.6
PTGER2	-5.707	2.98E-15	-1.618
SMARCA4	-4.269	1.45E-13	
STAT1	-2.088	1.59E-15	-2.766
STAT3	-4.897	1.80E-14	
TREM1	-2.097	2.85E-15	-1.936
RABL6	-5.032	7.89E-23	1.39

The table shows upstream regulators identified by IPA® to explain the observed gene expression changes in HME50-5E dataset relative to HME50 based on the information using IPA® Knowledge Base. The upstream regulators are predicted to be activated (orange) or inhibited (blue) based on the activation Z-score; fold changes observed for probesets corresponding to overexpressed or under-expressed genes in the uploaded dataset are colored red or green respectively.

IPA® Core Analysis of altered gene expression in HME50-hTERT relative to HME50 cells: The canonical pathways (**Figure 25**) that contain significant numbers of genes from the observed genes differentially regulated between HME50-hTERT and HME50 are given in **Table 16**. IPA® predicted an activated state for *CDKN1A, NUPR1* while inhibited state for *RABL6, ERBB2, CSF2, CCND1, VEGF, HGF, PTGER2, E2F1, ESR1* upstream regulators. Although *TP53* was predicted as the topmost transcriptional factor based on the direction of expression of 216 genes in the dataset, due to a statistical threshold set, its activation state was not predicted (z-score 1.411, p-value 4.86E-39). The biological functions strongly associated with the altered gene regulation pattern observed to decrease in HME50-hTERT were cell viability, proliferation and survival of tumor cells, migration of cells and growth of lesion whereas increased organismal death increased (z-score 5.916, p-value 6.73E-11). The top disease and functions associated with the HME50-hTERT were cell cycle, cell function and maintenance, tissue development and morphology.
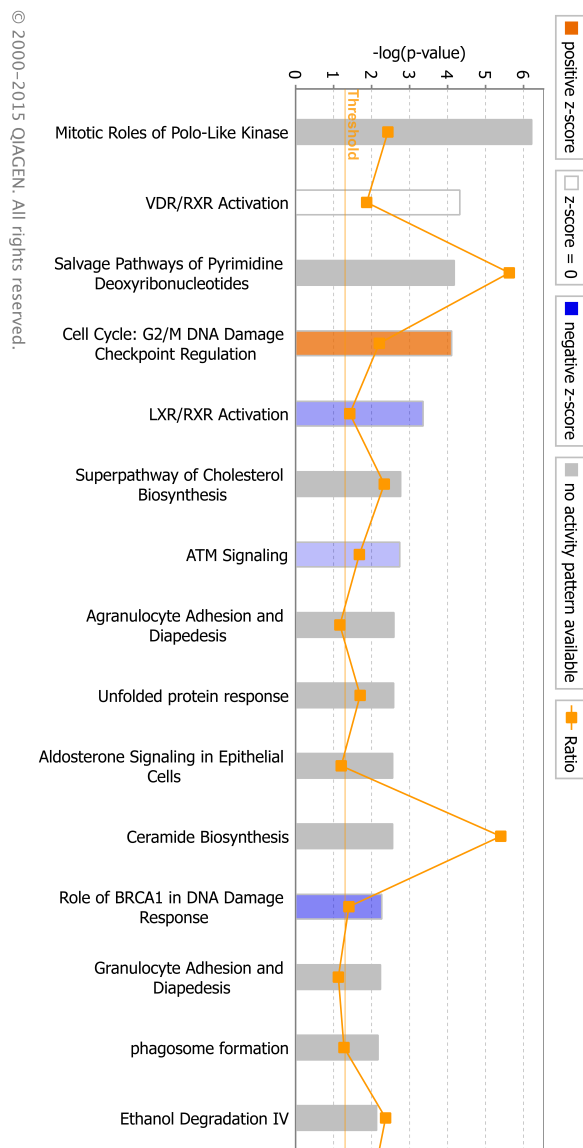
Figure 25. Canonical pathways enriched in HME50-hTERT relative to parental HME50. The bar graph depicts the 15 most significant canonical pathways associated with gene expression differences observed in HME50-hTERT relative to HME50. The significant p-values based on right-tailed Fisher Exact Test are denoted as –log(p-value) on X-axis and ratios are displayed as line graph indicating number of molecules from dataset relative to total number of molecules in that pathway; orange and blue bars denote activated and inhibited state respectively.

Table 16. IPA® predicted upstream regulators responsible for observed gene expression changes in HME50-hTERT relative to HME50.

GENE	HME50-hTERT vs HME50		
	z-score	p-value	FC
TP53	1.1411	4.86E-39	
CDKN1A	2.737	1.32E-32	
NUPR1	8.25	4.01E-18	1.735
Irgm1	4.99	7.69E-22	
CDKN2A	3.892	6.01E-15	
RABL6	-5.06	4.69E-31	-1.36
ERBB2	-2.063	3.23E-29	-1.43
CCND1	-3.171	2.93E-26	
VEGF	-4.751	6.85E-23	
HGF	-4.236	2.17E-20	
PTGER2	-4.764	3.95E-20	-1.63
E2F1	-4.683	4.10E-20	-1.36
ESR1	-3.95E+00	2.41E-19	-2.027
CSF2	-6.325	6.35E-28	
MITF	-5.104	1.00E-15	1.882
FOXM1	-4.886	1.86E-15	-9.891
TBX2	-4.807	1.08E-14	

The table shows upstream regulators identified by IPA® to explain the observed gene expression changes in HME50-hTERT dataset relative to HME50 based on the information using IPA® Knowledge Base. The upstream regulators are predicted to be activated (orange) or inhibited (blue) based on the activation Z-score; fold changes observed for probesets corresponding to overexpressed or under-expressed genes in the uploaded dataset are colored red or green respectively.

IPA® Core Analysis of altered gene expression in HMET relative to HME50

cells: The enriched canonical pathways based on significantly different genes expressed

in HMET cell line provide an overview of cellular processes that are different relative to

parental HME50 are given in **Figure 26**. Based on the gene expression changes observed

in HMET, *TNF*, *HGF*, *VEGF*, *CSF2* and *MAPK1* are predicted (**Table 17**) to be in an

activated state whereas *TP53*, estrogen receptor (z-score -2.742, p-value 7.82E-19), *FAS*

(z-score -2.04, p-value 1.57E-17), *NUPR1*, Alpha Catenin (z-score -2.59, p-value 2.96E-

13) and *AHR* (z-score -2.123, p-value 3.0E-13) are predicted to be inhibited; the z-score

and p-values are in **Table 17**. Moreover, cell movement (z-score 3.0, p-value 1.51E-06),

migration (z-score 2.67, p-value 3.27E-07), EMT (z-score 2.28, p-value 1.06E-07),

inflammation (z-score 2.2, p-value 2.93E-10) were observed to be increased biological

functions in HMET while decrease in proliferation of epithelial cell lines (z-score -2.4, p-

value 5.22E-10) and cell adhesion (z-score -2.0, p-value 8.75E-08) was predicted

biological state for HMET relative to HME50 cells. Interestingly, EGCG was predicted to

be in an inhibited state based on direction of expression of 64 molecules in the HMET

versus HME50 comparison. The mechanistic network for EGCG is depicted in **Figure

27.** The IPA® Core Analysis for other comparisons namely HME50-hTERT vs. HME50-

5E, HMET vs. HME50-5E, HMET vs. HME50-hTERT and, HMET vs. all non-

malignant HME50 cell lines were carried and the resulting canonical pathways are

depicted in **Figures 28 – 31**. In conclusion, IPA® predicted transcriptional changes in

agreement with the properties of each HME50 cell line as characterized.
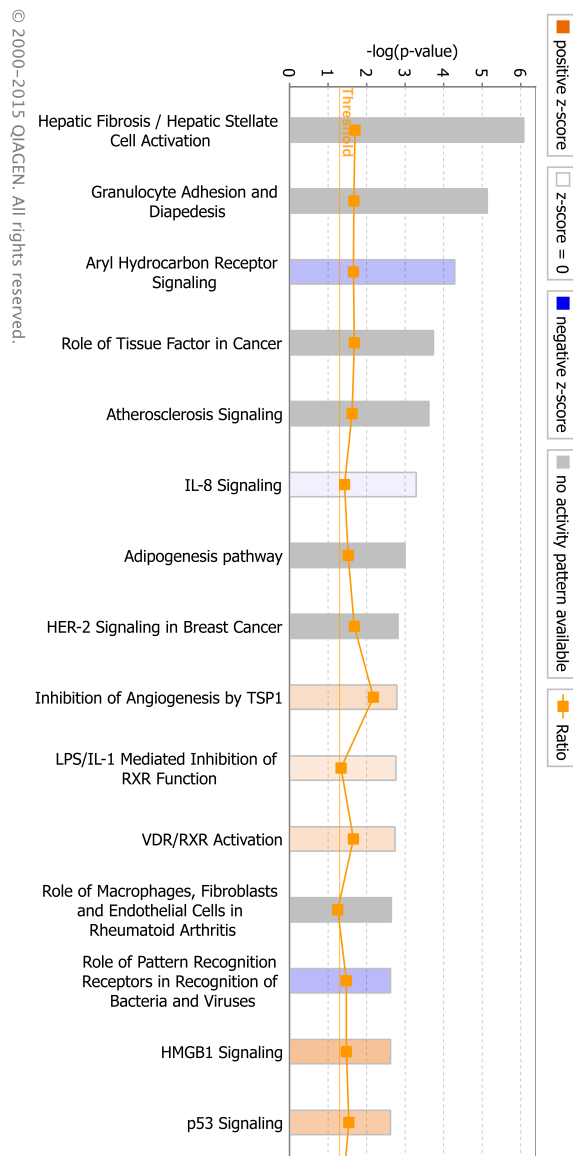
Figure 26. Canonical pathways enriched in HMET relative to parental HME50. The bar graph depicts the 15 most significant canonical pathways associated with gene expression differences observed in HMET relative to HME50. The significant p-values based on right-tailed Fisher Exact Test are denoted as –log(p-value) on X-axis and ratios are displayed as line graph indicating number of molecules from dataset relative to total number of molecules in that pathway.

114

Table 17. IPA® predicted upstream regulators responsible for observed gene expression changes in HMET relative to HME50.

GENE	HMET vs HME50		
	z-score	p-value	FC
TNF	2.26	4.61E-44	
HGF	2.47	1.44E-22	
VEGF	3.282	5.32E-19	
CSF2	3.899	8.70E-17	5.614
MAPK1	2.777	3.47E-16	
SNAI1	3.787	7.66E-08	2.645
FOXM1	3.663	5.97E-09	1.936
E2F1	3.258	3.35E-10	1.891
POU5F1	3.121	3.19E-04	
MMP3	3.011	2.57E-01	90.519
RABL6	2.785	7.20E-09	1.5
IL17A	2.193	2.33E-15	
TP53	-2.846	2.81E-38	2.168
NUPR1	-6.191	2.24E-15	-6.72
SPDEF	-4.059	2.21E-08	
RB1	-3.754	1.07E-08	
ROCK2	-3.503	4.74E-07	
RBL2	-3.079	1.93E-03	-1.45
SREBF2	-3.034	7.01E-03	-1.488
NUPR1	-6.191	2.24E-15	

The table shows upstream regulators identified by IPA® to explain the observed gene expression changes in HMET dataset relative to HME50 based on the information using IPA® Knowledge Base. The upstream regulators are predicted to be activated (orange) or inhibited (blue) based on the activation Z-score; fold changes observed for probesets corresponding to overexpressed or under-expressed genes in the uploaded dataset are colored red or green respectively.
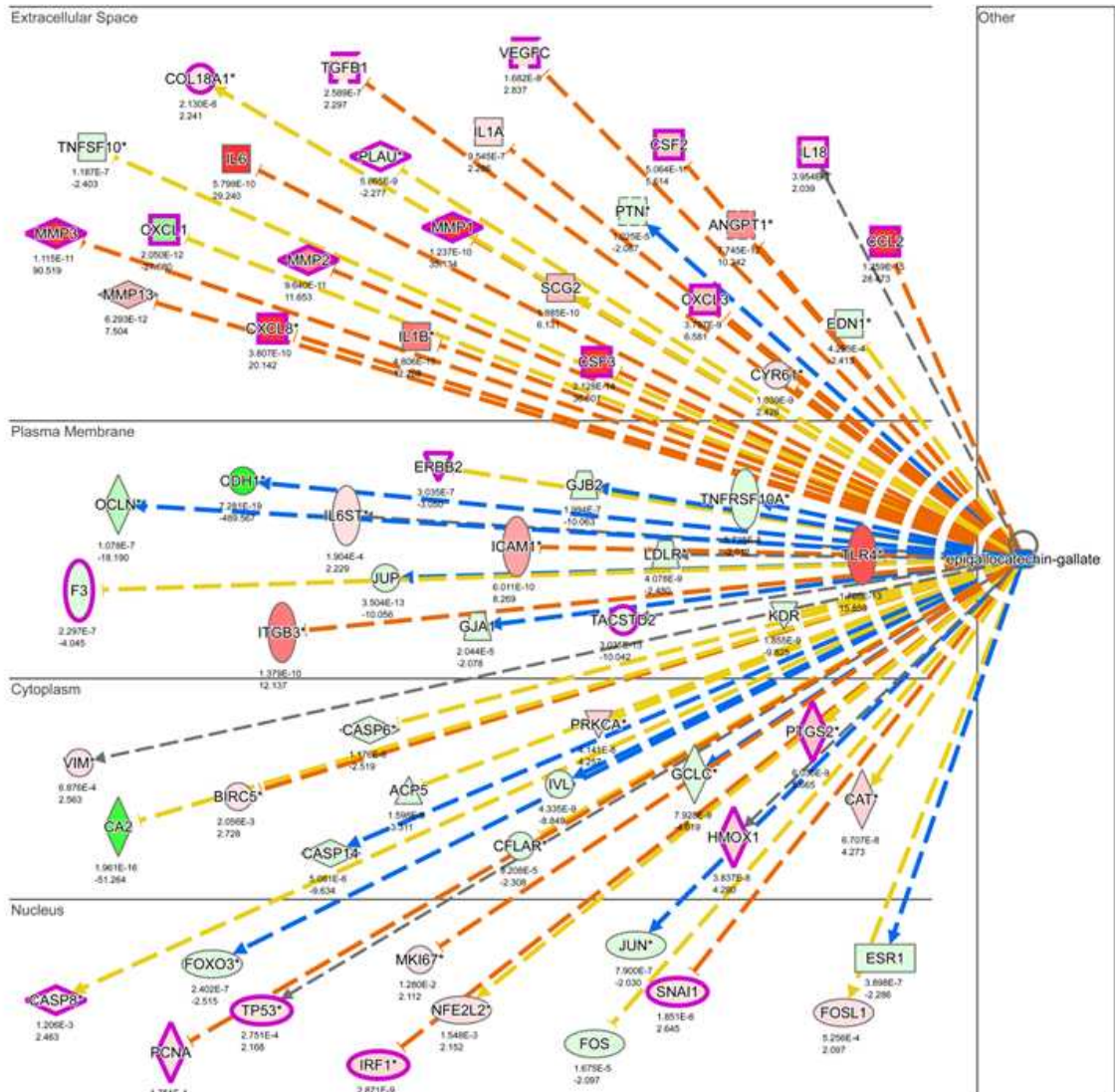
Figure 27. Inhibited state of EGCG is predicted based on connected genes and upstream regulators in HMET cell line. The mechanistic network depicts plausible gene and upstream regulator targets that are deregulated to elicit gene expression pattern that is affected given inhibited state of EGCG. This also indicates the treatment of EGCG may impact the targets and affect the gene expression change observed in HMET cell line. The up-regulated and down-regulated molecules observed in dataset are colored red and green respectively; higher intensity reflects extreme fold change. Orange dashed lines denote interaction leads to activation; blue dashed lines denote interaction results in inhibition; yellow lines represent inconsistent findings for downstream molecule prediction state.
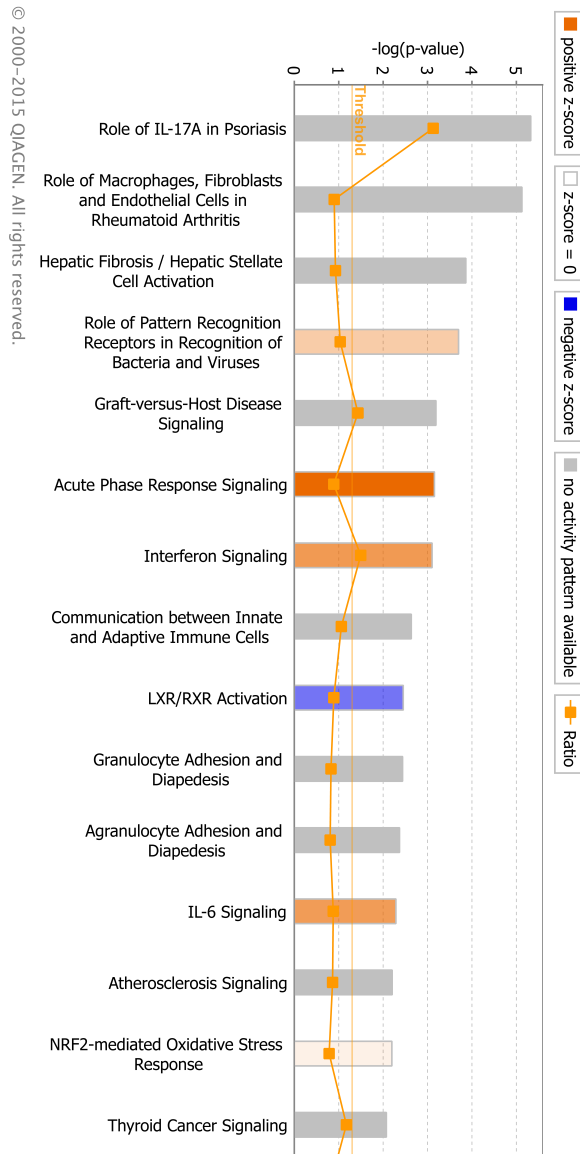
Figure 28. Canonical pathways enriched in HME50-hTERT relative to HME50-5E.
The top 15 canonical pathways associated with gene expression differences observed in HME50-hTERT relative to HME50-5E. The significant p-values based on right-tailed Fisher Exact Test are denoted as –log(p-value) on X-axis and ratios are displayed as line graph indicating number of molecules from dataset relative to total number of molecules in that pathway.

118
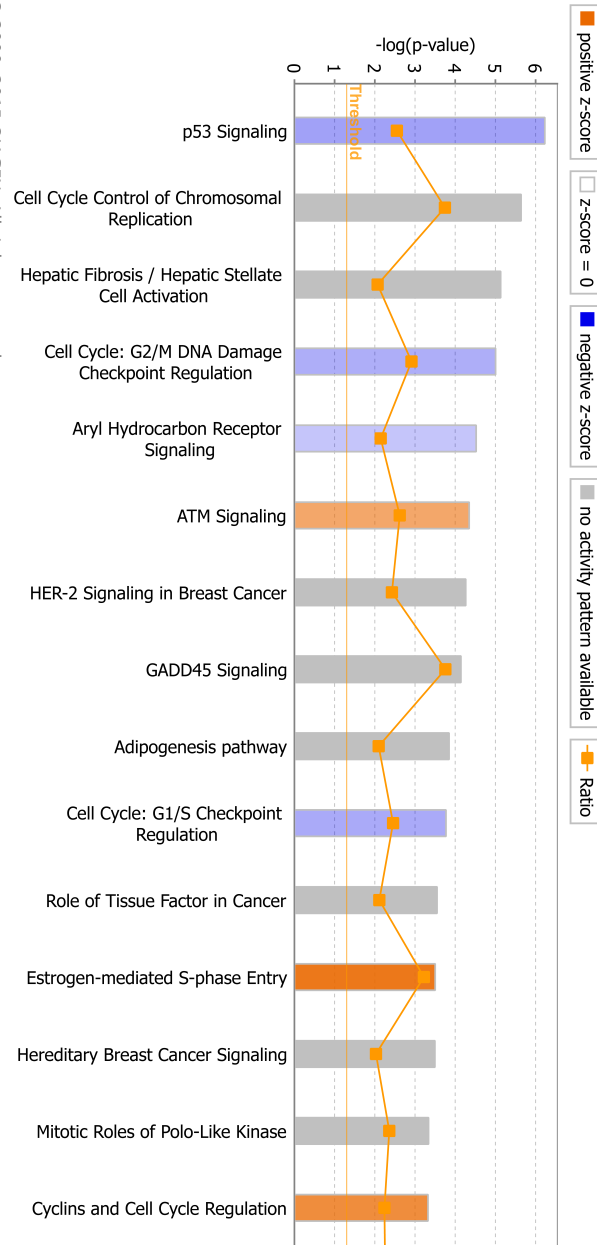
Figure 29. Canonical pathways enriched in HMET relative to HME50-5E. The top 15 canonical pathways associated with gene expression differences observed in HMET relative to HME50-5E. The significant p-values based on right-tailed Fisher Exact Test are denoted as –log(p-value) on X-axis and ratios are displayed as line graph indicating number of molecules from dataset relative to total number of molecules in that pathway.
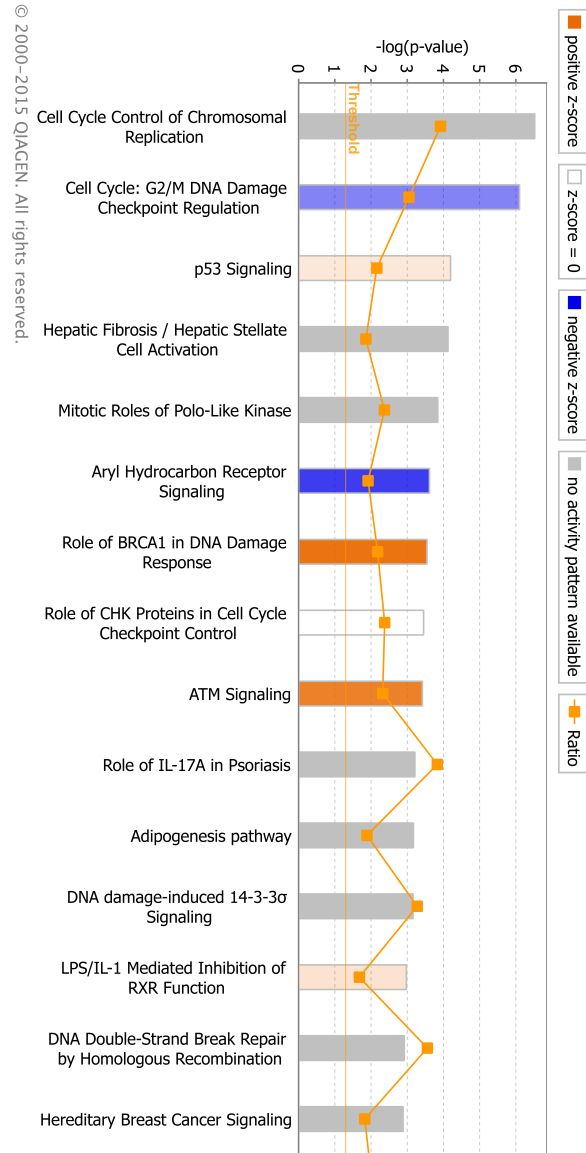
Figure 30. Canonical pathways enriched in HMET relative to HME50-hTERT. The top 15 canonical pathways associated with gene expression differences observed in HMET relative to HME50-hTERT. The significant p-values based on right-tailed Fisher Exact Test are denoted as –log(p-value) on X-axis and ratios are displayed as line graph indicating number of molecules from dataset relative to total number of molecules in that pathway.
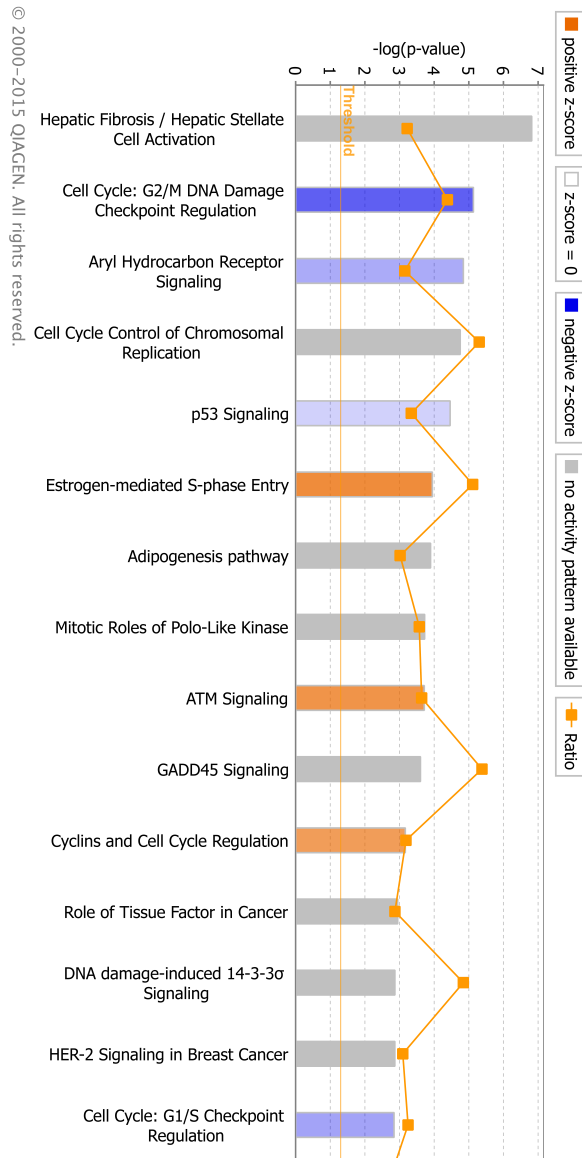
Figure 31. Canonical pathways enriched in HMET relative to non-malignant HME50 cell lines. The top 15 canonical pathways associated with gene expression differences observed in HMET relative to the non-malignant HME50, HME50-5E and HME50-hTERT cell lines. The significant p-values based on right-tailed Fisher Exact Test are denoted as –log(p-value) on X-axis.

3.3.6. Genes involved in EMT program during cancer progression are deregulated in malignant HMET cell line

The deregulated EMT hallmarks can confer invasive, migratory and stem-cell like characteristics that potentiate a complex metastatic cascade which are dependent on various molecular interactions and environmental context. To identify genes known to be involved in EMT process that are differentially expressed in HMET relative to non-malignant HME50 cell lines, comprehensive EMT signatures were appended from the literature consisting of genes from dbEMT database, core EMT signature reported by Taube et al. and literature search to generate a list of 522 known EMT players. To identify the genetic network affected in HMET that switches the EMT process ON, these 522 genes in the list (**Appendix 3**) were used to filter the ANOVA results and 328 genes were identified to be deregulated in HMET relative to non-malignant HME50 cell lines. A volcano plot (**Figure 32**) was invoked to visualize the p-values and fold-changes of these important players in the EMT process based on gene-list derived from Taube et al., 2010 and dbEMT [120, 121] (**Appendix 3**). Down-regulation of main epithelial markers namely *CDH1*, *KRT5*, *KRT6A*, *KRT6B*, *TP63*, *ITGB4*, *DSG3*, *ITGB6*, *SPINT2*, *SPINT3* and *DLGAP5* whereas up-regulation of mesenchymal and breast cancer markers such as *VIM*, *MGP*, *RHOBTB3*, *ZEB1*, *SRGN*, *MMP1*, *MMP2*, *MMP3*, *BIN1*, *CDH11*, *CDH2*, *FBN1*, *COL3A1*, *NID1* and *HAS2* (**Appendix 3**) was observed upon microarray data analysis.

Volcano plot of HMET vs Non-malignant HME50 cell lines
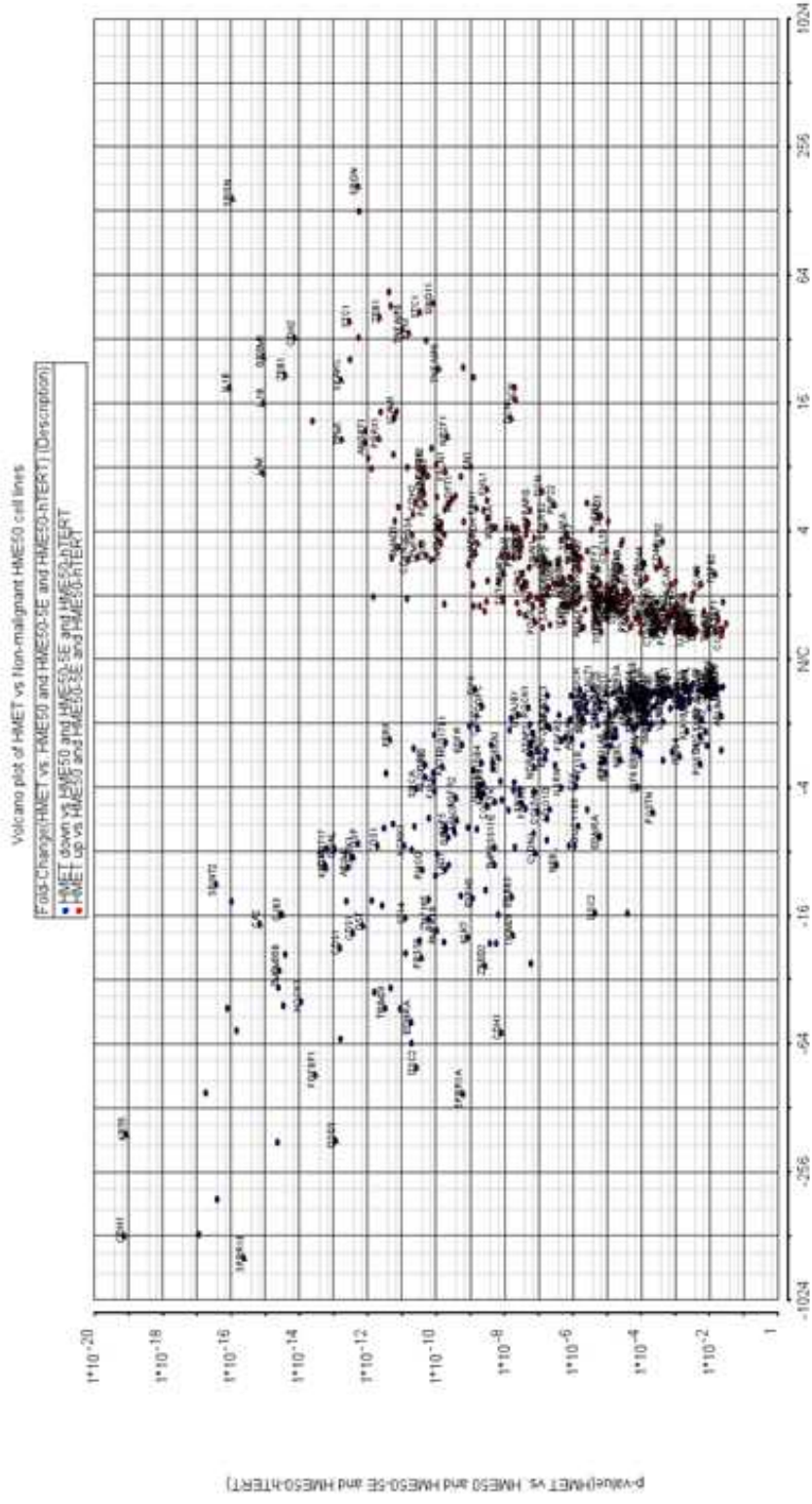
Figure 32. Volcano plot of most significant players involved in epithelial-to-mesenchymal transition program. Volcano plot of EMT genes displayed statistically significant expression differences in HMET samples relative to non-malignant HME50 cell samples. Fold changes are specified on X-axis and the p-value to the corresponding contrast displayed on Y-axis. Each dot represents a gene symbol annotated to the Probeset ID and is colored based on fold-change, up-regulated and down-regulated genes in HMET samples are colored red (positive fold changes on right hand side of X-axis) and blue (negative fold changes on left hand side of X-axis) respectively.

3.4. Identification of drug targets in malignant HMET cell line by pathway analyses and analyses of their pharmacological action in 3D *in vitro* phenotypic assays

3.4.1. IPA® Network Analysis revealed specific targets for pharmacological agents

IPA® pathways and Path designer was used to visualize the networks of differentially regulated genes in HMET cells relative to HME50-hTERT. The gene network was generated and 'add' tool used to find and include APR-246, EGCG, Rapamycin and Resveratrol in the pathway. Next, using the overlay tool the fold change and p-values observed for the expression data were attributed to the genes in the network and all molecules (genes and drugs) were connected to incorporate the knowledge of drug targets present in the Ingenuity® Knowledge Base.

Molecule activity predictor was also used to analyze the effect of drugs on genes and the function epithelial-to-mesenchymal transition (**Figure 33**). The direction of expression of molecules causally affected by each of these drugs relevant to the phenotype 'epithelial and mesenchymal transition' changed in response to activating each of the drug. This further corroborated the evidence for target of EGCG since the molecules that are inhibited by EGCG changed direction in the network after activating EGCG as well as other investigated drugs. For examples, the expression of *TP53*, *TP63* increased whereas the expression of *MMP2*, *SNAI1* and *ZEB1* decreased. The genes differentially regulated in HMET were enriched in the canonical p53 pathway (ratio: 34/98, p-value: 6.35E-05, Z-score: 0.4) and the mechanistic network depicted *TP53* to be in an inhibited state. In addition to this knowledge, network connecting the genes

enriched in EMT process, were the basis for exploring the targets that are affected by pharmacological agents used namely APR-246, EGCG, rapamycin and resveratrol. The known heterozygous *TP53* status, in addition to IPA® predicted TP53 inhibition state was motivation for targeting mutant *TP53* with APR-246.
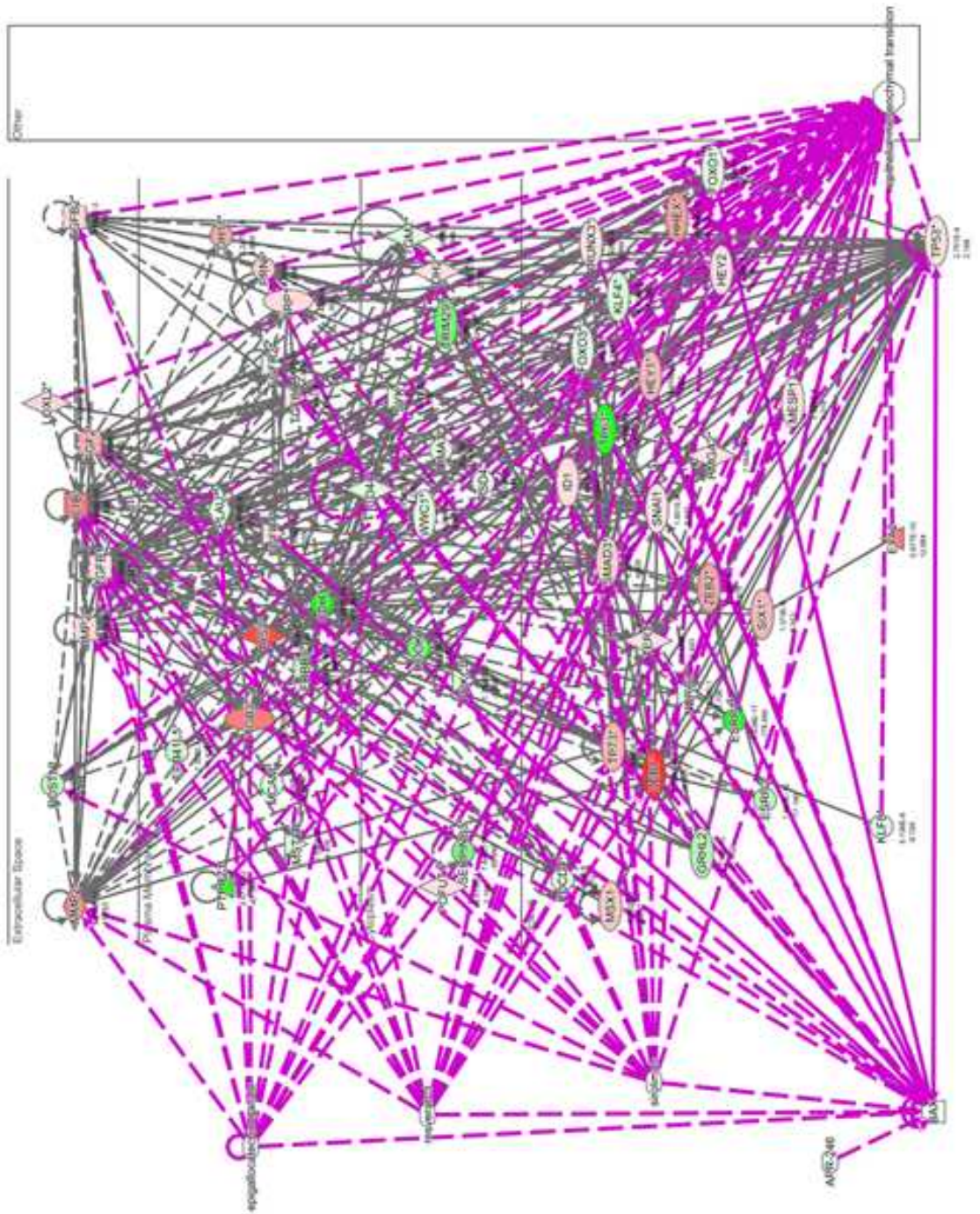
Figure 33. Evidence for pharmacological agents that target the activated state of epithelial-to-mesenchymal transition in HMET. IPA® identified the complex interactive network involved in epithelial to mesenchymal transition in HMET relative to HME50-hTERT based on genes that are deregulated and can be pharmacologically targeted by agents of interest to our laboratory such as APR-246, EGCG, rapamycin and resveratrol. Based on the experimentally observed evidence in IPA® Knowledge base, 62 genes involved in EMT contribute to activation of this functional process in HMET cells (p-value 1.06E-07; z-score 2.3). Red and green colored molecules represent up-regulated and down-regulated genes respectively.

3.4.2. IPA® Network Analysis predicted potential pharmacological agents that target

canonical pathways

The top regulated network comprising of differentially regulated genes in HMET

relative to non-malignant HME50-hTERT were used to generate and overlay expression

values and pharmacological agents of interest (**Figure 34**). The resulting network was

then subjected to molecule activity predictor to visualize effect of said agents on targets

in the network. For instance, EGCG activation leads to down-regulation of *APP* and

vimentin in the network and this information can also be used to identify targets that can

be used as endpoints.
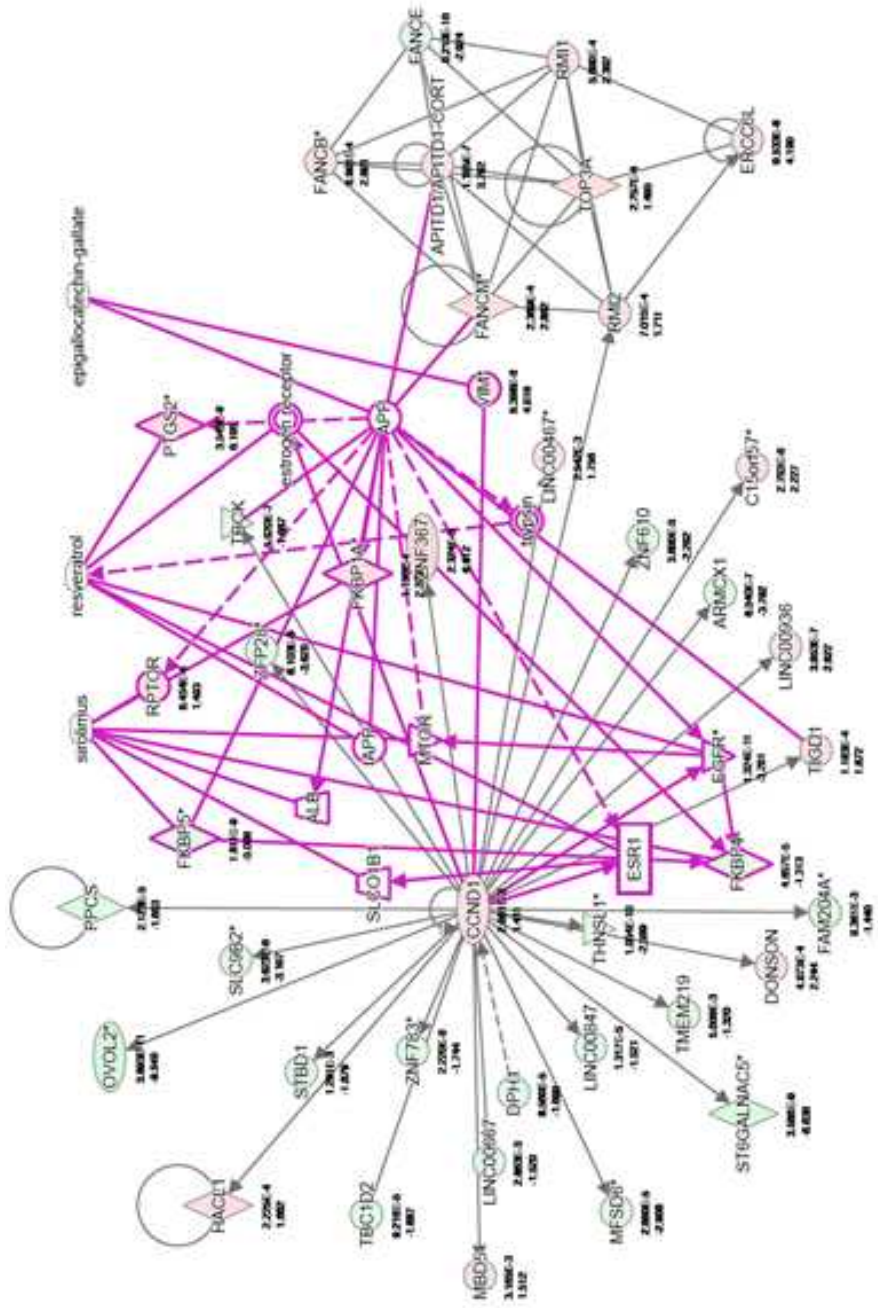
Figure 34. IPA® predicts targets for tested pharmacological agents. IPA® predicts

the deregulated genes in network that would be targeted in HMET cells relative to

HME50-hTERT based on the experimentally observed evidence for APR-246, EGCG,

Rapamycin and Resveratrol in IPA® Knowledge base.

3.4.3. Reversion of HMET invasive structures to organized acinar structures in 3D

culture by select pharmacological agents

The activated wild-type *TP53* functions can result in cell cycle arrest, senescence

and/or apoptosis in response to oncogenic stimuli that is conducive to carcinogenic

progression [122, 123]. This was the motivation for using small molecule PRIMA-1 that

can reactivate mutant p53 and restore its tumor suppressor function. The methylated

version of PRIMA-1, APR-246 was found to be safe and potent inducer of antitumor

effects in clinical trial [124]. Based on previous reports by Herbert et al. [48], PRIMA-1

and APR-246 were ideal pharmacological agents to be tested in the HME50 cell

progression series given the heterozygous mutant p53 background coupled with useful

readout of observing the effect of the agent in 3D Matrigel® culture for proof-of-concept

experiment. Based on IPA® Core Analysis, we were also interested in the potential

chemopreventive effects of EGCG and resveratrol that are predicted to target deregulated

networks in HMET cell line by each of the agents. After treatment, we observed changes

in morphologies of HME50-hTERT, HME50-TR and HMET cells grown in overlay as

well as embed cultures. The confluency (70-75%) of cell lines was maintained such that

3D culture yields consistent sizes of acini and heterogeneity in acini sizes is minimal. It

was observed that overlay method generally resulted in development of slightly larger

acini in shorter duration than the embed method (data not shown). However, the

properties and morphologies of each cell line remain constant for acini formed in both

methods, i.e. all cell lines give rise with >95% acini within the appropriate size range.

Initially, PRIMA-1 (5μM) treatment for 96 hours in 3D overlay culture led to reduction

of acini sizes in HMET cell line alone (**Figure 35-36**). These results prompted treatment

of HME50 cell lines with APR-246 (20μM) for 10 days in 3D overlay culture and image acquisition using stereomicroscope (**Figure 37-38**). We observed the size distribution of acini changed for HME50-TR cell line, close to 60% (n=1500) of the acini formed measured <50μm in diameter after treatment with APR-246. Drastic changes were also observed for HMET cell line where 91% of the acini adopted spherical morphology after treatment with concomitant reduction in formation of irregular stellate colonies (**Figures 37-38**).

In 3D embed cultures, spherical acini (31-50μm) with hollow lumens for HME50-hTERT; larger acini (50-80μm) with hollow lumens for HME50-5E cells; very large acini (75-100μm) with filled lumen for HME50-TR; and stellate HMET (>100μm) disorganized acini were regarded the baselines for experimental analysis. The average acini size of HME50-hTERT, HME50-TR and HMET are 35μm, 70μm and >100μm respectively. The morphologies of HME50 cell lines distinctly change after treatment with APR-246, EGCG and rapamycin in 3D Matrigel embed culture for 10 days (**Figure 39**). HME50-TR acini formed were smaller in size and lumen is observed after treatment. Most striking phenotypic changes were evident in the tumorigenic HMET cell line, as the formation of mass like acini is observed in contrast to stellate-like morphology typical of HMET.
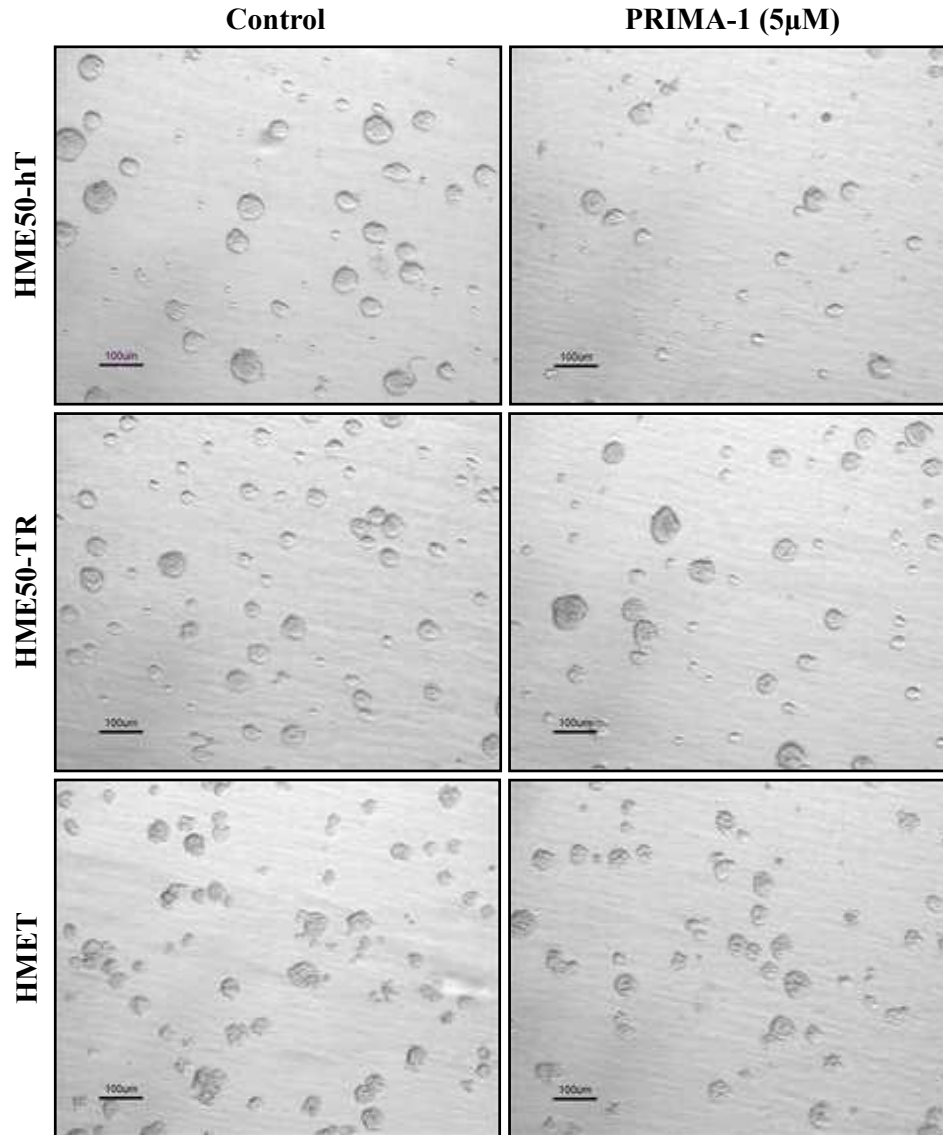
Figure 35. Mutant p53-binding small molecule PRIMA-1 affects acini size in 3D culture. Effect of PRIMA-1 treatment on acini size Matrigel™ overlay culture To study effect of PRIMA-1 on acini size, the stereomicroscope images at 10X magnification (scale 100µm) were captured for size quantification of were performed using Adobe™ software. Mutant p53 binding small molecule PRIMA-1 affects acini size in 3D culture (**Figure 36** for quantification).
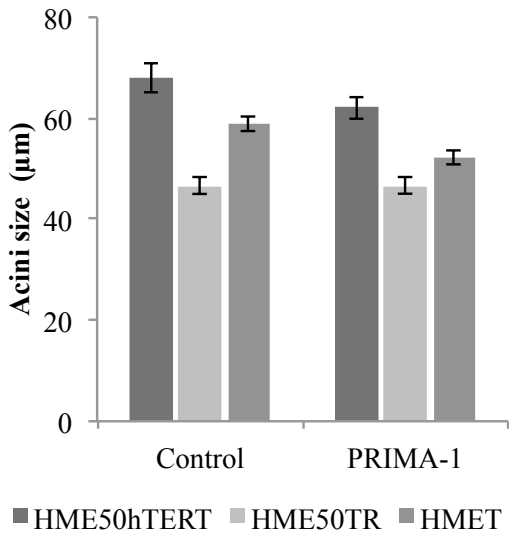
Figure 36. HMET acini size decreases with PRIMA-1 (5µM) treatment. The average

sizes of HMET acini treated with PRIMA-1 for 96 hours differ significantly (***P <

0.001, unpaired, two-tailed t-test; 99% confidence interval). Error bars ± SE (n=500)
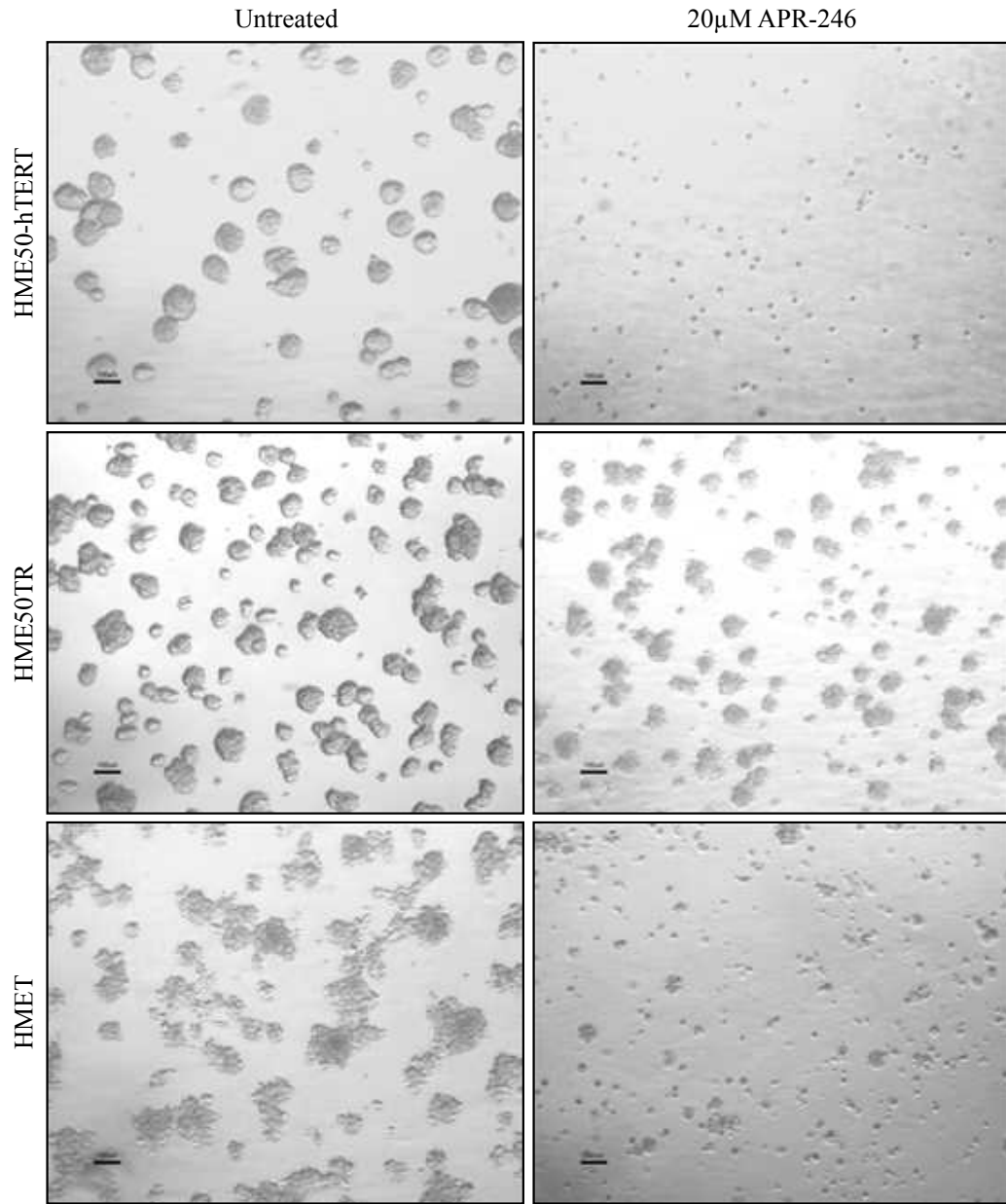
Figure 37. Mutant p53 reactivating APR-246 affects acini size in 3D culture.

HME50-hTERT, HME50TR and HMET cells were grown in 3D overlay culture for 10 days (left column) or treated with (right) of 20μM APR-246 in culture media. The drug was supplemented beginning day 0, 3 hours post seeding and every other day during media changes. Representative images acquired using Nikon SMZ1500 microscope, scale bar, 100μm.

b Effect of APR-246 treatment on size of HME50-hTERT acini, n=1500

c Effect of APR-246 treatment on size of HME50-TR acini, , n=1500

d Observed HMET acini morphologies control, n=1543

Effect of APR-246 treatment on HMET acini Morphologies, n=1541
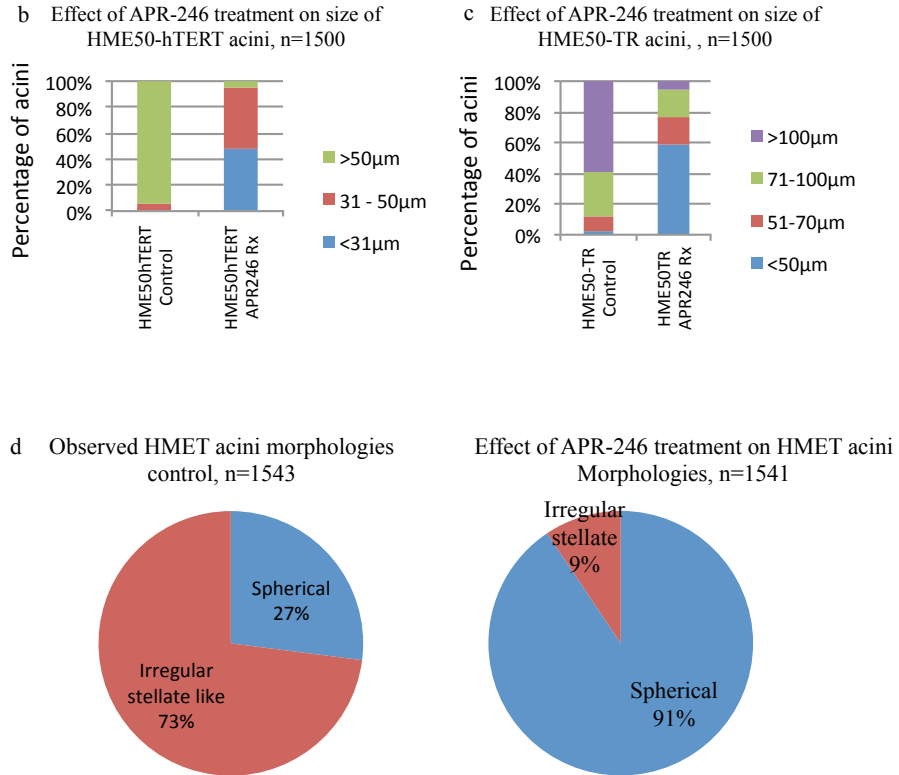
Figure 38. Mutant p53 reactivating APR-246 affects acini size in 3D culture.

Histogram showing the percentage of acini distributed over size categories for (b) HME50-hTERT (c) HME50TR and pie-chart plotted for morphological categories (d) HMET as percentage of population for cells grown in 3D overlay culture for 10 days in presence or absence of APR-246 treatment.
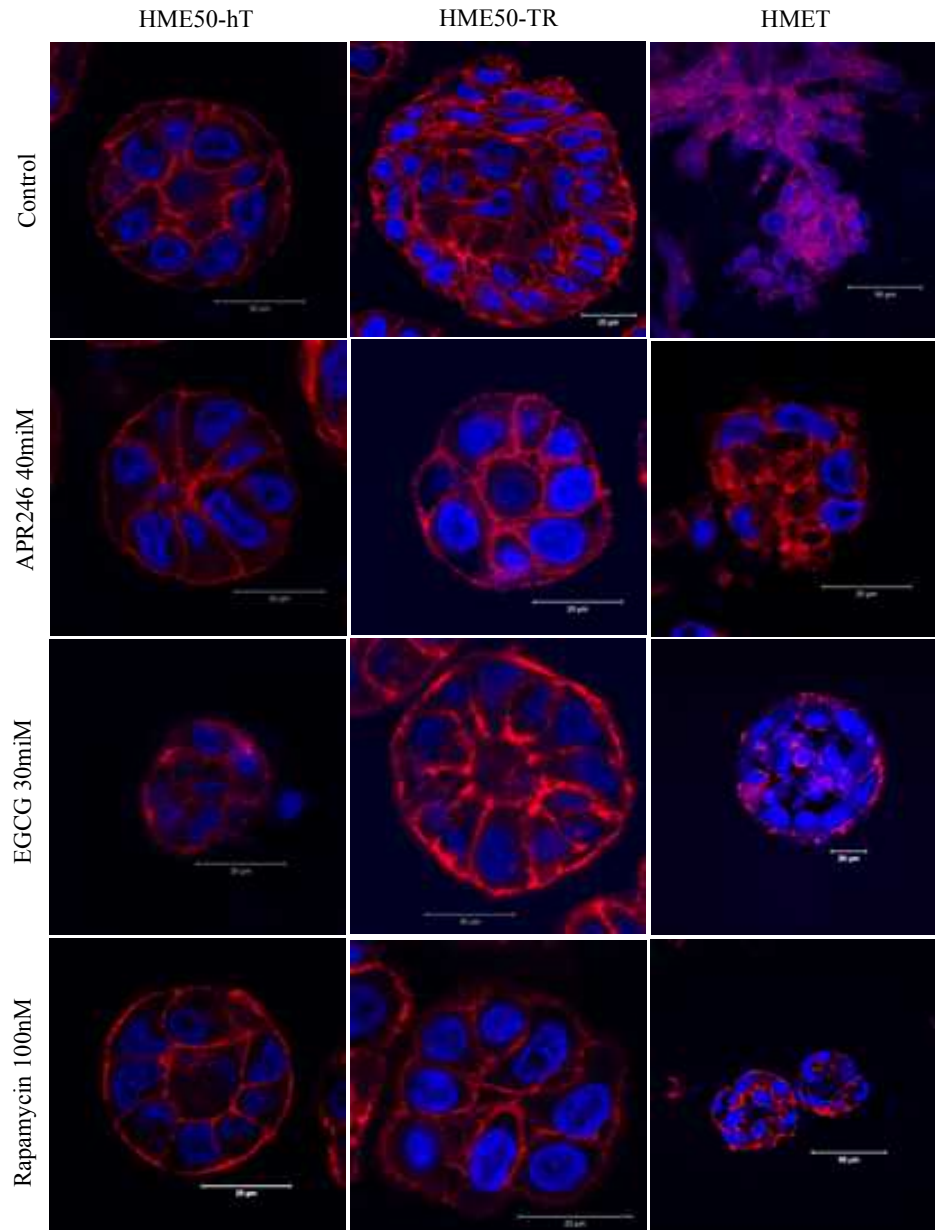
Figure 39. Effect of pharmacological agents on acinar morphologies of HME50 cell lines. HME50 cell lines embedded in 3D Matrigel grown for 10 days untreated (top row) or in the presence of APR-246 (40μM), or EGCG (30μM) or rapamycin (100nM) as indicated. Representative confocal images of center z-stacks of HME50-hTERT (left column), HME50-TR (center) and HMET (right) are shown; scale bar, 20μm.
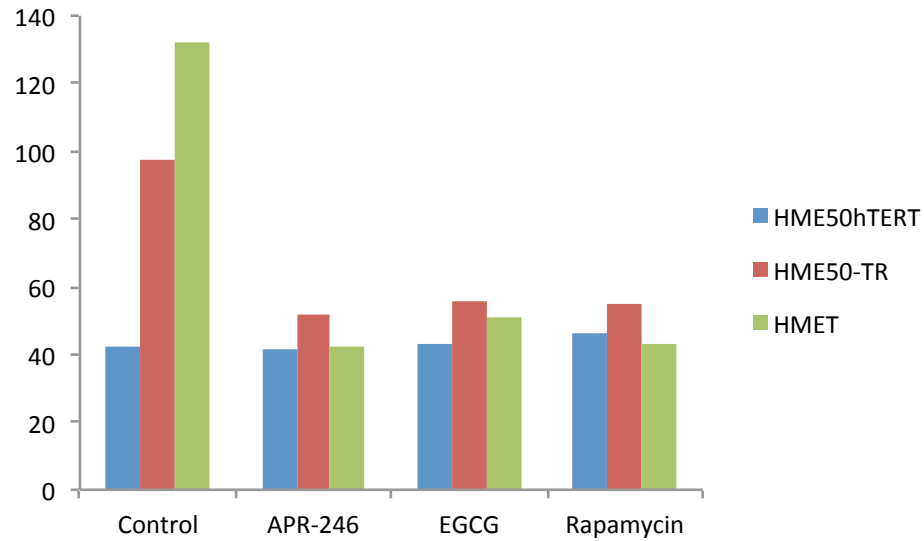
Figure 40. Quantification of the acini size of HME50 cell lines. The acini diameters were quantified after treatment of HME50-hTERT, HME50-TR and HMET 3D embedded cultures with APR-246 (40µM), or EGCG (30µM) or Rapamycin (100nM) for 10 days (20 acini per group). No statistically significant effects on HME50-hTERT acini size were observed; however both HME50-TR and HMET cells showed decreases in acini sizes (pairwise ANOVA comparison with Bonferroni's post-test correction p-value < 0.0001).

CHAPTER FOUR: Discussion

With the recent advances in diagnostic screening and treatment modalities, breast cancer is being detected at earlier stages including benign and precursor lesions. The precursor breast lesions such as atypical ductal hyperplasia (ADH) or ductal cancer *in situ* (DCIS) may not be prerequisites for linear progression to an invasive breast cancer. However, when such ADH or DCIS lesions are diagnosed, surgical excision with radiation therapy is the common clinical management practice [125-130]. In the case of individuals genetically predisposed to the development of cancer such as in Li-Fraumeni syndrome and hereditary breast and ovarian cancer (HBOC) syndrome families, the diagnosis of benign, hyperplastic or precancerous lesions presents a dilemma due to limited chemopreventive and management options [131]. This problem is amplified in the cases of germline mutation carriers who carry variants of unknown significance, thereby increasing the complexity of genetic counseling, recommendations for risk reduction and clinical decisions [132]. The pre-malignant, benign and high-risk disease states are currently not well represented at the preclinical stage [133]. This limitation can be partly addressed by using primary cells and cell lines established from unaffected tissues of patients with increased cancer risk due to inherited or familial cancer predisposition. In this study, we present a model of *in vitro* multistep cancer progression characterized by gene expression profiling to investigate the altered pathways that can be targeted to prevent the progression of premalignant lesions to an invasive stage using the Li-Fraumeni syndrome HME50 cell series.

The HME50 cell series demonstrated unique gene expression patterns that correlated with unique *in vitro* growth phenotypes. Immortalization is the first step during malignant progression and a hallmark of cancer [134, 135]. Toward the goal of establishing malignant progression *in vitro*, immortalization of HME50 cells by retroviral transduction of hTERT was achieved in the first step. The hTERT driven immortalization alone without simultaneous inhibition of *TP53* or other tumor suppressors such as p16 [136] was sufficient to generate HME50-hTERT cell line that is capable of indefinite growth in monolayer culture. Chromosomal analysis using spectral karyotyping (SKY) of HME50-hTERT cells did not exhibit gross chromosomal alterations (data not shown). Other techniques of immortalization such as viral SV40 large and small T antigen, HPV-E6 and E7 mediated immortalization approaches are associated with profound weakening of DNA damage response and gross chromosomal aberrations. These methods of immortalization may facilitate the acquisition of new properties and selection of immortalized cells that do not represent the characteristics of parental source [137, 138]. The inhibition of p53 along with overexpression of telomerase has been implemented in immortalized cell line generation [139]. Due to inherent the *TP53* missense mutation harbored by HME50 patient, a crucial step of inhibiting a major tumor suppressor (e.g. *TP53* or *RB*) essential for the malignant transformation of normal human cells was bypassed unlike in other models requiring viral manipulation [140-142]. This facilitated the use of HME50hTERT cell line as a renewable and genetically stable, non-malignant control with minimal genetic manipulation encompassing introduction of hTERT. Apart from inconsequential differences in gene expression profiles of HME50-hTERT and HME50 cells, the immortalization step resulted in a stable renewable resource that

resembled the parental HME50 cells. In addition, the phenotype of HME50-hTERT cells in 3D monolayer also exhibited normal morphogenesis and a clear lumen which is a hallmark of non-malignant acini. Therefore, hTERT mediated immortalization of primary cells derived from cancer susceptibility syndrome patients undergoing cancer preventative surgeries can be an invaluable source of genetically stable non-tumorigenic cell lines that represent high-risk individuals in preclinical studies. In addition to the spectral karyotyping (SKY) analysis performed in our laboratory, this conclusion is further supported by the microarray data analysis described herein. Using hierarchical clustering analysis, we found that HME50-hTERT closely resembled parental HME50 gene expression profile. Apart from minor gene expression changes, this suggests that although telomerase expression is a hallmark of cancer, the HME50-hTERT maintained the growth characteristic and gene expression profile similar to HME50.

The spontaneously immortalized HME-50E cell line was established from a clone of cells that escaped crises [110]. Aneuploidy is a hallmark of cancer, which is known to confer breast epithelial cells unique features that may promote breast disease [143-145]. HME50-5E cell line although near tetraploid, lacked tumorigenic potential as observed by failure of xenografted HME50-5E cells to form tumors in athymic mice (data not shown) as well as formation of spherical acini with hollow lumens in 3D culture. The differences between aneuploid non-malignant HME50-5E and aneuploid malignant HMET can be exploited to further understand changes associates with breast cancer progression.

In the HME50 progression series, transformation of immortalized HME50-hTERT cells was achieved by simultaneous expression of telomerase reverse transcriptase and a constitutively activated HRas protein. The successive immortalization and constitutive expression of active mutant HRas12 as sufficient to drive the transformation of non-malignant HME50-hTERT to transformed, pre-invasive HME50-TR cells. The transformed HME50-TR cells express HRas, hTERT and inherently carry heterozygous dominant negative mutant *TP53*. With minimal genetic manipulations compared to other breast cell line models such as MCF10-AT or human mammary epithelial isogenic cell line [60, 141, 146] as well as other cell types reported [134, 147-149], these transformed HME50-TR cells, supplemented with extracellular matrix (Matrigel®) were xenografted in athymic mice resulting in their progression to the malignant stage phenotype. It is interesting to note that during inoculation in athymic mice, extracellular matrix was initially essential for successful xenografting of pre-invasive HME50-TR cells even in presence of constitutively active HRasV12 and mutant p53. The resulting tumor lesion was subsequently established as the tumorigenic cell line HMET which exhibited malignant potential independent of extracellular matrix supplement during xenografting process.

The impact of sequential genetic manipulations of HME50 primary cells is evident in the PCA plot which showed that indeed the four HME50 cell lines displayed unique gene expression patterns that allow them to cluster into distinct groups. The HMET cell line has the most distinct gene expression pattern and separated farthest as a cluster. Consistent with the exploratory PCA plot, the hierarchical clustering of HME50

cell lines underscored the drastic differences in gene expression pattern of malignant HMET cells in comparison with the non-malignant HME50 cell lines. Moreover, although aneuploid, the HME50-5E gene expression pattern resembled closely to the non-malignant parental HME50 cells and immortalized HME50-hTERT than the tumorigenic HMET cell line. After filtering the 391 gene identifiers that vary the most across all cell lines, the PCA scatter plot showed HMET cell line grouped separately than the rest of the HME50 cell lines. To identify interesting genes in the clusters grouped in dendrogram, we performed ANOVA analysis on gene set in each cluster and compared each of the cell line compared.

HME50-5E cells although non-malignant are aneuploid and yet more closely resembled HME50 cells than tumorigenic HMET cells. However, we observed up-regulation of 21 genes (**Figure 19**; azure cluster) in HME50-5E relative to HME50 parental cells, in contrast to HME50-hTERT in cluster colored with azure branches on the dendrogram. This group showed overexpression of the *WNT1* inducible signaling pathway (WISP) *WISP3* gene that belongs to connective tissue growth factor family in HME50-5E relative to parental HME50 cells. *WISP3* is overexpressed in colon cancers and is relevant to malignant transformation. On the other hand, HME50-hTERT closely resemble the gene expression pattern of HME50 cell but interestingly, in contrast to HME50-5E cells, show expression differential expression of genes involved in tryptophan degradation, VDR/RXR activation and IL-10 signaling are up-regulated relative to HME50 cell line grouped in cluster with brown branches in dendrogram. Vanin-1 is epithelial surface anchored pantetheinase that hydrolyzed pantetheine to

vitamin B5 and anti-oxidant cysteamine which plays key roles in regulation of essential metabolic pathways and inhibition of invasion [150]. The up-regulation of kynureninase (*KYNU*) involved in tryptophan degradation and *HTATIP2* oxidoreductase with tumor suppressive functions was detected uniquely in HME50-hTERT.

The most distinguishing changes observed were those between the HMET and HME50 non-malignant cells, as shown in the cluster with grey and orange branches (**Figure 19**). The 50 genes in the grey cluster down-regulated in HMET compared to non-malignant group are regulated by *ZEB1*. The EMT transcriptional switches *ZEB1* and *ZEB2* were observed up-regulated in HMET clustered in the group with orange branches. The gene encoding for transcriptional regulator Epitheilum specific Ets Homolgoous factor (EHF) was also predicted to be inhibited based on the expression of its downstream targets of the observed down-regulation. Moreover, the drastic negative fold changes observed for epithelial and ECM markers such as *CDH1*, *TP63*, *ESRP1*, *DOCK8*, *EREG*, *DSC2*, *FBN2*, *ITGB6*, *AREG*, *EREG*, *EDNRA* and *DSG3* in the grey cluster alludes to processes involved in morphology, disease and acquisition of migratory potential. On the other extreme of the grey cluster, we observed 148 genes up-regulated in HMET relative to non-malignant HME50 cells that are relevant to the EMT program and invasive potential. This group comprised of drastic positive fold changes in genes that regulate ECM signaling and remodeling, mediate angiogenesis and protease secretion such as *MMP1*, *MMP3*, *MMP2*, *ZEB1*, *STC1*, *SRGN*, *EDIL3*, *MGP*, *FOXG1*, *CDH2*, *CDH11* and *NID2*. Also, cyclins E2, Cyclin A and chromosome replication related proteins such as *CACS5*, *BUB1* that promote S-phase cell cycle entry and ensure sustained proliferation

were increased. This cluster analysis provided a snapshot of overall differences present in HME50 cell lines and for comprehensive analysis, we next performed 2-way ANOVA to identify genes and pathways involved in HME50 progression.

Consistent with the hTERT induced immortalization step, HME50-hTERT cells exhibit down-regulation of anaphase promoting complex subunit 7 (*ANAPC7*) relative to HME50 cells. Also, similar to HME50-5E, the HME50-hTERT relative to HME50 cells, exhibits down-regulation of crucial molecular players in the mitotic polo-like signaling pathway in addition to *PLK2*. Interestingly, Aurora kinase A (*AURKA*) is down-regulated in HME50-hTERT cells along with *TOP2A* (30 fold down-regulation), *CCNB1*, *CCNB2*, *CDC25C*, *CHEK1*, *CKS2*, *CKS1B* and *PLK1* which are components of cell cycle G2/M DNA damage checkpoint regulation. The down-regulation of these genes indicates activation of G2/M DNA damage checkpoint that maintains genomic instability and prevents malignant transformation of cells by allowing for DNA repair to occur before cells progress into M phase. The proto-oncogene *MDM2* is up-regulated two-fold in HME50-hTERT cells and one of its functions is ubiquitination of p53 thereby initiating its proteasomal degradation. Inhibition of cyclin B1 also blocks cell entry into mitosis ensuring DNA damage check before progression into M phase. It is remarkable to note that a sub-group of human breast cancers that exhibit amplification of both *TOP2A* and HER-2/neu have shown favorable response to anthracycline therapy [151, 152]. This suggests that transformed or malignant cells evolved from HME50-hTERT may not respond well to anthracyclines such as doxorubicin given lack of *ERBB2* amplification and down-regulation of *TOP2A*. The down-regulated *ATM* and *BRCA1* DNA damage

signaling in HME50-hTERT can be attributed to the down-regulated of *BLM*, *CDK1*, *BARD1*, *CDC25C*, *BRIP1*, *FANCD2*, *RFC5*, *RAD51*, *CHEK1* and *BRCA2* genes. *ATM* is the key regulator of DNA breaks and signaling activation of checkpoints, DNA repair and apoptosis. Among other important downstream substrates, the tumor suppressors p53, *CHEK2* and *BRCA* amplify and initiate the DNA repair and G2 cell cycle arrest or inhibition of S phase progression. Similar to HME50-5E, the upstream *CDKN1A* is predicted to be active whereas *ERBB2* is inhibited. Additionally, the transcription factor *FOXM1* is down-regulated (~10 fold) relative to HME50 cells. *FOXM1* is a master regulator of cell cycle progression and it involved in cell proliferation and regulates the expression of cell cycle regulators such as cyclin B1, Cyclin D1 and it is also crucial in DNA damage checkpoint response and elevated *FOXM1* expression has been documented in various breast cancer data sets. *FOXM1* can enhance DNA repair and mediate resistance mechanisms [153]. Down-regulated *FOXM1* coupled with lack of *TOP2A* and *ERBB2* over amplification suggested that the *FOXM1-TOP2A-ERBB2* axis might not be a good target for tumors that evolve from HME50-hTERT and preserve the direction of expression through the malignant progression. Moreover, in HME50-hTERT, cell differentiation and death were decreased while the proliferation, migration and malignant transformation increased as predicted by IPA® functional analysis. These changes can be attributed to the inherent heterozygous *TP53* missense [M133T] mutation and addition of hTERT indicating the HME50-hTERT cell line is primed for malignant transformation. Based on these results, we conclude that well characterized and properly maintained HME50-hTERT cells that carry relatively few gene expression and

chromosomal alterations compared to parental HME50 cells are ideal non-malignant controls for mechanistic *in vitro* studies.

Upon comparing the gene expression profile of aneuploid HME50-5E with the parental HME50 cells, IPA® predicted the upstream regulator *ERBB2* and *CCND1* to be inhibited whereas *CDKN1A*, *KRAS* and *NUPR1* to be activated. The mitotic polo-like kinase pathway is an essential cell cycle regulator and cytokinesis promoter; 17 of the 63 major players such as *CDK1*, *CCNB1*, *CCNB2*, *PLK1*, *PLK2*, *PLK4* and *RAD21* are down-regulated in HME50-5E while the stress-activated protein kinases/Jun amino terminal kinases (*SAPK/JNK*) pathway suppressors anaphase promoting complex and *PPM1L* are up-regulated in HME50-5E. It is of interest that although over-expression of *CCNB1* in conjunction with inactivation of *TP53* has been documented, *CCNB1* and *CCNB2* are in fact, down-regulated in HME50-5E. This possibly alludes to the fine balance that allows HME50-5E cell lines to growing indefinitely in culture despite genomic alterations, as has been reported in *TP53* deficient spontaneously immortalized cells. Interestingly, HME50-5E gene expression data shows down-regulation of claudins family members namely *CLDN1*, *CLDN3*, *CLD4*, *CLD8*, and *CLDN23* that are integral membrane proteins and components of tight junctions that play a major role in cell polarity, permeability if epithelia, and signal transduction programs. Additionally, adhesion molecules, cell surface receptors and extracellular matrix proteins such as fibronectin (*FN1*), matrix metalloproteases namely *MMP7*, *MMP1*, *MMP10*, *ITGA4* and inflammatory cytokines namely *CXCL1*, *CXCL14*, *CXCL11*, *CXCL3*, *CXCL10* are down-regulated relative to HME50 indicating altered adhesion and reduced inflammatory

receptor signaling. The IPA® results also indicated decreased cell migration, survival, proliferation, differentiation, metastasis, vasculogenesis, angiogenesis and decreased chromosomal alignment in HME50-5E cell line. On the other hand, in HME50-5E, apoptosis, keratosis and inflammatory response were an increased function.

The HME50 series comprises of both spontaneously immortalized HME50-5E and hTERT immortalized HME50-hTERT that have ability of indefinite proliferation cell lines that provided an opportunity to compare the molecular players involved in both spontaneous and hTERT immortalization events. By pathway analysis of differential gene expression in HME50-hTERT relative to HME50-5E cell lines, we observed up-regulation of *S100A7* transcript relative to spontaneously immortalized HME50-5E cell line. The members of the S100 family of proteins are known to be expressed variably in different subtypes as well as in different stages of breast cancer and can be affected by the hormone receptor status [154]. In an ERα-negative context, *S100A7* is associated with aggressive disease and can induce ductal hyperplasia *in vivo*; on the other hand, *in vitro*, *S100A7* can activate pro-survival signaling to promote anchorage independent growth. Additionally, *S100A8* and *S100A9* were also up-regulated in HME50hTERT cells. The proteoglycan Versican (*VCAN*) protein is an important ECM component central to epithelial morphogenesis and it is involved in cell adhesion, proliferation and migration. Down-regulation of *VCAN* in HME50-hTERT suggests perturbed intercellular and ECM signaling in addition to up-regulation of TERT (~4 fold) and kallikrein 5 peptidase which is involved in desquamation of epidermis. This suggests retroviral hTERT mediated immortalization of parental HME50 cells with hetereozygous *TP53* background can

149

initiate priming of the malignant progression cascade. Notwithstanding the *S100A7* and *VCAN* expression in HME50-hTERT, IPA® predicted activated Oncostatin M signaling which indicates inhibition of proliferation of tumor cells as well as increased immune response with chemotaxis.

In the HMET microarray gene expression dataset, we detected the transmembrane collagen type XVII alpha1 (*COL17A1*) gene encoding the structural component of hemidesmosomes that mediates epithelial interaction with basement membrane is down-regulated along with fibrillar forming *COL5A2* that binds to thrombospondin and heparin sulfate, *COL2A1*, and down-regulation of structural constituent of extracellular matrix *COL12A1* that confers tensile strength in HMET relative to HME50. Down-regulation of these collagens was accompanied with up-regulation of *COL3A1*, *COL1A1*, *COL5A1*, *COL6A2*, *COL13A1*, *COL6A1*, *COL4A1* and *COL4A2* indicating collagen expression involved in fibrosis and ECM remodeling. The up-regulation of immunoregulatory and inflammatory molecules intereukin 6 (*IL6*), *CCL2*, *CXCL8*, *CXCL3* and *TL4*; matrix metallopeptidases *MMP1*, *MMP2* and *MMP13*; growth factors *IGFBP4*, *PGF*, *IGF2*, *IGFBP3*, *FGF2*, *PDGFRA*, *FGFR1* and *TGFB2* collectively hint to an aggressive EMT phenotype. The aryl hydrocarbon receptor signaling pathway plays an important role in modifying immune responses and in the major stages of tumorigenic cascade and is predicted to be inhibited in HMET cells. The knowledge of functional role of *AHR* in tumor proliferation is currently incomplete as conclusions are based on the cell lines/model used and mechanism being investigated. The AHR signaling depends on the microenvironment context of tumor and can contribute to inflammatory signaling through

various mechanisms, however, the evidence for direct for transcriptional regulation of IL6 by AHR has been established [155]. Interestingly proliferation of epithelial cell lines and adhesion of breast cancer was predicted to be decreased; whereas migration, epithelial-to-mesenchymal transition and inflammation predicted to be increased. In conclusion, the malignant nature of HMET can be attributed to the gain of epithelial-to-mesenchymal transition capability which confers cell motility, stem-cell like plasticity, apoptotic resistance, loss of cellular communication and chemoresistance.

Since HME50-hERT cell line was used as a non-malignant control in lieu of primary HME50 cells, we compared the differential gene expression pattern of malignant HMET with HME50-hTERT cells. The up-regulation of members of multiprotein DNA replication initiation complex that includes *CDC45*, *MCM5*, *MCM2*, *MCM7*, *MCM6*, *ORC1*, *ORC6*, *CDC6* in addition to *TOP2A*, *AURKA*, *CDK1*, *PCNA*, *BIRC5* and down-regulation of stratifin (14-3-3 Sigma) protein in HMET relative to HME50-hTERT indicated increased proliferation and decreased G2/M DNA damage checkpoint response. These pathways are accompanied with evidence of activated ATM signaling and *BRCA1* mediated DNA damage response in HMET cells. The predicted activation and observed up-regulation of *CCND1*, *HGF*, *RABL6*, Vegf, *TBX2*, *FOXM1*, *MYC* and *HRAS* complemented with inhibition of *TP53*, *CDKN1A*, *RB1*, *CTNNA1* and *SMARCB1* fueled the malignant progression of HMET. The HMET malignant gene expression signature was consistently observed when comparing the non-malignant HME50 cell lines as the pathways and gene expressions differences are preserved irrespective of the differences between the non-malignant HME50 cell lines. This was also supported by GSEA result,

which indicated HMET cell line to have a mesenchymal phenotype whereas HME50, HME50-5E and HME50-hTERT to exhibit a basal-like gene expression signature.

Taking both the general morphology/phenotype characterizations and the gene expression profiling of the HME50 cell series together, we were able to utilize these cells for the experimental therapeutics research. The reversion of the invasive phenotype of breast cancer cells by select agents, as well as the disruption of normal acinar architecture by providing oncogenic stimuli has been shown by the pioneer work of Bissell and Brugge groups [68, 71, 109, 156-159]. These studies were instrumental in cataloging stages of breast acinar *in vitro* morphogenesis as well as identifying novel targets for pharmacological intervention. Cell lines are indispensable tools for understanding biology, drug discovery and development and this underscores the need of additional diverse cell line progression systems that can recapitulate complexity and molecular heterogeneity of breast cancer for preclinical research [160-162]. The growth properties in monolayer, three-dimensional culture and the gene expression profiles of HME50 progression series together suggest that HME50 cell lines can be used as preclinical three-dimensional system that can be easily manipulated for pharmacological testing with phenotypic reversion as an endpoint parameter. The distinct properties of increasingly transformed HME50-hTERT, HME50-5E, HME50-TR and HMET derived by genetic manipulations of HME50 parental source are well defined by their gene expression profiles and are evident in 3D culture. In this study, we assessed and distinguished the cell lines based on their size, morphology and ability of lumen clearing and used the size and morphology as endpoint parameters to test effect of various agents. The immortal

HME50-hTERT cells grown embedded in 3D Matrigel culture exhibited a spherical morphology with average diameter of 50μm, growth arrest, robust cell-cell communication and radially organized cells around a clear lumen similar to the non-malignant HMT-3522 S1 described by Kenny et al. [159]. In contrast to the non-malignant HME50-hTERT acini, the transformed, pre-malignant HME50-TR cells exhibit dis-organized acini that lack hollow lumen by day 10 of morphogenesis; however, they exhibited robust cell-cell adhesion with larger spherical acini of average 78μm in diameter. On the other extreme of the spectrum, malignant HMET cells lacked acinar organization and displayed invasive stellate morphology similar as malignant cell lines that lack E-cadherin such as MDA-MB-231 [159]. Until day 3 after embedding HMET cells in Matrigel, HMET acini are not easily distinguishable than their non-malignant counterparts by phase contrast microscopy. However, by day 5, the HMET acini grow considerably larger than other HME50 cell lines and exhibit invasive projections and motility evident by bridges between multiple acini visualized by phase contrast and confocal microscopy. By day 10, HMET acini can be observed as stellate like structures >100μm in size comprising of multiple colonies and invasive protrusions. Owing to these easily distinguishable characteristics of HME50 cell lines in culture and guided by the gene expression profiling data, the HME50 progression series can be further used to explore the biological differences between different stages of cancer progression *in vitro*.

The analysis of HME50 gene expression profiles and identification of the differentially regulated gene sets identified among these cell lines, enables further hypothesis enabled studies on novel genes of interest and effect of pharmacological

agents by utilizing acinar morphology as endpoint parameter. The known heterozygous mutant *TP53* status of HME50 cells provided a known target for small molecule inhibitor APR-246 for proof-of-concept study. Our goal was to determine whether the changes in HME50 acini size and morphology can be used as a readout to predict targeting of deregulated pathways as has been reported previously for HMT-3522 progression series [68, 71, 72, 156, 163]. We observed, APR-246 treatment resulted in reversion of stellate morphology of HMET cells in majority of the acini (p-value<0.001) and did not affect the size or viability of non-malignant HME50 cells. This provided a confirmation that changes in size and morphology of HMET acini can be used as readout parameter to test pharmacological agents. Next, by analysis of deregulated gene expression, the potent polyphenol in green tree extract, EGCG was predicted as an inhibited chemical drug using IPA® analysis based on the knowledge of gene expression pattern of the downstream targets. We used this information to treat the non-malignant, pre-malignant and malignant HMET cells embedded in 3D culture and then studied the effect on acinar morphology. We found EGCG resulted in drastic change in acinar size and morphology of HMET cells as well as clearing of lumens in pre-invasive HME50-TR cell by confocal imaging. Similarly, this system was also used to test effect of resveratrol and rapamycin and their targets were identified using IPA® Network Analysis. The HME50 progression system thus provides an opportunity to study pharmacological intervention using chemopreventive or novel chemotherapeutic agents on malignant cells possessing invasive potential as well as non-malignant controls arising both arising from a common source. This strategy eliminates the bias arising from employing normal control and malignant cell lines from a different source and genetic background and provides a

154

precise experimental control due to well-defined series of manipulations used. Unlike

other routinely used well-characterized models established from reduction

mammoplasties or fibrocystic disease, utilizing the tissue donated by predisposed

individuals during preventative surgeries provides an opportunity to dissect novel tumor

suppressor mutations as well as variants of unknown significance in familial cancers

allowing hypothesis testing on a model systems derived from different individuals. This

is especially pertinent in cases where genotype/phenotype correlation and differential

impact of cancer predisposing *TP53* missense mutations in development of various

cancers as observed in Li-Fraumeni syndrome families needs to be ascertained.

CHAPTER FIVE: Concluding remarks and future directions

5.1. Analysis of the molecular drivers of tumor reversion in the HME50 cell progression series and potential for novel chemopreventive agent testing

Many seminal studies have insinuated the capacity of cancer cells to differentiate and reprogram into normal cells that are capable of maintaining the refurbished normal state and are reviewed in [164, 165]. The reversion models propose that the cancer cells can turnaround to the non-malignant state either by (i) a single step loss or inhibition of key transforming events for example loss of oncogenic viruses or gene [166], (ii) multistep molecular changes that revert the malignant phenotype back to normal due to suppression of major oncogenic cue e.g. role of cues from 3D Matrigel® and integrin blocking to regulate the oncogenic pathway [71] or (iii) gradual rewiring of the complex molecular circuitry in a cancer cells that was originally fuelled by multitude of oncogenic stimuli (accumulation of mutations, loss of tumor suppressors and tumor promoting microenvironment as coconspirators) [164]. This third model is intriguing because Telerman and colleagues have proposed that the oncogenic mutation or tumor promoting signals are not lost but in fact effector genes downstream of the oncogenic signal activate alternative mechanisms responsible for reprogramming cells back to native state [164]. Telerman and colleagues developed comprehensive models (leukemia, breast, colon, melanoma and lung cancer cell lines) and used negative selection and identified about 300 major effectors of reprogramming pathways in cancer cells that mainly are regulated by *SIAH1* (Seven In Absentia Homolog 1), *PSEN1* (Presenilin), *STEAP3* (Six-

Transmembrane Epithelial Antigen of Prostate 3; previously called *TSAP6*) and *TPT1* (Tumor protein, Translationally-controlled) that inhibit tumor progression [167-169]. This work by Telerman, Oren, and colleagues using the same cell models further showed *SIAH1* and *STEAP3* are up-regulated in revertants after induction of wild-type p53 function whereas, the *PS1* and *TPT1* are repressed in presence of wild-type p53 function suggesting selection pressures in mutant *TP53* context in cell lines derived from cancers originating in different tissues. Also, all revertants expressed increased levels of the p53 direct transcriptional target, *CDKN1A* (p21). The SIAH proteins are ubiquitin ligases involved in apoptotic program and are known to affect numerous important players implicated in cancer progression such as *APC*, *NUMB*, *NOTCH* and can modify the transcriptional landscape by guiding degradation of transcriptional regulators. Previously, overexpression of *SIAH1* has shown to cause tumor reversion phenotype in MCF7 breast cancer cell line, and myelomonocytic leukemia U937 cell line [164]. Also, the *STEAP3* protein is up-regulated by *TP53* and has a p53-dependent pro-apoptotic function. On the other hand, the inhibition of Presenilin 1 was observed in tumor reversion models [170] and promotion of apoptosis in these models signified PS1 repression is crucial for the reversion process. The translationally controlled tumor-protein, *TPT1* is an anti-apoptotic factor that also regulates several cellular functions and its inhibition in tumor cells resulted in apoptosis [168, 169, 171]. The expression changes of *SIAH1*, *STEAP3*, *PSEN1*, *TPT1* and *CDKN1A* are relevant parameters to the HME50 progression model as plausible endpoints in drug response studies in a heterozygous mutant p53 setting. The HME50 microarray data showed *SIAH1, STEAP3, PSEN1, TPT1, CDKN1A* are all down-regulated in HMET (statistically significant <-1.5 fold down-regulation) relative to

157

HME50-hTERT. One future goal is to pharmacologically target mutant p53 and study the gene expression changes of associated with the resumption of wild-type p53 function using qRT-PCR analysis. Also, since the 3D culture system gives an unparalleled experimental control and ability to study cellular function in a physiologically pertinent setting, we would extend the pharmacological testing using EGCG, resveratrol and rapamycin to see the effect of treatment on the gene expression changes and the concomitant changes observed with the acinar morphologies in 3D culture. Although reversion of the malignant phenotype is a novel intervention to block metastatic progression [172], the caveat is that the "reversion" of malignant phenotype may be transient and thus it would need to be ensured the tumor revertants have disarmed the tumorigenic arsenal after complete reversion. To ensure this in the 3D culture, a more comprehensive gene signature will need to be examined in addition to the re-establishment of apical-basal polarity and loss of mesenchymal markers *in vitro* (**Figure 41**). Since neomorphic *TP53* mutants can have distinct gain-of-function properties, it is imperative to test the *TP53*-related gene functions in models that represent the heterogeneity in appropriate tissue- and mutation-context. To address this issue, primary cells derived from tissues surgically excised as a preventative measure for individuals harboring distinct mutations in tumor suppressors (e.g. a *BRCA1/2* mutation and *TP53* mutation) can be used to as preclinical models. For instance, the Herbert laboratory has developed a cell line derived from another LFS patient [48] with heterozygous truncating germline *TP53* mutation that can be used as 3D culture models (**Figure 42**) for mechanistic and pharmacological testing to identify mutation based differences.

a Models

p53 dependent
TSAP1–8; TSIP1–2 (10 genes)

p53 independent
TSAP9–22; TSIP3 (15 genes); 263 genes by differential cDNA display, MEGASORT and MPSS

b Pathways

SIAH1 E3 ubiquitin ligase
KID, GluR1, MYB, NCOR, TRAF2, T-STAR, α-Tubulin, Phospholipase C
PHD, HIF, HIF1α, HIPK2, TIEG1, BAG1, PML
DCC, CTIP, ORF45, APC, β-Catenin, NUMB, Notch, Synphilin 1, PML–RAR

PS1 γ-secretase
APP, APLP1, APLP2, APOER2, ERBB4, GluR3, LRP1, SCNB2, ADAM10, ADAM17
CD44, CSF1
DCC
EphrinB1, EphrinB2, HLA-A2, GHR, p75^NTR, NRADD
Notch, Jagged, Delta
E-cadherin, N-cadherin, γ-Protocadherin
Syndecan 3, Nectin 1α, Tyrosinase, TYRP1, TYRP2

TSAP6
NIX
MYT1

TSAP6 controls the secretion of TCTP

TCTP
EFIBβ, EFIA, CHFR, PLK1, BCL-X₁, MCL1, Tubulin, Ca²⁺, OCT4, NANOG

c
Upregulation of SIAH1
Dimming of signalling pathways

Downregulation of PS1
Dimming of signalling pathways

Upregulation of TSAP6
Control of exosome secretion

Downregulation of TCTP
Reprogramming of tumour cell growth and phenotype

159

Figure 41. Molecular players in tumor reversion. Based on the above schematic adapted from Telerman et. al., 2009 [164] which depicts the molecular circuitry involved in tumor reversion, we identified genes deregulated in HMET that may be involved in both p53 dependent (e.g. *TSAP1-8*; *TSIP1-2*) and p53 independent pathways (e.g. *TSAP9-22*; *TSIP3*). Microarray data analysis showed *SIAH1* E3 ubiquitin ligase, *PSEN1*, *STEAP3* and *TPT1* were down-regulated in HMET and their substrates may be up-regulated or down-regulated in HMET relative to HME50-hTERT as indicated by red and green boxes respectively.

The tumor suppressive functions of wildtype p53 confer cancer cell sensitivity to chemotherapy and radiation therapy. Since TP53 mutations are observed in sporadic as well as inherited human cancers, pharmacological agents that can exploit the compromised p53 pathway, for example in the context of germline mutant TP53 carriers in LFS families are required to overcome chemoresistance and cancer progression. Apart from the novel p53 targeting approaches such as chemoradiation, gene therapy, small-molecule reactivation of mutant p53 (e.g. PRIMA-1) and small peptides, a unique promising agent is FDA approved antidiabetic drug metformin (a biguanide derivative for treatment of type II diabetes) which selectively targets cancer cells lacking normal p53 function [173, 174]. Based on epidemiological studies, it was observed that metformin use reduces cancer risk and mortality in Type II diabetics. Although the mechanism still remains unclear, preclinical data has showed anti-proliferative properties of metformin on different cancer cell lines, anti-cancer effect in animal models and beneficial effect in colon and breast cancers in a clinical trial [175-181].

Amongst other effects, metformin modulates mitochondrial function by decreasing oxidative phosphorylation and preclinical research suggests metformin to be particularly effective as an anti-cancer and anti-proliferative agent in context of p53-deficient background [173]. Since, no approved chemopreventive agents currently exist for LFS patients, the tolerability and the effect of metformin on circulating *IGF-1*, insulin and *IGFBP3* is being determined in an interventional pilot study (NCI identifier: NCT01981525).

A registry of clinical trials with metformin conducted for prevention of cancer in human participants is available online on Clinical trials database (www.ClinicalTrials.gov) and these have been recently reviewed by Kasznicki et al. [174]. Targeting the HME50 cell line progression series in 3D culture is our future goal to establish its effect on malignant progression and determine plausible endpoint markers for proposed mechanism of action of metformin in cancer.

The mutations in *TP53* are frequent in triple-negative breast cancers (TNBCs) and probably are one of the most important drivers of TNBC [44, 182, 183]. Since HME50 cell lines carry mutant *TP53* as exhibit Estrogen Receptor negative status (unpublished data), the triple-negative breast cancer (TNBC) Subtyping of HME50 cell lines using microarray gene expression data was performed using a web-based prediction TNBCtype subtyping tool as per instructions [184]. This TNBCtype tool analysis, using the gene expression data, assigned the LFS HME50 cell lines to a distinct TNBC subtype with corresponding correlation coefficient and P-value from 1000 permutations.

For the TNBCtype analysis, the samples with correlation coefficient >0.1 and P-value <0.05 were assigned a TNBC subtype (**Table 18**); briefly, non-malignant HME50-hTERT are predicted to have gene expression signature that corresponds to Basal-like 2 (BL2) subtype whereas, malignant HMET cell lines correspond to the mesenchymal stem-like (MSL) subtype. These results implied that the gene unique gene expression signatures reflect characteristics of distinct TNBC subtypes and that the HME50 cell lines may respond differentially to chemotherapeutic agents [30, 184, 185]. For instance, based

on the preliminary microarray data analyses using the TNBCtype subtyping tool, we can expect the HME50-hTERT cell line to model BL2 TNBC subtype and respond to cisplatin whereas, the HMET cell line show characteristics of TNBC MSL subtype and respond to PI3K/mTOR inhibitors.

Recently, RNA and DNA profiling of about 200 TNBC tumor samples by Burstein et al. [186] defined TNBC subtypes and subtype-specific biomarkers and targets characterized by distinct molecular profiles. This study confirmed four stable TNBC subtypes namely luminal androgen receptor (LAR), mesenchymal (MES), basal-like immunosuppressed (BLIS) and basal-like immune activated (BLIA). These findings substantiate TNBC heterogeneity as well as provide subtype-specific molecular signatures that can be utilized clinically to identify TNBC subtypes. The Burstein study addresses limitations of the Lehmann's TNBC subtyping signature [30] which doesn't readily segregate basal-like 1 and basal-like 2 based on hierarchical clustering and the Burstein analysis unlike TNBCtype tool doesn't rely on inclusion of samples with Estrogen Receptor (ER), Progesterone Receptor (PR) and Human Epidermal growth factor Receptor 2 (HER2) immunohistochemistry data. However, a user-friendly TNBC subtyping tool has not yet been made available by the Burstein group. Both TNBC subtyping studies carried out by Lehmann et al. and Burstein et al. demonstrate molecular heterogeneity in TNBC tumors that can be divided into distinct stromal, immune and basal signaling components. These studies also support the possibility of targeting specific TNBC subtypes using approved drugs such as Androgen Receptor (AR) antagonists for treatment of AR or LAR subtype [30, 185, 186].

The LFS HME50 cell lines that arise from a common background, exhibit unique features (karyotypes, growth characteristics in monolayer, anchorage independent soft agar growth, tumorigenic potential in mouse xenograft) and distinct gene expression profiles facilitating the study of cancer mechanisms. Furthermore, gene expression profiling can reveal the predominantly deregulated signaling pathways and genetic alterations that drive the oncogenic proliferation of the tumor cells. Thus, the analyses of LFS HME50 gene expression data using bioinformatics can improve its utility as a preclinical model of high-risk breast cancer by addressing some aspects of molecular heterogeneity.

Table 18. The TNBCtype tool predicted TNBC subtype to each of the HME50 cell lines with corresponding correlation coefficient and permutation P-value.

Sample	Cell Line		Subtype	Correlation	P-value
BH02002.CEL	HME50	BL2	**Basal-like 2**	0.217	0.028
BH02003.CEL	HME50	BL1	**Basal-like 1**	0.152	<0.001
BH02004.CEL	HME50	BL1		0.118	<0.001
BH02007.CEL	HME50-5E	LAR	**Luminal Androgen Receptor-like**	0.118	0.002
BH02008.CEL	HME50-5E	LAR		0.134	0.001
BH02009.CEL	HME50-hT	BL2	**Basal-like 2**	0.132	0.025
BH02010.CEL	HME50-hT	BL2		0.217	0.002
BH02011.CEL	HME50-hT	BL2		0.122	0.023
BH02012.CEL	HME50-hT	BL2		0.126	0.039
BH02013.CEL	HMET	MSL	**Mesenchymal Stem Cell-like**	0.101	<0.001
BH02015.CEL	HMET	MSL		0.164	<0.001

5.2. Endpoint parameters for future chemopreventive agent testing

Based on the aforementioned differences in characteristics of cells in 3D culture relative to monolayer culture, it will be imperative to analyze gene expression changes and protein localization to ascertain drug response and tumor reversion. Based on the gene expression data and pharmacological agents used, in addition to the direct transcriptional targets of *TP53*, our main endpoints for analysis are the cell adhesion markers E-cadherin, gap junction channel proteins, *DSG3*, *DSC3*; the mesenchymal markers vimentin, N-cadherin, *ZEB1* and *SNAI1*; extracellular matrix modeling proteins metalloproteases (*MMP1*, *MMP2*, *MMP3*); the tumor reversion markers *SIAH1*, *STEAP3*, *PS1* and *TPT1*. A feedback loop between *TP53* and *TPT1* is a crucial switch for tumor reversion in context of wildtype *TP53* function [171], and hence it will be an important endpoint in examining the effect of pharmacological of *TP53* reactivating agents such as kevetrin and APR-246. Moreover, the restoration of apical-basal polarity markers based on their localization in 3D Matrigel® will serve as validation of complete phenotypic reversion that accompanies the change in morphology.

Healthy breast tissue donated by volunteers with no clinical history of breast cancer at Susan G. Komen Tissue Bank at Indiana University Simon Cancer Center is available to researchers. The Herbert laboratory has derived both primary epithelial and stromal cells that can represent a non-malignant state in 3D Matrigel cultures for preclinical studies (**Figure 43**). Use of these primary cells can alleviate our absolute reliance on cell lines derived from fibrocystic breast disease as "normal" equivalents in

166

preclinical studies. In conclusion, the HME50 series model is based on high-risk but clinically healthy tissue sourced from the contralateral breast of LFS patient undergoing surgery and this averted the use of tissue from reduction mammoplasty or fibrocystic breast disease or normal breast epithelium in contrast to previously reported progression models [60-62, 141, 187-189]. It is important to note that in this study, we have chosen the commonly used extracellular matrix for *in vitro* 3D culture of breast epithelial cells is Matrigel® - a commercially available reconstituted basement membrane matrix preparation from extracellular matrix protein-rich mouse tumor. Since Matrigel® preparation is not completely defined (or precisely reproducible) and biological variations present in different batches may introduce skewed or gene expression changes that may not be easily reproducible in highly sensitive and specific transcriptome profiling experiments. For this reason, the gene expression profiling of HME50 cell lines for initial characterization was carried out using cells harvested from monolayer culture and the data was extrapolated for pharmacological treatments in 3D cultures.

RNA-Sequencing of HME50 cell series is a future goal as it will allow further profiling transcriptome and unbiased discovery of novel, rare or low-abundance transcripts and gene level changes occurring through the progression series with high specificity and sensitivity as compared to microarray based approach. The RNA-sequencing approach will allow validation of microarray profiles along with in depth genetic analysis of specific genes of interest that contribute to the distinct phenotypes of HME50 cell lines in 3D cultures during malignant progression. A future goal of this study is to manipulate deregulated genes predicted by gene expression profiling, by

revolutionary genome editing techniques and investigate the phenotypic reversion in 3D culture. Shalem et al. [190] have recently reviewed the use of CRISPR (clustered regularly interspaced short palindromic repeat)-Cas9 (RNA-guided nuclease) system for genome-scale screen that enables transcriptional perturbations that inhibit (CRISPRi) or activate (CRISPRa) gene expression in a functional genomic screen. Cas9 is proficient and programmable tool that can be used for specific targeting of genomic loci using the guide RNA in large-scale functional genetic screening of cultured mammalian cells. The conglomerate of revolutionary next-generation sequencing, CRISPR-Cas9 genome editing, and 3D cultures systems together allow high-throughput scalability with unprecedented experimental control to decipher genotype-phenotype correlation. Recently, normal and tumor 3D organoids derived from patients with colorectal carcinoma were shown to reflect mutational changes as patient tissue source; these 3D organoid cultures were also used in a high throughput drug screen to establish causal links between drug effect and genetic landscape [191]. Given that the patient-derived cell lines such as HME50 cell lines are a renewable resource, hypothesis-driven experiments can be scalable to ascertain the genotype-phenotype correlations in high-throughput genetic screens. With the known genetic information, both forward (phenotype-to-genotype) and hypothesis-driven reverse (genotype-to-phenotype) genetic screens employing RNA-guided CRISPR/Cas9 nuclease system combined with genome-scale guide RNA libraries for unbiased, phenotypic screens can be used. This is an exciting prospect to model diverse *TP53* mutations observed in Li-Fraumeni families as well as sporadic cancers.

This study highlights the utility of patient-derived tissue samples for developing cell lines and cell progression series that can represent cancer predisposition due to unique gene mutations in preclinical studies. Additionally, the analysis of HME50 series microarray data enabled the characterization of the altered gene expression patterns and deregulated cellular processes that reflect malignant progression *in vitro*. We expect the use of this system facilitates mechanistic studies that provide clues to inhibit or reverse malignant progression.

Figure 42. Cytoskeletal actin staining of IUSM Li-Fraumeni patient-derived epithelial cell line. The cytoskeletal actin staining of human mammary epithelial cell line derived from IUSM Li-Fraumeni Syndrome patient harboring a truncating *TP53* germline mutation. After the acini underwent morphogenesis for 10 days embedded in 3D Matrigel culture, actin staining and confocal imaging was performed. The representative centermost z-stack of the acini stained with AF594 conjugated Phalloidin and Hoechst staining showed radially organized IUSM LFS patient-derived epithelial cells surrounding a clear lumen and acinar size (51µm) similar to MCF10A [109] and HME50-hTERT non-malignant controls. Scalebar 10µm.

AF 594 Phalloidin||DAPI
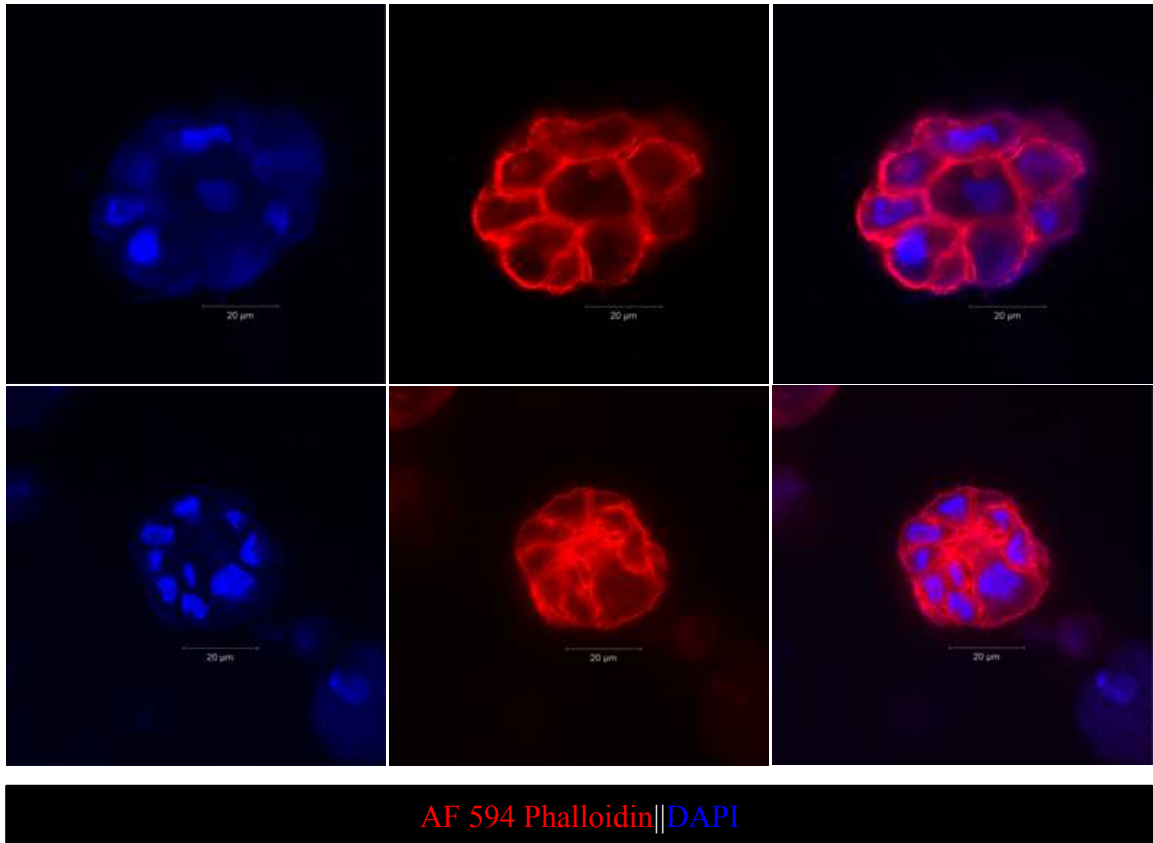
Figure 43. Cytoskeletal actin staining of primary human mammary epithelial acini.

The cytoskeletal actin staining of K-HME 496 acini grown for 10 days embedded in 3D

Matrigel culture was performed followed by confocal imaging. KTB496 breast epithelial

cells were derived from breast tissue of a healthy (no breast disease or cancer history)

female volunteer donor at the Susan G. Komen for the Cure® Tissue Bank (KTB) at the

IU Simon Cancer Center [192]. The centermost z-stack of K-HME 496 acini stained with

AF594 conjugated Phalloidin and Hoechst staining showed radially organized HME496

epithelial cells surrounding a clear lumen and acinar size (51µm) similar to MCF10A

[109] and HME50-hTERT non-malignant controls. Scalebar 20µm.

171

Patients suspected of inherited cancer predisposition upon consultation and recommendation of clinician and clinical geneticists can undergo genetic testing. Based on individual case, the gene expression profiling and next-generation sequencing methods (microarray, gene panel, whole-exome, whole-genome sequencing) can be employed either independently or in combination that can help identify and validate the genetic variants and followed by traditional Sanger sequencing and complex segregation analysis. Briefly, diagnostic analysis of blood samples, clinical histopathological analysis of normal, preinvasive and tumor tissue samples along with *in vivo* imaging data (mammograms, MRI, X-rays) can provide phenotypic classification; whereas, genotypic and molecular stratification can be carried out by genomics (next-gen sequencing, genetic panels, copy number and epigenetic classification), transcriptomics (mRNA and miRNA signatures and pathway analysis), proteomics (e.g. immunohistochemistry, liquid chromatography and mass spectroscopy), and metabolomics. The combined systems biology and next-generation sequencing approach used in tandem with diagnostic analysis of biopsied sample to establish tumor pathophysiology and disease phenotype can provide morphological, histological and molecular data useful for patient stratification. Biobanking of patient-derived normal and tumor samples will provide a renewable of phenotypically stable and biologically diverse tissue resource. Along with biobanking and development of patient-derived xenograft (PDX) mouse models, the biopsied material can also digested for expansion as primary cells and organoids, immortalized cell lines and 3D cultures to enable functional assays and screening tools to test drug efficacy and toxicity in preclinical trials as well as personalized medicine. Following this integrated genotypic and phenotypic stratification approach, predictions

on patient clinical outcome, response and resistance signatures, drug toxicities and personalized therapeutic strategies can be determined. A brief summary of steps involved in the characterization of the established HME50 progression series is highlighted in red along with future directions depicted in **Figure 44**. The pedigree analysis, polymerase chain reaction-single strand conformation polymorphism (PCR-SCCP) analysis [110], and Sanger sequencing was performed that ascertained inherited germline TP53 mutations in the proband – a 31 year old woman with undergoing surgery for breast cancer (unpublished data). The normal predisposed breast tissue obtained from this patient during the prophylactic surgery was digested and primary HME50 cells were cultured. Next, primary HME50 cells were immortalized and progressed to a malignant stage through sequential genetic manipulations followed by gene expression profiling and characterizing the phenotypic changes observed in 3D Matrigel® cultures. To use the HME50 cell lines as a screening tool, we further show proof-of-concept experiment to link the mutant p53 context with the use of PRIMA-1 and APR-246, which are known to reactivate mutant p53. Next, the characterized changes in morphology and acini size was used as a phenotypic readout to study the action of agents such as EGCG, Resveratrol and rapamycin in a gene-expression guided approach. In conclusion, we propose the use of patient-derived cells from genetically predisposed individuals as renewable cell lines, perform gene expression profiling to identify targets and signaling pathways that can be targeted to prevent or revert the cancer progression by novel chemopreventive agents. Genetic testing along with patient-derived preclinical models can facilitate surveillance, medical management decisions, prophylactic surgeries as well as opportunities for chemoprevention.
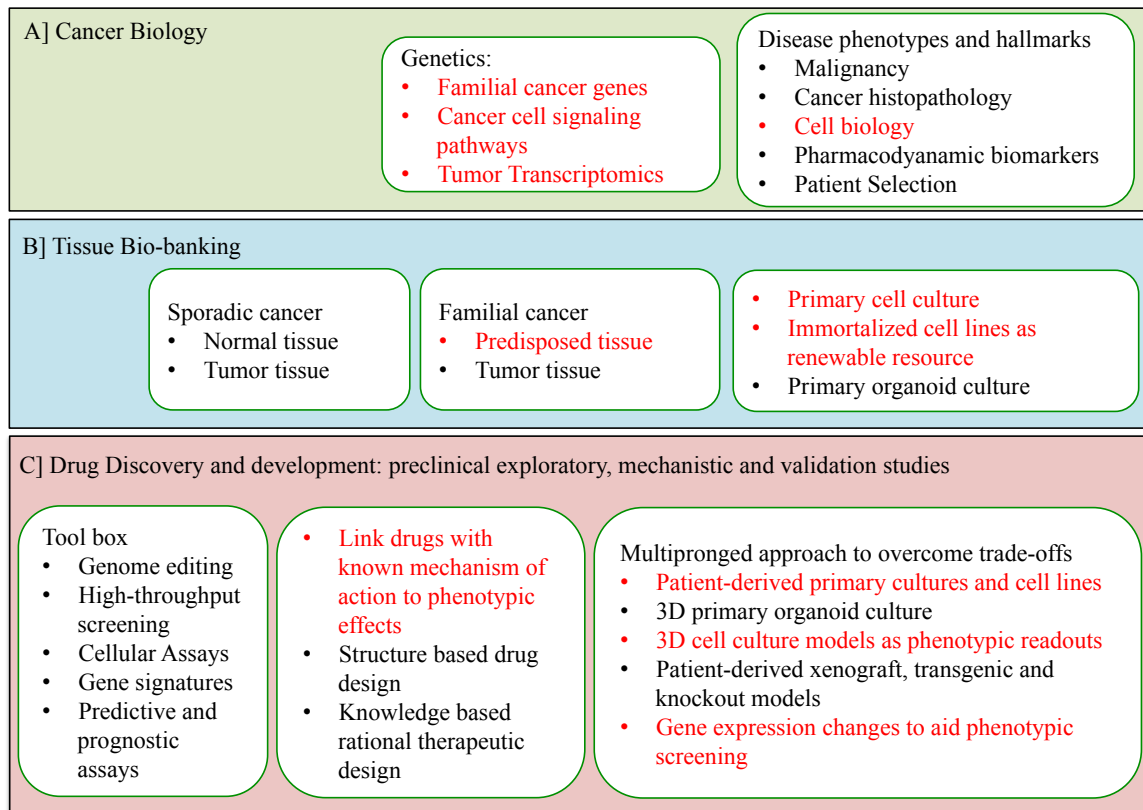
Figure 44. Summary of current research and proposed future objectives to integrate cancer genetics, tissue biobanking and phenotypic screening in Herbert Laboratory.

Preclinical models of cancer are necessary to understand cancer biology, mimic cancer phenotypes, drug discovery and translate fundamental research to the clinic. Multipronged approach with an integrated approach of systems biology is necessary to overcome caveats associated with each breast cancer model in addition.

Appendix 1: QC for microarray

Please see attached supplemental file for supplementary figures and supporting

information on QC metrics generated using Partek® Genomics Suite software.

Appendix 2: Lists of differentially expressed genes in HME50 cell lines

Please see attached supplemental file for lists of differentially expressed genes as

identified in this dissertation.

Appendix 3: Comprehensive epithelial-to-mesenchymal transition gene list used

Please see attached supplemental file for supporting information on epithelial and

mesenchymal gene list used.

REFERENCES

1.   Nagy R, Sweet K, Eng C: **Highly penetrant hereditary cancer syndromes**. *Oncogene* 2004, **23**(38):6445-6470.
2.   Li F, Fraumeni JF, Jr.: **Soft-tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome?** *Ann Intern Med* 1969, **71**:747-752.
3.   Li FP, Fraumeni JF, Jr.: **Rhabdomyosarcoma in children: epidemiologic study and identification of a familial cancer syndrome**. *J Natl Cancer Inst* 1969, **43**(6):1365-1373.
4.   Nichols KE, Malkin D, Garber JE, Fraumeni JF, Jr., Li FP: **Germ-line p53 mutations predispose to a wide spectrum of early-onset cancers**. *Cancer Epidemiol Biomarkers Prev* 2001, **10**(2):83-87.
5.   Nichols KE, Malkin D: **Genotype Versus Phenotype: The Yin and Yang of Germline TP53 Mutations in Li-Fraumeni Syndrome**. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2015, **33**(21):2331-2333.
6.   Malkin D, Li FP, Strong LC, Fraumeni JF, Jr., Nelson CE, Kim DH, Kassel J, Gryka MA, Bischoff FZ, Tainsky MA *et al*: **Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms**. *Science* 1990, **250**(4985):1233-1238.
7.   Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA *et al*: **Mutational landscape and significance across 12 major cancer types**. *Nature* 2013, **502**(7471):333-339.
8.   Leroy B, Fournier JL, Ishioka C, Monti P, Inga A, Fronza G, Soussi T: **The TP53 website: an integrative resource centre for the TP53 mutation database and TP53 mutant analysis**. *Nucleic acids research* 2013, **41**(Database issue):D962-969.
9.   Powell E, Piwnica-Worms D, Piwnica-Worms H: **Contribution of p53 to metastasis**. *Cancer discovery* 2014, **4**(4):405-414.
10.   Solomon H, Madar S, Rotter V: **Mutant p53 gain of function is interwoven into the hallmarks of cancer**. *Journal of Pathology* 2011, **225**(4):475-478.
11.   Borresen-Dale AL: **TP53 and breast cancer**. *Human mutation* 2003, **21**(3):292-300.
12.   Lalloo F, Varley J, Ellis D, Moran A, O'Dair L, Pharoah P, Evans DGR: **Prediction of pathogenic mutations in patients with early-onset breast cancer by family history**. *The Lancet* 2003, **361**(9363):1101-1102.
13.   Olivier M, Goldgar DE, Sodha N, Ohgaki H, Kleihues P, Hainaut P, Eeles RE: **Li-Fraumeni and Related Syndromes: Correlation between Tumor Type, Family Structure, and TP53 Genotype**. *Cancer research* 2003, **63**:6643–6650.
14.   Kuperwasser C, Hurlbut GD, Kittrell FS, Dickinson ES, Lauciríca R, Medina D, Naber SP, Jerry DJ: **Development of Spontaneous Mammary Tumors in BALB/c p53 Heterozygous Mice**. *The American journal of pathology* 2000, **157**(6):2151-2159.
15.   Olive KP, Tuveson DA, Ruhe ZC, Yin B, Willis NA, Bronson RT, Crowley D, Jacks T: **Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome**. *Cell* 2004, **119**(6):847-860.

16. Dittmer D, Pati S, Zambetti G, Chu S, Teresky AK, Moore M, Finlay C, Levine AJ: **Gain of function mutations in p53**. *Nat Genet* 1993, **4**(1):42-46.

17. Oren M, Rotter V: **Mutant p53 gain-of-function in cancer**. *Cold Spring Harbor perspectives in biology* 2010, **2**(2):a001107.

18. Muller PA, Vousden KH: **Mutant p53 in cancer: new functions and therapeutic opportunities**. *Cancer cell* 2014, **25**(3):304-317.

19. Lee DF, Su J, Kim HS, Chang B, Papatsenko D, Zhao R, Yuan Y, Gingold J, Xia W, Darr H *et al*: **Modeling familial cancer with induced pluripotent stem cells**. *Cell* 2015, **161**(2):240-254.

20. Bougeard G, Renaux-Petel M, Flaman JM, Charbonnier C, Fermey P, Belotti M, Gauthier-Villars M, Stoppa-Lyonnet D, Consolino E, Brugieres L *et al*: **Revisiting Li-Fraumeni Syndrome From TP53 Mutation Carriers**. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2015, **33**(21):2345-2352.

21. Durno CA, Aronson M, Tabori U, Malkin D, Gallinger S, Chan HS: **Oncologic surveillance for subjects with biallelic mismatch repair gene mutations: 10 year follow-up of a kindred**. *Pediatric blood & cancer* 2012, **59**(4):652-656.

22. Finch AP, Lubinski J, Moller P, Singer CF, Karlan B, Senter L, Rosen B, Maehle L, Ghadirian P, Cybulski C *et al*: **Impact of oophorectomy on cancer incidence and mortality in women with a BRCA1 or BRCA2 mutation**. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2014, **32**(15):1547-1553.

23. **NCCN clinical practice guidelines in oncology: Genetic/familial high-risk assessment—breast and ovarian, Version 1.2014** [http://www.nccn.org]

24. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS *et al*: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications**. *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**(19):10869-10874.

25. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S *et al*: **Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors**. *Genome biology* 2007, **8**(5):R76.

26. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA *et al*: **Molecular portraits of human breast tumours**. *Nature* 2000, **406**(6797):747-752.

27. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z *et al*: **Supervised risk predictor of breast cancer based on intrinsic subtypes**. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2009, **27**(8):1160-1167.

28. Prat A, Parker JS, Fan C, Perou CM: **PAM50 assay and the three-gene model for identifying the major and clinically relevant molecular subtypes of breast cancer**. *Breast cancer research and treatment* 2012, **135**(1):301-306.

29. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, He X, Perou CM: **Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer**. *Breast cancer research : BCR* 2010, **12**(5):R68.

30.   Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, Pietenpol JA: **Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies**. *The Journal of clinical investigation* 2011, **121**(7):2750-2767.

31.   Ellis MJ, Perou CM: **The genomic landscape of breast cancer as a therapeutic roadmap**. *Cancer discovery* 2013, **3**(1):27-34.

32.   van de Vijver M, He YD, van 't Veer L, Dai H, Hart AH, Voskuil DW, Schreiber G, Peterse JL, Roberts C, Marton MJ *et al*: **A gene-expression signature as a predictor of survival in breast cancer**. *N Engl J Med* 2002, **347**(25):1999-2009.

33.   Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner F, Walker M, Watson D, Park T *et al*: **A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer**. *The New England journal of medicine* 2004, **351**(27):2817-2826.

34.   Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DSA, Nobel AB, Veer LJvt, Charles M. Perou PD: **Concordance among Gene-Expression–Based Predictors for Breast Cancer**. *The New England Journal of Medicine* 2006.

35.   Perou CM, Borresen-Dale AL: **Systems biology and genomics of breast cancer**. *Cold Spring Harbor perspectives in biology* 2011, **3**(2).

36.   Weigelt B, Baehner FL, Reis-Filho JS: **The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective of the last decade**. *The Journal of pathology* 2010, **220**(2):263-280.

37.   Chia SK, Bramwell VH, Tu D, Shepherd LE, Jiang S, Vickery T, Mardis E, Leung S, Ung K, Pritchard KI *et al*: **A 50-gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen**. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2012, **18**(16):4465-4472.

38.   Perreard L, Fan C, Quackenbush JF, Mullins M, Gauthier NP, Nelson E, Mone M, Hansen H, Buys SS, Rasmussen K *et al*: **Classification and risk stratification of invasive breast carcinomas using a real-time quantitative RT-PCR assay**. *Breast cancer research : BCR* 2006, **8**(2):R23.

39.   Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, Nordgren H, Farmer P, Praz V, Haibe-Kains B *et al*: **Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis**. *J Natl Cancer Inst* 2006, **98**(4):262-272.

40.   Loi S, Piccart M, Sotiriou C: **The use of gene-expression profiling to better understand the clinical heterogeneity of estrogen receptor positive breast cancers and tamoxifen response**. *Crit Rev Oncol Hematol* 2007, **61**(3):187-194.

41.   Hunter K: **Host genetics influence tumour metastasis**. *Nature reviews Cancer* 2006, **6**(2):141-146.

42.   Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, Cronin M, Baehner F, Watson D, Bryant J *et al*: **Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer**. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2006, **24**(23):3726-3734.

43. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT *et al*: **Gene expression profiling predicts clinical outcome of breast cancer**. *Nature* 2002, **415**(6871):530-536.

44. Cancer Genome Atlas N: **Comprehensive molecular portraits of human breast tumours**. *Nature* 2012, **490**(7418):61-70.

45. Drukker CA, Bueno-de-Mesquita JM, Retel VP, van Harten WH, van Tinteren H, Wesseling J, Roumen RM, Knauer M, van 't Veer LJ, Sonke GS *et al*: **A prospective evaluation of a breast cancer prognosis signature in the observational RASTER study**. *International journal of cancer Journal international du cancer* 2013, **133**(4):929-936.

46. de Snoo F, Bender R, Glas A, Rutgers E: **Gene expression profiling: decoding breast cancer**. *Surgical oncology* 2009, **18**(4):366-378.

47. Kopelovich L, Herbert B-S: **Heritable one-hit events defining cancer prevention?** *Cell Cycle* 2014, **12**(16):2553-2557.

48. Herbert BS, Chanoux RA, Liu Y, Baenziger PH, Goswami CP, McClintick JN, Edenberg HJ, Pennington RE, Lipkin SM, Kopelovich L: **A molecular signature of normal breast epithelial and stromal cells from Li-Fraumeni syndrome mutation carriers**. *Oncotarget* 2010, **1**(6):405-422.

49. Shoemaker RH: **The NCI60 human tumor cell line anticancer drug screen**. *Nature Reviews Cancer* 2006, **6**:813-823.

50. Shoemaker RH, Wolpert-DeFilippes MK, Kern DH, Lieber MM, Makuch RW, Melnick NR, Miller WT, Salmon SE, Simon RM, Venditti JM *et al*: **Application of a human tumor colony-forming assay to new drug screening**. *Cancer research* 1985, **45**(5):2145-2153.

51. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP, Tong F *et al*: **A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes**. *Cancer cell* 2006, **10**(6):515-527.

52. Sos ML, Michel K, Zander T, Weiss J, Frommolt P, Peifer M, Li D, Ullrich R, Koker M, Fischer F *et al*: **Predicting drug susceptibility of non-small cell lung cancers based on genetic lesions**. *The Journal of clinical investigation* 2009, **119**(6):1727-1740.

53. Macconaill LE, Garraway LA: **Clinical implications of the cancer genome**. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2010, **28**(35):5219-5228.

54. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehar J, Kryukov GV, Sonkin D *et al*: **The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity**. *Nature* 2012, **483**(7391):603-607.

55. Lin WM, Baker AC, Beroukhim R, Winckler W, Feng W, Marmion JM, Laine E, Greulich H, Tseng H, Gates C *et al*: **Modeling genomic diversity and tumor dependency in malignant melanoma**. *Cancer research* 2008, **68**(3):664-673.

56. Dry JR, Pavey S, Pratilas CA, Harbron C, Runswick S, Hodgson D, Chresta C, McCormack R, Byrne N, Cockerill M *et al*: **Transcriptional pathway signatures predict MEK addiction and response to selumetinib (AZD6244)**. *Cancer research* 2010, **70**(6):2264-2273.

57.   Chin L, Hahn WC, Getz G, Meyerson M: **Making sense of cancer genomic data**. *Genes & development* 2011, **25**(6):534-555.

58.   Chin L, Andersen JN, Futreal PA: **Cancer genomics: from discovery science to personalized medicine**. *Nature medicine* 2011, **17**(3):297-303.

59.   Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, Greninger P, Thompson IR, Luo X, Soares J *et al*: **Systematic identification of genomic markers of drug sensitivity in cancer cells**. *Nature* 2012, **483**(7391):570-575.

60.   Dawson PJ, Wolman SR, Tait L, Heppner GH, Miller FR: **MCF10AT: a model for the evolution of cancer from proliferative breast disease**. *The American journal of pathology* 1996, **148**(1):313-319.

61.   Santner SJ, Dawson PJ, Tait L, Soule HD, Eliason J, Mohamed AN, Wolman SR, Heppner GH, Miller FR: **Malignant MCF10CA1 cell lines derived from premalignant human breast epithelial MCF10AT cells**. *Breast cancer research and treatment* 2001, **65**(2):101-110.

62.   Briand P, Petersen OW, Van Deurs B: **A new diploid nontumorigenic human breast epithelial cell line isolated and propagated in chemically defined medium**. *In Vitro Cell Dev Biol* 1987, **23**(3):181-188.

63.   Souter LH, Andrews JD, Zhang G, Cook AC, Postenka CO, Al-Katib W, Leong HS, Rodenhiser DI, Chambers AF, Tuck AB: **Human 21T breast epithelial cell lines mimic breast cancer progression in vivo and in vitro and show stage-specific gene expression patterns**. *Laboratory investigation; a journal of technical methods and pathology* 2010, **90**(8):1247-1258.

64.   Clarke R: **The role of preclinical animal models in breast cancer drug development**. *Breast cancer research : BCR* 2009, **11 Suppl 3**:S22.

65.   Landis M, Lehmann B, Pietenpol J, Chang J: **Patient-derived breast tumor xenografts facilitating personalized cancer therapy**. *Breast cancer research : BCR* 2013, **15**(1):201.

66.   Sharma SV, Haber DA, Settleman J: **Cell line-based platforms to evaluate the therapeutic efficacy of candidate anticancer agents**. *Nature reviews Cancer* 2010, **10**(4):241-253.

67.   Yamada KM, Cukierman E: **Modeling tissue morphogenesis and cancer in 3D**. *Cell* 2007, **130**(4):601-610.

68.   Petersen O, Rønnov-Jessen L, Howlett A, Bissell M: **Interaction with basement membrane serves to rapidly distinguish growth and differentiation pattern of normal and malignant human breast epithelial cells**. *Proceedings of the National Academy of Sciences of the United States of America* 1992, **89**(19):9064-9068.

69.   Bissell MJ, Radisky D: **Putting tumours in context**. *Nature reviews Cancer* 2001, **1**(1):46-54.

70.   Bissell MJ, Radisky DC, Rizki A, Weaver VM, Petersen OW: **The organizing principle: microenvironmental influences in the normal and malignant breast**. *Differentiation* 2002, **70**(9-10):537-546.

71.   Weaver VM, Petersen OW, Wang F, Larabell CA, Briand P, Damsky C, Bissell MJ: **Reversion of the malignant phenotype of human breast cells in three-**

dimensional culture and in vivo by integrin blocking antibodies**. *The Journal of cell biology* 1997, **137**(1):231-245.

72. Wang F, Weaver VM, Petersen OW, Larabell CA, Dedhar S, Briand P, Lupu R, Bissell M: **Reciprocal interactions between b1-integrin and epidermal growth factor receptor in three-dimensional basement membrane breast cultures: A different perspective in epithelial biology**. *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**:14821-14826.

73. Abbott A: **Cell culture: biology's new dimension**. *Nature* 2003, **424**(6951):870-872.

74. Gurski LA, Jha AK, Zhang C, Jia X, Farach-Carson MC: **Hyaluronic acid-based hydrogels as 3D matrices for in vitro evaluation of chemotherapeutic drugs using poorly adherent prostate cancer cells**. *Biomaterials* 2009, **30**(30):6076-6085.

75. Gorlach A, Herter P, Hentschel H, Frosch PJ, Acker H: **Effects of nIFN Beta and rIFN gamma on growth and morphology of two human melanoma cell lines: comparison between two- and three- dimensional culture**. 1994.

76. Chignola R, Schenetti A, Andrighetto G, Chiesa E, Foroni R, Sartoris S, Tridente G, Liberati D: **Forecasting the growth of multicell tumour spheroids: implications for the dynamic growth of solid tumours**. *Cell Prolif* 2000, **33**(4):219-229.

77. Santini MT, Rainaldi G, Romano R, Ferrante A, Clemente S, Motta A, Indovina PL: **MG-63 human osteosarcoma cells grown in monolayer and as three-dimensional tumor spheroids present a different metabolic profile: a 1H NMR study**. *FEBS letters* 2004, **557**(1-3):148-154.

78. Yamazaki D, Kurisu S, Takenawa T: **Involvement of Rac and Rho signaling in cancer cell motility in 3D substrates**. *Oncogene* 2009, **28**(13):1570-1583.

79. Gurski LA, Petrelli NJ, Jia X, Farach-Carson MC: **3D Matrices for Anti-Cancer Drug Testing and Development**. *Oncology Issues* 2010.

80. Fischbach C, Chen R, Matsumoto T, Schmelzle T, Brugge JS, Polverini PJ, Mooney DJ: **Engineering tumors with 3D scaffolds**. *Nature methods* 2007, **4**(10):855-860.

81. Smalley KS, Lioni M, Herlyn M: **Life isn't flat: taking cancer biology to the next dimension**. *In Vitro Cell Dev Biol Anim* 2006, **42**(8-9):242-247.

82. Thiery JP, Acloque H, Huang RY, Nieto MA: **Epithelial-mesenchymal transitions in development and disease**. *Cell* 2009, **139**(5):871-890.

83. Lamouille S, Xu J, Derynck R: **Molecular mechanisms of epithelial-mesenchymal transition**. *Nature reviews Molecular cell biology* 2014, **15**(3):178-196.

84. Gonzalez D, Damian M: **Signaling mechanisms of the epithelial-mesenchymal trasitiion**. *Science* 2014, **7**.

85. Puisieux A, Brabletz T, Caramel J: **Oncogenic roles of EMT-inducing transcription factors**. *Nature cell biology* 2014, **16**(6):488-494.

86. Craene BD, Berx G: **Regulatory networks defining EMT during cancer initiation and progression**. *Nature reviews Cancer* 2013.

87. Tsai JH, Yang J: **Epithelial-mesenchymal plasticity in carcinoma metastasis**. *Genes & development* 2013, **27**(20):2192-2206.

88. Langerod A, Zhao H, Borgan O, Nesland JM, Bukholm IR, Ikdahl T, Karesen R, Borresen-Dale AL, Jeffrey SS: **TP53 mutation status and gene expression profiles are powerful prognostic markers of breast cancer**. *Breast cancer research : BCR* 2007, **9**(3):R30.

89. Chang CJ, Chao CH, Xia W, Yang JY, Xiong Y, Li CW, Yu WH, Rehman SK, Hsu JL, Lee HH *et al*: **p53 regulates epithelial-mesenchymal transition and stem cell properties through modulating miRNAs**. *Nature cell biology* 2011, **13**(3):317-323.

90. Brosh R, Rotter V: **When mutants gain new powers: news from the mutant p53 field**. *Nature reviews Cancer* 2009, **9**(10):701-713.

91. Molchadsky A, Rivlin N, Brosh R, Rotter V, Sarig R: **p53 is balancing development, differentiation and de-differentiation to assure cancer prevention**. *Carcinogenesis* 2010, **31**(9):1501-1508.

92. Rivlin N, Brosh R, Oren M, Rotter V: **Mutations in the p53 Tumor Suppressor Gene: Important Milestones at the Various Steps of Tumorigenesis**. *Genes & cancer* 2011, **2**(4):466-474.

93. Muller PA, Vousden KH, Norman JC: **p53 and its mutants in tumor cell migration and invasion**. *The Journal of cell biology* 2011, **192**(2):209-218.

94. Yu X, Vazquez A, Levine AJ, Carpizo DR: **Allele-specific p53 mutant reactivation**. *Cancer cell* 2012, **21**(5):614-625.

95. Selivanova G: **Therapeutic targeting of p53 by small molecules**. *Seminars in cancer biology* 2010, **20**(1):46-56.

96. Zandi R, Selivanova G, Christensen CL, Gerds TA, Willumsen BM, Poulsen HS: **PRIMA-1Met/APR-246 induces apoptosis and tumor growth delay in small cell lung cancer expressing mutant p53**. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2011, **17**(9):2830-2841.

97. Selivanova G, Wiman KG: **Reactivation of mutant p53: molecular mechanisms and therapeutic potential**. *Oncogene* 2007, **26**(15):2243-2254.

98. Hait WN: **Anticancer drug development: the grand challenges**. *Nature reviews Drug discovery* 2010, **9**(4):253-254.

99. Voskoglou-Nomikos T, Pater JL, Seymour L: **Clinical predictive value of the in vitro cell line, human xenograft, and mouse allograft preclinical cancer models**. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2003, **9**(11):4227-4239.

100. Boven E, Winograd B, Berger DP, Dumont MP, Braakhuis BJ, Fodstad O, Langdon S, Fiebig HH: **Phase II preclinical drug screening in human tumor xenografts: a first European multicenter collaborative study**. *Cancer research* 1992, **52**(21):5940-5947.

101. Fiebig HH, Maier A, Burger AM: **Clonogenic assay with established human tumour xenografts: correlation of in vitro to in vivo activity as a basis for anticancer drug discovery**. *Eur J Cancer* 2004, **40**(6):802-820.

102. Shamir ER, Ewald AJ: **Three-dimensional organotypic culture: experimental models of mammalian biology and disease**. *Nature reviews Molecular cell biology* 2014, **15**(10):647-664.

103. Lama R, Zhang L, Naim JM, Williams J, Zhou A, Su B: **Development, validation and pilot screening of an in vitro multi-cellular three-dimensional**

cancer spheroid assay for anti-cancer drug testing. *Bioorganic & medicinal chemistry* 2013, **21**(4):922-931.

104. Justice BA, Badr NA, Felder RA: **3D cell culture opens new dimensions in cell-based assays**. *Drug discovery today* 2009, **14**(1-2):102-107.

105. Kenny HA, Lal-Nag M, White EA, Shen M, Chiang CY, Mitra AK, Zhang Y, Curtis M, Schryver EM, Bettis S *et al*: **Quantitative high throughput screening using a primary human three-dimensional organotypic culture predicts in vivo efficacy**. *Nature communications* 2015, **6**:6220.

106. Edmondson R, Broglie JJ, Adcock AF, Yang L: **Three-dimensional cell culture systems and their applications in drug discovery and cell-based biosensors**. *Assay and drug development technologies* 2014, **12**(4):207-218.

107. Suggitt M, Bibby MC: **50 years of preclinical anticancer drug screening: empirical to target-driven approaches**. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2005, **11**(3):971-981.

108. Lee GY, Kenny PA, Lee EH, Bissell MJ: **Three-dimensional culture models of normal and malignant breast epithelial cells**. *Nature methods* 2007, **4**(4):359-365.

109. Debnath J, Muthuswamy SK, Brugge JS: **Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures**. *Methods* 2003, **30**(3):256-268.

110. Shay JW, Tomlinson G, Piatyszek MA, Gollahon LS: **Spontaneous in vitro immortalization of breast epithelial cells from a patient with Li-Fraumeni syndrome**. *Mol Cell Biol* 1995, **15**(1):425-432.

111. Morales CP, Holt SE, Ouellette M, Kaur KJ, Yan Y, Wilson KS, White MA, Wright WE, Shay JW: **Absence of cancer-associated changes in human fibroblasts immortalized with telomerase**. *Nat Genet* 1999, **21**(1):115-118.

112. Gollahon LS, Shay JW: **Immortalization of human mammary epithelial cells transfected with mutant p53 (273his)**. *Oncogene* 1996, **12**(4):715-725.

113. Yi X, Tesmter VM, Savre-Train I, Shay JW, Wright WE: **Both Transcriptional and Posttranscriptional Mechanisms Regulate Human Telomerase Template RNA Levels**. *MOLECULAR AND CELLULAR BIOLOGY* 1999, **19**(6):3989-3997.

114. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing**. *Journal of the Royal Statistical Society* 1995, **57**(1):289-300.

115. Benjamini Y, Yekutieli D: **The control of the false discovery rate in multiple testing under dependency**. *The Annals of Statistics* 2001, **29**(4):1165-1188.

116. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *et al*: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles**. *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**(43):15545-15550.

117. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP: **Molecular signatures database (MSigDB) 3.0**. *Bioinformatics* 2011, **27**(12):1739-1740.

118. Charafe-Jauffret E, Ginestier C, Monville F, Finetti P, Adelaide J, Cervera N, Fekairi S, Xerri L, Jacquemier J, Birnbaum D *et al*: **Gene expression profiling of breast cell lines identifies potential new basal markers**. *Oncogene* 2006, **25**(15):2273-2284.

119. Kobayashi S, Shimamura T, Monti S, Steidl U, Hetherington CJ, Lowell AM, Golub T, Meyerson M, Tenen DG, Shapiro GI *et al*: **Transcriptional profiling identifies cyclin D1 as a critical downstream effector of mutant epidermal growth factor receptor signaling**. *Cancer research* 2006, **66**(23):11389-11398.

120. Zhao M, Kong L, Liu Y, Qu H: **dbEMT: an epithelial-mesenchymal transition associated gene resource**. *Scientific reports* 2015, **5**:11459.

121. Taubea JH, Herschkowitz JI, Komurov K, Zhou AY, Gupta S, Yang J, Hartwell K, Onder TT, Gupta PB, Evans KW *et al*: **Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudinlow and metaplastic breast cancer subtypes**. *PNAS* 2010.

122. Horn HF, Vousden KH: **Coping with stress: multiple ways to activate p53**. *Oncogene* 2007, **26**(9):1306-1316.

123. Vousden KH, Lu X: **Live or let die: the cell's response to p53**. *Nature reviews Cancer* 2002, **2**(8):594-604.

124. Lehmann S, Bykov VJ, Ali D, Andren O, Cherif H, Tidefelt U, Uggla B, Yachnin J, Juliusson G, Moshfegh A *et al*: **Targeting p53 in vivo: a first-in-human study with p53-targeting compound APR-246 in refractory hematologic malignancies and prostate cancer**. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2012, **30**(29):3633-3639.

125. Duffy SW, Agbaje O, Tabar L, Vitak B, Bjurstam N, Bjorneld L, Myles JP, Warwick J: **Overdiagnosis and overtreatment of breast cancer: estimates of overdiagnosis from two trials of mammographic screening for breast cancer**. *Breast cancer research : BCR* 2005, **7**(6):258-265.

126. El-Sayed ME, Rakha EA, Reed J, Lee AH, Evans AJ, Ellis IO: **Predictive value of needle core biopsy diagnoses of lesions of uncertain malignant potential (B3) in abnormalities detected by mammographic screening**. *Histopathology* 2008, **53**(6):650-657.

127. Provenzano E, Pinder SE: **Pre-operative diagnosis of breast cancer in screening: problems and pitfalls**. *Pathology* 2009, **41**(1):3-17.

128. Polyak K: **Is breast tumor progression really linear?** *Clinical cancer research : an official journal of the American Association for Cancer Research* 2008, **14**(2):339-341.

129. Costa A, Zanini V: **Precancerous lesions of the breast**. *Nature clinical practice Oncology* 2008, **5**(12):700-704.

130. Allred DC, Wu Y, Mao S, Nagtegaal ID, Lee S, Perou CM, Mohsin SK, O'Connell P, Tsimelzon A, Medina D: **Ductal carcinoma in situ and the emergence of diversity during breast cancer evolution**. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2008, **14**(2):370-378.

131. Clark AS, Domchek SM: **Clinical management of hereditary breast cancer syndromes**. *Journal of mammary gland biology and neoplasia* 2011, **16**(1):17-25.

132. Ripperger T, Gadzicki D, Meindl A, Schlegelberger B: **Breast cancer susceptibility: current knowledge and implications for genetic counselling**. *European journal of human genetics : EJHG* 2009, **17**(6):722-731.

133. Marino-Enriquez A, Fletcher CDM: **Shouldn't we care about the biology of benign tumours?** *Nature Reviews Cancer* 2014, **14**(11):701-702.

134. WC H, CM C, AS L, RL B, MW B, RA W: **Creation of human tumour cells with defined genetic elements.** *Nature* 1999, **400**(6743):2523-2534.

135. Hanahan D, Weinberg RA: **Hallmarks of cancer: the next generation**. *Cell* 2011, **144**(5):646-674.

136. Herbert BS, Wright WE, Shay JW: **p16(INK4a) inactivation is not required to immortalize human mammary epithelial cells**. *Oncogene* 2002, **21**(51):7897-7900.

137. Toouli CD, Huschtscha LI, Neumann AA, Noble JR, Colgin LM, Hukku B, Reddel RR: **Comparison of human mammary epithelial cells immortalized by simian virus 40 T-Antigen or by the telomerase catalytic subunit**. *Oncogene* 2002, **21**(1):128-139.

138. Gudjonsson T, Villadsen R, Ronnov-Jessen L, Petersen OW: **Immortalization protocols used in cell culture models of human breast morphogenesis**. *Cellular and molecular life sciences : CMLS* 2004, **61**(19-20):2523-2534.

139. Yang G, Rosen DG, Mercado-Uribe I, Colacino JA, Mills GB, Bast RC, Jr., Zhou C, Liu J: **Knockdown of p53 combined with expression of the catalytic subunit of telomerase is sufficient to immortalize primary human ovarian surface epithelial cells**. *Carcinogenesis* 2007, **28**(1):174-182.

140. T K, Foster SA, Koop JI, McDougall JK, Galloway DA, Klingelhutz AJ: **BothRb/p16INK4a inactivation and telomerase activity are required to immortalize human epithelial cells**. *Nature* 1998, **396**.

141. Ulbricht U, Sommer A, Beckmann G, Lutzenberger M, Seidel H, Kreft B, Toschi L: **Isogenic human mammary epithelial cell lines: novel tools for target identification and validation. Comprehensive characterization of an isogenic human mammary epithelial cell model provides evidence for epithelial-mesenchymal transition**. *Breast cancer research and treatment* 2013, **138**(2):437-456.

142. Boehm JS, Hession MT, Bulmer SE, Hahn WC: **Transformation of human and murine fibroblasts without viral oncoproteins**. *Mol Cell Biol* 2005, **25**(15):6464-6474.

143. Ottesen GL, Christensen IJ, Larsen JK, Kerndrup GB, Hansen B, Andersen JA: **DNA aneuploidy in early breast cancer**. *British journal of cancer* 1995, **72**(4):832-839.

144. Shackney SE, Singh SG, Yakulis R, Smith CA, Pollice AA, Petruolo S, Waggoner A, Hartsock RJ: **Aneuploidy in breast cancer: a fluorescence in situ hybridization study**. *Cytometry* 1995, **22**(4):282-291.

145. Yildirim-Assaf S, Coumbos A, Hopfenmuller W, Foss HD, Stein H, Kuhn W: **The prognostic significance of determining DNA content in breast cancer by DNA image cytometry: the role of high grade aneuploidy in node negative breast cancer**. *J Clin Pathol* 2007, **60**(6):649-655.

146. Kendall SD, Linardic CM, Adam SJ, Counter CM: **A network of genetic events sufficient to convert normal human cells to a tumorigenic state**. *Cancer research* 2005, **65**(21):9824-9828.
147. Schinzel AC, Hahn WC: **Oncogenic transformation and experimental models of human cancer**. *Front Biosci* 2008, **13**:71-84.
148. Elenbaas B: **Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells**. *Genes & development* 2001, **15**(1):50-65.
149. Lundberg AS, Randell SH, Stewart SA, Elenbaas B, Hartwell KA, Brooks MW, Fleming MD, Olsen JC, Miller SW, Weinberg RA *et al*: **Immortalization and transformation of primary human airway epithelial cells by gene transfer**. *Oncogene* 2002, **21**(29):4577-4586.
150. **Pantetheinase activity of membrane-bound Vanin-1: lack of free cysteamine in tissues of Vanin-1 deficient mice**. *FEBS letters* 2000, **483**:149-154.
151. Scandinavian Breast Group T, Tanner M, Isola J, Wiklund T, Erikstein B, Kellokumpu-Lehtinen P, Malmstrom P, Wilking N, Nilsson J, Bergh J: **Topoisomerase IIalpha gene amplification predicts favorable treatment response to tailored and dose-escalated anthracycline-based adjuvant chemotherapy in HER-2/neu-amplified breast cancer: Scandinavian Breast Group Trial 9401**. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2006, **24**(16):2428-2436.
152. Nitiss JL: **Targeting DNA topoisomerase II in cancer chemotherapy**. *Nature reviews Cancer* 2009, **9**(5):338-350.
153. Lam EW, Brosens JJ, Gomes AR, Koo CY: **Forkhead box proteins: tuning forks for transcriptional harmony**. *Nature reviews Cancer* 2013, **13**(7):482-495.
154. Bresnick AR, Weber DJ, Zimmer DB: **S100 proteins in cancer**. *Nature reviews Cancer* 2015, **15**(2):96-109.
155. Murray IA, Patterson AD, Perdew GH: **Aryl hydrocarbon receptor ligands in cancer: friend and foe**. *Nature Reviews Cancer* 2014, **14**(12):801-814.
156. Wang F, Hansen RK, Radisky D, Yoneda T, Barcellos-Hoff MH, Petersen OW, Turley EA, Bissell MJ: **Phenotypic reversion or death of cancer cells by altering signaling pathways in three-dimensional contexts**. *J Natl Cancer Inst* 2002, **94**(19):1494-1503.
157. Debnath J, Mills K, Collins N, Reginato M, Muthuswamy S, Brugge J: **The role of apoptosis in creating and maintaining luminal space within normal and oncogene-expressing mammary acini**. *Cell* 2002, **111**(1):29-40.
158. Debnath J, Brugge JS: **Modelling glandular epithelial cancers in three-dimensional cultures**. *Nature reviews Cancer* 2005, **5**(9):675-688.
159. Kenny PA, Lee GY, Myers CA, Neve RM, Semeiks JR, Spellman PT, Lorenz K, Lee EH, Barcellos-Hoff MH, Petersen OW *et al*: **The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression**. *Molecular oncology* 2007, **1**(1):84-96.
160. J TC, V A, B B, Kinzler KW: **Use of isogenic human cancer cells for high-throughput screening and drug discovery**. *Nat Biotechnol* 2001, **10**:940-945.

161. Wilding JL, Bodmer WF: **Cancer cell lines for drug discovery and development**. *Cancer research* 2014, **74**(9):2377-2384.
162. Weinstein JN: **Drug discovery: Cell lines battle cancer**. *Nature* 2012, **483**(7391):544-545.
163. Weaver VM, Lelievre S, Lakins Jn, Chrenek MA, Jones JCR, Giancotti F, Werb Z, Bissell MJ: **B-integrin-dependent formation of polarized threedimensional architecture confers resistance to apoptosis in normal and malignant mammary epithelium**. *Cancer cell* 2002, **2**:205-216.
164. Telerman A, Amson R: **The molecular programme of tumour reversion: the steps beyond malignant transformation**. *Nature reviews Cancer* 2009, **9**(3):206-216.
165. Bissell MJ, Hines WC: **Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression**. *Nature medicine* 2011, **17**(3):320-329.
166. Huettner CS, Zhang P, Van Etten RA, Tenen DG: **Reversibility of acute B-cell leukaemia induced by BCR-ABL1**. *Nat Genet* 2000, **24**(1):57-60.
167. Telerman A, Tuynder M, Dupressoir T, Robaye B, Sigaux F, Shaulian E, Oren M, Rommelaere J, Amson R: **A Model for Tumor Suppression Using H-1 Parvovirus**. *Proceedings of the National Academy of Sciences of the United States of America* 1993, **90**(18):8702-8706.
168. Tuynder M, Susini L, Prieur S, Besse S, Fiucci G, Amson R, Telerman A: **Biological models and genes of tumor reversion: cellular reprogramming through tpt1/TCTP and SIAH-1**. *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**(23):14976-14981.
169. Tuynder M, Fiucci G, Prieur S, Lespagnol A, Geant A, Beaucourt S, Duflaut D, Besse S, Susini L, Cavarelli J *et al*: **Translationally controlled tumor protein is a target of tumor reversion**. *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(43):15364-15369.
170. Roperch JP: **Inhibition of presenilin 1 expression is promoted by p53 and p21WAF-1 and results in apoptosis and tumor suppression.** . *Nature Med* 1998, **4**:835–838.
171. Amson R, Pece S, Lespagnol A, Vyas R, Mazzarol G, Tosoni D, Colaluca I, Viale G, Rodrigues-Ferreira S, Wynendaele J *et al*: **Reciprocal repression between P53 and TCTP**. *Nature medicine* 2012, **18**(1):91-99.
172. Brock A, Krause S, Ingber DE: **Control of cancer formation by intrinsic genetic noise and microenvironmental cues**. *Nature reviews Cancer* 2015, **15**(8):499-509.
173. Buzzai M, Jones RG, Amaravadi RK, Lum JJ, DeBerardinis RJ, Zhao F, Viollet B, Thompson CB: **Systemic treatment with the antidiabetic drug metformin selectively impairs p53-deficient tumor cell growth**. *Cancer research* 2007, **67**(14):6745-6752.
174. Kasznicki J, Sliwinska A, Drzewoski J: **Metformin in cancer prevention and therapy**. *Annals of translational medicine* 2014, **2**(6):57.
175. Evans JMM, Donnelly LA, Emslie-Smith AM, Alessi DR, Morris AD: **Metformin and reduced risk of cancer in diabetic patients**. *Bmj* 2005, **330**.

176. Ben Sahra I, Le Marchand-Brustel Y, Tanti JF, Bost F: **Metformin in cancer therapy: a new perspective for an old antidiabetic drug?** *Molecular cancer therapeutics* 2010, **9**(5):1092-1099.

177. Libby G, Donnelly LA, Donnan PT, Alessi DR, Morris AD, Evans JM: **New users of metformin are at low risk of incident cancer: a cohort study among people with type 2 diabetes**. *Diabetes Care* 2009, **32**(9):1620-1625.

178. Chong CR, Chabner BA: **Mysterious metformin**. *The oncologist* 2009, **14**(12):1178-1181.

179. Kunnumakkara AB, Memmott RM, Maier CR, Kawabata S, Quinn B, Gills J, Dallos M, Hollander CM, Dennis PA: **Efficacy of metformin in two mouse models of Li-Fraumeni syndrome**. In: *AACR 102nd Annual Meeting 2011 2011; Orlando, FL*: Proceedings of the 102nd Annual Meeting of AACR; 2011.

180. Walcott FL, Hwang PM, Wang P-y, Savage SA, Mai P, Steinberg SM, Pollak MN, Dennis PA, Fojo AT: **Design of a phase I chemoprevention study of metformin and Li-Fraumeni syndrome (LFS)**. In: *AACR Special Conference: Cancer Susceptibility and Cancer Susceptibility Syndromes: 2014; San Diego, CA*; 2014.

181. Morales DR, Morris AD: **Metformin in cancer treatment and prevention**. *Annual review of medicine* 2015, **66**:17-29.

182. Turner N, Moretti E, Siclari O, Migliaccio I, Santarpia L, D'Incalci M, Piccolo S, Veronesi A, Zambelli A, Del Sal G *et al*: **Targeting triple negative breast cancer: Is p53 the answer?** *Cancer treatment reviews* 2013.

183. Kriegsmann M, Endris V, Wolf T, Pfarr N, Stenzinger A, Loibl S, Denkert C, Schneeweiss A, Budczies J, Sinn P *et al*: **Mutational profiles in triple-negative breast cancer defined by ultradeep multigene sequencing show high rates of PI3K pathway alterations and clinically relevant entity subgroup specific differences.** *Oncotarget* 2014, **5**(20):9952-9965.

184. Chen X, Li J, Gray WH, Lehmann BD, Bauer JA, Shyr Y, Pietenpol JA: **TNBCtype: A Subtyping Tool for Triple-Negative Breast Cancer**. *Cancer informatics* 2012, **11**:147-156.

185. Lehmann BD, Pietenpol JA: **Identification and use of biomarkers in treatment strategies for triple-negative breast cancer subtypes**. *The Journal of pathology* 2014, **232**(2):142-150.

186. Burstein MD, Tsimelzon A, Poage GM, Covington KR, Contreras A, Fuqua SA, Savage MI, Osborne CK, Hilsenbeck SG, Chang JC *et al*: **Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer**. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2015, **21**(7):1688-1698.

187. Rizki A, Weaver VM, Lee SY, Rozenberg GI, Chin K, Myers CA, Bascom JL, Mott JD, Semeiks JR, Grate LR *et al*: **A human breast cell model of preinvasive to invasive transition**. *Cancer research* 2008, **68**(5):1378-1387.

188. Briand P, Nielsen KV, Madsen MW, Petersen OW: **Trisomy 7p and malignant transformation of human breast epithelial cells following epidermal growth factor withdrawal**. *Cancer research* 1996, **56**(9):2039-2044.

189. Price DJ, Avraham S, Feuerstein J, Fu YG, Avraham HK: **The invasive phenotype in HMT-3522 cells requires increased EGF receptor signaling**

through both PI 3-kinase and ERK 1,2 pathways. *Cell Communication and Adhesion* 2002, **9**(2):87-102.

190. Shalem O, Sanjana NE, Zhang F: **High-throughput functional genomics using CRISPR–Cas9**. *Nature Reviews Genetics* 2015, **16**(5):299-311.

191. van de Wetering M, Francies HE, Francis JM, Bounova G, Iorio F, Pronk A, van Houdt W, van Gorp J, Taylor-Weiner A, Kester L *et al*: **Prospective derivation of a living organoid biobank of colorectal cancer patients**. *Cell* 2015, **161**(4):933-945.

192. Sauder CA, Koziel JE, Choi M, Fox MJ, Grimes BR, Badve S, Blosser RJ, Radovich M, Lam CC, Vaughan MB *et al*: **Phenotypic plasticity in normal breast derived epithelial cells**. *BMC Cell Biol* 2014, **15**(20):20.

CURRICULUM VITAE
Amruta Rajendra Phatak

EDUCATION

Indiana University, Indianapolis, IN
      Major: Medical and Molecular Genetics, Doctor of Philosophy (2015)
      Minor: Life Sciences

University of Mumbai, India
      M.S., Biotechnology (2007)
      B.S., Biotechnology (2005)

Additional Training/Certificate courses
  Center for the Business of Life Sciences, Kelley School of Business
      - Certificate in the Business of Life Sciences (April 2015)

  The Wharton School, University of Pennsylvania:
      - An Introduction to Marketing (Jan 2014)
      - An Introduction to Financial Accounting (Jan 2014)

  University of California, San Diego:
      - Drug Discovery, Development and Commercialization (Aug 2013)

PROFESSIONAL EXPERIENCE

University of Mumbai, Birla College, India (2008-2009)
      Lecturer, Dept. of Biotechnology

Geo-Chem Laboratories (Rajkot) P. Ltd, Mumbai, India (Sept-Dec 2007)
      Trainee Microbiologist (Inspection, analytical testing and Certification Company
      for drugs, pharmaceutical and agricultural products)

Padmaja Aerobiologicals (P) Ltd, Navi Mumbai, India (Jul-Sept 2007)
      Trainee Microbiologist (Analytical lab for environmental monitoring; drug,
      microbial & cosmetic analysis)

University of Mumbai, Birla College, India (2006 – 2007)
      Student Researcher (Dept. of Biotechnology under guidance of Dr. M. Sharon)

LEADERSHIP

Laboratory Assistant
Susan G. Komen for the Cure® Tissue Bank (2011-present)

Volunteer Tour Guide
Butler University College of Pharmacy and Health Sciences (2011-2012)

Organizing Committee Member
Birla College of Arts, Science and Commerce (Feb 2009)
Member of organizing committee for 'Frontiers in Biotechnology: Proteomics, Genomics and Nanobiotechnology' a two-day National Seminar sponsored by University Grants Commission.

SKILLS AND TECHNIQUES

- **3D cell culture**: culture of primary and immortalized cell lines and their phenotypic and molecular characterization in 3D lrECM culture

- **Bioinformatics**: microarray data analysis and pathway analysis tools (such as Ingenuity® Pathway Analysis, Oncomine, Partek GS) and publicly available databases and tools (such as UCSC genome browser, GSEA, GeneCards, CCLE)

- **Cellular biology**: Isolation and culture of human breast epithelial cells, human breast tissue sample processing, establishing and maintaining primary cell culture and cell lines

- **Computer skills**: Microsoft word, Excel, PowerPoint, Photoshop, Illustrator, ImageJ, Prism

- **Microscopy**: light microscopy, stereomicroscopy, and fluorescent/laser confocal microscopy

- **Molecular biology**: recombinant DNA techniques, western blotting, qRT-PCR, Telomeric Repeat Assay Protocol (TRAP)

MEMBERSHIPS IN PROFESSIONAL SOCIETIES

- Associate member of American Association of Cancer Research (**2013-present)**
- Healthcare Financial Management Association (**2014-present)**
- Charter Member and VP of Education Scientific Toasters at IUSM (2014-2015)

AWARDS

- Poster award and short invited talk, World Preclinical Congress 2015
- Student Fellowship, 3D Cellular and Preclinical Models in Oncology, WPC 2015
- Honorable mention, poster presentation, IUSCC Annual Research Day (2013)
- Indiana University School of Medicine Fellowship **(2009-2011)**

PRESENTATIONS

- "Modeling cancer predisposition: a patient-derived lrECM 3D cell culture platform to study disease mechanisms and drug resistance" World Preclinical Congress, Boston, MA (June 2015). **(poster and talk)**

- "Li-Fraumeni syndrome patient-derived LFS50 progression cell series: an experimental model for breast cancer prevention research" AACR Special Conference on Cancer Susceptibility and Cancer Susceptibility Syndromes, San Diego, CA (Jan 2014). **(poster)**

- "Li-Fraumeni patient-derived LFS50 progression cell series as an experimental model for breast cancer prevention research" Indiana University Melvin and Bren Simon Cancer Center's Cancer Research Day, IUSM, USA (May 2013). **(poster)**

- "Li-Fraumeni syndrome patient-derived LFS50 progression cell series as an experimental model for breast cancer prevention research" Third International Symposium on breast cancer prevention: Models for Breast Cancer Prevention: from innovation to action, Purdue University, West Lafayette, IN, USA (Oct 2012). **(poster)**

- "Apical polarity as an architectural assessment of breast cancer risk to study the epigenetic modifications that protect the breast epithelium under the influence of nutrients" Third International Symposium on breast cancer prevention: Models for Breast Cancer Prevention: from innovation to action, Purdue University, West Lafayette, IN, USA (Oct 2012). **(poster)**

- "Use of Candida glabrata and Rhodococcus terrae as Nanofactory for Biosynthesis of Gold Nanoparticles" UGC Sponsored Two Day National Seminar on Frontiers in Biotechnology: Proteomics, Genomics and Nanobiotechnology, Mumbai (Feb 2009). **(poster)**

- "A Simple Methodology Developed for Recovery of Gold and Silver Nanoparticles from Industrial Waste Using Pseudomonas stutzeri" International Convention on Environmental Crisis, Mount Abu, India (2007). **(poster)**

- "Removal of Heavy Metal Pollutants from Water and converting it into Valuable Nano-product by Fusarium oxysporum" International Convention on Environmental Crisis, Mount Abu, India (2007). **(poster)**

PUBLICATIONS

- **Amruta R. Phatak**, Chirayu Goswami, Yunlong Liu, David Leffel, Brittney-Shea Herbert*, Alan Bale*, Levy Kopelovich*. Global gene expression patterns in histologically normal tissues of BCNS subjects reveal "single-hit" effects of PTCH and are influenced by rapamycin. **In preparation.**

- **Amruta R. Phatak**, Yunlong Liu, Brittney-Shea Herbert. Gene expression profiling of Li-Fraumeni syndrome patient-derived LFS50 cell progression series and its characterization as preclinical 3D culture model for breast cancer prevention research. **In preparation.**

- S. Khatri*, P. Ambilwade*, **A. Phatak***, Sharon M., "Toxicity versus Biocompatibility of Carbon Nanotubes" in Carbon Nano Forms and Applications, First edition, Sharon M., and Sharon M., United States: McGraw-Hill, 2010, Chapter 27 (* equal contribution)

- **A. Phatak**. Use of Yeasts and Actinomycetes as Nanofactories for Biosynthesis of Nanomaterials; archived at Birla College, University of Mumbai (March 2007).