

12-2017

A Comprehensive Model and Modulation of Cellular Signaling Involved in Early Mammary Development and Aggressive Cancer Using a Novel Recombinant Protein of the G3 Domain of Laminin-5

Kathryn Adrian Elliott
Clemson University

Follow this and additional works at: https://tigerprints.clemson.edu/all_dissertations

Recommended Citation

Elliott, Kathryn Adrian, "A Comprehensive Model and Modulation of Cellular Signaling Involved in Early Mammary Development and Aggressive Cancer Using a Novel Recombinant Protein of the G3 Domain of Laminin-5" (2017). *All Dissertations*. 2071.
https://tigerprints.clemson.edu/all_dissertations/2071

This Dissertation is brought to you for free and open access by the Dissertations at TigerPrints. It has been accepted for inclusion in All Dissertations by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.

A COMPREHENSIVE MODEL AND MODULATION OF CELLULAR SIGNALING
INVOLVED IN EARLY MAMMARY DEVELOPMENT AND AGGRESSIVE
CANCER USING A NOVEL RECOMBINANT PROTEIN OF THE G3 DOMAIN OF
LAMININ-5

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Biological Sciences

by
Kathryn Adrian Elliott
December 2017

Accepted by:
Dr. Thomas Scott, Committee Chair
Dr. Zhicheng Dou
Dr. Scott Pratt
Dr. William Bridges

ABSTRACT

The mammary gland is a unique and specialized epidermal organ; mammary organogenesis begins in the embryo but is not fully complete until puberty. As such, formation of the mammary gland depends on temporally and spatially regulated developmental steps that require coordination of multiple biological and cell signaling processes; many of which have parallels with cancer development. Research describing the events that occur between birth and puberty is lacking and little is known about human breast development of youth. Since mammary gland development requires a coordinated balance between cell growth, proliferation, and apoptosis, it is critical to understand which signaling pathways are utilized to relay developmental signals, and how these pathways and their targets interact and cooperate with age. Additionally, interactions between integrin molecules and their laminin ligands, especially Laminin-5 (Ln-5; also known as Laminin-332), regulate multiple facets of both embryonic development and tumor growth, invasion, and metastasis. $\alpha 6\beta 4$ integrin serves as a marker to detect distant metastases in the early stages of specific malignancies and $\beta 4$ integrin overexpression has been found in basal-like breast cancers, correlating with aggressiveness to institute a prognostic $\beta 4$ signature that increases with tumor grade. The mechanism $\alpha 6\beta 4$ integrin utilizes to modulate oncogenic signaling through association with Ln-5 molecules in the ECM is the basis for the recombinant protein (rG3, the third of five G domains of Ln-5) produced for the work reported in this dissertation. Here, it is shown there are specific transcriptional differences and a unique interaction of a gene set over time that contributes to postnatal mammary gland development, and this model

clearly shares similarities and signaling pathways with oncogenic development. Especially important are pathways of the adaptive and innate immunities, ECM remodeling and integrin interactions, and extrinsic and intrinsic TP53-mediated apoptosis, greater understanding of which could lead to early detection of potential tumorigenic growth and identification of potential treatment avenues. Presented is a comprehensive model of early mammary development along with several panels of biomarkers that possess a role in normal mammary development, are involved in aggressive cancers, and are affected by apoptosis induced by rG3 treatment. rG3 has proven to be a valuable tool to study apoptotic pathways and the crosstalk among those pathways.

DEDICATION

I would like to dedicate this dissertation to my advisor, mentor, and friend, Dr. Thomas R. Scott. The success of my work is in large part due to his support, encouragement, and patience. During my time at Clemson University, I have grown substantially as a scientist and on a personal level; Dr. Scott deserves credit for both of those achievements. I will be forever grateful for the generosity he has shown me and the confidence he has in me. Dr. Scott, I share in the success of this work with you.

ACKNOWLEDGMENTS

Without several other important individuals, my success would not have been possible. I am grateful to my committee for their advice and sharing their passion for scientific research. Thank you, Drs. Dou, Pratt, and Bridges, for serving as excellent role models and investing your valuable time in me. Additionally, I'd like to express my gratitude to Dr. Vijay Shankar for patiently sharing his expertise with me and helping me become a more versatile scientist.

To my family, especially Dr. Robert Schalkoff, Leslie Schalkoff, and future-doctor Chrissie Schalkoff for their unwavering support and guidance. Thank you for raising me to be strong and independent, and for providing the solid groundwork for my achievement.

To my husband, Blake Elliott, for his encouragement, loyalty, and providing a stable and steady foundation during this wavering and sometimes unsteady process. You have shared in my joys and my many failures; thank you for remaining by my side always.

To Summer Chandler, for her wisdom, laughs, comfort, and friendship. Without you, hard decisions would have been impossible and football season would have been unbearable. I am a better person for knowing you.

I am indebted to all of you and immensely grateful for your love and support; I humbly share this success with you.

TABLE OF CONTENTS

	Page
TITLE PAGE	i
ABSTRACT	ii
DEDICATION	iv
ACKNOWLEDGMENTS	v
LIST OF TABLES	viii
LIST OF FIGURES	x
 CHAPTER	
I. MAMMARY DEVELOPMENT	1
Overview	1
Significance	2
Embryonic mammary gland development	2
Postnatal mammary gland development	6
Human breast development	6
Signaling pathways and the ECM	7
Immune involvement	8
Cytokeratins, integrin molecules, and laminins	10
The swine model	13
Novel techniques	14
Hypothesis	16
 II. BREAST CANCER	16
Overview	16
Basal-like vs. TNBC	18
TNBC complications	18
Signaling pathways in TNBC	19
Integrin molecules in cancer	22
Wnt in cancer	24
Macrophages in cancer	26
Mammary stem cells	27

	TNBC cell lines.....	29
	TNBC treatment.....	30
	rG3 overview	31
	Laminin-5.....	31
	rG3 production history.....	33
	Hypothesis.....	34
III.	MATERIALS AND METHODS.....	35
	Animal sampling.....	35
	Necropsy	36
	Collection for RNA isolation and histology	36
	Contingency analysis	37
	RNA isolation and sequencing: tissue	38
	Functional analysis.....	40
	rG3 production, isolation, and purification.....	41
	Western Blotting: rG3.....	44
	Cell culture.....	45
	Primary cell culture and fibroblast isolation.....	46
	Subculture and rG3 treatment	47
	RNA isolation and sequencing: cells	49
	RNA-seq data analysis.....	50
	Western Blotting: cells.....	51
	Viability assay.....	52
	Statistical analysis.....	53
IV.	RESULTS AND DISCUSSION.....	53
	Chapter 1 results and discussion	53
	Chapter 2 results and discussion.....	149
V.	SUMMARY AND CONCLUSIONS	194
	APPENDICES	198
	A: PRC Top 50 positive gene transcripts.....	198
	B: PRC Top 50 negative gene transcripts.....	201
	C: rG3 treated MDA-MB-231 gene transcript shift changes.....	204
	REFERENCES	231

LIST OF TABLES

Table	Page
1 Phenotype of mammary gland biopsies for RNA-seq	40
2 QC values for swine mammary tissue RNA.....	58
3 CLA indicates genes cluster into functional groups	66
4 Common gene IDs identified using DFA	69
5 Reactome pathway characterization of DFA genes	70
6 Gene name reference for genes identified using DFA.....	76
7 PRC positive weights.....	111
8 PRC negative weights.....	113
9 Summary of PRC trends	116
10 Functional roles of top ten network modules.....	123
11 Reactome pathway summary	126
12 Gene families contained in each module (cluster).....	127
13 Reactome analysis of top 15 pathways utilized by fibroblasts <i>in vitro</i> after 24 hr growth.....	158
14 Top 75 Reactome pathways utilized by differentially expressed genes – rG3 treated fibroblasts.....	160
15 Shift change values for rG3 treated fibroblasts.....	163
16 Reactome pathways utilized by MDA-MB-231 cells after 24 hrs growth <i>in vitro</i>	166
17 Top 75 Reactome pathways utilized by differentially expressed genes in rG3 treated MDA-MB-231s.....	169

List of Tables (Continued)

Table		Page
18	Top 75 pathways utilized by upregulated genes in MDA-MB-231 cells after rG3 treatment	172
19	Top 75 pathways utilized by downregulated genes in MDA-MB-231 cells after rG3 treatment	174

LIST OF FIGURES

Figure	Page
1 H&E visualization of developmental index levels.....	38
2 Swine growth rates during sampling.....	54
3 Formalin fixed, paraffin embedded, H&E stained sections of swine mammary tissue at a) 4 days, b) 8 days, c) 11 days, d) 18 days, e) 25 days, and f) 32 days of age demonstrating histological differences in mammary tissue morphology over time.....	55
4 Weighted scores can quantify histological trends in mammary tissue collected over time.....	57
5 PCA presents large patterns of variability within gene expression data contributing to specific ordination trends	60
6 RDA confirms that variability in pigs does exist, but does not contaminate variability over time.....	62
7 CLA confirms trends are present in the transcriptomic data set that correlate to time	64
8 DFA of mammary development over time indicates temporal regulation during 4-18 days vs 25-39 days after birth.....	75
9 PRC indicates there are specific trends and interactions between sets of genes that have developmental and biological significance over time	94
10 WGCNA demonstrates functional modules in early mammary gland development.....	120
11 MCODE algorithm identifies highly correlated gene networks within RNA-seq dataset	122
12 Functional annotation of MCODE clusters from WGCNA demonstrates biological significance	125

List of Figures (Continued)

Figure		Page
13	rG3 Western Blot comparison between in-house and Vanderbilt.....	150
14	Evaluation of hsp-60 effect on MDA-MB-231 cell viability	152
15	MTT results for MDA-M-231, MCF-10A, and primary swine fibroblast cell line after 24 hr treatment in vitro with 0 – 30µg/mL rG3.....	153
16	Western Blotting for p53 after treatment with 20ug/mL rG3 in vitro using MCF-10A, MDA-MB-231, and primary swine fibroblast lysate.....	154
17	Western Blotting for integrin β 4 <i>in vitro</i> using MCF-10A, MDA-MB-231, and primary swine fibroblast lysate.....	155
18	Panther biological Process analysis; primary fibroblasts after 24 hr growth <i>in vitro</i>	159
19	Panther Biological Processes utilized by MDA-MB-231 transcriptome after 24 hrs growth <i>in vitro</i>	167
20	Panther Biological Process utilized by top genes expressed with negative zscore (downregulated) in MDA-MB-231 cells in vitro	178
21	Panther Pathway utilized by top genes expressed with positive z-score (upregulated) in MDA-MB-231 cells in vitro.....	179
22	Panther Biological Process utilized by top genes expressed with positive zscore (upregulated) in MDA-MB-231 cells in vitro	180
23	Integrin transcript expression in MDA-MB-231 cells at 0 and after 24 hours of in vitro treatment.....	181
24	α 6 β 4 integrin modulation and signaling in MDA-MB-231 cells with and without rG3 treatment.....	183

List of Figures (Continued)

Figure		Page
25	Apoptosis modulation and signaling in MDA-MB-231 cells with and without rG3 treatment.....	192
26	TP53 signaling modulation and signaling in MDA-MB-231 cells with and without rG3 treatment.....	193

CHAPTER ONE: MAMMARY DEVELOPMENT

Overview

The mammary gland is a unique and specialized epidermal organ in that mammary organogenesis begins in the embryo but is not fully complete until puberty and pregnancy. As such, formation of the mammary gland depends on temporally and spatially regulated developmental efforts that require coordination of multiple biological and cell signaling processes; many of which have parallels with cancer development ¹. Three main stages of mammary organogenesis exist: embryonic, pubertal, and reproductive ². Of importance, research describing the events that occur between birth and puberty is lacking. Similarly, little is known about human breast development ³. Much of the research regarding genetic interactions and mammary organogenesis cell signaling has been performed in the mouse model, but initial data from humans indicates the same pathways are utilized in both models ⁴. Since mammary gland development requires a coordinated balance between cell growth, proliferation, and apoptosis ⁵, the current challenges in the field are to determine how mammary cell identity is decided, which signaling pathways are utilized to relay developmental signals, and how these pathways and their targets interact and cooperate throughout time ⁴. This research aims to generate a model from transcriptomic data of postnatal mammary gland development over time in the swine system to provide insight into postnatal activity and cellular signaling mechanisms utilized for mammary growth.

Significance

Research in the field of mammary organogenesis is significant largely due to the suspected links between development and invasive cancer and highly conflicting views of researchers on the presence of developmental activity between birth and puberty. It is the perspective of many that after birth, the mammary gland exists as a rudimentary organ that remains quiescent and dormant until puberty, when development is truly completed⁶⁻⁹. However, reports also suggest that there is growth occurring from birth to puberty, ranging from allometric growth of the gland¹⁰ to significant proliferation, branching, and expansion¹¹⁻¹⁴. If epithelial cell proliferation in the mammary gland can persist for weeks after birth, detection and characterization of stromal influences on epithelial patterning in this phase of development are integral^{2,10}. Also, there are conflicting views of post-natal/pre-pubertal mammary gland development that need to be resolved. Despite advances in the clinical treatment of breast cancer, the disease is still prevalent due to incomplete understanding of mechanisms involved in normal mammary development and physiology¹⁵. In humans especially, the formation of the breast in the embryo and early postnatal development contains some events akin to early breast carcinogenesis¹⁶. Identifying and characterizing the signaling pathways responsible for mammary development and patterning over time, and their downstream effects, can advance our understanding of developmental biology and could have momentous implications for breast cancer treatment¹⁷.

Embryonic mammary gland development

Although the mammary gland is not functional at birth and the majority of organogenesis occurs postnatally under hormonal control, development begins as a ventral

skin derivative in the embryo independently of hormones and is very well characterized^{4,10}. Embryonic mammary gland development requires precise communication between different cell types, specifically those of the epidermis, epithelium, and mesenchyme, and is regulated both temporally and spatially by numerous signaling pathways, which will be discussed¹⁸⁻²¹. There are three distinct developmental steps during mammary gland organogenesis in the embryo, each requiring activation of specific genes and signaling pathways: the formation of bilateral milk lines, formation of placodes along the milk lines, and lastly, formation of mammary buds in the mesenchyme⁴. The initial step, formation of the bilateral milk lines, generates an embryonic region that will be committed to mammary cell fate²². The gland will eventually form in locations along two lines that run from the axilla to the groin of the animal, known as the bilateral milk lines. These lines begin to develop around embryonic day 10-11 of gestation in mice²². At this point, Wnt/ β -catenin signaling activity localizes along these lines, and then quickly just to the positions of future mammary buds, making Wnt signaling pathway members molecular markers of the milk lines^{23,24}. The mammary line is marked by the presence of Wnt10b and the Wnt signaling reporter T-cell factor optimal promoter (TOP) β -galactosidase (GAL)²³. Formation of the mammary line and expression of Wnt10b is the earliest detectable ectodermal event in mammary gland organogenesis in the mouse model²⁴ and within these lines, the previously single-layered ectodermal cells of the embryo become morphologically columnar and multilayered, forming a definitive ridge that is more mature than the primitive epidermis and rises above and below this layer^{22,24}.

During the second stage of embryonic mammary organogenesis, mammary milk line cells migrate to the locations of future mammary buds along the lines^{4,18,25}. When the mammary cells have arrived at these specific locations, five pairs of small lens-shaped

placodes begin to form around E11.5 in mice; placodes are multilayered ectodermal structures that will eventually become epithelial bulbs and are also under the control of WNT/ β -catenin signaling and the transcriptional mediator Lymphoid Enhancer Binding Factor 1 (LEF1) ^{24,26,27}. In humans, the mammary lines form during the first trimester and contain only one pair of placodes ³. Several other signaling mechanisms promote placode formation, growth, and positional direction along the milk line, including Parathyroid hormone-related protein (PTHrP), fibroblast growth factor 10 (FGF10), Polo-like kinase 2 (Plk2), and bone morphogenetic protein 4 (BMP4) ^{28–32}. During the third step, around E13.5 in the mouse, the placodes begin to increase in size and cells within each placode invade the underlying dermal tissue to form mammary buds ^{1,10,22–24}. These buds are bulb-shaped structures containing mammary epithelial cells that are connected to the embryonic epidermis by a stalk. Neighboring cells begin to condense, concentrically surrounding the buds and becoming more densely packed and slightly more organized than other epidermal tissue, forming what is known as the initial mammary mesenchyme ^{28,29}. The mammary mesenchyme aids in continued early development by maintaining epithelial cell identity, secreting growth factors and other key proliferative signals, and supporting ductal invasion ^{2,20}. By E16, the mammary epithelial cells have proliferated and elongated the mammary bud to form a small sprout that is primed to invade the fat precursor (an immature collection of pre-adipocytes at this stage) ^{2,22}. The mammary bud is now a rudimentary mammary ductal tree that will contain multi-laminate club-shaped terminal end buds (TEBs) at their distal ends responsible for leading proliferation out through the adipose and stroma, further branching to generate the primary ductal tree ³³. TEBs are supported by dense stroma containing fibroblasts, macrophages, mast cells, and eosinophils ^{34–36} as they are forming.

Basophils, T cells, and B cells are not detected³⁷. In contrast to mice, the pre-pubertal mammary glands of humans and swine contain a greater degree of ductal branching and terminal ductal units (TDUs) are present instead of TEBs. The TDUs reside in a collagenous stroma and have a higher degree of inter and intra-lobular heterogeneity than TEBs³⁸. By E18 in mice and 21 weeks in humans, the TEBs or TDUs, respectively, divide rapidly to generate a rudimentary ductal tree with lumen that begins to embed in the immature fat pad^{39,40}. At the conclusion of embryonic development, the framework has been established for pubertal expansion. Existing is a rudimentary bi-layered ductal tree with an inner luminal layer of epithelial cells facing the central cavity and an outer layer of basal (mainly myoepithelial, termed from high concentrations of actin microfilaments and α smooth muscle actin present) cells that are adhered to the basement membrane (BM) predominantly through integrin-mediated interactions^{3,41–43}. It is widely accepted that during embryonic, and postnatal, mammary gland development, crosstalk and interactions between stromal, epithelial, luminal, and myoepithelial cells are critical for normal development and expansion of the gland⁴⁴. This signaling relies heavily upon the extracellular matrix (ECM), ECM receptors, and ECM degrading enzymes such as matrix metalloproteinases (MMPs), cytokines, and growth factors⁴⁵. At birth, it is believed that growth halts until puberty but conflicting work has been published on this matter. Shortly after birth, adipocytes densely packed in the fat pad begin to mature and become capable of supporting further epithelial morphogenesis^{46,47}. Increased recognition has been given to the significance of the stroma and ECM in both development and disease in the past decade⁴⁸ which will be discussed subsequently.

Postnatal mammary development

Postnatal mammary growth in the mouse can be classified into five periods: neonatal (1-3 weeks of age), juvenile (3-4 weeks of age), pre-pubertal (4-5 weeks of age), pubertal (5-7 weeks of age), and mature virgin (7-9 weeks of age) ⁴⁹. In these phases, mammary epithelial ducts activate invasive cell signaling to penetrate the surrounding mesenchyme populated by fibroblasts and adipocytes ^{9,10}. This activity is known as branching morphogenesis, which is a fundamental process in development utilized to increase epithelial surface area in organs such as the lung, kidney, salivary gland, and mammary gland. Branching morphogenesis consists of iterative steps of epithelial bud formation, extension, branching and remodeling; the process has been thoroughly reviewed ⁵⁰.

Human breast development

A considerably smaller amount of information regarding human breast development exists, but the bulk of knowledge can be found in the review by Howard and Gusterson ³. Comprehensive studies of mammary gland development in humans have not been executed, in comparison with rodent models, due to constraints on tissue availability and ethical concerns, so the best that can be reported is parallels and obvious differences in developmental stages between humans and rodents ⁵¹. Similar comparisons can identify gaps that need to be addressed in a more appropriate model animal ³. Importantly, in humans, the infant mammary gland completes involution once separated from maternal hormones and these changes are similar to events that can be found in breasts of post-menopausal women ³. Collection of mammary tissue from humans, and relevant animal models, across ages during

all developmental stages would be a valuable tool to understand developmental variations and disease implications ³.

Signaling pathways and the ECM

A multitude of common cell signaling pathways and molecular factors integral in mammary development across species have been detected and thoroughly reviewed, including: Human epidermal growth factor receptor 2 (ErbB-2), fibroblast growth factor (FGF), Wingless-related integration site (Wnt), hedgehog, GATA3, Hepatocyte growth factor like protein (HGFL), nuclear factor kappa-light-chain-enhancer of activated B cells (NFkB), Phosphatidylinositol-4,5-bisphosphate 3-kinase/ Protein kinase B (PI3K/AKT), Janus kinase/Signal Transducer and Activator of Transcription (JAK-STAT), T-box transcription factor (TBX), epidermal growth factor receptor (EGFR or ErbB-1), B-Interferon Gene Positive-Regulatory Domain I (Blimp-1) and insulin-like growth factor (IGF) ^{5,52-71}. Importantly, the ECM, immune system, and cytoskeleton and constituted molecular factors have been implicated to influence both embryonic and postnatal mammary organogenesis. The ECM and stroma, including growth factors, cytokines and proteinases within these niches, have long been known for morphogenesis ⁷² and are essential for development of a functional mammary gland ^{7,73} where these also play important roles in breast cancer development (reviewed by ⁷⁴⁻⁷⁸). The branching morphogenesis of the mammary gland is controlled partially by the basement membrane (BM), interacting epithelial and mesenchymal cell populations, and cell-cell/cell-matrix adhesions, but this crosstalk is not comprehensively understood ⁷⁹. It is accepted that post-natal stroma (previously embryonic mesenchyme) influences cell fate and differentiation throughout mammalian adulthood ^{80,81}. After birth,

mammary epithelial cells are tightly connected to both each other and surrounding stroma, relying on tight adhesion to transmit biochemical signals, allow cell migration, and function normally ^{7,41,82}. Interestingly, epithelial cell movement utilizes many of the same molecules that modulate activity of the actomyosin network, including Ras-related C3 botulinum toxin substrate (Rac), rho-associated, coiled-coil-containing protein kinase (ROCK) and myosin light chain kinase, implying that actomyosin contractility possesses a critical role in mammary gland morphogenesis ⁸³.

Immune involvement in mammary development

Another seemingly unrelated organ system that compasses a major portion of prenatal and postnatal mammary development is the immune system ^{11,84-88}. The well-known cytokines interleukin-4 (IL-4) and IL-13 have been shown to support differentiation and maturation of luminal epithelial cells ⁸⁹, and macrophages, in addition to their roles in pathogen clearing and wound healing, are important regulators of branching morphogenesis during embryonic and post-natal development of the mammary gland ^{36,90,91}. Other innate immune cells have been found in pre-pubertal mammary tissue, likely influencing development prior to puberty ¹¹. Macrophages are abundant cells in nipple aspirates from reproductive age women ⁹² and taken together with previous reports of this cell type localizing with mammary epithelial cells ³⁶, it can be inferred that immune cells play roles throughout all stages of mammary development. Macrophages can be visualized lining the mammary ducts where they secrete growth factors, chemokines, and cytokines to support epithelial cell proliferation ⁹¹. Previous work has shown that the cytokines IL-4 and IL-13 are critical for promoting the differentiation and maturation of luminal epithelial cells ⁸⁹,

reciprocally, data from genetically modified mouse models show that Colony stimulating factor 1 (CSF1,) Transforming growth factor β 1 (TGFB1) and C-C Motif Chemokine Ligand 2 (CCL2), all secreted by mammary epithelial cells, regulate macrophages during mammary development^{34,93–97}. This relationship creates a positive developmental loop, wherein mammary epithelial cells signal to macrophage populations to secrete factors that provide support and promotion of further epithelial development⁹³. In humans, high amounts of CSF1 protein and mRNA were found in normal and lactating breast epithelium^{98–100}. In response to IL-4/IL-13 stimulation, macrophages can produce members of the MMP family *in vitro*^{101,102}. In the context of mammary gland development, MMPs function in ECM remodeling to allow ductal branching into the fat pad^{14,103–107}. Importantly, MMP3 is an extracellular regulator of the Wnt signaling pathway and an effector of epithelial stem cell function¹⁰⁸, and MMP11 has a paracrine function during development that can be exploited to promote tumor progression, providing another link between development, immune molecules, and malignancy¹⁰⁹. MMP14 and MMP15 also regulate adipocyte fate determination in the developing mammary gland¹⁴. In addition to their roles during normal mammary organogenesis, macrophages are often present in the breast tumor microenvironment, where they correlate with reduced overall survival^{110–113}. The importance of immune cells in mammary gland development is well known^{34,35}, but more research is needed to elucidate the molecular signaling that guides them within the gland, and largely unexplored are the roles of chemokines and the adaptive immune system^{86,114}. Accordingly, understanding the function of immune molecules in normal development and their concert with other signaling mechanisms can provide insight into these roles during cancer development⁹¹.

Cytokeratins, integrin molecules, and laminin molecules in mammary development

Several other molecules that are commonly implicated in early mammary development are cytokeratins, integrin molecules, and laminins. Cytokeratins can be found in most epithelial cells where they function as intermediate filament proteins. In the mouse mammary gland, K5 and K14 are markers of basal/myoepithelial cells and K8 and K18 are luminal cell markers^{115,116} and K5 and K14 are co-expressed during the first few postnatal weeks of development¹¹⁷. The role of integrin molecules in mammary development has been thoroughly reviewed by Ramovs *et al*¹¹⁸. As previously mentioned, remodeling of cell–cell and cell–ECM interactions is critical in both normal mammary development and oncogenic development; integrin molecules and laminins are both significant components of the ECM. Integrin molecules function to sense and integrate signals between the ECM and cytoskeleton, influencing intracellular signaling and cell proliferation, migration, and differentiation¹¹⁹. Integrin molecules can also respond to intracellular signals (inside-out signaling) and mediate cell adhesion and transmission of contractile forces for migration^{119,120}. More than 150 proteins have been detected to interact with integrin molecules at sites of adhesion, known as adhesomes^{121–123}. Interacting proteins at the adhesome include cytoskeletal components, enzymes, and adaptor proteins that control downstream signaling cascades and regulate proliferation, differentiation, and migration of cells¹²⁴. Cells must utilize integrin-mediated adhesion to progress through the G1/S cell cycle checkpoint¹²⁵ through signaling using mitogen-activated protein kinases/extracellular signal-regulated kinase (MAPK/ERK), PI3K/Akt, and the small GTPase Rac pathways¹²⁴. Two of the most critical integrin molecules for mammary branching morphogenesis are integrin molecules β 1

and $\beta 4$ ^{126,127}. Both $\beta 1$ and $\beta 4$ are involved throughout mammary morphogenesis ¹²⁸, but it is important to note that $\beta 1$ can form integrin heterodimers with $\alpha 1$, $\alpha 2$, $\alpha 3$, $\alpha 5$ and $\alpha 6$ chains, whereas $\beta 4$ can only associate with $\alpha 6$ ^{119,129}, appointing $\beta 4$ as an indicator of $\alpha 6$ presence. $\beta 1$ integrin, while associated with laminin molecules, has been shown to transmit signals to mammary epithelial cells during pregnancy and lactation ¹³⁰. $\beta 4$ is more controversial; although its contribution to epithelial biology has been investigated for over two decades, its role in mammary gland biology is unclear and confusing ¹³¹. $\beta 4$ is expressed mainly on the basal surface of epithelial ducts and like $\beta 1$, is an adhesion receptor for laminins ¹³². This association allows formation and mediation of stable hemidesmosomes (HDs) on the basal cell surface that link the cellular cytoskeleton with the BM ¹³³. $\beta 4$ and its exclusive partner $\alpha 6$ are organized primarily in basal/myoepithelial cells or basal luminal cells, appearing to prefer the basal plasma membrane surface where they provide an anchor to the BM ^{131,134}. The upregulation of integrin expression by several growth hormones and growth factors in the bovine model proved that integrin $\alpha 6\beta 4$ was critical for mammary development ¹³⁴, but studies in mice suggest $\alpha 6\beta 4$ integrin is actually not essential for proper branching morphogenesis *in vivo* ¹³⁵. In immortalized mammary epithelial cell lines such as MCF10A, $\alpha 6\beta 4$ integrin is critical in the formation of HDs and branching morphogenesis is disrupted without their presence ¹³⁶. Additionally, studies using three-dimensional cultures seem to prove $\alpha 6\beta 4$ is necessary for mammary gland development and function ¹³⁷. Li *et al* hypothesize that $\beta 4$ contributes to mammary gland development by sustaining Parathyroid hormone-related protein (PTHrP) expression and enabling PTHrP signaling, but the mechanism and function of $\alpha 6\beta 4$ integrin in mammary organogenesis is still poorly

understood compared to other molecules¹³¹. $\alpha 6\beta 4$ integrin involvement in cancer will be discussed in Chapter 2.

Laminin molecules in mammary development

Through basally located integrin molecules and laminins, direct attachment of mammary epithelial cells to the ECM can occur^{138,139}. Laminin-binding integrin molecules commonly are found in focal adhesions (FAs) and HDs, which are two adhesion complexes that form mechanical links between the ECM and the actomyosin cytoskeleton¹⁴⁰. Laminins are large heterotrimeric ECM glycoproteins and major components of the BM¹⁴¹. Four integrin molecules recognize laminins as their extracellular ligands: $\alpha 3\beta 1$, $\alpha 6\beta 1$, $\alpha 7\beta 1$ and $\alpha 6\beta 4$ (reviewed in¹⁴²). Laminin-5 can precisely interact with integrin $\alpha 6\beta 4$ in hemidesmosomes to support stromal adhesion¹³³. The first study to implicate Laminin molecules in mammary morphogenesis was completed by Barcellos-Hoff *et al* using Matrigel⁷³. Since then, a study with MCF-10A cells *in vitro* confirmed Laminin-5 and $\alpha 6\beta 4$ integrin are needed for hemidesmosome assembly and branching morphogenesis¹³⁶. Using the murine model, it was confirmed that both laminin and integrin molecules are required to support migration and cellular traction through the stromal, therefore critical in normal mammary gland development¹⁴³. Although evidence is strong for a supportive role of integrin molecules and laminins in mammary development, much remains to be discovered regarding the localization of these complexes and how they truly interact¹¹⁸.

The swine model of mammary development

Currently lacking in the field of mammary biology is continuous developmental data collected from the same live animal over time, which would allow a deeper understanding of developmental processes involved in the period between birth and puberty. Previous attempts commonly used necropsy samples from different animals at different ages ¹⁴⁴ which does not completely control for variability between animals. Additionally, much previous research utilizes a mouse or rat model because despite many anatomical and physical differences with humans, these small mammals can be manipulated easily ². Due to tissue availability constraints, comprehensive studies of human breast development do not exist. The best that could be accomplished is to draw parallels between the human and the mouse mammary gland ³ and studying the infant breast has been difficult and inconclusive ¹⁴⁵. Difficulty in maintaining cultures of primary human cells and continuous human mammary epithelial cells also prove controversial, as exemplified by the MCF-10A human mammary epithelial line which will be discussed ¹⁴⁶. Moreover, *in vitro* and *ex vivo* model systems have not been able to truly represent mammary gland architecture or stromal cell interactions; reviewed by ⁷⁹. Anatomically and physiologically, the swine model is more closely related to that of the human. Sullivan *et al* conducted a comprehensive review of the swine model for wound healing and concluded it is an excellent tool for the evaluation of therapeutic agents destined for use in human wounds ¹⁴⁷. Miller & Ulrey concluded the swine model was appropriate to study human nutrition ¹⁴⁸ while Meurens *et al* concluded that domestic pigs are related enough to humans in terms of anatomy, genetics and physiology that they represent a perfect model to study various microbial infectious diseases, vaccine development, and contribute significant knowledge to improve human health ¹⁴⁹. The use of swine biomedical models has

been well-reviewed by Lunney in 2007; some highlights of the swine model's suitability for human disease include its similarity in size and physiology, organ development, disease progression, allowance of deliberately timed studies and imaging of vessels and organs using clinical technology, and collection of repeated samples¹⁵⁰. It is valuable to have the ability to study multiple pigs from the same litter as well as access to numerous swine cell lines developed from a broad range of tissues¹⁵⁰. Especially useful for genomic and transcriptomic studies is the existence of the swine genome sequence that shows high homology with humans, resulting in the development of advanced genetic and proteomic tools¹⁵⁰. Lastly, the immune system of humans differs considerably from that of the mouse; the swine model may address the limitation of the mouse model to study the immune system in a living animal. With the assessment of key immune regulators and markers in swine, it is now viewed as a better model of human innate immunity and disease than rodents¹⁵¹.

Novel techniques to study mammary development

Nearly all previous studies of postnatal mammary gland development have used microarrays¹⁵². Recently, next-generation sequencing (NGS) technology has emerged as a revolutionary tool in genomics due to its high throughput, reasonable cost, and genome-wide investigation capabilities^{153,154}. One application of NGS, and a major improvement over microarrays, is RNA sequencing (RNA-Seq), which uses NGS technology to sequence, map, and quantify mRNA transcripts^{155,156}. It can be used to directly survey RNA content of mammalian cells and has been used effectively for human samples¹⁵⁷. Technical reviews of RNA-Seq conclude that this technology is an effective way to study the transcriptome of a cell or animal^{153,154,158–160}. Successful use of RNA-Seq to analyze transcriptomes of the sheep

mammary gland ¹⁶¹, the lactating bovine mammary gland ¹⁶², MDA-MB-231 mammary carcinoma cells ¹⁶³, B-cell lymphoma biopsies compared with normal biopsies in dogs ¹⁶⁴ and whole mouse ¹⁶⁵ have been published. In 2014, Zhao *et al* predicted that the RNA-Seq platform will become the predominant tool for transcriptome analysis after demonstrating the benefits over microarrays by performing a T-cell transcriptome comparison using both methods ¹⁶⁶. They concluded RNA-Seq was superior in detecting low abundance transcripts, able to better differentiate biologically critical isoforms, allowed for the detection of more differentially expressed genes, and avoided many of the technical issues present in microarrays ¹⁶⁶. Mantonione *et al*, also in 2014, compared RNA-Seq and microarrays for whole genome profiling and concluded RNA-Seq has a bright future in bioinformatics ¹⁶⁷. There are numerous software packages available to use with the technology ^{168,169} and it has been fruitful for both coexpression analysis ¹⁷⁰ and miRNA expression analysis during mammary gland development ¹⁷¹.

Despite the complexity of RNA-sequencing for gene expression, studies often only use a small number of biological replicates due to tissue availability, invasiveness, or cost limitations. This poses a statistical problem that can be mitigated with the use of specialized statistical techniques, particularly multivariate statistics to assess dependency of inherent variabilities on each other, to extract the most value out of the data ¹⁷². These techniques are currently commonly used in microbial and community ecology; multivariate techniques for exploratory analysis ¹⁷³ and application of multivariate statistics in microbial ecology ¹⁷⁴ have been expertly reviewed. Benefits compared to traditional statistical methods are seen from utilization in community ecology ^{175,176}, microbial ecology ¹⁷⁷, gut microbiology ^{178,179}, and endometriosis ¹⁸⁰.

Chapter 1 Hypothesis: There are specific transcriptional differences and unique interaction of gene set over time that play critical roles in postnatal mammary gland development

Based on our current knowledge of mammary gland development in the post-natal period up to puberty, the hypothesis for this study is that there are specific transcriptional differences and unique interactions of gene sets over time that contribute to postnatal mammary gland development, which have not been fully documented. Since we have high throughput data from mammary development over time in a post-natal swine model analyzed by RNA-Seq, it is both novel and appropriate to use multivariate statistical techniques to identify trends although not currently commonly used in developmental biology. Here, we use a clinically relevant swine model to repeatedly sample mammary tissue throughout early post-natal development in the same animal using minimally invasive technology. In this manuscript, we will provide novel information about the postnatal but pre-pubertal gene expression in swine, insight into the unique developmental trends over time characterized by gene sets, novel pathways utilized and support of currently known pathways in embryonic and pubertal development, and for the first time (to our knowledge) show the transcriptome of mammary development over time in the same animal.

CHAPTER TWO: BREAST CANCER

Overview

Breast cancer is a heterogeneous disease and is currently classified into subtypes based on similar molecular features¹⁸¹. There currently exist five molecular breast cancer subtypes: luminal A, luminal B, normal breast-like, HER2, and basal-like^{182–186}. Of the five

subtypes, basal-like breast cancer (also sometimes known as triple negative breast cancer (TNBC)) is the deadliest form of breast cancer^{181,187}. TNBC lacks the hormone receptors (HRs) present on the cell surface of other cancer varieties (estrogen (ER), progesterone (PR), and human epidermal growth factor 2 (HER2))¹⁸⁸. In the U.S., TNBC represents 12% of all breast cancers and is more prevalent in pre-menopausal women, African-American women, and women with the breast cancer 1 (BRCA1) gene mutation^{181,187,189}. For the purpose of this dissertation, the definition of TNBC will be a group of malignant tumor cells that lack the expression of ER, PR and HER2.

In contrast with TNBC, approximately 70% of breast cancers do possess ER and PR on their cellular surface¹⁹⁰. Termed luminal cancers, they can additionally be classified based on HER2 status to represent the luminal A and B subtypes¹⁸⁵. ER positive breast cancers represent the majority of HR positive cases and the estrogen cell surface receptor has long been a major target for therapy with relatively good prognosis¹⁹¹. HER2 overexpression occurs in about 15–20% of luminal cases and also holds a favorable projection¹⁹². In contrast to the well characterized HR positive subtypes of breast cancer, there remains a rare case with a variable phenotype termed Inflammatory Breast Cancer (IBC)¹⁹³. This is an extremely aggressive form titled by the extremely swollen condition of IBC patients¹⁹⁴. This swollen presentation, due to cancerous cells obstructing lymph vessels underneath the skin, resembles inflammation^{195,196}. IBC occurs in approximately 1-5% of breast cancers in the U.S¹⁹⁴. Interestingly, although IBC is a heterogeneous disease with the ability to express all (or none) of the HRs, the tumors are frequently HR negative, resemble TNBC, and are very difficult to treat^{197,198}.

Basal-like vs TNBC

Basal-like breast cancer is a subtype with defining features that has some overlap with characteristics of TNBC. Some adopt the two terms interchangeably and claim basal-like breast cancer is composed of almost entirely TNBC tumors^{187,199–203}. However, other researchers believe this usage is misleading and that TNBC and basal-like tumors do have significant differences^{182,204–210}. Expression of basal markers such as CK5, CK6, CK14, CK17, and EGFR are associated with a higher histological and morphological grade, presence of BRCA1 mutation, and poor prognosis when compared to TNBC tumors without basal phenotype^{209,211,212}. If basal-like tumors were classified based on IHC expression of basal markers, 10-30% of HR positive tumors would be termed basal-like, leading to this discrepancy^{213,214}; TNBC tumors that do express basal markers are a subgroup with an especially poor outcome^{209,211}. These disparities, combined with the fact that both types are distinct classes with high overlap²⁰⁴, expose the need for a more defined set of biomarkers for each class.

TNBC complications

Despite its grave short-term prognosis and a 300% higher recurrence rate than other types of breast cancer, there are presently no FDA-approved targeted therapies for TNBC^{215–217}. An improved clinical outcome is possible for patients suffering from other breast cancers through the use of targeted therapies toward defective cellular signaling pathways assuming the cells express common surface receptors; this is demonstrated by the success of trastuzumab, pertuzumab, and lapatinib in HER2-overexpressing breast cancer^{218–221}. TNBC patients have a dismal prognosis from the sheer lack of available targeted therapies, and

sadly, most cases eventually develop therapeutic resistance and capitulate to the cancer²²². In addition to the absence of targeted therapies, TNBC lacks a definitive panel of biomarkers for phenotypic correlation²²⁰. If a panel existed, it could aid in one of the most critical intervention steps in cancer development: early disease detection. Without any predictive factors, mammography remains the most powerful advanced detection approach, but often presents over diagnoses as the method cannot distinguish between threatening and non-threatening cancers. Because of the shortcoming in detection, approximately 5%-30% of women are misdiagnosed and undergo unnecessary and costly surgeries, procedures, and undue agony²²³. TNBC is a diverse and complicated set of diseases, each consisting of extended intratumoral heterogeneity and a distinct molecular subtype, resulting in unpredictable responses to chemotherapy²²⁴. Inherent heterogeneity has severe consequences related to treatment resistance even when targeted therapeutics are administered²²⁵. Targeted therapy for TNBC has been inadequate due to the lack of specific targets and partial understanding of the pathways involved^{226,227}.

Signaling pathways in TNBC

One of the most important and well known mutations of TNBC and basal-like breast cancer is that of the BRCA1 (BRest CAncer 1) and BRCA2 (BRest CAncer 2) genes and associated pathways²²⁸⁻²³⁰. Tumors from BRCA1 mutation carriers show dysfunctional cell cycle progression²³¹⁻²³⁴ and express low levels of Cyclin-dependent kinase inhibitor 1B (p27), high levels of S-Phase Kinase Associated Protein 2 (Skp2), cyclin E, and caspase-3 in comparison with sporadic tumors and tumors with BRCA2 mutations^{232,234}. Other genes found mutated with or without BRCA1 and BRCA2 mutations are ATM serine/threonine

kinase (ATM) and tumor protein 53 (TP53); these have a potential role in TNBC metastasis through interaction with BRCA1, BRCA2, and additional proteins that will be discussed subsequently^{183,235–238}. An inherited TP53 mutation causes a condition known as Fraumeni-Li syndrome, which has been proven to predispose women to breast cancer at an early age^{235,238}.

A second pathway critical in cell cycle maintenance and cellular proliferation (and a possible contributor to oncogenic development) is the Phosphatidylinositol 3-kinase/mammalian target of rapamycin (PI3K/mTOR) pathway^{239–241}. PI3K/mTOR is a common focus of clinical research; it is one of the most frequently altered pathways in human breast tumors and controls central aspects of cancer, which include cell proliferation, survival, metabolism, and genomic stability²⁴². It has also become a desirable target for therapy^{243,244}. Mutations are commonly found in the PIK3CA gene in TNBC²⁴⁵ with dysfunction in supporting genes of this pathway including AKT3, Phosphatase and tensin homolog (PTEN), and Tuberous Sclerosis 1 (TSC1)^{245,246}. Functional PI3K molecules transduce upstream signals from receptor tyrosine kinases (RTKs) and G protein-coupled receptors (GPCRs) by phosphorylating the 3'-hydroxyl group of the inositol ring of phosphatidylinositol-4,5-bisphosphate (PI-4,5-P2) to generate phosphatidylinositol-3,4,5-trisphosphate (PIP3)^{247,248}. PIP3 then can recruit cytosolic proteins to the plasma membrane to promote activation or co-localization with other proteins^{249–251}. AKT, a protein with multiple roles in metabolism, growth, and cancer, is able to bind PIP3^{252,253} and activate a wide range of target proteins. The 3' phosphatase PTEN is a lipid phosphatase that converts PIP3 back to PIP2, blocks and regulates PI3K signaling²⁵⁴, and has been characterized as a major tumor suppressor in humans^{255,256}. The PI3K-AKT pathway regulates cell cycle

progression at the G1/S transition through inhibition of degradation of cyclin D1 and Myc, two proteins that drive S phase entry^{257,258}. Specifically, FOXO transcription factors inhibit cyclin D1 expression and promote the cell cycle regulators p27kip1 and p130Rb2²⁵⁹. When phosphorylated by AKT, FOXO supports cyclin D1 expression and inhibits expression of cell cycle inhibitors. Human tumors often show overexpression of cyclin D1 or Myc and reduction in expression of p27kip1, which allows cancerous cells to proliferate^{260,261}. p53 is a critical tumor suppressor protein involved in cell cycle regulation and apoptosis; in normal cells p53 is negatively regulated by the p53-binding protein Mouse double minute 2 homolog (MDM2), which blocks the apoptotic properties of p53 and targets it for ubiquitination^{262,263}. p53 selectively regulates cell cycle arrest, DNA repair, and apoptosis to protect genomic stability and prevent tumor formation²⁶⁴. Instinctively, p53 is often mutated in metastatic tumors; somatic p53 mutations occur in greater than 50% of tumors and almost every cancer type^{235,265,266}. PI3K-AKT is critical in the survival, growth, and proliferation of normally functioning cells; inhibitors of this pathway would require the tumor to be more sensitive than normal tissue which complicates therapeutic development²⁶⁷.

Additionally, several critical AKT-independent PI3K signaling pathways exist, including the PDK1-mTORC2-SGK axis, Rac signaling, and the TEC family kinases²⁶⁸. It is important to note that the Rac proteins (RAC1, 2, and 3) are a subfamily of the Rho GTPase family that function to remodel the actin cytoskeleton²⁶⁹ to promote invasion and metastasis²⁷⁰. Additional pathways potentially exhibiting causal involvement in TNBC are the RAS/RAF/MEK pathway, including amplifications or mutations in Fibroblast Growth Factor Receptor 1 (FGFR1), Insulin-Like Growth Factor 1 Receptor (IGF1R), B-Raf Proto-Oncogene, Serine/Threonine Kinase (BRAF), KRAS Proto-Oncogene, GTPase (KRAS),

HRas Proto-Oncogene, GTPase (HRAS), and Dual Specificity Phosphatase 4 (DUSP4)^{245,246}. Furthermore, cell cycle checkpoints, including amplifications or deletions of RB Transcriptional Corepressor 1(RB1), Cyclin Dependent Kinase 6 (CDK6), Cyclin D1 (CCND1), and Cyclin D2 (CCND2) can lead to tumorigenesis when dysfunctional^{239,246,271–276}. Other pathways commonly studied are the Janus kinase/signal transducers and activators of transcription (JAK/STAT) pathway (amplification in JAK2), and the Notch pathway (including amplification or mutation in WNT and β catenin)^{61,276–283}.

Integrin molecules in cancer

As discussed in Chapter 1, the interactions between integrin molecules and their laminin ligands, especially laminin-5, regulate multiple facets of embryonic development. However, they also effect cell proliferation, tumor growth and the capability of human breast tumors to metastasize^{284–290}. During progression from tumor growth to metastasis, specific integrin signals enable cancer cells to detach from neighboring cells, re-orientate their polarity during migration, and survive and proliferate in foreign areas, similar to their function in normal mammary development²⁹¹. In mature normal tissue, laminin-5 is primarily involved in adhesive function; however, in cancer, it has a more complex role in cell migration and tumor invasion²⁹². The most common integrin molecules that function as laminin receptors on cancer cells are $\alpha 2\beta 1$, $\alpha 3\beta 1$, $\alpha 6\beta 1$, and $\alpha 6\beta 4$ ²⁹³. Integrin heterodimers comprised of the $\alpha 6$ subunit can pair with either the $\beta 1$ or $\beta 4$ subunit, while integrin $\beta 4$ subunit can only pair with the $\alpha 6$ subunit, making $\beta 4$ an indicator of the presence of $\alpha 6\beta 4$ ²⁹¹. Increased expression of integrin $\alpha 6$ has been observed in invasive ductal carcinomas²⁹⁴, skin squamous cell carcinomas^{295,296}, oral squamous cell carcinomas²⁹⁷, ovarian carcinoma²⁹⁸,

prostate tumors²⁹⁹, bladder cancer³⁰⁰, pancreatic cancer³⁰¹ and colorectal cancer³⁰². Furthermore, a range of studies implicate $\alpha 6\beta 4$ integrin as a contributor to oncogenic development in breast cancer^{137,303–305}. In advanced breast carcinomas, $\alpha 6\beta 4$ integrin localizes to the leading edge of invasive cells, cooperating with ErbB2/3 and associating with F-actin to promote metastasis, which leads to deregulation of adhesion, liberation from hemidesmosomes, loss of cell polarity, and hyper-proliferation³⁰⁶. In contrast with development, $\alpha 6\beta 4$ is not limited to the basal layer of epithelial cells in cancer and in tumors it can be found in multiple cell layers of the tissue^{285,289}. Many normal cellular signaling and ECM pathways containing integrin receptors are deregulated in cancer to allow cellular migration, as evidenced by differential integrin expression among normal breast, simple hyperplasia, atypical hyperplasia, ductal carcinoma *in situ* (DCIS) and invasive tumors³⁰⁷. $\alpha 6\beta 4$ integrin is thought to remodel the ECM through interaction with the PI3K and RhoA cell signaling pathways previously discussed and induction of traction forces generated by the actomyosin cytoskeleton^{308,309}. Additionally, $\alpha 6\beta 4$ can induce the production of MMP1 and MMP2^{310,311}, which as discussed in Chapter 1, drive normal mammary development in conjunction with macrophages. $\alpha 6\beta 4$ supports PI3K activation through different receptor tyrosine kinases (RTKs) which is necessary for cell survival³¹². PI3K signaling induced by integrin molecules can also activate the transcription factors STAT3 and c-Jun³¹³. Moreover, $\alpha 6\beta 4$ serves as a marker to detect distant metastases in the early stages of specific malignancies^{289,314} and $\beta 4$ overexpression was found in basal-like breast cancers, correlating with aggressiveness to institute a prognostic $\beta 4$ signature²⁸⁸ that increases with tumor grade^{289,315}. Integrin $\beta 1$ and $\beta 4$ signaling also leads to resistance to treatments including radiotherapy, chemotherapy, hormone therapy and targeted therapies¹³⁷. Notably, $\alpha 6\beta 4$

integrin may modulate oncogenic signaling through association with Laminin molecules in the ECM.

Wnt in cancer

The Wnt gene family consists of 19 Wnt genes encoding closely related proteins and are the most critical family of signaling molecules in embryonic and postnatal development³¹⁶. As described in Chapter 1, formation of the mammary line and expression of Wnt10b is the earliest detectable ectodermal event in mammary gland organogenesis in the mouse model²⁴ and the Wnt family continues to support placode development and TEB expansion throughout mammary organogenesis^{24,26,27}. In fact, expression of Wnt2, Wnt4, Wnt5a, Wnt5b, Wnt6, Wnt7b, and Wnt10b regulate growth and differentiation in the developing mammary gland; in other tissues, Wnt proteins influence analogous processes including cell fate determination, cell growth, apoptosis, maturation, and differentiation³¹⁷. In human breast tumors, overexpression of WNT2, WNT4, WNT5A, WNT7B, WNT10B, and WNT13 has been detected^{318–322}. Specifically, WNT2 is normally expressed in mammary stroma while in breast carcinomas is present in both the epithelium and the stroma³²⁰. Defining the functional significance of Wnt overexpression in breast cancer is difficult due to additional oncogenic mutations found in downstream signaling components in the pathway²⁸⁰. Recently, data suggest that Wnt signaling can promote stem cell self-renewal^{283,323–326}.

Cell surface receptors for the Wnt proteins are composed of a transmembrane domain plus one low density lipoprotein receptor related protein, either LRP5 or LRP6^{316,327–329}. A critical downstream signaling component, β catenin, aids in cell adhesion and transcriptional regulation. It is cytosolically degraded in the absence of Wnt signaling and in the presence of

Wnt is translocated to the nucleus to form a complex with LEF-1 family members, allowing regulation of target genes such as c-Myc and cyclin D1^{317,326,327}. This is classified as the canonical Wnt signaling pathway^{54,316,327}. It is important to note that many Wnt and β catenin target genes are tumor suppressors or oncogenes; in colon cancer, 85% of tumors have a mutation in the APC protein, a negative Wnt regulator that normally destabilizes β catenin³²⁷. Elevated β catenin activity has been shown in many breast cancer cell lines and in the nucleus and cytosol of close to 60% of human breast cancer tissue samples, correlating with poor prognosis³³⁰. In addition, elevated levels have also been found in tumor lysates confirmed by Western Blotting³³¹. Transgenic mouse models containing stabilized expression of β catenin in their mammary gland result in tumor formation^{332,333}. Importantly, many researchers suspect Wnt signaling in breast cancer may be influenced by other members of the Wnt pathway that have not yet been comprehensively studied³³⁴. Several studies have reported altered expression of regulators of the Wnt pathway in breast cancer at the RNA and protein level^{335,336}. Of significance is the crosstalk between Wnt and the PI3K signaling pathways in breast cancer; mutations in or inactivation of PTEN occurs in a significant portion of breast cancers³³⁷. When PTEN is silent, it acts as a PI3K suppressor and leads to the hyperactivation of AKT and nuclear accumulation of β catenin³³⁸. Upon binding growth factors such as Insulin, IGF1, and FGF1, the PI3K/AKT pathway may contribute to the stimulation of β catenin signaling³³⁹. An important downstream signaling molecule in the PI3K pathway, p53, may rely on β catenin degradation to function³¹⁷. The p53 gene is mutated in 20-40% of human breast cancers and can trigger β catenin degradation; this may be a method utilized by the cell to prevent excessive β catenin levels³⁴⁰. Loss of TP53 function in cancer would then result in overexpression and stability of β

catenin and increased expression of transcriptional targets- potentially oncogenes³¹⁷. Finally, kinases of the NF- κ B signaling pathway (I κ B kinases IKK α and IKK β) can regulate β -catenin transcription and activity, further supporting the large network of signaling pathways converging with the Wnt pathway³⁴¹. Data currently indicate activation of the Wnt signaling pathway is frequently one of the most critical signaling abnormalities in breast cancer^{276,283,318,327,334,342,343} and may be a source of novel therapeutic targets for treatment of TNBC³⁴⁴.

Macrophages in cancer

As reviewed in Chapter 1, MMPs secreted by macrophages function in ECM remodeling to allow ductal branching into the mammary fat pad^{14,103–107}. One MMP, MMP3, serves as an extracellular regulator of the Wnt signaling pathway and effector of epithelial stem cell function¹⁰⁸, offering a link between the Wnt pathway and the immune system that can impact adult epithelial cell function¹⁰⁸. Similar to normal mammary organogenesis, the stroma and ECM play a vital role in tumorigenesis³⁴⁵. There are conflicting opinions on whether macrophages present in the tumor stroma environment are involved in tumor rejection or promotion of tumor growth and invasion³⁴⁶, but a detailed review by Lin *et al*³⁴⁷ provides significant evidence that macrophages support tumorigenesis in a comparable manner to their function in normal mammary organogenesis. CSF-1, a regulator of macrophage recruitment, is frequently overexpressed in human breast cancer tissues and correlates with poor prognosis^{348,349}. Furthermore, several clinical studies provide strong evidence in a causal role for CSF-1 in breast cancer progression and metastasis³⁴⁷. The current hypothesis implies CSF-1 produced by tumors, in conjunction with other cytokines

(especially TGFB1 and CCL2), recruits macrophages to the tumor microenvironment and simultaneously inhibits development of immune cells that may present tumor antigens to T-cells. This strategy provides an attractive site of therapeutic intervention ³⁴⁷. The role of TGFB1 in breast cancer is not fully understood; data show that it can both suppress tumor development and promote tumor invasion ³⁵⁰ and possess a pro-tumorigenic effect at later stages of cancer progression where it modulates EMT ^{351,352}. CCL2, secreted by mammary epithelial cells during development, recruits macrophages and can be found in high levels in tumor tissue and surrounding stromal cells ³⁵³, where its expression is highly correlated with macrophage infiltration ³⁵⁴⁻³⁵⁷. It is evident that mammary epithelial cell interactions with macrophages, mediated by a range of cytokines, can significantly affect the risk of breast cancer by supporting tumor development ³⁴⁷. Future studies should investigate the specific signaling pathways utilized in mammary gland development that involve cytokines, macrophages, and if dysfunctional, promote tumor growth and metastasis.

Mammary stem cells

Given that normal stem cells, mammary stem cells (MaSCs), and cancer stem cells all share the ability to self-renew, differentiate, initiate anti-apoptotic pathways, increase membrane transporter activity, and migrate, understanding mammary stem cell profiles and their role in normal development aids in identifying the mechanisms involved in carcinogenesis. These properties make mammary stem cells ideal candidates for the initial events that drive tumor growth and metastasis ³⁵⁸. Adult mammary stem cells are older, and they have had more exposure to damage, leaving them susceptible to accumulate cancerous mutations. It is believed that the deregulation of self-renewal is a key event in cancer

progression, and many of the dysfunctional pathways seen in breast cancer are just reactivated developmental pathways^{57,76,281,286,359–362}. It is hypothesized that MaSCs could be the original cancer-initiating cell for at least some classes of breast cancer^{81,325,358,363,364}. Liu *et al* characterized breast cancer stem cells as a subpopulation of mammary stem cells containing the cell surface marker phenotype of ESA+/CD44+/CD25-/low and found that if these cells were isolated and purified from primary human breast cancers, only 100 were required to form tumors in NOD/SCID mice in contrast with tens of thousands of unpurified cells³⁶⁵. Each time this purified population of cells was passaged, the cells were capable of generating both differentiated cells and a stem cell population, consistent with the original stem cell theory³⁶⁵. These cancer stem cells were better able to form colonies under low-adherence conditions and also displayed increased resistance to chemical compounds³⁶⁵. Liu *et al* also identified multiple normal developmental pathways involved and extensive interaction between them, including Hedgehog, Notch, and Wnt, which were discussed in Chapter 1. The hedgehog signaling pathway mediates communication between the epithelial cells and stroma during mammary duct development and is also involved in MaSC self-renewal^{279,365,366}. The notch signaling pathway contains proteins commonly found in a variety of stem and progenitor cells, functioning to modify cell fate and differentiation along lineages and regulating asymmetric cell divisions³⁶⁵. Many of the notch genes can function as proto-oncogenes, which are normal genes that could become an oncogene if specific mutations are induced³⁶⁵. Wnt signaling plays a role in self-renewal of MaSCs³²⁶. Liu *et al* used mammosphere cultures to demonstrate the interaction between these different pathways and found that activation of the notch pathway leads to subsequent activation of the hedgehog pathway, and that deregulation of this positive feedback loop could induce cancer³²⁶. Like

TNBC, it is known that mammary stem cells lack ER/PR/HER2 receptors suggesting that they may be the original cell that undergoes a mutation and leads to carcinogenesis and tumor formation in basal breast cancers³⁶⁴. Additionally, the regulatory pathways utilized by MaSCs such as notch ligands, p63, and Wnt pathway are also known to be involved in basal like tumors and are associated with poor prognosis^{278,367–371}. Breast cancers that are positive for ER, PR, and Erb2 most likely arise in a committed progenitor cell population and are generally easier to treat³⁷². Recent research has shown a linear relationship in the number of cancer stem cells in breast tumors and the aggressiveness of the cancer and likelihood that it will spread^{361,365}.

TNBC cell lines

Proper *in vitro* models for breast cancer are necessary for accurate and reproducible results. In this dissertation, three cell lines are used: MDA-MB-231, MCF-10A, and primary swine fibroblasts. MDA-MB-231 cells and primary swine fibroblasts are used for RNA-seq due to complications with the MCF-10A cells which will be discussed. Frequently used in breast cancer research are the human cancer cell line MDA-MB-231 and the normal human mammary epithelial cell line MCF-10A. Both of these cell lines express some level of integrin $\alpha 6$ and varied expression levels of integrin $\beta 4$ ^{284,312,373–377}. The MDA-MB-231 line is an invasive breast cancer cell line that lacks expression of estrogen (ER) α , progesterone (PR) and human epidermal growth factor receptor 2 (HER2), and is classified as a TNBC line³⁷⁸. In addition to showing high expression of integrin $\alpha 6$ and integrin $\beta 4$, the $\alpha 6\beta 4$ complex is necessary for the tumorigenic properties of these cells³⁰³. The MCF-10A cell line, a normal human mammary epithelial cell model, is commonly used as a normal control, but recent

research suggests that these cells may not represent phenotypically normal epithelial cells¹⁴⁶ as these cells exhibit a basal phenotype but also share features of mesenchymal cell populations. In order to alleviate this concern, swine primary fibroblasts were used as a normal mammary cell type in development as they do have a role in the regulation of differentiation³⁷⁹.

TNBC treatment

As mentioned, TNBC patients have a poor prognosis and therapeutic outcome in comparison to HR positive breast cancer because of the aggressive nature and lack of targetable cell surface receptors. Current molecular targets include EGFR, vascular endothelial growth factor (VEGF), Poly (ADP-ribose) polymerase (PARP), PI3K, MEK, heat shock protein 90 (HSP90), and histone deacetylases (HDAC)^{216,380}. TNBC has been shown to express EGFR at a higher level in comparison to other subtypes³⁸¹. EGFR is involved in tumor growth and survival, and inhibition of this molecule via monoclonal antibodies or small molecule inhibitors induces cell cycle arrest, initiation of apoptosis, and inhibition of angiogenesis, invasion, and metastasis^{382–384}. Tumors require a blood vessel system in order to obtain nutrients for metastasis, growth, and invasion; one of the most important proangiogenic factors and mediators of this activity is VEGF (Hoeben *et al.*, 2004). VEGF expression within tumors in TNBC is higher than non-TNBC tumors³⁸⁶. Currently, VEGF can be blocked using specific antibodies and small molecule oral tyrosine kinase inhibitors³⁸⁷. Further investigation of the cellular signaling pathways utilized by TNBC could provide novel therapeutic targets.

rG3- Overview

As discussed, the interactions between integrin molecules and their laminin ligands, especially Laminin-5 (Ln-5; also known as Laminin-332), regulate multiple facets of both embryonic development and tumor growth, invasion, and metastasis in many carcinoma types^{284–287,289,290,292,294,308,388–391}. Additionally, integrin heterodimers comprised of the integrin $\beta 4$ subunit can only pair with the $\alpha 6$ subunit, making $\beta 4$ an indicator of the presence of $\alpha 6\beta 4$ ²⁹¹. $\alpha 6\beta 4$ integrin serves as a marker to detect distant metastases in the early stages of specific malignancies^{289,314} and $\beta 4$ integrin overexpression was found in basal-like breast cancers, correlating with aggressiveness to institute a prognostic $\beta 4$ signature²⁸⁸ that increases with tumor grade^{289,315}. The mechanism $\alpha 6\beta 4$ integrin utilizes to modulate oncogenic signaling through association with Ln-5 molecules in the ECM is the basis for the recombinant protein (rG3) produced in this dissertation.

Laminin-5

Ln-5 is a trimeric glycoprotein present in the ECM and basal lamina; each laminin molecule forms a cruciform structure with three chains, α , β , and γ , linked by disulfide bonds; five α , three β , and three γ chains have been identified^{132,388}. These chains combine to form at least 15 laminin isoforms that are uniquely distributed in tissues and carefully regulated during early development^{392,393}. Ln-5 can be degraded by proteases during tissue remodeling; the $\gamma 2$ subunit is proteolyzed by MMP-2 to promote cell migration³⁹⁴. Within the Ln-5 subunit, distinct functional domains are able to bind with other ECM molecules and integrin cell surface receptors³⁹⁵. Importantly, the globular (G) domain at the carboxy-terminal of the laminin α chain interacts with high affinity for the $\alpha 6\beta 4$ integrin^{396,397} and the complex

promotes hemidesmosome formation^{398,399} and stromal adhesion¹³³. The G3 domain is essential for the unique biological activity of Ln-5⁴⁰⁰; deletion of the G3 domain in recombinant forms of Ln-5 leads to a drastic loss of cell adhesion and motility⁴⁰¹. Studies have shown the G3 domain to contain at least two sites that regulate cell adhesion and motility in different ways, and activities of Ln-5 are largely dependent on these sites³⁸⁸. When cleaved during proteolytic processing, the G4/G5 fragment that is released from the G3 domain also functions to support cell migration³⁹⁷.

Ln-5 is associated with each of the seven defined cancer stages: mutation of normal cells, cancer cell proliferation, infiltration/invasion, migration, metastasis, angiogenesis, and additional cell proliferation in the metastatic lesion⁴⁰² and its expression is altered at the transcriptional, translational, and posttranslational levels in many types of cancer^{403,404}. Notably, overexpression of Ln-5 is controlled by additional molecules of the Wnt pathway; in colorectal carcinoma, β -catenin upregulation activates the laminin γ 2 subunit resulting in enhancement of tumor development and cancer cell invasion^{405,406}. High levels of Ln-5 are correlated with poor prognosis in cervical cancer, pancreatic carcinoma, hypopharyngeal cancer, urinary bladder urothelial cancer, small-sized lung adenocarcinoma, malignant glioma, gastric cancer, squamous cell carcinoma (SCC) of the tongue, colorectal adenoma and hepatocellular carcinoma⁴⁰⁷⁻⁴¹⁴. The accepted consensus is that Ln-5 is overexpressed in the leading edge of tumorigenic cells³⁹⁵. Although Ln-5 is best known as an epithelial BM component, it has also been reported in the BM of blood vessels along with α 6 β 4 integrin on endothelial cells^{415,416} and is thought to play critical roles in angiogenesis⁴¹⁷.

The role of Ln-5 in tumor invasion and metastasis makes it a popular target for cancer therapy and several synthetic peptides have been developed to inhibit Ln-5 expression or

mitigate its biological function in tumors⁴¹⁸⁻⁴²⁰. To develop effective reagents that limit ability of Ln-5 to induce cellular proliferation through association with $\alpha 6\beta 4$ integrin, a better understanding is needed of signaling cascades regulated by this adhesion³⁹⁵. Therefore, we have developed a recombinant form of the G3 domain of Ln-5 (rG3) in order to assess Ln-5/ $\alpha 6\beta 4$ integrin signaling and the critical pathways utilized in tumorigenic cells.

rG3 production history

The G3 domain of the rat Ln-5 $\alpha 3$ -chain was expressed in a prokaryotic system engineered with chaperone proteins, allowing for soluble protein expression in a cold temperature environment⁴²¹. Turner confirmed recombinant rG3 had a molecular weight (MW) of approximately 27 kDa based on the G3 input sequence and the addition of a histidine tag for purification and detection⁴²¹. After expression, Turner purified the recombinant protein using chromatography and performed adhesion and proliferation assays on three cancer cell lines (MDA-MB-231, MDA-MB-435 and MCF-7). For all cell lines, rG3 demonstrated dose-dependent adhesion and an inhibitory effect on proliferation and produced an apoptotic phenotype with elevated caspase levels present after varied hours of rG3 treatment⁴²¹. Further studies by Turner showed altered levels of phosphorylated AKT and reduced cell viability through pro-apoptotic signaling induced by the binding of rG3 to the α subunit of $\alpha 6\beta 4$ integrin⁴²¹. These results induced the activity of rG3 on the PI3K/AKT pathway to be studied in depth⁴²². Results from those studies showed significant differences in protein expression of PI3K/AKT pathway members in MDA-MB-231 after rG3 treatment compared to untreated cells; higher levels of AKT and phosphorylated AKT were seen in untreated cells, indicating the inhibitory effect the rG3 protein has on this pathway. Both

IKK $\alpha\beta$ and the phosphorylated IKK β catalytic subunit were expressed at a significantly higher level in untreated cells, as were the levels of phosphorylated nuclear NF κ B. Importantly, the greatest change in expression was seen in p53; in treated cells, it was present at higher levels than in the untreated cells, confirming Turner's conclusions that rG3 is able to significantly reduce cell viability of tumorigenic cells through specific PI3K/AKT signaling events ⁴²².

Hypothesis: There are transcriptional similarities in normal development and TNBC that can be affected through rG3 signaling pathways

As evidenced in Chapter 1, normal mammary development requires finely coordinated signaling between mammary epithelial cells, stromal molecules, and the ECM in order to produce a fully functional gland. This involves drastic tissue remodeling, degradation of the ECM, immune system regulation, and EMT- all key processes also implicated in TNBC. Although TNBC manifests in the mammary epithelium, surrounding cells play important roles in tumor growth and metastasis. Many critical signaling factors are shared between normal mammary development and cancer, especially those of the Wnt pathway, integrin and laminin association, ECM, PI3K pathway, and macrophage functions. Here, transcriptomic data collected over time in a post-natal swine model is analyzed by RNA-Seq from Specific Aim 1 to compare trends with the MDA-MB-231 TNBC cell line and primary swine fibroblasts. It is suspected that the postnatal but pre-pubertal gene expression in swine utilizes a unique set of developmental genes over time that is shared closely with cancer development.

Additionally, rG3 production will be optimized and the resulting recombinant protein will be characterized to determine suitability and purity for treatments; additional protein complexes in the sample will be identified by mass spectrometry. Next, dose-dependency of rG3 will be determined on MDA-MB-231 cells and a non-tumorigenic cell line and cell-signaling pathways and potential biomarkers affected by rG3 treatment *in vitro* will be compared to biomarkers in Chapter 1. It is hypothesized that signaling pathways effected by rG3 will share similarities with those present in early mammary development and tumor development, generating a panel of significant biomarkers for further study.

CHAPTER THREE: MATERIALS AND METHODS

Animal sampling

In order to identify signals involved in early mammary development, a fine needle biopsy sampling method was developed to collect mammary tissue of multiple species (bovine, ovine, and porcine) across different pre-pubertal ages (U.S. patent application number 62/256,416). A 14-gauge spring activated needle was used (Jorgenson Labs Inc. J0528) with a dual setting specimen notch of 8 mm or 16 mm depending on animal age and size. According to AUP 2015-064 approved 10/2015, each animal was manually restrained for no longer than about 10 minutes. The sample site was cleaned with a triple wash of chlorhexidine (2%; Durvet) and β dine (Purdue Products). A body weight was taken during each sampling and recorded. Based on weight, 1-4 CCs of Lidocaine (2%) was administered at the sample site via intradermal and subcutaneous routes and a core tissue sample was then removed from the parenchyma of the mammary gland below the nipple using the biopsy

needle. Mammary gland locations were alternated week to week to decrease risk of tissue damage. The animal was evaluated for 5 minutes following the procedure to ensure any bleeding had stopped from the biopsy sample site. If bleeding was observed, pressure was applied. After being released from the restraint the animal returned to a normal diet and housing with daily observation for 3-5 days for observation of any signs of infection. During the sampling procedure, the animal was closely monitored for any signs of pain. Post-sampling, tissue was handled according to downstream methods below.

Necropsy

Per AUP 2015-064 approved 10/2015, two euthanized animals were used to measure how deep the mammary tissue expands into the body and to compare necropsied tissue sample with live biopsy samples to confirm there is low or no variability in sampling live animals with the needle. Intravenous euthanasia was performed according to accepted practices under Godley-Snell Research Center veterinary supervision. After death was confirmed, a tissue sample was removed from the parenchyma of the mammary gland either using the biopsy needle for a small core sample or a sterile scalpel for a larger sample representative of past sampling technology. Post-sampling, tissue was handled according to downstream methods below.

Collection for RNA isolation

If RNA was to be extracted from the tissue, after sampling, each biopsy was placed directly into a sterile 1.5 mL microcentrifuge tube containing 1mL of RNAlater (Thermo Fischer Scientific) and stored at -20°C until RNA isolation.

Collection for histology

If the sample was to be processed histologically, each biopsy was immediately placed in a tissue cassette (VWR) with 0.67mm pore openings and fixed in 10% neutral buffered formalin (supplier) for 18-24 hours at room temperature. Fixed tissues were then dehydrated in ethanol, cleared in xylene, embedded in paraffin blocks, sectioned to a thickness of 6-8 μ using a rotary microtome (Leica Microsystems), mounted onto glass slides, and dried overnight on a slide warmer at 37°C. Slides were then visualized for morphology with hematoxylin and eosin (H&E).

Contingency analysis and developmental index

A contingency analysis was conducted for two necropsied animals to confirm that the mammary biopsy procedure produces a tissue sample that is representative of the results obtained by necropsy. In the contingency analysis, H&E stained tissue sections obtained by each method were visually assigned a development index (DI) score. The development index can be seen in Fig. 1. As illustrated in Fig. 1, cells with a maturity index of 0 are a cluster of cells without lumen and 1 ring layer, cells with a maturity index of 1 present as a cluster of cells without lumen and more than 1 ring layer of cells, cells with a maturity index of 2 present as loosely organized cells with lumen, cells with a maturity index of 3 present as semi-organized cells with lumen, cells with a maturity index of 4 present as dense and organized cells with lumen, cells with a maturity index of 5 present as dense and organized cells with lumen and branching. Tissue was prepared histologically according to previous section and development index scores were statistically compared between the two methods

using JMP Pro 12 software (SAS, Cary, NC).

Sequential H&E in swine mammary tissue collected over time proves ability of method to study time course of development

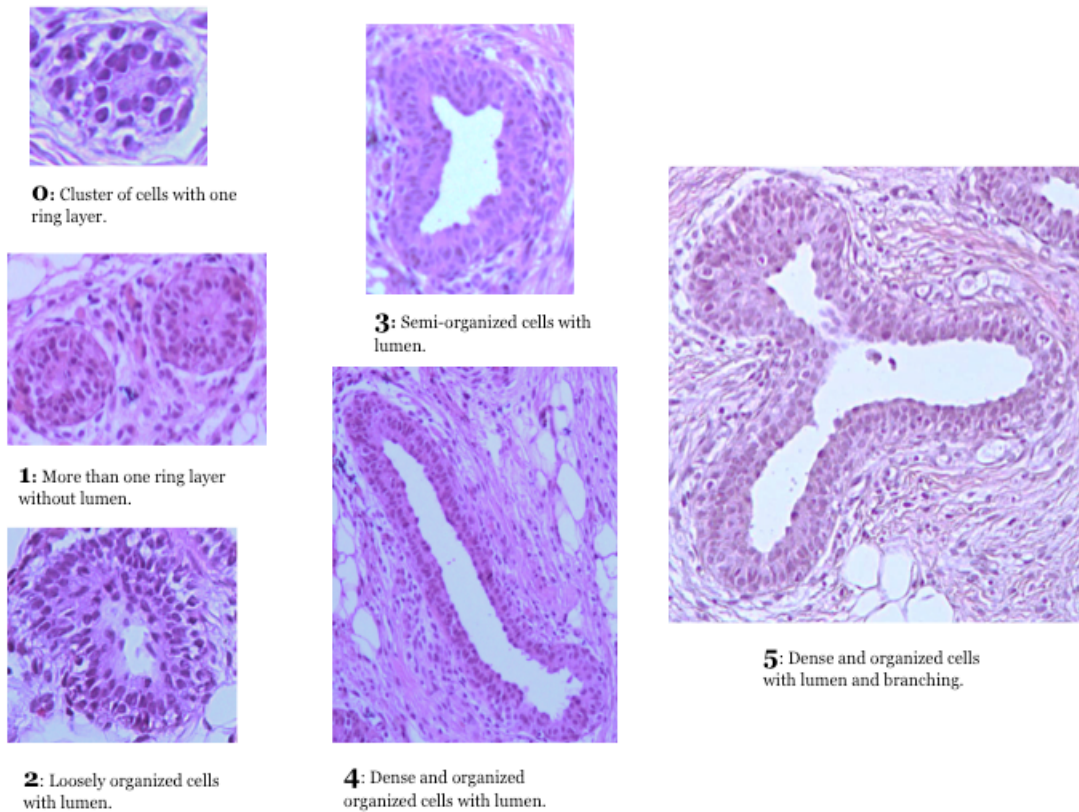


Figure 1: H&E stained formalin fixed, paraffin embedded sections of swine mammary tissue demonstrate developmental index levels 0-5.

RNA isolation

RNA sequencing was performed with RNA isolated from the tissue using an RNeasy

Kit (Qiagen Inc., Valencia, CA) according to manufacturer's instructions and a TissueRuptor (Qiagen Inc., Valencia, CA) according to manufacturer's instructions. After disruption with the TissueRuptor, the RNA samples were also treated with an on-column DNase (Qiagen Inc., Valencia, CA) to remove any possible genomic DNA contamination. Isolated RNA samples were quantified and integrity validated on a Nano Drop 1000 Spectrophotometer (Thermo Scientific, Waltham, MA) and Bioanalyzer 2100 (Agilent Technologies) and stored in 1.5mL microcentrifuge tubes at -20°C until library preparation.

Library preparation and RNA sequencing

Each mammary tissue sample was normalized to a standard input concentration of total RNA and an Illumina compatible sequencing library was prepared robotically on a Microlab STAR (Hamilton) with the TruSeq stranded total RNA library prep kit (Illumina) following the manufacturer's recommended procedures (Illumina). The resulting sequencing libraries were assessed for size on a 2100 Bioanalyzer (Agilent) and sequence data collected on 1 lane of an Illumina HiSeq2500 with a 2x125 bp PE read on high-output mode (Hollings Cancer Center, MUSC) up to a depth of at least 22 million reads. Raw sequence reads were assessed for run quality with the FastQC analysis package (Babraham Bioinformatics) and then preprocessed to remove adapter and low quality bases with the Trimmomatic software package⁴²³. Trimmed reads were aligned to the Sscrofa 10.2 genome using gsnap⁴²⁴. SAMtools was used to generate sorted and indexed bam files⁴²⁵. Subread's featureCounts was used to identify paired-end reads that were concordantly assigned to known genes in the correct order of transcription^{426,427}. Raw read counts were normalized for each technique using MATLAB (MathWorks) and RStudio version 1.0.136 (The R Foundation).

From the resulting dataset, more advanced statistical analysis methods were utilized using MATLAB and R. Preprocessing of redundant variables was completed using correlation analysis (Caret R) with a cutoff of 0.95. Details for RNA-seq input are shown in Table 1.

Table 1: Phenotype of mammary gland tissue biopsies submitted for RNA sequencing shows age ranges.

Sample labels	Tissue source	Phenotype (age in days)	Age in weeks	Pig ID
1	Swine mammary	4.0	0.6	1-8
5	Swine mammary	11.0	1.6	1-8
6	Swine mammary	18.0	2.6	1-8
11	Swine mammary	25.0	3.6	1-8
13	Swine mammary	32.0	4.6	1-8
16	Swine mammary	39.0	5.6	1-8
3	Swine mammary	11.0	1.6	1-13
7	Swine mammary	18.0	2.6	1-13
9	Swine mammary	25.0	3.6	1-13
12	Swine mammary	32.0	4.6	1-13
15	Swine mammary	39.0	5.6	1-13
2	Swine mammary	8.0	1.1	1-10
4	Swine mammary	11.0	1.6	1-10
8	Swine mammary	18.0	2.6	1-10
10	Swine mammary	25.0	3.6	1-10
14	Swine mammary	32.0	4.6	1-10
17	Swine mammary	39.0	5.6	1-10

Functional analysis

From the resulting transcriptomic dataset, in depth biological analysis was performed to determine functional significance of transcriptomic trends in the data using Panther

(Thomas Lab at the University of Southern California), Genesis⁴²⁸, Cytoscape⁴²⁹, Reactome^{430,431} for pathway analysis, NCBI (U.S. National Library of Medicine), and GeneCards (Weizmann Institute of Science).

Recombinant G3 (rG3) Protein production, isolation, and purification

In Turner's work, the G3 cDNA was cloned into a vector including an ampicillin resistance gene to allow for selection of the plasmid and an N-terminal 6X-histidine tag was added⁴³². ArcticExpress™ cells were used for expression of rG3⁴²¹. Once cells were induced for protein expression and harvested, they were prepared for separation by SDS-PAGE to confirm rG3 expression. Once soluble rG3 expression was confirmed, a culture was prepared to yield a larger volume of the recombinant protein for purification and the fraction was applied to a G50 Sephadex column for size exclusion separation. As previously demonstrated, the molecular weight of rG3 including the histidine tag is ~27 kDa⁴²¹. To determine which fractions contained the protein, an OD reading was taken, and to confirm the identity of the expressed soluble protein as rG3, a Western Immunoblot was performed using an anti-polyHistidine monoclonal antibody⁴²¹.

For this dissertation, production was modified to get the purest form of rG3. ArcticExpress™ cells were still used for expression. Transformed cells were cultured by streaking for isolation on LB agar plates containing ampicillin (62.5µl 100mg/mL stock per 125 mL agar) and gentamicin (250 µl 10mg/mL stock per 125 mL agar) and incubating at 37°C overnight until colonies were clearly identified the next day. One isolated colony was chosen from the plate and added to 10mL sterile LB broth with 5µl ampicillin and 20µl gentamicin into a 125mL flask, which was incubated overnight at 37°C with shaking at 250

RPM. The next day, 2.5mL of the overnight culture was added to 247.5 mL sterile LB broth without antibiotics and incubated at 30°C with shaking at 250 RPM. The OD was checked periodically until the culture reached absorbance of approximately 0.4, or the exponential growth phase. At this point, cells were induced for protein expression with IPTG (1mM) and incubated at 11.5°C with shaking at 250 RPM for 24 hours. The flask was transferred to a sterile bottle and spun to pellet at 5250 RPM for 10 minutes at 4°C. The pellet was either snap-frozen in liquid nitrogen and sent to Vanderbilt (MSRC Proteomics Laboratory, Vanderbilt Medical University, Nashville, TN) for further purification using their equipment or purified in house using the Ni-NTA Fast Start purification kit for His-tagged proteins (Qiagen, Germantown, MD).

At Vanderbilt, the pellet was lysed using an Emulsiflex C3 homogenizer (Avestin) and spun down in an ultracentrifuge. After filtering, the lysate was applied to a Roche cOmplete nickel column (Roche Life Science) using a GE AKTExpress purification system (GE Healthcare Life Sciences). The system applied the lysate to the column and the column was washed with binding buffer until a stable UV baseline was achieved. A gradient elution of 0mM Imidazole and increasing to 500mM Imidazole was used to elute into different wells of a 96-well deep well plate based on single, separate peaks. The wells were examined via coomassie staining to find wells containing protein of the appropriate size. Protein was quantified using a NanoDrop 1000 Spectrophotometer (Thermo Scientific).

Mass spectrometry analysis of the purified protein components after separation by SDS-PAGE was performed by the MSRC Proteomics Laboratory (Vanderbilt University, Nashville, TN). All MS/MS samples were analyzed using Sequest (Thermo Fisher Scientific, San Jose, CA, USA; version 27, rev. 12) and X! Tandem (The GPM, thegpm.org; version

CYCLONE (2010.12.01.1)). Sequest was set up to search the ecoli-uniprot-012309-200901_rev_4292_20151111_rev database assuming the digestion enzyme trypsin. X! Tandem was set up to search the ecoli-uniprot-012309-200901_rev_4292_20151111_rev database (unknown version, 9378 entries) also assuming trypsin. Sequest was searched with a fragment ion mass tolerance of 0.00 Da and a parent ion tolerance of 2.5 Da. X! Tandem was searched with a fragment ion mass tolerance of 0.50 Da and a parent ion tolerance of 2.5 Da. Glu->pyro-Glu of the n-terminus, ammonia-loss of the n-terminus, gln->pyro-Glu of the n-terminus, oxidation of methionine and carbamidomethyl of cysteine were specified in X! Tandem as variable modifications. Oxidation of methionine and carbamidomethyl of cysteine were specified in Sequest as variable modifications. Scaffold (version Scaffold_4.4.8, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 99.0% probability by the Peptide Prophet algorithm⁴³³. Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm⁴³⁴. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

For in house purification using the Ni-NTA Fast Start purification kit for His-tagged proteins (Qiagen, Germantown, MD) according to the manufacturer's protocol with the addition of a liquid nitrogen freeze-thaw cycle for cell lysis and 4°C incubation for binding. A cleanup process to eliminate salts and smaller fragments was conducted with Pall spin columns (10 K) according to the manufacturer's protocol.

Western Blotting – rG3

In preparation for protein separation by SDS-PAGE, purified rG3 and pure HSP60 (Prospec Bio, Israel) samples were added to an equal volume of Laemmli Buffer with 5% 2-mercaptoethanol. Samples were mixed and loaded onto a 10-20% Ready Gel Precast Gel (Bio-Rad Laboratories, Inc., Hercules, CA) together with a protein ladder (Precision Plus Protein™ Kaleidoscope™ Prestained Protein Standards #1610375). The gel was run in the Ready Gel System (Bio-Rad Laboratories, Inc., Hercules, CA) with running buffer (15g Tris Base, 72g glycine, 5g SDS per 1L) at 200 V for 30 minutes. Meanwhile, a sheet of Immuno-Blot™ PVDF Membrane and two pieces of 14 x 11 cm filter pads (Bio-Rad Laboratories) were soaked in chilled Western transfer buffer (3.03g Tris Base, 14.4g glycine per 1L) for least 30 minutes. Following electrophoresis, the gel was washed in cold Western transfer buffer for 30 min. The filter pads, gel, and PVDF membrane were arranged in a layered fashion, and protein transfer was completed at 100 V for 45 minutes using a Mini Trans-Blot® Cell (BioRad Laboratories). Non-specific proteins were blocked with 5% non-fat dry milk in TBS at room temperature on a shaker, followed by three washes of 5 minutes each in TBS with 0.5% Tween (Sigma, St. Louis, MO) (TTBS) added. An overnight primary antibody incubation at 4°C was performed with the following antibodies (insert anti-his and hsp ab specs) at dilutions according to manufacturer's instructions (Cell Signaling Technology, Inc.). Following primary antibody incubation, a 15 min wash with TTBS was performed to remove unbound antibody. Next, a secondary anti-rabbit IgG HRP-linked antibody (Cell Signaling Technology®, Inc. #7074) was added for 1 hr at RT. Protein detection was obtained through the use of the HR Peroxidase Visualization Kit (Vector

Laboratories, Burlingame, California).

Cell Culture

MDA-MB-231, MCF-10A, and primary fibroblast cells were used for biological assays. The MDA- MB-231 cells (ATCC HBT-26) are a metastatic breast cancer cell line obtained from a 51-year old female patient, derived from a pleural effusion removed on October 17, 1974. MDA-MB-231 cells were maintained in Dulbecco's Modified Eagles' Medium (DMEM) with 4.0 mM L-glutamine and 4500 mg/L glucose (HyClone, Logan, Utah). The culture medium was supplemented with 10% bovine growth serum, penicillin G (100 units/ml)/streptomycin sulfate (100 µg/ml) and sodium pyruvate (0.11 µg/ml). MCF-10As (ATCC CRL-10317) are a non-tumorigenic human mammary epithelial cell line. MCF10A cells were maintained in DMEM/ high glucose supplemented with 5% horse serum, 20 ng/mL human epidermal growth factor (Sigma Chemical Co.), 0.05 mg/mL hydrocortisone (Sigma Chemical Co.), 100 ng/mL of cholera toxin (Calbiochem), 10 µg/mL of insulin (Sigma Chemical Co.), and penicillin G (100 units/ml)/streptomycin sulfate (100 µg/ml). The primary fibroblasts were obtained from swine mammary tissue in our laboratory; details are below. All cell lines were grown in 75 cm² tissue culture flasks (Corning inc., Corning, NY) and maintained at 37°C and 5% CO₂. To prevent overcrowding of cells, they were routinely passaged using Cell stripper (Corning, NY) or 1X Trypsin/EDTA (HyClone, Logan, Utah), which removed the cells from the bottom of the flask. Before cells were transferred into a new culture flask, they were washed at least one time with fresh culture medium and centrifuged at 400 x g for 5 min at 23°C.

Primary cell culture and fibroblast isolation:

After collection in the field, fresh biopsy tissue was immediately placed in a 1.5 mL microcentrifuge tube containing warm primary cell culture medium. Culture medium for mammary epithelial cells, per 100mL of medium, consisted of: 4ug/mL insulin (40ul from 10mg/mL stock), 10ng/mL EGF (5ul from 200ug/mL stock), 10% Fetal Bovine Serum (10mL), 1% Penicillin-Streptomycin solution (Cellgro #30-002-CI) (1mL), and Hyclone DMEM/high glucose media (Thermo #SH300222) to volume. In the lab, immediately before digestion, collagenase solution was prepared which consisted of (per 10mL of solution): 5% Fetal Bovine Serum (500ul), 1mg/mL Type 1 collagenase (EMD Millipore, 234153) (10mg), and Hyclone DMEM/high glucose media (Thermo #SH300222) to volume. All media was sterile filtered using vacuum filtration. Biopsy samples were minced using sterile supplies into pieces smaller than 1mm³. Tissue pieces were transferred into a sterile dish and washed until clear of blood with sterile PBS. Clean pieces were transferred into a small autoclaved beaker with the collagenase type 1 solution and incubated at 37°C O/N with gentle shaking to stir. Following O/N incubation, the tissue fragments were disrupted with a sterile 1mL pipette and spun to pellet at 820 RPM and 23°C for 5 minutes. Liquid and fat were removed with a sterile pipette and organoids pelleted at the bottom were resuspended in fresh media and plated in T25 flasks with 10 mL culture media. After several days of incubation at 37°C and 5% CO₂ for several days, the cells were adhered and ready for differential trypsinization.

To separate mammary epithelial cells from fibroblasts, differential trypsinization was performed on adhered cells. Media was aspirated and flasks were gently washed with warm culture media. 2-3 mL of trypsin was added and the flasks were incubated at 37°C and 5% CO₂ for 2-3 minutes or until fibroblasts began to lift off. Fresh media was added to

deactivate the trypsin and this layer was removed and transferred to a new culture flask with fresh media. After 2-3 days at 37°C and 5% CO₂, the fibroblasts were ready for subculture.

Subculture: rG3 treatment and cell collection for RNA isolation

Once flasks were confluent, medium was aspirated and each flask was washed with 5-7 mL of warm, sterile PBS. For human cell lines, 3-4 mL of Cell Stripper was added to each flask with incubation at 37°C with 5% CO₂ for 5 min or until most cells are detached. For primary mammary epithelial cells, trypsin was added to form a thin layer on top of the cells with incubation at 37°C with 5% CO₂ for 25-20 minutes or until the cells were detached. The flask was tapped lightly on the counter to loosen cells and the cell suspension was transferred into a 15 mL tube. For human cell lines, sterile PBS was added and for primary mammary cells, sterile media was added to get to 10-12 mL in the tube. Tubes were centrifuged at 820 RPM for 5 min at 23°C. Supernatant was discarded and pellets were washed again with either sterile PBS or media, vortexing to mix. Tubes were centrifuged a second time at 820 RPM for 5 min at 23°C and pellet resuspended in 1mL fresh culture medium. Cells were counted by adding 180µl trypan blue to 20 µl cells in media and loading 20µl of mix onto a hemacytometer. Cells were plated at 1x10⁶ cells per well in a 6-well plate in 1mL fresh media and let adhere overnight at 37°C and 5% CO₂. After overnight adhesion, all wells were treated with either 20µg/mL rG3 diluted in fresh media or equal amounts media for controls. Cells were then collected immediately after addition of rG3 for time 0 followed by 3, 6, 9, 12, and 24 hours after treatment. To collect cells from well, culture medium was removed from the well of interest and well was gently washed with sterile PBS. 400ul cell stripper (for human lines) or trypsin (for primary lines) was added to well and

plates incubated at 37°C and 5% CO₂ for 10 minutes. Cells were added to a sterile 1.5 mL microcentrifuge tube along with 500ul PBS, spun at 12,000 rpm for 2 minutes at room temperature. Supernatant was removed and wash repeated with 500ul PBS and another 2-minute spin at 12,000 rpm. Pellet was resuspended in 100ul RNA later and stored at -20° C for RNA isolation.

Subculture: Cell collection for Western Blotting

Once flasks were confluent, medium was aspirated and each flask was washed with 5-7 mL of warm, sterile PBS. For human cell lines, 3-4 mL of Cell Stripper was added to each flask with incubation at 37°C with 5% CO₂ for 5 min or until most cells are detached. For primary mammary epithelial cells, trypsin was added to form a thin layer on top of the cells with incubation at 37C with 5% Co₂ for 25-20 minutes or until the cells were detached. The flask was tapped lightly on the counter to loosen cells and the cell suspension was transferred into a 15 mL tube. For human cell lines, sterile PBS was added and for primary mammary cells, sterile media was added to get to 10-12 mL in the tube. Tubes were centrifuged at 820 RPM for 5 min at 23°C. Supernatant was discarded and pellets were washed again with either sterile PBS or media, vortexing to mix. Tubes were centrifuged a second time at 820 RPM for 5 min at 23°C and pellet resuspended in 1mL fresh culture medium. Cells were counted by adding 180µl trypan blue to 20 µl cells in media and loading 20µl of mix onto a hemacytometer. Cells were plated at 1x10⁶ cells per well in a 6-well plate in 1mL fresh media and let adhere overnight at 37°C and 5% CO₂. After overnight adhesion, all wells were treated with either 20µg/mL rG3 diluted in fresh media or equal amounts media for controls. Cells were then collected immediately after addition of rG3 for time 0

followed by 3, 6, 9, 12, and 24 hours after treatment. Before each collection, fresh lysis buffer was prepared using 100 µl of cold 10X lysis buffer (Cell Signaling #9803), 890 µl sterile water, and 10 µl of protease inhibitor cocktail (Thermo Scientific #78410) and kept on ice. To collect cells from well, culture medium was removed from the well of interest and well was gently washed with cold sterile PBS. 100 µl of lysis buffer was added directly on the well and plates were incubated on ice for 5 minutes. After incubation, wells were scraped and cell suspension collected in a 1.5 mL microcentrifuge tube. Tubes were spun at 12,000 rpm at 4°C for 15 min after which supernatant was transferred to a fresh tube and stored at -20° C until Western Blotting.

RNA isolation for cells

RNA sequencing will be performed with RNA isolated from the cells using an RNeasy Kit (Qiagen Inc., Valencia, CA) according to manufacturer's instructions. The RNA samples will be treated with an on-column DNase (Qiagen Inc., Valencia, CA) to remove any possible genomic DNA contamination. Isolated RNA samples will be quantified and integrity validated on a Nano Drop 1000 Spectrophotometer (Thermo Scientific, Waltham, MA) and Bioanalyzer 2100 (Agilent Technologies).

RNA seq of cells

Quality control checks, library preparation, and sequencing of RNA collected from cells was performed by Vanderbilt University Medical Center. Total RNA quality was assessed using the 2100 Bioanalyzer (Agilent). At least 200ng of DNase-treated total RNA with a RNA integrity number greater than 6 was used to generate polyA (mRNA) enriched

libraries using TruSeq Stranded mRNA sample kits with indexed adaptors (Illumina). Library quality was assessed using the 2100 Bioanalyzer (Agilent) and libraries were quantitated using KAPA Library Quantification Kits (KAPA Biosystems). Pooled libraries were subjected to 75 bp paired-end sequencing according to the manufacturer's protocol (Illumina HiSeq3000). Bcl2fastq2 Conversion Software (Illumina) was used to generate de-multiplexed Fastq files.

RNA-seq data analysis

To analyze differential gene expression using RNA-sequencing data, three unique and independent methods were utilized: shift change with normalization by control, shift change with normalization by treatment, and statistically significant differences between treated samples and untreated samples assuming normal data distribution and calculating z-scores for each individual gene. Formulas for shift change are below. For shift change with normalization by control, $((\text{Raw count of treated sample} - \text{raw count of control sample}) / (\text{raw count of control sample}))$ was used. For shift change with normalization by treated sample, $((\text{Raw count of treated sample} - \text{raw count of control sample}) / (\text{raw count of treated sample}))$ was used.

RNA seq shift change values between rG3 treated and untreated MDA-MB-231 cells were used to curate pathways using PathVisio and WikiPathways^{435–438}; raw counts for rG3 treated and untreated MDA-MB-231 cells and primary swine fibroblasts were normalized and prepared into heat maps in Genesis⁴²⁸. Blue represents downregulation (negative shift change) or low expression, as noted in figure legends. Red represents upregulation (positive shift change) or high expression, as noted in figure legends. In PathVisio, gray indicates no

shift change; in Genesis heat maps, gray indicates gene or gene ortholog not found in dataset. For comparison of human and swine genes, swine gene orthologs were included if possible using OrthoDB ⁴³⁹.

Western blotting for cells

In preparation for protein separation by SDS-PAGE, cell lysate samples were added to an equal volume of Laemmli Buffer with 5% 2-mercaptoethanol. Samples were mixed and loaded onto a 10-20% Ready Gel Precast Gel (Bio-Rad Laboratories, Inc., Hercules, CA) together with a protein ladder (Precision Plus Protein™ Kaleidoscope™ Prestained Protein Standards #1610375). The gel was run in the Ready Gel System (Bio-Rad Laboratories, Inc., Hercules, CA) with running buffer (15g Tris Base, 72g glycine, 5g SDS per 1L) at 200 V for 30 minutes. Meanwhile, a sheet of Immuno-Blot™ LF-PVDF Membrane and two pieces of 14 x 11 cm filter pads (Bio-Rad Laboratories) were soaked in chilled Western transfer buffer (3.03g Tris Base, 14.4g glycine per 1L) for least 30 minutes. Following electrophoresis, the gel was washed in cold Western transfer buffer for 30 min. The filter pads, gel, and PVDF membrane were arranged in a layered fashion, and protein transfer was completed at 100 V for 45 minutes using a Mini Trans-Blot® Cell (BioRad Laboratories). Non-specific proteins were blocked with 5% non-fat dry milk in TBS at room temperature on a shaker, followed by three washes of 5 minutes each in TBS with 0.5% Tween (Sigma, St. Louis, MO) (TTBS) added. An overnight primary antibody incubation at 4C was performed with p53 (Cell Signaling Technology, Inc.), beta4 integrin (Abcam, Cambridge, MA, Cambridge, MA), and beta actin (Cell Signaling Technology, Inc.) at dilutions according to manufacturer's instructions (Cell Signaling Technology, Inc.). Following primary antibody incubation, a 15

min wash with TTBS was performed to remove unbound antibody. Next, a secondary anti-rabbit IgG HRP-linked antibody (Cell Signaling Technology®, Inc. #7074) was added for 1 hr at RT. Protein detection was obtained through the use of the Pierce ECL Plus Chemifluorescent kit (Thermo Fischer), visualized on the Typhoon FLA 7000 (GE HealthSciences), and quantified using Image Studio Lite (Licor).

Viability Assay

To assay for cellular activity (viability), replicate MTT (3-[4,5-Dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide) assays were performed with purified rG3 and appropriate controls. Different dilutions of rG3 were made in fresh culture medium. One hundred μ l of each concentration was added in triplicate to wells of a treated Falcon 96-well tissue culture plate (Falcon, Franklin Lakes, NJ). All cell lines were collected following the same method described previously. Cells were plated at 1×10^4 per well (in 100 μ l), and the plate was incubated for 24 hr at 37°C in a humidified incubator (5% CO₂). Cells were either treated immediately with rG3, or allowed to adhere for 6-18 hr before treatment. Four hr prior to the end of incubation, 20 μ l of MTT (Sigma-Aldrich Co., St. Louis) was added to all wells. At the end of the 24 hr incubation, the culture supernatant was removed from the wells and 150 μ l of dimethylsulfoxide (DMSO) were added to solubilize cells. The plate was shaken for 15 min and absorbance values for MTT were recorded using a BioTek plate reader at dual wavelengths of 570/640 nm. Optical Density (OD) values of dissolved MTT were used to create a stimulation index (SI), which was calculated using the ratio of treatment OD \div control OD.

Statistical Analysis

Statistical models were developed to relate responses (including stimulation index, differential gene expression, and protein expression) to experimental factors (including rG3 dosage, treatment, and cell type). Analysis of Variance (ANOVA) was used to test significance of the experimental factors on responses. When ANOVA indicated the responses were significantly affected by the experimental factors, Tukey HSD and Student's t tests were used accordingly. All statistical tests used $\alpha = 0.05$ and were performed using the Statistical Analysis System (SAS, Research Triangle Park, NC) and/or JMP Pro 12 software (SAS, Cary, NC).

CHAPTER FOUR: RESULTS AND DISCUSSION

Chapter 1 results and discussion

Mammary tissue sampling procedure causes minimal adverse effects

Growth curves for animals sampled are shown in Fig. 2. Numerical representation for each sample indicates the litter and animal, respectively. For example, a numerical representation such as "1-10" would indicate the tenth animal of the first litter with the assigned number having no significance except as a system to ensure consistency and monitor the animals. Weights were taken before each sampling period weekly, and animals maintained a similar growth curve throughout sampling, gaining a similar amount of weight throughout the study.

Several studies have implicated adverse effects of repeated sampling of the mammary gland^{440,441}, especially when bluntly dissecting the tissue requiring wound closure⁴⁴¹. It is necessary to utilize a sampling technique that is minimally invasive, yet still tissue specific.

The results presented in this figure demonstrate that the animals gained a similar and healthy amount of weight over the course of sampling, and along with histological results below that show tissue specificity, indicate minimal or no effects of the sampling procedure causing adverse effects to health.

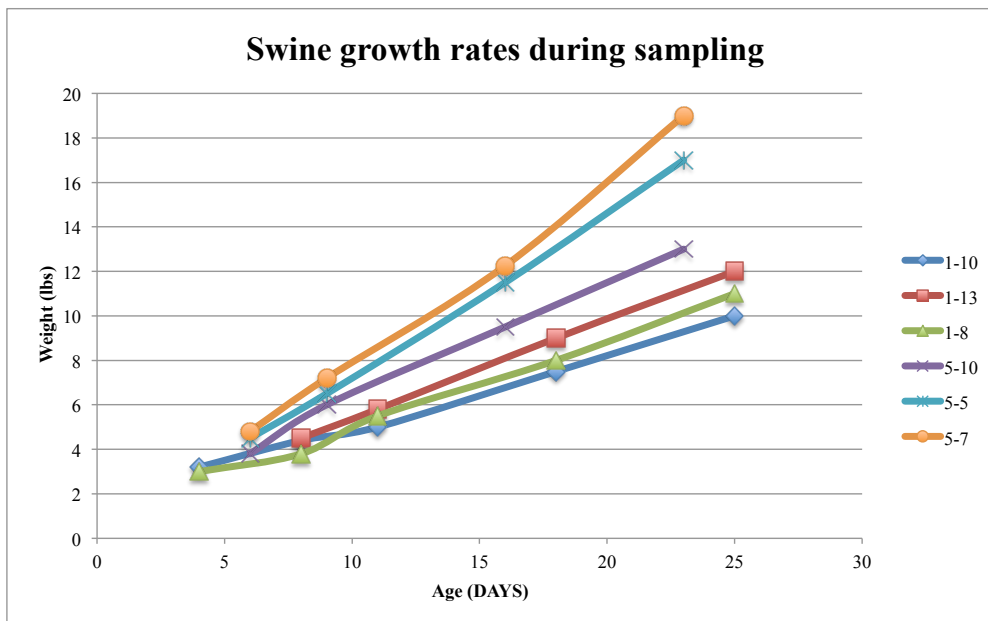


Figure 2. Swine growth rates during sampling demonstrate that the animals gained a similar and healthy amount of weight over the course of sampling. “1-10” indicates litter 1, animal 10. Animals were weighed before each tissue biopsy collection.

Sequential H&E in swine mammary tissue collected over time proves ability of method to study time course of development

Figure 3 represents samples taken from the same pig at 4 days, 8 days, 11 days, 18 days and 32 days, respectively. Samples are formalin fixed, paraffin embedded, and H&E stained to show histological differences in mammary tissue morphology over time. These samples demonstrate normal differentiation of mammary tissue^{442,443} over time with no noted

biasing or alteration of the tissue differentiation due to repeated sampling compared to the first sampling taken. This indicates the fine needle biopsy method is adequate to sequentially monitor mammary growth in a single animal at the cellular level without detriment to the animal or alteration of the morphology or differentiation.

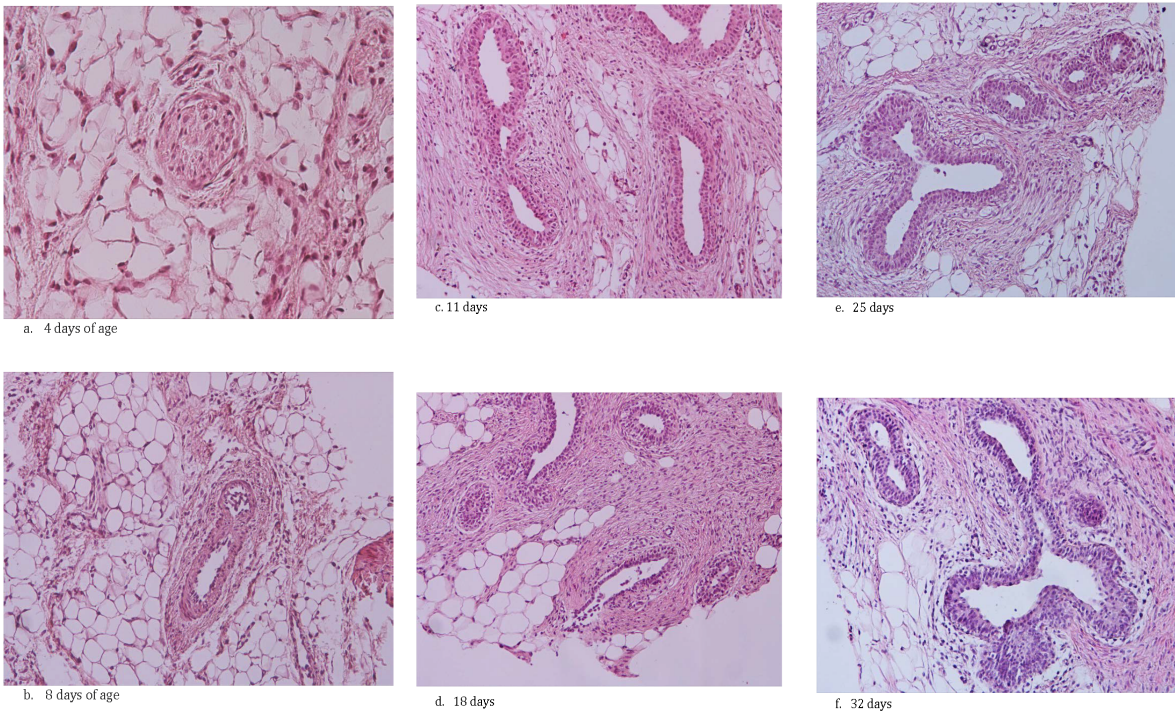


Figure 3: Formalin fixed, paraffin embedded, H&E stained sections of swine mammary tissue at a) 4 days, b) 8 days, c) 11 days, d) 18 days, e) 25 days, and f) 32 days of age demonstrating histological differences in mammary tissue morphology over time. These samples show normal differentiation of mammary tissue with no noted biasing or alteration of the tissue differentiation due to repeated sampling compared to the first sampling taken.

Necropsy and fine needle biopsy comparison utilizing novel development index confirms biopsy method is histologically representative of larger core sampling

The development maturity index can be seen in Fig 1. As illustrated, tissue areas with a developmental index of 0 are a cluster of cells without lumen and 1 ring layer, areas with a

developmental index of 1 present as a cluster of cells without lumen and more than 1 ring layer of cells, areas with a developmental index of 2 present as loosely organized cells with lumen, areas with a developmental index of 3 present as semi-organized cells with lumen, areas with a developmental index of 4 present as dense and organized cells with lumen, and areas with a developmental index of 5 present as dense and organized cells with lumen and branching. This developmental index was used to score mammary tissue sections obtained by both live fine needle biopsy and tissue necropsy post-euthanasia. A contingency analysis was conducted for pigs 19-12 and 20-7, at the age of 4 weeks to confirm that the mammary biopsy procedure produces a tissue sample that is representative of, and no different than, the results obtained by necropsy. After tissue sections collected by both techniques were assigned a numerical developmental index, a contingency analysis was performed. For pig 19-12, a $p \geq 0.3267$ indicates that any differences in the samples are likely due to chance and therefore the null hypothesis is correct and there is no statistical difference in the responses for the two methods. For pig 20-7, a $p \geq 0.4835$ also confirms tissue sections obtained via the fine needle biopsy technique are statistically equivalent to the tissue sections obtained by necropsy. Once confirmed that the sampling method is appropriate, data obtained could be confidently interpreted through RNA sequencing and H&E.

Weighted scores over time can quantify histological trends in mammary tissue collected over time

In an effort to quantify mammary development using histology, several weighted scoring methods were evaluated. These took into account the developmental index score of each tissue area as well as the total count of tissue areas and percent of tissue compared to stroma, connective, and adipose tissue.

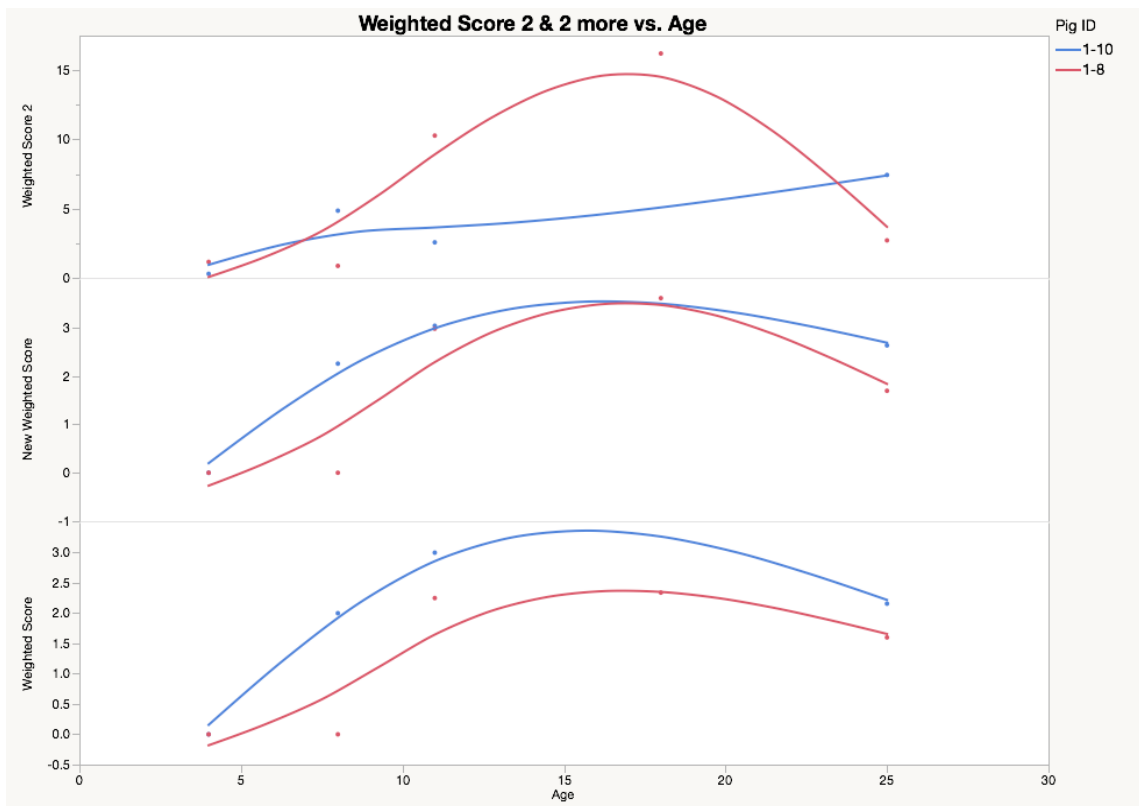


Figure 4: Weighted scores calculated for Pig IDs 1-10 and 1-8 over time. Weighted Score 1: $(\text{Index } 0 * 0 + \text{Index } 1 * 1 + \text{Index } 2 * 2 + \text{Index } 3 * 3 + \text{Index } 4 * 4 + \text{Index } 5 * 5) / (\text{Index } 0 + \text{Index } 1 + \text{Index } 2 + \text{Index } 3 + \text{Index } 4 + \text{Index } 5)$. New Weighted Score: $\text{Weighted Score} * (1 + \text{Total Count for week} / \text{total count for sampling period})$. A split is seen at 18 DOA that can be used for discriminate analysis using Weighted Score 1 and Weighted Score 2. Weighted Score 3 attempted to take into account the percentage of mammary tissue in comparison to adipose and stroma, but did not represent the histological results as well as the first two methods.

Formulas used were:

1. Weighted Score: $(\text{Index 0} * 0 + \text{Index 1} * 1 + \text{Index 2} * 2 + \text{Index 3} * 3 + \text{Index 4} * 4 + \text{Index 5} * 5) / (\text{Index 0} + \text{Index 1} + \text{Index 2} + \text{Index 3} + \text{Index 4} + \text{Index 5})$
2. New Weighted Score: $\text{Weighted Score} * (1 + \text{Total Count for week} / \text{total count for sampling period})$

It proved difficult to quantify consistently over time, but a split is seen at 18 DOA in Fig. 4 that can be used for discriminate analysis using Weighted Score 1 and Weighted Score 2. Weighted Score 3 attempted to take into account the percentage of mammary tissue in comparison to adipose and stroma, but did not represent the histological results as well as the first two methods.

Mammary biopsy technique produces tissue samples with high quality and quantity for downstream RNA isolation and sequencing

Table 2 shows quantity, quality, and integrity values of RNA isolated from swine mammary tissue collected via the fine needle biopsy technique.

Table 2: Quality Control for swine mammary tissue samples submitted for RNA sequencing

Sample labels	Tissue source	Phenotype (age in days)	Other descriptor- Pig ID	Quality (260/280)	Total mass of RNA (ug)	RNA integrity number (RIN)
1	Swine mammary	4.0	1-8	2.11	5.04	10
2	Swine mammary	8.0	1-10	2.15	1.47	10

3	Swine mammary	11.0	1-13	2.12	1.34	10
4	Swine mammary	11.0	1-10	2.16	2.41	10
5	Swine mammary	11.0	1-8	2.15	2.98	9
6	Swine mammary	18.0	1-8	2.13	2.10	9.7
7	Swine mammary	18.0	1-13	2.13	1.57	9.8
8	Swine mammary	18.0	1-10	2.18	0.56	10
9	Swine mammary	25.0	1-13	2.17	1.22	9.9
10	Swine mammary	25.0	1-10	2.29	0.81	10
11	Swine mammary	25.0	1-8	2.13	2.97	8.8
12	Swine mammary	32.0	1-13	2.14	1.85	NA
13	Swine mammary	32.0	1-8	2.13	2.97	9.1
14	Swine mammary	32.0	1-10	2.14	1.85	9.5
15	Swine mammary	39.0	1-13	2.27	0.26	10
16	Swine mammary	39.0	1-8	2.40	1.16	10
17	Swine mammary	39.0	1-10	2.17	0.95	9.9

These high quality and quantity values indicate the fine needle biopsy sampling, method and intermediate processing procedures are adequate to generate a tissue sample enabling accurate library preparation, RNA sequencing, and data analysis. RNA integrity and quality is vital to reproducible RNA sequencing results, and contaminated or degraded mRNA can have a detrimental effect on downstream analysis⁴⁴⁴⁻⁴⁴⁷. RIN is calculated using a software algorithm developed by Schroeder *et al*; values range from 1-10 with 10 possessing the least amount of RNA degradation and being ideal for transcriptome analysis⁴⁴⁸. There was an issue with sample quantity for sample 12, so an RIN could not be established, but based on the NanoDrop values for quality and quantity, it was safe to proceed with library preparation for all samples. Once it was confirmed through the above results that the fine needle biopsy was an appropriate method of sampling, data could be confidently interpreted from RNA sequencing output.

PCA presents large patterns of variability within the gene expression data contributing to specific ordination trends

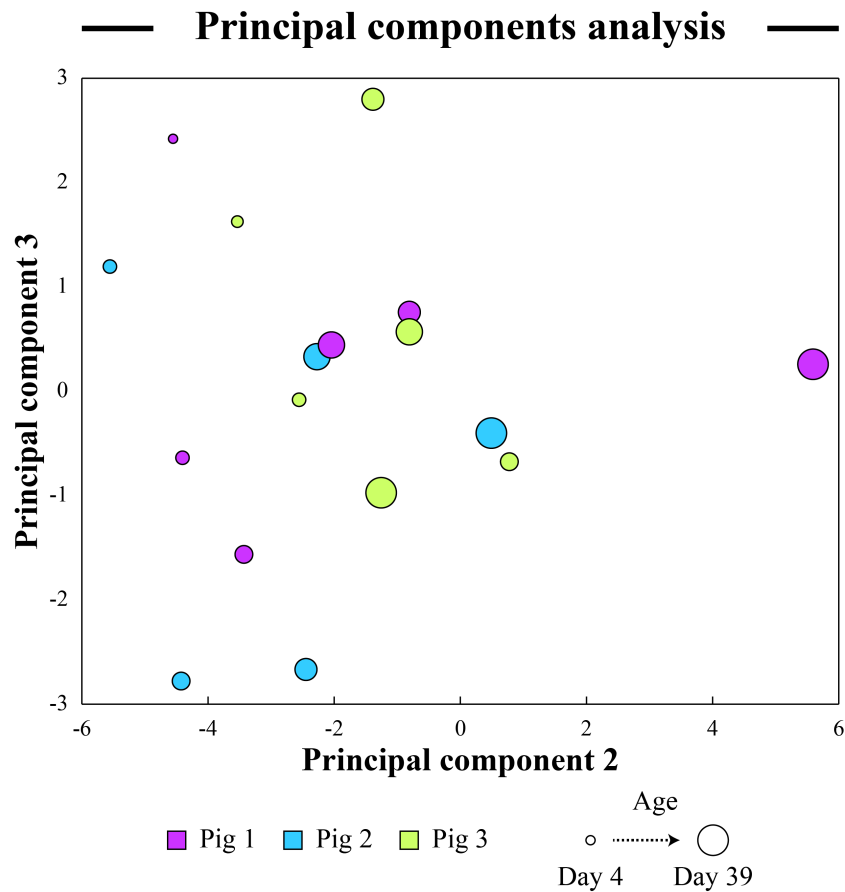


Figure 5: PCA plot for PC2 vs PC3 of developing swine mammary gland transcriptome. Dot color corresponds to Pig ID; size of dot corresponds to age (days). PCA presents large patterns of variability within gene expression data contributing to specific ordination trends.

PCA is an exploratory technique used at the beginning of study to confirm whether there are any trends in data. The color of the dots in Fig. 5 corresponds to pig ID, and size of the dot represents the time at which the biopsy sample was collected (age; 4 days smallest – 39 days largest). As shown in the figure, the younger samples are on the top left corner and as

one moves away toward the right, the sample dot sizes become larger indicating older pigs were sampled.

PCA is commonly used as a simple visualization tool to summarize data variance and present the dominant factors contributing to variance in a large data set ^{174,449,450}. When dealing with multiple variables, as are present in large scale gene expression and transcriptomics studies such as the present, relationships between variables (here, gene expression and age) can be quickly assessed. PCA assumes that the data set contains linear combinations of variables and that large variances have important dynamics, such as contribution to a trend over time; this exploratory technique is appropriate for the large transcriptomic data set ^{450,451}. The first PC axis, PC1, of the PCA output represents the largest gradient of variability in the data set ¹⁷⁴ with the PC2 axis containing the second largest, followed by PC3 until all of the data set variability has been accounted for. Here, variables were displayed in the first three PC dimensions and PC2 and PC3 were found to contain the most important gradients. Fig. 5 shows significant variability due to age of the animal across the PC2 axis and little organization by animal (Pig ID) across the PC3 axis. This suggests animal age is the dominant gradient in the data set that effects the data distribution. It also suggests there is a functional pattern within the variability of the RNA-sequencing data that corresponds to time of sample collection. This supports the hypothesis that there are unique signaling events occurring over time; PCA confirms the transcriptomic data set allows for evaluation of mammary development throughout time without interference from variability between animals used.

RDA confirms that variability in pigs does exist, but does not contaminate variability over time (age)

The color of the dots in Fig. 6 corresponds to pig ID, and size of the dot represents the time at which the biopsy sample was collected (age; 4 days smallest – 39 days largest).

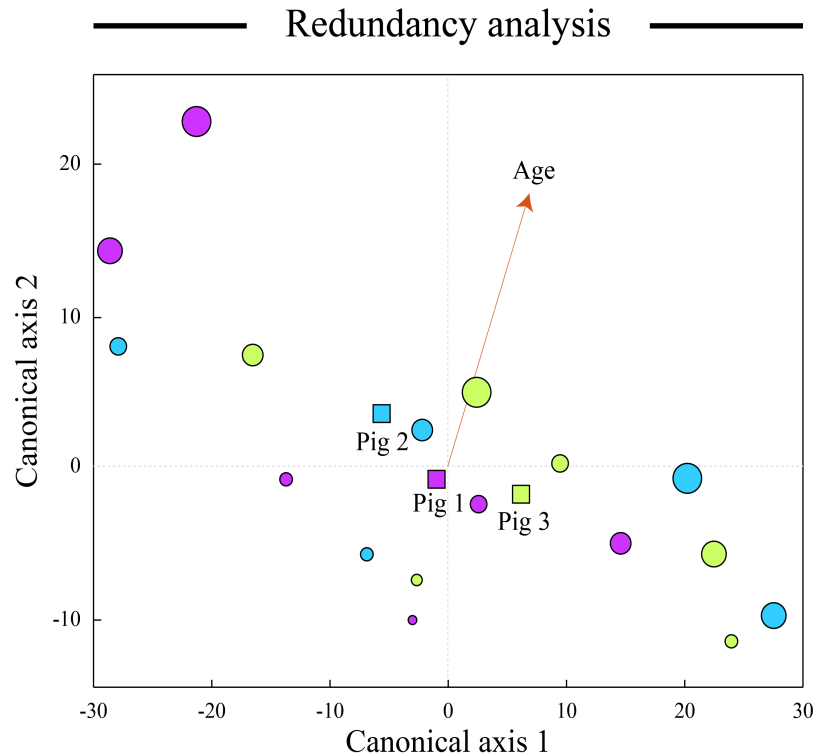


Figure 6: RDA of swine mammary transcriptome. Color of dots corresponds to Pig ID; size of dot corresponds to pig age (size increases with age). RDA shows orthogonal variabilities in the two axes (canonical axis 1 representing Pig ID and canonical axis 2 representing time/age) indicating that the variabilities are independent and very low redundancy is present in the data set.

Redundancy analysis (RDA) is a constrained ordination technique that analyzes how much variation in one set of variables (i.e. pig ID) explains variation in a second set of variables (i.e. time/age).^{174,452–455} In multivariate statistics, RDA is the equivalent of linear regression; it assumes a linear dependence between the response variables on the explanatory variables

and is appropriate for use with the transcriptomic data set. RDA is widely used in community ecology and has been reviewed in depth by Paliy *et al* ¹⁷⁴; it is not currently used to assess trends in mammary development. For the transcriptomic data set collected from different pigs over a period of time, RDA shows orthogonal variabilities in the two axes (canonical axis 1 representing Pig ID and canonical axis 2 representing time/age) indicating that the variabilities are independent and very low redundancy is present in the data set. Therefore, trends in gene expression can be attributed mainly to age with little contamination from variability in Pig ID, providing higher confidence in the dataset.

CLA confirms trends are present in the transcriptomic data set correlating to time

The panels in Fig. 7 show k-means clustering for genes present in the RNA sequencing data set. Two panels are for downward trend (downregulated expression over time; clusters 19 and 13), one panel for upward trend (upregulated expression over time; cluster 16), and two panels are shown that have a periodic or sinusoidal trend and one with no trend or change across time (clusters 2 and 10). The objective of a cluster analysis (CLA) is to separate variables into groups based on the similarity of their scores, generating functional clusters; the variables contained in each cluster are more similar to each other than variables in other groups (minimizing within-group distances and maximizing between-group distances) ¹⁷⁴. Like PCA, DCA, CCA, and RDA, CLA is an exploratory technique to confirm trends in a data set and is becoming a popular choice for use in high-throughput microarray gene expression data, suggesting that it also could be a valuable tool in transcriptomic data sets from RNA-seq. CLA assumes data independence; the RDA results in Fig. 6 suggest CLA is appropriate for use. K-means clustering was utilized in Genesis ⁴²⁸ to group all variables

defined by RNA-seq counts into 20 individual clusters (k=20) based on the different variability patterns over time in the dataset.

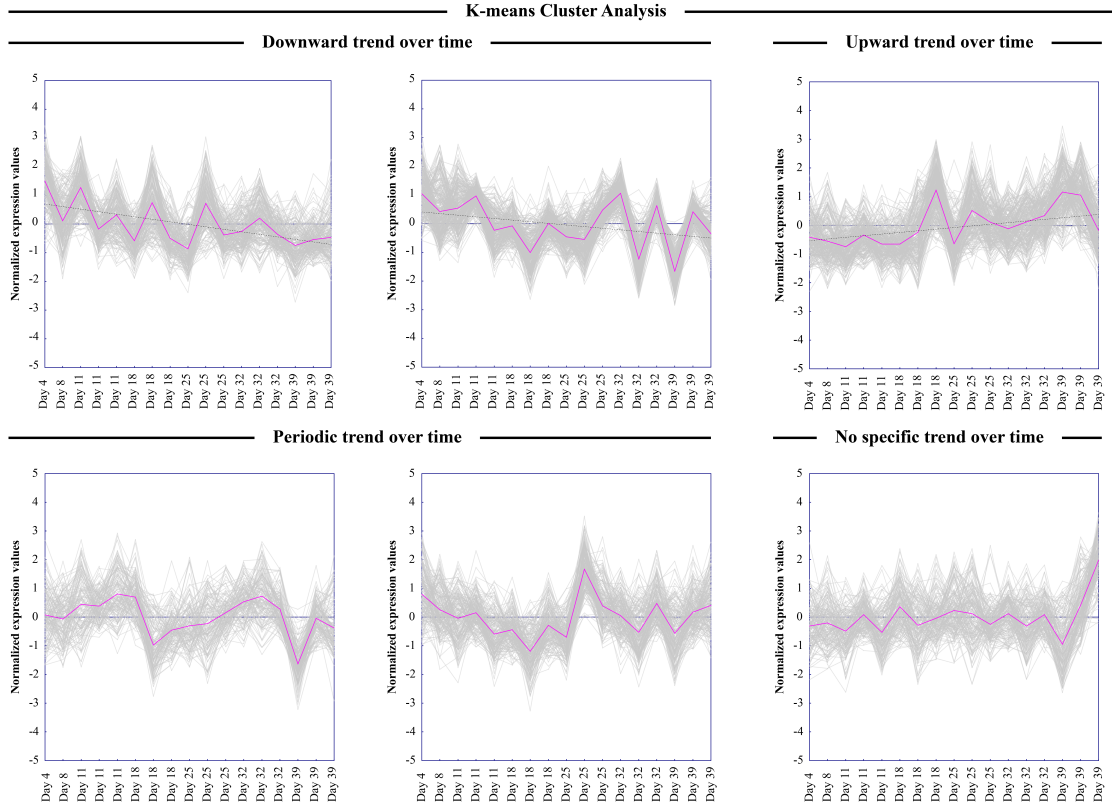


Figure 7: K-means cluster analysis in Genesis. Two panels are for downward trend (downregulated expression over time; clusters 19 and 13), one panel for upward trend (upregulated expression over time; cluster 16), and two panels are shown that have a periodic or sinusoidal trend and one with no trend or change across time (clusters 2 and 10).

In the literature, K-means clustering has been used to: define groups of genera that are modified by the faecal microbiota transplantation (FMT) in patients with *Clostridium* difficile-associated disease¹⁷⁹, reveal molecular cascades underlying adult neurogenesis⁴⁵⁶, dissect the multicellular ecosystem of metastatic melanoma⁴⁵⁷, identify novel gene expression signatures of heart failure⁴⁵⁸, tie macrophage polarization to growth rates of

intracellular bacteria ⁴⁵⁹, and reveal cell type-specific transcriptional signatures at the maternal-fetal interface ⁴⁶⁰. However, most of these studies use single-cell RNA-seq, and k-means clustering not yet been used to analyze developmental gene expression data in the mammary gland. K-means clustering of the RNA-seq data set created clusters of genes containing distinct patterns over time, which can be further dissected into biologically significant functions and pathways.

Clusters containing genes with distinct patterns over time are dispersed into various functional biological categories. Reactome pathway analysis is shown in Table 3 for clusters with downward trend (downregulated expression over time; clusters 19 and 13), upward trend (upregulated expression over time; cluster 16), and periodic or sinusoidal trend and one with no trend or change across time (clusters 2 and 10). K-means clustering of our RNA-seq data set in Genesis generated twenty clusters of genes containing distinct patterns over time, which were then further analyzed for functional significance and pathways using Reactome. All clusters revealed signaling pathways expected to be present in early mammary development, including immune system, metabolism, signal transduction, cell cycle regulation, gene transcription, and protein translation, as shown in Table 3. Interestingly, gene expression contributing to the **upward trend** in cluster 16 utilized signal transduction pathways in conjunction with those of the **immune system**, innate immune system, cytokine signaling, interleukin signaling, and the adaptive immune system. Clusters 13 and 19, displaying downward trends over time, contained genes that utilized mainly metabolic, gene expression, transcription and translation, and cell cycle pathways. Clusters 2 and 10, displaying periodic trends over time, contained genes that utilized a mix of pathways. Clusters containing distinct patterns of gene expression over time indicate that the

transcriptomic data can be partitioned based on specific developmental trends present, many of which contain immune system components, which will be further elucidated with more advanced methods.

Table 3: CLA indicates genes cluster into functional groups with similar trends over time

Pathway Rank	Cluster 2 (periodic)	Cluster 10 (periodic)	Cluster 13 (downward)	Cluster 16 (upward)	Cluster 19 (downward)
1	Metabolism of proteins	Metabolism	Metabolism	Signal Transduction	Metabolism
2	Metabolism	Signal Transduction	Signal Transduction	Immune System	Metabolism of proteins
3	Signal Transduction	Immune System	Cell Cycle	Metabolism	Immune System
4	Cell Cycle, Mitotic	Metabolism of proteins	Gene expression (Transcription)	Innate Immune System	Post-translational protein modification
5	Cell Cycle	Innate Immune System	Metabolism of proteins	Cytokine Signaling in Immune system	Metabolism of lipids
6	Post-translational protein modification	Post-translational protein modification	RNA Polymerase II Transcription	Developmental Biology	Metabolism of RNA
7	Cellular responses to external stimuli	Generic Transcription Pathway	Cell Cycle, Mitotic	Signaling by Interleukins	Disease
8	Metabolism of RNA	RNA Polymerase II Transcription	Generic Transcription Pathway	Axon guidance	Gene expression (Transcription)
9	Gene expression (Transcription)	Gene expression (Transcription)	Immune System	Adaptive Immune System	RNA Polymerase II Transcription
10	Cellular responses to stress	Adaptive Immune System	Transcriptional Regulation by TP53	Metabolism of proteins	Organelle biogenesis and maintenance
11	Immune System	Cytokine Signaling in Immune system	Cell Cycle Checkpoints	Signalling by NGF	Cell Cycle, Mitotic
12	Mitochondrial translation initiation	Developmental Biology	Metabolism of RNA	Hemostasis	Cell Cycle

13	Mitochondrial translation elongation	Transport of small molecules	Metabolism of lipids	Signaling by GPCR	HIV Infection
14	Mitochondrial translation termination	Metabolism of lipids	Post-translational protein modification	Gene expression (Transcription)	Infectious disease
15	Mitochondrial translation	Signaling by Interleukins	M Phase	Membrane Trafficking	Transcriptional Regulation by TP53
16	Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	Adherens junctions interactions	RHO GTPase Effectors	Vesicle-mediated transport	Cytokine Signaling in Immune system
17	The citric acid (TCA) cycle and respiratory electron transport	Collagen degradation	Signaling by Rho GTPases	Signaling by Rho GTPases	Generic Transcription Pathway
18	Translation	Cell-cell junction organization	DNA Repair	Generic Transcription Pathway	Signal Transduction
19	Developmental Biology	Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	Cellular responses to stress	RNA Polymerase II Transcription	HIV Life Cycle
20	Generic Transcription Pathway	Degradation of the extracellular matrix	Cellular responses to external stimuli	Fc epsilon receptor (FCERI) signaling	Translation

Discriminant function analysis of mammary development over time indicates temporal regulation during 4-18 days (phase one) vs. 25-39 (phase two)

Based on histological trends seen in Fig. 4, the RNA-seq dataset was partitioned into two groups representing early postnatal prepubertal (4 days after birth through 18 days after birth, hereafter referred to as phase one) and late postnatal prepubertal development (25 days after birth through 39 days after birth, hereafter referred to as phase two). Discriminant function analysis using three different widely accepted methods showed significant biological differences in the two groups, summarized in Table 4.

The objective of discriminant function analysis (DFA) is to find linear combinations of observed variables that maximize the grouping of objects into separate classes; DFA assumes multivariate normal distribution and homogeneity of variance and covariance⁴⁶¹. Here, three DFA methods were used to isolate two significantly different groups of transcriptomic variables based on a split in time. Early development and late development were chosen and defined as 4-18 days and 25-39 days, respectively (19-24 days fell in between two sampling weeks). In DFA, the numeric weights (discriminant coefficients) can be used to define the relative importance of each predictor variable¹⁷⁴ (in our case, gene) to the observed discrimination between two classes of objects (early phase one vs. late phase two). DFA is appropriate for extremely large data sets and has been successfully employed to: predict clinical outcomes from microarray data⁴⁶², study reprogramming of cancer metabolism⁴⁶³, identify lipodomic biomarkers for NAFLD⁴⁶⁴, compare hepatocellular carcinoma miRNA expression profiles⁴⁶⁵, discriminate gene expression profiles of plants to melatonin⁴⁶⁶, and assess sex differences in the human skeletal muscle transcriptome⁴⁶⁷. However, this method has not been used to study early mammary gland development.

Therefore, three individual DFA methods were utilized to characterize the two phases seen initially in histological analysis: partial least-squares (PLS), support vector machine (SVM), and random forest (RF). PLS assumes linear relationships within the data while SVM and RF can be used with any distribution; the basis of these DFA techniques has been reviewed in depth by Paliy *et al*¹⁷⁴. Here, all three methods were utilized to find the most significant differences in gene expression during early and late phases of postnatal/pre-pubertal development.

Table 4: Common genes identified in mammary transcriptome using three DFA methods

GENE ID	SVM	PLS	RF
COL8A1	SVM	PLS	RF
DTX3L	SVM	PLS	RF
MAP3K14	SVM	PLS	RF
SLC29A4	SVM	PLS	RF
TBC1D8	SVM	PLS	RF
ANP32A	PLS		RF
ARHGAP9	SVM	PLS	
ARL16	SVM		RF
ATP8		PLS	RF
BHLHE40	SVM	PLS	
CDYL2	SVM		RF
DENND5B	SVM	PLS	
GPAT3		PLS	RF
GRB10	SVM	PLS	
HSPA4L		PLS	RF
IF2		PLS	RF
ITIH4	SVM		RF
KCND2		PLS	RF
MAP3K8	SVM	PLS	
NT5DC1		PLS	RF
PAX8	SVM	PLS	
PCGF1	SVM	PLS	
PCGF5	SVM	PLS	
PRUNE2	SVM		RF
RAP1GAP2	SVM	PLS	
RBP5		PLS	RF

RHOBTB2	SVM	PLS	
RNF217	SVM	PLS	
SLC47A2	SVM	PLS	
THPO	SVM	PLS	
VAMP5	SVM		RF

Table 5: Reactome pathway characterization of genes identified using DFA

GENE ID	SVM	PLS	RF	Reactome Pathway
COL8A1	SVM	PLS	RF	Assembly of collagen fibrils and other multimeric structures Collagen biosynthesis and modifying enzymes Collagen formation Extracellular matrix organization
DTX3L	SVM	PLS	RF	Adaptive Immune System Antigen processing- Ubiquitination and Proteasome degradation Class I MHC mediated antigen processing and presentation Immune System
MAP3K14	SVM	PLS	RF	Adaptive Immune System CD28 co-stimulation CD28 dependent PI3K/Akt signaling CLEC7A (Dectin-1) signaling Costimulation by the CD28 family C-type lectin receptors (CLRs) Cytokine Signaling in Immune system Dectin-1 mediated noncanonical NF- κ B signaling Immune System Innate Immune System NIK-->noncanonical NF- κ B signaling TNF receptor superfamily (TNFSF) members mediating non-canonical NF- κ B pathway TNFR2 non-canonical NF- κ B pathway
SLC29A4	SVM	PLS	RF	SLC-mediated transmembrane transport Transmembrane transport of small molecules Transport of nucleosides and free purine and pyrimidine bases across the plasma membrane Transport of vitamins, nucleosides, and related molecules

TBC1D8	SVM	PLS	RF	Unknown
ANP32A	PLS		RF	Gene Expression HuR (ELAVL1) binds and stabilizes mRNA Regulation of mRNA stability by proteins that bind AU-rich elements
ARHGAP9	SVM	PLS		Immune System Innate Immune System Rho GTPase cycle Signal Transduction Signaling by Rho GTPases
ARL16	SVM		RF	
ATP8		PLS	RF	Formation of ATP by chemiosmotic coupling Metabolism Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins. The citric acid (TCA) cycle and respiratory electron transport
BHLHE40	SVM	PLS		BMAL1-CLOCK,NPAS2 activates circadian gene expression Circadian Clock
CDYL2	SVM		RF	
DENND5B	SVM	PLS		Membrane Trafficking Vesicle-mediated transport
GPAT3		PLS	RF	Fatty acid, triacylglycerol, and ketone body metabolism Glycerophospholipid biosynthesis Metabolism Metabolism of lipids and lipoproteins Phospholipid metabolism Synthesis of PA Triglyceride Biosynthesis
GRB10	SVM	PLS		Axon guidance Developmental Biology Insulin receptor signaling cascade IRS activation

				RET signaling Signal attenuation Signal Transduction Signaling by Insulin receptor Signaling by SCF-KIT
HSPA4L		PLS	RF	Cellular response to heat stress Cellular responses to stress Regulation of HSF1-mediated heat shock response
IF2		PLS	RF	Cap-dependent Translation Initiation Eukaryotic Translation Initiation Gene Expression GTP hydrolysis and joining of the 60S ribosomal subunit Metabolism of proteins Translation
ITIH4	SVM		RF	Hemostasis Platelet activation, signaling and aggregation Platelet degranulation Response to elevated platelet cytosolic Ca ²⁺
KCND2		PLS	RF	Cardiac conduction Muscle contraction Neuronal System Phase 1 - inactivation of fast Na ⁺ channels Potassium Channels Voltage gated Potassium channels
MAP3K8	SVM	PLS		Activated TLR4 signaling Adaptive Immune System CD28 co-stimulation CD28 dependent PI3K/Akt signaling Costimulation by the CD28 family Cytokine Signaling in Immune system Immune System Innate Immune System Interleukin-1 signaling MAP kinase activation in TLR cascade MAP3K8 (TPL2)-dependent MAPK1/3 activation MyD88 cascade initiated on plasma membrane MyD88 dependent cascade initiated on endosome MyD88-independent TLR3/TLR4 cascade MyD88-Mal cascade initiated on plasma membrane Signaling by Interleukins Toll Like Receptor 10 (TLR10) Cascade

				Toll Like Receptor 2 (TLR2) Cascade Toll Like Receptor 3 (TLR3) Cascade Toll Like Receptor 4 (TLR4) Cascade Toll Like Receptor 5 (TLR5) Cascade Toll Like Receptor 7/8 (TLR7/8) Cascade Toll Like Receptor 9 (TLR9) Cascade Toll Like Receptor TLR1-TLR2 Cascade Toll Like Receptor TLR6-TLR2 Cascade Toll-Like Receptors Cascades TRAF6 mediated induction of NFkB and MAP kinases upon TLR7/8 or 9 activation
NT5DC1		PLS	RF	
PAX8	SVM	PLS		
PCGF1	SVM	PLS		Developmental biology
PCGF5	SVM	PLS		Developmental biology
PRUNE2	SVM		RF	
RAP1GAP2	SVM	PLS		Immune System;Rap1 signaling; Adaptive Immune System
RBP5		PLS	RF	
RHOBTB2	SVM	PLS		Signal Transduction; Signaling by Rho GTPases;Rho GTPase cycle
RNF217	SVM	PLS		Adaptive Immune System Antigen processing- Ubiquitination and Proteasome degradation Class I MHC mediated antigen processing and presentation

				Immune System
SLC47A2	SVM	PLS		SLC-mediated transmembrane transport Transmembrane transport of small molecules Transport of glucose and other sugars, bile salts and organic acids, metal ions and amine compounds
THPO	SVM	PLS		Hemostasis; Platelet activation, signaling and aggregation; Platelet Aggregation (Plug Formation)
VAMP5	SVM		RF	

Heatmap visualization of top discriminant genes

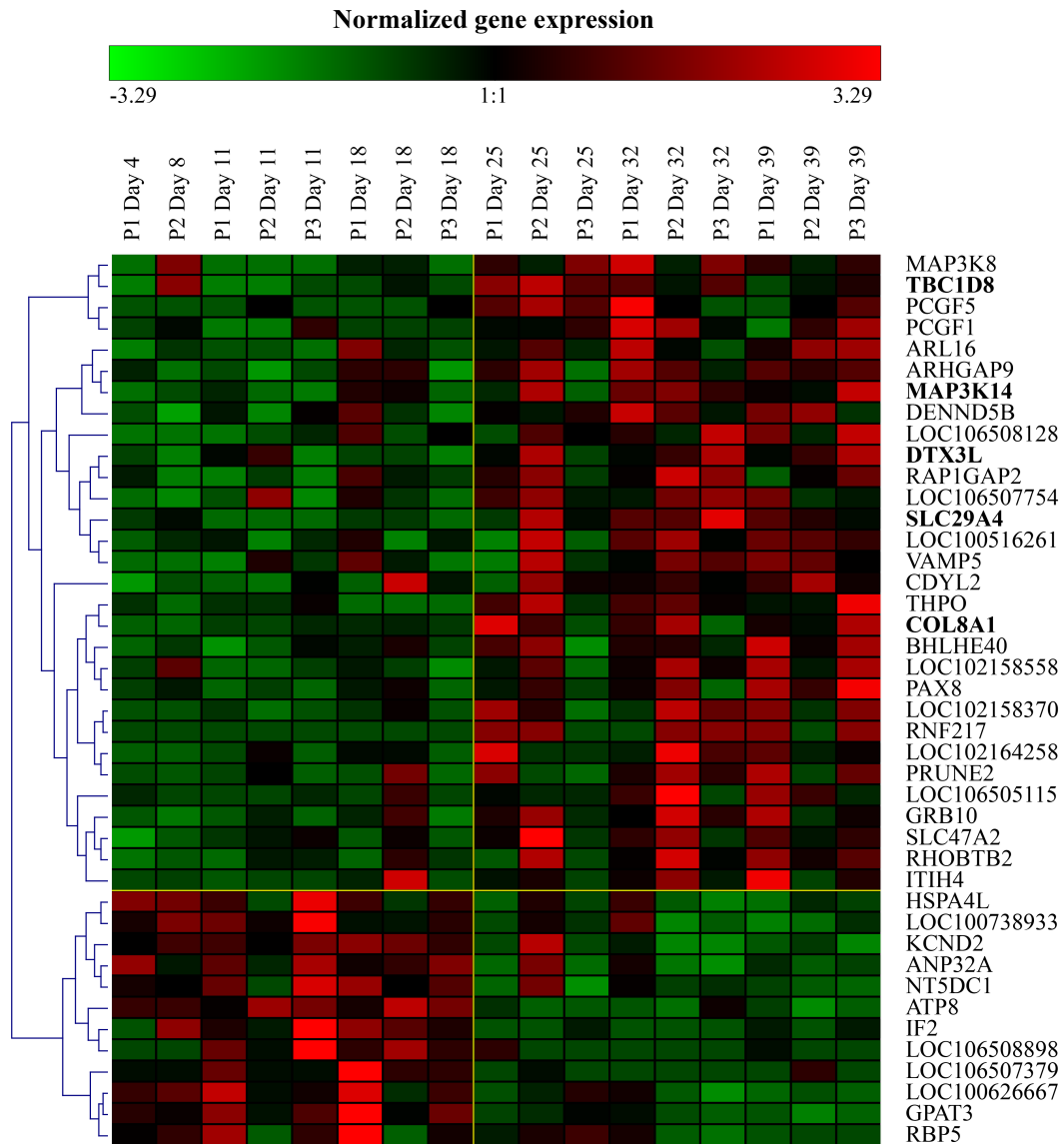


Figure 8: DFA of mammary development over time indicates temporal regulation during 4-18 days of age vs 25-39 days of age: heatmap visualization performed in Genesis of top discriminant genes. Genes in bold were identified using all three DFA methods; others were identified using two out the three methods.

Table 6: Gene name reference for genes identified during DFA

GENE ID	SVM	PLS	RF	Gene Name, Gene Symbol, Ortholog
COL8A1	SVM	PLS	RF	Collagen α -1(VIII) chain;COL8A1;ortholog
DTX3L	SVM	PLS	RF	E3 ubiquitin-protein ligase DTX3L;DTX3L;ortholog
MAP3K14	SVM	PLS	RF	Mitogen-activated protein kinase kinase kinase 14;MAP3K14;ortholog
SLC29A4	SVM	PLS	RF	Equilibrative nucleoside transporter 4;SLC29A4;ortholog
TBC1D8	SVM	PLS	RF	TBC1 domain family member 8;TBC1D8;ortholog
ANP32A	PLS		RF	Acidic leucine-rich nuclear phosphoprotein 32 family member A;ANP32A;ortholog
ARHGAP9	SVM	PLS		Rho GTPase-activating protein 9;ARHGAP9;ortholog
ARL16	SVM		RF	ADP-ribosylation factor-like protein 16;ARL16;ortholog
ATP8		PLS	RF	ATP synthase protein 8;MT-ATP8;ortholog
BHLHE40	SVM	PLS		Class E basic helix-loop-helix protein 40;BHLHE40;ortholog
CDYL2	SVM		RF	Chromodomain Y-like protein 2;CDYL2;ortholog
DENND5B	SVM	PLS		DENN domain-containing protein 5B;DENND5B;ortholog
GPAT3		PLS	RF	Glycerol-3-phosphate acyltransferase 3;AGPAT9;ortholog
GRB10	SVM	PLS		Growth factor receptor-bound protein 10;GRB10;ortholog
HSPA4L		PLS	RF	Heat shock 70 kDa protein 4L;HSPA4L;ortholog
IF2		PLS	RF	Eukaryotic translation initiation factor 5B;EIF5B;ortholog
ITIH4	SVM		RF	Inter- α -trypsin inhibitor heavy chain H4;ITIH4;ortholog
KCND2		PLS	RF	Potassium voltage-gated channel subfamily D member 2;KCND2;ortholog
MAP3K8	SVM	PLS		Mitogen-activated protein kinase kinase kinase 8;MAP3K8;ortholog
NT5DC1		PLS	RF	5'-nucleotidase domain-containing protein 1;NT5DC1;ortholog
PAX8	SVM	PLS		Paired box protein Pax-8;PAX8;ortholog
PCGF1	SVM	PLS		Polycomb group RING finger protein 1;PCGF1;ortholog
PCGF5	SVM	PLS		Polycomb group RING finger protein 5;PCGF5;ortholog
PRUNE2	SVM		RF	Protein prune homolog 2;PRUNE2;ortholog

RAP1GAP2	SVM	PLS		Rap1 GTPase-activating protein 2;RAP1GAP2;ortholog
RBP5		PLS	RF	Retinol-binding protein 5;RBP5;ortholog
RHOBTB2	SVM	PLS		Rho-related BTB domain-containing protein 2;RHOBTB2;ortholog
RNF217	SVM	PLS		Probable E3 ubiquitin-protein ligase RNF217;RNF217;ortholog
SLC47A2	SVM	PLS		Multidrug and toxin extrusion protein 2;SLC47A2;ortholog
THPO	SVM	PLS		Thrombopoietin;THPO;ortholog
VAMP5	SVM		RF	Vesicle-associated membrane protein 5;VAMP5;ortholog

Discriminate analysis of gene expression during early (Day 4 – Day 18) and later (Day 25 – Day 29) postnatal mammary development reveals temporal regulation of important signaling events

Discriminate analysis between the mammary transcriptome during 4-18 days of age (DOA) (phase one) and 25-39 DOA (phase two) implicated important biological differences that are temporally regulated during early postnatal mammary development and support the two phases suggested by histology (see histology Fig. 3). As shown in Table 4, five known genes were found using all three discriminate analysis methods (SVM, PLS, and RF): COL8A1, DTX3L, MAP3K14, SLC29A4, and TBC1D8. 26 known genes were found in two out of the three discriminate analysis methods: ANP32A, ARHGAP9, ARL16, ATP8, BHLHE40, CDYL2, DENND5B, GPAT3, GRB10, HSPA4L, IF2, ITIH4, KCND2, MAP3K8, NT5DC1, PAX8, PCGF1, PCGF5, PRUNE2, RAP1GAP2, RBP5, RHOBTB2, RNF217, SLC47A2, THPO, and VAMP5. Genes found using all three or two out of the three methods were deemed significant and will be discussed along with their upregulation or

downregulation during phase one (and respectively, downregulation and upregulation during the phase two).

Discriminate analysis data was assembled into a heat map using hierarchical clustering in Genesis⁴²⁸. As shown in Fig. 8., genes identified using all three discriminate methods are bolded; generation using three separate methods indicates the highest significance of differential gene expression between the two phases. Differential expression of all five highly significant genes was **upregulated during phase two of postnatal mammary gland development** in comparison to their downregulated expression immediately after birth in phase one. Importantly, the large majority (75%) of discriminately expressed genes identified with all three DA methods were expressed highly during phase two with significantly lower expression during phase one. Overall, common Reactome pathways utilized by the top 31 known differentially expressed genes were (from most often to least): Immune System, Adaptive Immune System, Signal Transduction, SLC-mediated transmembrane transport, Transport of small molecules, Innate Immune System, CD28 dependent PI3K/Akt signaling, CD28 co-stimulation, Rho GTPase cycle, Costimulation by the CD28 family, BMAL1:CLOCK,NPAS2 activates circadian gene expression, Transport of glucose and other sugars, bile salts and organic acids, metal ions and amine compounds, Antigen processing: Ubiquitination & Proteasome degradation, Circadian Clock, Class I MHC mediated antigen processing & presentation, Signaling by Rho GTPases, and Cytokine Signaling in Immune system. Greater than 50% of the pathways utilized by differentially expressed genes were direct components of the immune system, indicating significant impact of both innate and adaptive immune system factors in regulation of postnatal developmental events. The immune system's involvement in prenatal and postnatal mammary development

has been gaining acceptance ^{11,84–88}, and this data further confirms significant involvement, specifically suggesting critical roles in temporal regulation during the first two months after birth. Although the importance of immune cells in mammary gland development is well known ^{34,35}, the molecular signaling that guides them, the roles of chemokines, and the importance of the adaptive immune system remain unexplored ^{86,114}.

Pathway analysis was performed on each phase of discriminately expressed genes separately. Pathways utilized by the smaller subset of differentially expressed genes that are **upregulated during phase one** were (from most often to least): signal transduction, metabolism of proteins, lipids, and RNA, and organelle biogenesis. CLA analysis using k-means clustering in Fig. and Table 5 also indicated that gene expression contributing to the downward trends over time in Clusters 13 and 19 utilized mainly metabolic, gene expression, transcription and translation, and cell cycle pathways. In stark contrast, the larger subset of differentially expressed genes that were **upregulated during phase two** utilized mainly immune system pathways, including (from most often to least): immune system, adaptive immune system, innate immune system, CD28 dependent PI3K/AKT signaling, CD28 co-stimulation, co-stimulation by the CD28 family, Class I MHC mediated antigen processing and presentation, platelet activation, signaling, and aggregation, and cytokine signaling in immune system. These immune system pathways were **not** found in differentially expressed genes of phase one. Pathway analysis on the differentially expressed genes during early mammary development indicates that the adaptive immune system is highly utilized beginning in the period of 18-25 days after birth and remains active until at least 39 days after birth, during which it contributes to developmental trends. The adaptive immune system pathway is only preceded in use by differentially expressed genes by the generic immune

system category, which includes both adaptive and innate components. This is especially novel because adaptive immune roles in mammary development are currently unexplored. CLA analysis using k-means clustering in Fig. 7 and Table 3 also indicated that gene expression contributing to the **upward trend** in cluster 16 utilized signal transduction pathways in conjunction with those of the **immune system**, innate immune system, cytokine signaling, interleukin signaling, and the adaptive immune system, further supporting the notion that the immune system's role in early mammary development does increase over time during the period of 4-19 days after birth. Taken together, these results do indicate significant involvement and regulation of innate and adaptive immune system genes and components in early postnatal, but prepubertal, mammary gland development. Genes of functional significance will be discussed individually below.

Differentially expressed genes that are upregulated during Phase one: Day 4- Day 18 after birth

Differentially expressed genes that are suspected, as a result of DFA analysis, to hold important roles in mammary development will be individually discussed, beginning with the small subset that are **upregulated during phase one (from 4-18 days after birth) and expressed at lower levels from days 25-39**. As mentioned, pathway analysis initially indicated that these genes mainly utilize signal transduction and metabolic pathways and are not present in innate or adaptive immune system pathways. Still, several functionally significant genes are present in phase one. HSPA4L, Heat shock 70 kDa protein 4L, encodes a heat shock inducible chaperone protein that is highly expressed in leukemia cells⁴⁶⁸⁻⁴⁷². Current evidence supports a role of HSPA4L in normal spermatogenesis⁴⁷⁰, but little to no

data exist explaining its function in normal mammary development. Results suggest that this chaperone protein may play a role in very early protection of developing mammary cells from heat stress and aids in osmotolerance. KCND2, Potassium voltage-gated channel subfamily D member 2, encodes a voltage-gated potassium channel protein that is known to mediate transmembrane potassium transport in excitable membranes⁴⁷². Since voltage-gated potassium (Kv) possesses diverse functions including regulation of neurotransmitter release, heart rate, insulin secretion, neuronal excitability, epithelial electrolyte transport, smooth muscle contraction, and cell volume⁴⁷³, it is possible that this gene is upregulated during phase one in order to immediately regulate electrolyte transport of mammary epithelial cells after birth. ANP32A, or Acidic leucine-rich nuclear phosphoprotein 32 family member A, encodes a protein implicated in a number of cellular processes, including proliferation, differentiation, caspase-dependent and caspase-independent apoptosis, suppression of transformation (via acting as a tumor suppressor) and is aberrantly expressed in multiple types of cancers^{474–477}. ANP32A is a member of the ANP32 family, that contains a large group of proteins implicated in a wide range of functions: modulation of cell signaling, transduction of gene expression, hormone-receptor interactions, cell adhesion, cellular trafficking, early mammalian development, apoptosis signaling, and regulation of cytoskeleton morphology^{478,479}. The high expression of ANP32A during phase one in contrast to its downregulation in phase two suggests it supports one or more of these functions in very early postnatal mammary gland development. The function of NT5DC1, 5'-nucleotidase domain-containing protein 1, remains unknown⁴⁷² and to our knowledge no data exist to elucidate its role in early mammary development; this data suggest a role in the first few weeks after birth. ATP8, ATP synthase protein 8, encodes a protein that functions as

a subunit of mitochondrial ATP synthase⁴⁸⁰. ATP8 supports normal mitochondrial performance⁴⁸¹, therefore it is possible that this differentially expressed gene supports mitochondrial function in young mammary cells of phase one. IF2, or Eukaryotic translation initiation factor 5B, is a general transcription factor that supports initiation of translation by promotion of 60S ribosome subunit joining and pre-40s subunit proofreading^{482–486}. Recently, Lee et al uncovered a critical role for IF2 in specific developmental stages, including those of immature oocytes and ES cells, and during growth-factor deprivation of mammalian cells, inducing the transition to cell-cycle arrest⁴⁸³. Importantly, these researchers demonstrate that IF2/eIF5B controls and regulates cell-cycle transition and developmental stages⁴⁸³, which is likely akin to its function in early mammary development. Taken together, the subset of genes that are highly differentially expressed during phase one of postnatal mammary development likely promote signaling pathways of cellular biogenesis, metabolism, epithelial and stromal differentiation, and widespread expansion of the gland. Interestingly, functional characterization of the small subset of DFA genes found upregulated in phase one immediately after birth suggests that this time period of early mammary development is one of generic growth rather than advanced differentiation. The presence of genes aiding in early protection of developing mammary cells from heat stress, osmotolerance, regulation of electrolyte transport of mammary epithelial cells, modulation of cell signaling, transduction of gene expression, hormone-receptor interactions, cell adhesion, cellular trafficking, mitochondrial function, and cell-cycle transition provide a model for widespread growth focused on cellular activity in phase one, rather than differentiation with support from stromal and immune factors that is seen in phase two. It may be that immediately after birth, the mammary gland requires three to four weeks to adjust to growth

and expansion postnatally, after which advanced differentiation and maturation of the organ begins.

Differentially expressed genes that are upregulated during Phase two: Days 25-39 after birth

Next, the major subset of differentially expressed genes (those that are downregulated phase one and upregulated during later phase two) will be discussed, beginning with the five significant genes identified using all three DA methods. COL81A encodes type VIII collagen, a matrix protein widely expressed during tissue remodeling, wound repair, and cell migration; little is known about its functional mechanism, especially in mammary development^{472,487–496}. As discussed previously, postnatal mammary epithelial cells are tightly connected to both each other and surrounding stroma, relying on this both collagen and this stable adhesion to transmit biochemical signals, promote cell migration, and function normally^{7,41,82}. Cell migration is a carefully coordinated process during which attachment to matrices and subsequent release for translocation are regulated by MMPs, forming a physical path through the ECM⁴⁹⁷. This process is also commonly mediated by integrin and laminin molecules: Ln-5 can be degraded by proteases (including MMP-2 and MMP-9) during tissue remodeling to promote cell migration³⁹⁴. When cleaved during proteolytic processing, the G4/G5 fragment of Ln-5 that is released from the G3 domain also functions to support cell migration³⁹⁷. Hou *et al* investigated smooth muscle cell (SMC) migration via type VIII collagen and concluded its effects are mediated by $\alpha 2\beta 1$ and $\alpha 1\beta 1$ integrin receptors as well as MMP-2 and MMP-9 expression. This demonstrates a critical role for type VIII collagen in cellular invasion and migration⁴⁸⁸. As reviewed in Chapter 1, MMPs secreted by macrophages function in ECM remodeling to allow ductal branching into the mammary fat

pad^{14,103–107}, so it is not surprising that MMPS coordinate with type VIII collagen in this suggested model. During tumorigenesis, reduction of COL8A1 expression leads to decreased invasion of hepatocarcinoma cells⁴⁹⁸, indicating that type VIII collagen has a functional role in *both* cancer development and the migration and adhesion required for normal development. However, little data exist to elucidate the role of type VIII collagen in early postnatal mammary development or TNBC. This data show that COL8A1 is significantly upregulated during phase two of early postnatal mammary development, and taken together with the accepted role in general tissue remodeling, wound repair, cell migration, and integrin molecule interaction, it is hypothesized that type VIII collagen has a functional role in early postnatal mammary development and epithelial cell signaling in conjunction with MMPS and other immune system components.

The Deltex (DTX)-3-like E3 ubiquitin ligase (DTX3L), also known as B-lymphoma and BAL-associated protein (BBAP), is overexpressed in many high-risk, chemo-resistant, diffuse large B cell lymphoma (HR-DLBCL) subtypes⁴⁹⁹, is associated with both IFN-gamma and STAT1 signaling^{499–502}, and functions as a binding partner to ADP-ribosyltransferase-9 (ARTD9) (also known as B-aggressive lymphoma protein (BAL1))^{499,502}. These genes have also been found overexpressed in myeloma, Ewing tumors, head and neck cancers, and cervical cancer^{502–507}. Functionally, DTX3L responds to cellular exposure of DNA damaging agents and its inhibition is thought to increase the efficacy of DNA-targeting chemotherapeutic agents, but the mechanism utilized in the response is unclear⁵⁰⁸. In melanoma cells, DTX3L has been shown to mediate proliferation, invasion, and metastasis through FAK/PI3K/AKT signaling⁵⁰⁹, suggesting a similar role for this gene during tissue remodeling in the early mammary gland during phase two while upregulated because studies

investigating its function in mammary development do not exist.

The MAP3K14 gene, also known as NF- κ B-inducing kinase (NIK), encodes a serine/threonine protein-kinase that promotes activity of the noncanonical NF- κ B pathway to regulate cell proliferation and survival both normally and in tumorigenesis^{510,511}. NF- κ B is a well-characterized transcription factor family that influences cellular immune responses, survival, proliferation, angiogenesis, and metastasis⁵¹². NIK regulation occurs at the post-translational level through the translocon associated protein complex (TRAF), a cellular inhibitor of apoptosis⁵¹³, promoting differentiation, development, embryogenesis, and tumor cell invasion through phosphorylation of either IKK α and IKK β subunits^{514–519}. Additionally, Osman *et al* showed that both NIK and AKT are necessary for activation of NF- κ B, suggesting involvement of NIK in the PI3K/AKT pathway⁵²⁰. Importantly, NIK is overexpressed in basal breast cancer cell lines, leading to constitutive activation and uncontrolled tumor proliferation^{521–523}. NIK overexpression also increases the activity of breast cancer stem cells *in vitro* through the ERK pathway and is correlated with ALDH overexpression in a tissue microarray of *in vivo* breast cancer samples, implicating NIK in regulation of the breast cancer stem cell phenotype⁵²⁴. Since NIK is essential for the canonical NF- κ B pathway activation normally, inhibition of NIK could be a better therapeutic strategy in cancer. It is clear that NIK is involved in regulation of breast cancer stem cells, tumor invasion, and basal cell expansion, but the mechanism remaining poorly understood⁵²⁴. This data identifies NIK as a highly significantly differentially expressed gene that is upregulated during phase two of early postnatal mammary development, suggesting NIK functions to promote differentiation and proliferation of mammary epithelial cells

during this period in the immature gland and also may regulate growth of the mammary gland through apoptotic cell signaling pathways.

TBC1D8, TBC1 domain family member 8, which may act as a GTPase-activating protein for Rab family protein(s)⁴⁷² and SLC29A4, Equilibrative nucleoside transporter 4, which functions as a polyspecific organic cation transporter⁴⁷², have lesser characterized functions in development and to our knowledge data does not exist to elucidate their function. Since these two genes are highly upregulated during phase two of postnatal mammary development, our data suggest a role for them in this area.

Taken together, the presence of COL81A, DTX3L, MAP3K14 in the top subset of genes upregulated during phase two suggest contribution of the ECM, immune system, and tissue remodeling to differentiation and proliferation of mammary epithelial cells during this period in the immature gland. Importantly, in contrast with the generic cell growth model provided by discriminately expressed phase one genes, phase two appears to contain more advanced regulatory and differentiation activities along with the presence of stromal influence. Many of the phase two genes have functions in tissue remodeling, wound repair, and cell migration while relying on stable adhesions. It is known that cell migration is a carefully coordinated process, and this data suggests growth due to migration and differentiation begins in phase two rather than phase one, in part due to regulation by immune system components such as MMPs. Importantly, DFA analysis suggests that during phase two, proliferation, invasion, and maturation may begin to rely more heavily on FAK/PI3K/AKT signaling.

In addition to genes identified using all three discriminant analysis methods, several functionally important genes were generated by two of the three methods, which will be discussed individually, including: ARHGAP9, BHLHE40, GRB10, ITIH4, MAP3K8, PAX8, PCGF1, PCGF5, PRUNE2, RNF217, THPO. All of these genes were upregulated during phase two of early postnatal mammary development.

ARHGAP9 encodes Rho GTPase-activating protein 9, a G-protein modulator that is a member of the Rho-GAP family of GTPase activating proteins that has been implicated in regulating adhesion of hematopoietic cells to the extracellular matrix ⁵²⁵ and interacting with Erk2 and p38 α to sequester the MAP kinases in their inactive states ⁵²⁶, which is a novel mechanism of cross-talk between Rho GTPase and MAP kinase signaling. Given that matrix adhesion is a critical activity for cell migration and mammary development, it is hypothesized that this gene has a similar adhesion function in phase two of early postnatal mammary development.

BHLHE40 encodes Class E basic helix-loop-helix protein 40, which is a transcriptional repressor believed to be involved in control of circadian rhythm (through negative regulation of clock and clock-controlled genes), cell differentiation, proliferation, and apoptosis ^{527–533}. Also commonly known as Differentiated embryo-chondrocyte expressed gene 1 (DEC1), this gene has been shown to assist BHLHE41 in the suppression of EMT effectors and tumor cell invasion *in vitro* ⁵²⁹ and its loss may promote tumor progression in through upregulation of cyclin D1 ⁵³⁴. Conflictingly, data suggest its presence also promotes the invasiveness of breast cancer through downregulation of claudin-1 ⁵³¹. Overexpression of this gene has been reported to contribute to the cellular differentiation, proliferation, and apoptosis of various cancers ^{527–538} although the mechanism is poorly

understood. This data suggests a significant role for BHLHE40 during phase two, which could consist of either promotion of cellular differentiation and proliferation (positive role) or cell cycle regulation (negative/regulatory role). Given that cell cycle regulation was a hallmark of phase one genes, it is possible BHLHE41 in phase two contributes to differentiation of mammary epithelial cells, a focus of this larger subset of DFA results.

GRB10, Growth factor receptor-bound protein 10, is a common adapter protein which modulates interaction of several surface receptor kinases by interfering with downstream signaling pathways and receptor degradation⁴⁷². GRB10 is widely accepted to influence cell proliferation, apoptosis, and metabolism in varied cell types^{539–541}. It can bind to, and suppress signals from, activated RTKs including the insulin (INSR)⁵⁴² and insulin-like growth factor (IGF1R) receptors⁵⁴³ via delay and reduction of AKT1 phosphorylation. GRB10 has a widespread functional impact in both development and cancers- it has been shown to bind to the intracellular portion of Low-density lipoprotein receptor-related protein 6 (LRP6) and negatively influence Wnt signaling, as a negative regulator of the pathway⁵⁴⁴. Loss of GRB10 can also induce hyperproliferation and tumor development⁵⁴⁵ and its presence has also been found to negatively correlate with lactation performance⁵⁴⁶, illustrating its wide range of function. Interestingly, the phenotype resulting from GRB10 dysfunction is dependent on the parental origin of the mutation, a phenomenon known as genomic imprinting⁵⁴⁷. Genomic imprinting, reviewed in depth⁵⁴⁸, can influence developmental programming that links growth in early stages of life to overall health at adulthood. Expression of GRB10 is gained from the maternal chromosome in most tissues and presents a complex and tissue-specific expression pattern⁵⁴⁷. Data has identified imprinted GRB10 as a mediator of nutrient supply and demand in the postnatal period and a

negative regulator of growth-related pathways and insulin signaling in development ⁵⁴⁷; reduction in GRB10 expression results in widespread postnatal overgrowth and insulin sensitivity ^{549–552}. This data suggests that GRB10 is upregulated during phase two in order to perform similar negative regulatory roles in early postnatal mammary development.

ITIH4, Inter- α -trypsin inhibitor heavy chain H4, is a type II acute phase protein that is involved in the inflammatory trauma response ^{472,553}. Predominantly expressed in liver tissue, mRNA expression levels of this protein have been seen upregulated in response to administration of IL-6 ⁵⁵³, but data regarding the involvement of this protein in mammary development does not exist. This data indicates upregulation of ITIH4 during phase two, where it could be involved in the developing epithelial cells and the immune system in response to cytokines present in the postnatal mammary gland.

MAP3K8, mitogen-activated protein kinase kinase kinase 8, is a potent oncogene that encodes a member of the serine/threonine protein kinase family ⁴⁷². Widely expressed in skeletal muscle and the immune system ⁵⁵⁴, MAP3K8 is a tumor suppressor that can activate the ERK, JNK, and p38 MAPK pathways ⁵⁵⁵, support IL-12 production and Th cell differentiation ⁵⁵⁶, and possesses a critical role in promotion of cancer-associated inflammation in the TME when induced by IL-22 ⁵⁵⁷. IL-22 signaling can promote aggressiveness in human breast cancer through the induction of MAP3K8, leading to epithelial transformation; knockdown of MAP3K8 decreases tumorigenicity in MCF-7 cells ⁵⁵⁷. It is well-known that the MAP3K8 pathway components can have both tumorigenic or apoptotic functions that are dependent on the stimuli and the other stromal cells involved ⁵⁵⁸; its involvement in normal mammary development can either be hypothesized as regulatory or promotional or both, depending on the time period. Our data show that this gene is

upregulated during phase two and downregulated during phase one, leading us to believe that it could have a regulatory role in postnatal mammary development- possibly controlling growth after initial widespread metabolic and biogenesis events during phase one growth.

PAX8, Paired box protein 8, encodes a transcription factor that is known to control cellular fate and embryonic development of the central nervous system, eye, kidney, thyroid gland, and mesonephric gut organs ^{559–565} and has a critical role in maintaining the normal function of certain cells, especially epithelial cells, of these organs after birth ^{566,567}. Ozcan *et al* showed through an immunohistochemical study that PAX 8 expression occurs in a tissue and organ specific manner in both non-neoplastic adult tissue and transformation and metastasis ⁵⁶⁸. During phase two of postnatal mammary gland development, this transcription factor is likely upregulated to support and control normal epithelial cell differentiation and organ development.

Two polycomb group RING finger proteins are upregulated during phase two. PCGF1, Polycomb group RING finger protein 1, encodes a protein component of the polycomb group (PcG) multiprotein BCOR complex, which functions to maintain several genes in a transcriptionally repressed state and regulate developmental genes ⁴⁷². PCGF5, Polycomb group RING finger protein 5, is also a component of a PcG complex (PRC1-like) required to maintain the transcriptionally repressed state of HOX (homeobox) genes throughout development ⁵⁶⁹. Target genes repressed by PCGF1 include BCL6 and CDKN1A. CDKN1A functions as a regulator of cell cycle progression at G1 phase and its expression is also controlled by p53 in order for cell cycle arrest to occur if needed ^{570–572}. PCGF1, as part of a PcG complex, can promote cell cycle progression and can enhance proliferation by repressing various cell cycle regulators ^{572–579}. Importantly, PcG proteins have been

implicated in both cancer development and maintenance of embryonic and adult stem cells; it is possible that PCGF1 and PCGF5 play a role in stem cell pluripotency⁵⁷⁰. Recent data show that PCGF5 is expressed preferentially in hematopoietic stem cells (HSCs) and multipotent progenitors (MPPs) compared with differentiated cells⁵⁸⁰. Dupret *et al* showed through a zebrafish model that PCGF1 is critical in early development; zebrafish without normal PCGF1 gene function had significantly lower early developmental growth rates and later, signs of premature aging⁵⁸¹. PCGF1 and PCGF5 are likely functioning in early postnatal mammary development in a similar fashion as described above in order to regulate and promote growth of the mammary gland, specifically during phase two.

PRUNE2, Protein prune homolog 2, encodes a protein of the B-cell CLL/lymphoma 2 and adenovirus E1B 19 kDa interacting family⁴⁷², where it functions in apoptotic processes as tumor suppressor^{582–584}. However, little data exist for a role in mammary development. Similarly, RNF217, E3 ubiquitin-protein ligase RNF217, encodes a protein that interacts with a known anti-apoptotic protein, HAX-1, via its C-terminal RING finger motif, suggesting a role in apoptosis signaling. Our data suggest that presence of both PRUNE2 and RNF217 upregulation phase two supports their regulatory roles in growth and proliferation of the mammary gland, but further study is needed since literature is lacking for these genes.

THPO encodes the protein Thrombopoietin, the major cytokine that regulates platelet production, controls differentiation and proliferation of megakaryocyte progenitor cells, and maintains normal thrombopoiesis levels^{585,586}. This indispensable cytokine has been shown to activate the JAK/STAT⁵⁸⁷ and PI3K⁵⁸⁸ pathway as well as activate in an inside-out manner integrin $\alpha 2\beta 3$ ⁵⁸⁹. Matsumura *et al* proved that THPO induced differentiation of a human megakaryoblastic leukemia cell line through transcriptional activation of p21 by

STAT5⁵⁹⁰. THPO is vital for the regulation and maintenance of HSC quiescence and data suggests it supports stem cell renewal after bone marrow transplantation^{591,592}. The upregulation during Days 25-39 in the early postnatal mammary gland shown by our data suggest that THPO supports platelet production and differentiation in the developing mammary microenvironment.

Taken together, discriminate analysis data show temporal regulation of important signaling events during two phases of early postnatal, prepubertal mammary development. Differential expression of all five highly significant genes (revealed using all three methods) was **upregulated during phase two of postnatal mammary gland development** in comparison to their downregulated expression immediately after birth in phase one. Greater than 50% of the pathways utilized by differentially expressed genes were direct components of the immune system, indicating significant impact of both innate and adaptive immune system factors in regulation of postnatal developmental events. The immune system's involvement in prenatal and postnatal mammary development has been gaining acceptance^{11,84-88}, and our data further confirms significant involvement, specifically suggesting critical roles in temporal regulation during the first two months after birth.

Pathway analysis was performed on each phase of discriminately expressed genes indicated mainly metabolic, gene expression, transcription and translation, and cell cycle pathways utilized in phase one while the larger subset of differentially expressed genes that were **upregulated during phase two** utilized mainly immune system pathways that were **not** found in differentially expressed genes of phase one. Our pathway analysis on the differentially expressed genes during early mammary development indicates that the adaptive immune system is highly utilized beginning in the period of 18-25 days after birth and

remains active until at least 39 days after birth, during which it contributes to developmental trends. Additionally, our data and the assessment of key immune regulators and markers in swine by other researchers suggests this is a much better model of human innate immunity and disease than rodents ¹⁵¹. Accordingly, understanding the function of immune molecules in normal development and their concert with other signaling mechanisms can provide insight into these roles during cancer development ⁹¹.

DFA, in addition to providing novel insight regarding temporally regulated events of early postnatal mammary development, suggested that phase one and phase two can be defined as widespread biogenesis and refined differentiation, respectively. The presence of genes with advanced regulatory and immune function, as well as many with tumorigenic roles, in phase two is a unique finding; although growth is seen immediately after birth, maturation and differentiation seems to take three to four weeks to truly begin.

There are specific trends + interactions between sets of genes that have developmental and biological significance over time (PRC)

PRC analysis generated a list of genes that contribute functionally, both positively and negatively, over time (4-39 DOA) to our model of early postnatal mammary gland development. PRC identifies specific genes contributing functionally, and possibly interacting with each other, toward development over time, as shown in Fig. 9. Positive PRC gene transcripts increase in abundance with the curve over time; the top 20 are shown on the effect axis in green text. Conversely, the top 20 genes that are contributing negatively toward development over time, or decreasing in abundance with the curve, are shown in red text. Reactome pathway analysis plus manual pathway curation if necessary is indicated by the

colored box to the right of each Gene ID. In Tables 7 and 8, PRC weights are shown along with each Gene ID and family for the top 50 negative and positive genes. A greater PRC weight indicates greater contribution to the curve trend.

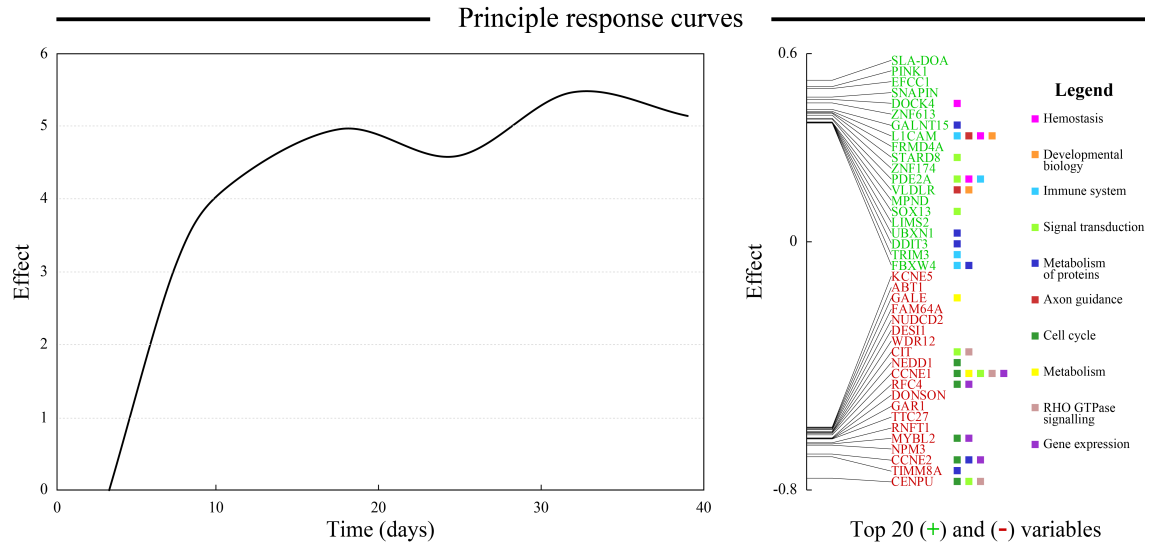


Figure 9: PRC identifies specific genes contributing functionally toward development over time (days of age); Positive PRC gene transcripts increase in abundance with the curve over time (top 20 are shown on the effect axis in green text). The top 20 genes that are contributing negatively toward development over time, or decreasing in abundance with the curve, are shown in red text. Reactome pathway analysis plus manual pathway curation if necessary is indicated by the colored box to the right of each Gene ID.

In an RNA-seq data set that contains a large number of response variables (genes), interpretation of RDA and CCA plots is complicated because interaction effects exist among explanatory variables⁵⁹³. Initially developed by van den Brink & ter Braak⁵⁹⁴, PRC can be used to study the effect of age (as a function of time) on gene expression in the early pre-pubertal/postnatal mammary gland. PRC is based on RDA with the same assumptions and is frequently used in ecological studies as resulting biological and chemical data sets from analysis of environmental trends are large and difficult to evaluate^{593–599}. PRC curves visually show changes over time in a format that is easy to read and are usually used with an

internal reference (overall mean, reference age, or reference year)^{594,596,600}. Time-dependent effects can be clearly isolated in PRC¹⁷⁴, making it a valuable tool to extract developmental trends from the large transcriptomic data set. This advanced technique has not yet been applied to data in the field of developmental or mammary biology. Shown in Fig. 9, the PRC plot of the transcriptomic data set illustrates a clear trend over the period of 4-39 DOA, indicating time-dependent effects are present. PRC identifies the set of positive genes contributing functionally toward development over time, increasing in abundance with the curve or contributing negatively toward development over time. Each gene has a unique weight, shown in Tables 6-7, that indicates its contribution to the curve, and therefore, its contribution to postnatal mammary development.

Tables 7 and 8 are a summary of the top 50 positively and negatively contributing genes to the PRC curve. It is hypothesized that the genes decreasing in abundance with the PRC curve are very early genes contributing to the phenotype at birth or prior; these could also be similar to phase one genes (upregulated during 4-18 DOA) discussed in the discriminant analysis section. Reactome pathway analysis indicated that the top pathways utilized by the top 50 genes negatively contributing to the developmental trend are: Cell Cycle, Mitotic, Cell Cycle, Metabolism, Metabolism of proteins, Signal Transduction, RHO GTPase Effectors, Signaling by Rho GTPases, Gene expression (Transcription), and Generic Transcription Pathway. A review of the literature and characterization using Panther⁶⁰¹⁻⁶⁰⁵ and GeneCards⁴⁷² indicates that the main biological functions of the negatively contributing genes are organelle biosynthesis, transcription and gene expression, cell cycle control, RNA binding and processing, DNA elongation and amplification, CDK interactions, and molecular chaperoning. In agreement with phase one DA genes, which mainly utilized signal

transduction and metabolic pathways, negative PRC genes are not present innate or adaptive immune system pathways.

Conversely, it is hypothesized that the positive PRC genes, those that increase in abundance with the PRC curve, are developmental genes, possibly similar to phase two DA genes discussed previously. These positive PRC genes theoretically function synergistically to significantly contribute to early postnatal mammary development over time. Reactome pathway analysis indicated that the top pathways utilized by the top 50 genes positively contributing to the developmental trend are: Immune System, Metabolism of proteins, Signal Transduction, Adaptive Immune System, Cytokine Signaling in Immune system, Post-translational protein modification, Axon guidance, Hemostasis, Developmental Biology, Interleukin-4 and 13 signaling, Signaling by Interleukins, and Innate immune system. A review of the literature and characterization using Panther tools^{601–605} and GeneCards⁴⁷² indicated that the main biological functions of the positively contributing PRC genes are: comprehensive immune system function, stem cell signaling, antigen presentation, protection from stress, inflammation, apoptosis and transcriptional repression, Integrin signaling, utilization of NFkB, Wnt, MAPK, and AKT pathways, cell migration and adhesion, epithelial and mesenchymal cell polarity and morphology, focal adhesion formation, adherens junction formation, and ECM cell spreading. Interestingly, a review of the literature indicated that 65% of the positive PRC genes can function as oncogenes or proto-oncogenes in comparison with only 22% of the negative PRC genes. As shown in Fig. 9, the top 20 positive and negative PRC genes have the potential to utilize more than one Reactome pathway (indicated by colored boxes). As briefly mentioned, our analysis showed the top 20 negative genes that decrease in abundance with the curve are not reported to utilize any immune system or

developmental biology pathways in contrast with high utilization by the positive genes (as characterized by Reactome). Conversely, the top 20 positive genes that increase in abundance with the curve utilize very few cell cycle and gene expression pathways in comparison by high utilization by the negative genes. Taken together, PRC and pathway analysis suggests that development and maturation of the mammary gland during the first 39 days after birth heavily relies on the immune system, immune components, integrin molecules, and the ECM, utilizing genes known to function in development and previously uncharacterized, novel genes that are also likely to function as oncogenes or proto-oncogenes.

Since the data illustrated that a majority of the top 50 positive PRC genes are functionally significant, linked with cell signaling pathways of biological interest, and have been implicated in various types of cancers, a select subset and implications for functional roles in early mammary development will be discussed herein. These are: SLA-DOA, PINK1, DOCK4, L1CAM, FRMD4A, STARD8, VLDLR, SOX13, LIMS2, UBXN1, DDIT3, HGF, DMAP1, and TRAF3. TRAF3 and HGF were not able to be shown due to figure size limitations. These genes are a chosen subset of the most significant functional genes in early mammary development uncovered through PRC analysis with critical, widespread, and/or novel roles in the process. A full characterization for the top 50 positive and top 50 negative PRC genes can be found in Appendix A and B.

First, a large portion of genes contributing to development over time were linked with the innate and/or adaptive immune system. SLA-DOA, also known as HLA class II histocompatibility antigen do α chain (HLA-DOA), modulates HLA class II restricted antigen presentation pathways in B-cells and is both an MHC Class II molecule chaperone and an adaptive immune system constituent ^{472,606,607}. As reviewed in Chapter 1, the role of

the adaptive immune system in early mammary development remains unexplored with an unknown what function, if any, in mammary organogenesis. An adaptive immune system protein with the highest positive PRC weight indicates significant involvement of this pathway in mammary development and also supports our previous findings of adaptive immune system involvement in development using disparate statistical methods including discriminant analysis (Fig. 8; Tables 5-6) and k-means clustering (Fig. 7). It has been previously established that MHC Class II molecules are frequently expressed in colorectal and breast carcinoma tumors with expression suspected to occur in response to hormones and/or cytokines⁶⁰⁸, but correlation with disease prognosis is unknown⁶⁰⁹ and little is recognized regarding roles of HLA-DOA in tumor response⁶⁰⁷. Although HLA-DOA is upregulated in B-Cell Lymphoma (BCL)^{610,611}, the mechanism is not elucidated and to our knowledge, no data exist supporting a role for this molecule in postnatal mammary gland development. Currently, it is believed that the role of immune cells in developmental processes is restricted to the innate immune system with the exception of lactation⁸⁸. This novel data suggests that both the innate and adaptive immune system have significant roles in developmental trends during the first two months of postnatal organogenesis. Additional immune system components are present in the top 25 positive PRC genes and will be subsequently discussed.

TRAF3 encodes a protein member of the TNF receptor associated factor (TRAF) family, an assembly of six proteins that mediate cellular signals through association with members of the TNF receptor (TNFR) superfamily⁴⁷². TRAF3 is the most obscure of the six and participates in a variety of signaling pathways in a range of cell types: it demonstrates critical roles in both B and T lymphocyte development, including B cell homeostasis,

development, and survival, isotype switching from IgM to IgG, B-cell malignancies, cytokine production, CD40 signal transduction, TLR activation, T cell function, thymocyte development, antiviral responses, Treg cell development, and NFkB/JNK activation ⁶¹². Its position in the top 25 positive PRC genes suggests one or more of these critical roles exists in early postnatal mammary development. TRAF3 may also be responsible for oncogenic consequences of several molecules, including the Epstein-Barr virus encoded latent infection membrane protein-1 (LMP1) ⁶¹³. Importantly, TRAF3 has been found to promote ubiquitination and proteasomal degradation of MAP3K14 ⁶¹⁴, one of the important phase two genes found using all three DA methods (Fig. 8), suggesting functional cooperation between developmental genes discovered in our transcriptomic dataset using more than one multivariate technique. Mouse models demonstrated that TRAF3 is critical for development of the marginal zone (MZ) B cell subset as well as general B-cell homeostasis and survival through mediation of the NF-κB2 pathway ^{615–617}. In addition to NF-κB2 activation, TRAF3 is thought to act as a pro-apoptotic factor and/or restrain additional pro-survival pathways ^{612,615–620}. Mutations inducing loss of function are seen in TRAF genes of multiple myeloma and some types of B cell lymphoma, indicating it plays a negative regulatory role in B lymphocytes, commonly as a tumor suppressor ^{619,621–624}. TRAF3 also regulates activity downstream of B cell receptors; it has been seen to negatively regulate CD40 and LMP1 ^{613,625}, bind to members of the TNFR subfamily including BAFFR, BLyS, TALL-1, TAC1, and APRIL (reviewed in ^{45,626–629}), and associate with CD120b/TNFR2 ⁶³⁰. In TLRs, TRAF3 has been shown to positively mediate antiviral TLR signals and IFN production in myeloid cell responses ^{631–633} while also functioning as a negative regulator of B cell TLR responses, including TLRs 3, 4, 7, and 9 ⁶³⁴. More recently, TRAF3 has been shown to be required for

multiple stages of T cell development, production of IgG1, T- cell function in vivo, Treg development, invariant natural killer T (iNKT) cell signaling⁶³⁵, and TCR/CD28 mediated signaling⁶¹⁸; it could be involved in these developmental events of the adaptive immune system in our model. TRAF expression by B and T lymphocytes and its function in these immune cell types indicates the versatility, criticality, and significance of this gene in cellular development, homeostasis, immune function, and apoptosis. It is suspected that TRAF3 may impact some or all of these processes during early postnatal mammary development, contributing to its role as a top 25 positive PRC gene. Importantly, its dual role as a negative regulator of TLR responses and positive regulator of myeloid responses places this gene in both innate and adaptive immune system categories, suggesting contributory activity of both systems during the first two months of postnatal mammary development.

HGF, hepatocyte growth factor, encodes a protein that binds to the hepatocyte growth factor receptor (HGFR) to regulate a wide range of processes, including cell growth, motility and morphogenesis in numerous cell and tissue types; additionally, it has roles in angiogenesis, tumorigenesis, tissue regeneration, and MET (HGFR) activation⁴⁷². As reviewed in Chapter one, the mammary gland is a dynamic tissue in which differentiation and apoptosis occur cyclically during early development, puberty, pregnancy, and lactation and most mammary branching and development occurs postnatally²². Mammary epithelial cell migration, growth, and proliferation depends heavily on stromal and ECM signals that support mammary gland morphogenesis⁴. HGF, produced by mesenchymal cells/mammary fibroblasts⁶³⁶, provides these specific signals that mainly influence cells of epithelial origin, acting as a paracrine mediator of epithelial interactions^{636,637}. HGF potentially functions as a paracrine mediator of mammary epithelial cell signaling and interaction in our early postnatal

mammary development model since the gene and its cellular receptor are carefully temporally regulated during mammary development to support ductal branching, invasiveness, and epithelial cell motility⁶³⁸. The receptor, a unique transmembrane tyrosine kinase, is expressed in surrounding mesenchyme during development; importantly, it is also a product of the c-met proto-oncogene expressed in epithelial cells, providing a mechanism for EMT processes during development⁶³⁹. In mature tissue types, c-met functions in wound healing and tissue regeneration^{640–642}; c-met and HGF are expressed in mesenchymal stem cells, suggesting involvement in hematopoiesis⁶⁴³. The presence of HGF, and theoretically its receptor, suggests involvement of EMT processes during postnatal mammary gland development. It is widely accepted that this naturally occurring cytokine plays a significant role in murine^{644–646} and chick embryogenesis^{647–649} collaborating with the multifunctional cytokine, TGF- β 1-3, to promote ductal morphogenesis. Interestingly, HGF is mitogenic for luminal cells but not myoepithelial cells and morphogenic to myoepithelial cells but not luminal cells^{636,644,650}. This suggests c-met/HGF signaling can control the function of luminal progenitors and modulate their fate during mammary development⁶⁵¹ and provides another possible role for the molecule in our model. Pursuant to our research focus, it has been shown *in vitro* that ductal morphogenesis induced by HGF is dependent on functional β 1 integrin molecules¹⁴³. The involvement of integrin molecules in normal mammary development and TNBC have been reviewed in Chapters 1 and 2. Especially interesting is the finding that β 4 integrin can be activated through phosphorylation by the c-met receptor for HGF, which subsequently recruits the tyrosine phosphatase Shp2 and Src in order to signal through the Ras–ERK cascade^{652,653}. HGF/c-met represent a novel signaling pathway that connects β 4 integrin to ERK activation in the mammary gland⁶⁵³ and the presence of HGF in the top 25

positive PRC genes confirms our suspicion of significant $\beta 4$ integrin involvement and EMT in early postnatal mammary development. From Chapter one, it is recalled that breast cancer, metastasis, invasion, and progression require epithelial cells to enhance their normal growth and survival capability. Frequently, deregulated HGF/c-met signaling is observed in cancer with c-met, $\beta 4$, and Src all commonly overexpressed (and coexpressed) ^{413,654,655}.

Constitutively active Src can upregulate HGF in breast carcinoma cells, promoting tumorigenesis ⁶⁵⁶. Also in breast cancer, c-met and PIK3CA mutations are frequent- dysregulation of c-met and PI3K significantly increased cell proliferation and invasion *in vitro* and *in vivo*, suggesting that response to MET-targeted therapy is dependent on an aberrant MET-HGF/PI3K axis ⁶⁵⁷. HGF/c-met signaling has recently been associated recently with specifically basal type breast cancers/TNBC, with mutations occurring in progenitor cells in the luminal mammary epithelium ^{658,659}. C-met overexpression has been associated with BRCA1 mutated tumors in a mouse model where BRCA1 and p53 were deleted in the mammary epithelium ⁶⁶⁰; taken together, these above findings suggest functional cooperation between HGF/c-met, $\beta 4$, and Src in order to advance cell survival and metastasis which raises the question of similar cooperation in early mammary development. Most of this information on HGF signal transduction has been derived *in vitro*, however, and has not been extrapolated to, or confirmed in the mammary gland *in vivo* ⁶³⁹ (and if it is, murine models are used). Swine models have not been frequently used to investigate functions of any of these genes in early postnatal development, adding further novelty to the approach.

SOX13, SRY-related HMG-box 13, encodes a member of the sox family of transcription factors, which are well-established regulators of developmental cell fate decisions that have roles in tissue homeostasis and regeneration and are also dysregulated in

cancers^{661–663}. The role of SOX13, however, is mostly undefined⁶⁶¹ and this gene is one of the lesser characterized positive PRC genes in our dataset. However, it has been identified as a gammadelta T-cell specific gene in the immune system- promoting development of this type of T-cell while opposing $\alpha\beta$ differentiation, potentially through inhibition of signaling by Wnt/T cell factor (TCF) pathway⁶⁶⁴.

In addition to immune involvement, protection of cells from DNA damage or dysfunction is critical in development and genes of this nature present in the top 25 positive PRC set. PINK1, PTEN induced putative kinase 1, also known as PARK6, is a highly-conserved Parkinson's disease-susceptibility gene whose product protects cells from mitochondrial dysfunction induced by stress and eradicates damaged mitochondria through mitophagy^{665–668}. PINK1 is induced by the tumor suppressor PTEN⁶⁶⁹, a significant component of the PI3K/AKT pathway subject to mutations²⁵⁵ PI3K/AKT and one of the most frequently mutated genes in breast cancer^{337,670}. In parallel with the discovery that PINK1 is induced by PTEN, increased expression of this gene was exposed in metastatic murine cancer cell lines⁶⁷¹ and loss of function mutations were discovered in several forms of Parkinson's disease⁶⁷². Extensive Parkinson's research has shown that signaling through PINK1 controls several other processes along with normal function in mitochondrial homeostasis and dynamics^{673,674}, including PI3K/AKT signaling, chemoresistance, and regulation of apoptosis in cancer cells⁶⁷⁵. The gene has been reported to promote contrasting oncogenic and tumor suppressor properties^{665,669,673,676–682}, largely through activation of the PI3K/AKT/mTORC signaling pathways^{675,680,681}. As mentioned in Chapter 1, AKT signaling activates FOXO transcription factors that subsequently inhibit cyclin D1 expression and promote the cell cycle regulators p27kip1 and p130Rb2²⁵⁹. When phosphorylated by AKT,

FOXO is inactive and supports cyclin D1 expression while inhibiting expression of cell cycle inhibitors. One specific FOXO, FOXO3a, is the only transcription factor to date shown to regulate PINK1 during nutrient deprivation or chemotherapeutic stress, promoting cell survival⁶⁸³. FOXO3a can also promote production of MMPs in support of tumorigenesis and invasiveness⁶⁸⁴; signaling during early mammary development also heavily relies on MMPs for expansion of the gland, suggesting in our model, PINK and FOXOs could collaborate to regulate gland expansion through MMP production. It is hypothesized that PINK1 may be activated by FOXO3a under nutrient deprivation as a form of short-term protection so that when stress is lifted, the system is ready to resume PI3K/AKT signaling⁶⁷⁵. Additionally, PINK1 loss increases the susceptibility of cells to DNA damage⁶⁸⁵, and in parallel with the major functional links of this gene and PI3K/AKT signaling needed for growth and survival, it can be hypothesized that PINK1 protects actively proliferating cells during early mammary development. PINK1, mitochondrial homeostasis, and cancer has been reviewed in depth by O’Flanagan *et al*⁶⁷⁵, who showed that PINK1 is critical for cell cycle progression, G2/M and G0/G1 checkpoints, cell division, growth, and stress resistance in cancer⁶⁸⁶. Importantly, PINK1 presents a tissue specific location in epithelial tissues⁶⁸⁷, suggesting it probably functions to protect and regulate mammary epithelial cells in the postnatal mammary gland. This data confirms PINK1 highly contributes as a positive PRC gene to mammary growth in the first two months after birth and suggests PINK1 is localized in mammary epithelial cells where it plays similar roles in stress resistance reported in the literature.

Another protective gene that positively contributes to developmental trends in the model is DDIT3. DDIT3, DNA damage induced transcript 3, encodes a regulator of the cellular stress response and is induced upon DNA damage, ER stress, hypoxia, and

starvation, triggering either apoptosis or cell cycle arrest^{472,688–708}. DDIT3 is also involved in differentiation and specialization of varied tissue and cell types^{709–713} and is versatile; it can be expressed either in the cytoplasm or nucleus with distinct functions^{688,714}. Nuclear DDIT3 controls cellular growth and can induce G1 cell cycle arrest if needed⁷¹⁵ while cytoplasmic DDIT3 mainly inhibits migration⁷¹⁶. Pursuant to our developmental model, DDIT3 is expressed during early embryogenesis⁷¹⁷ and functions as a protective Wnt repressor, inhibiting Wnt/T-cell factor (TCF) signaling⁷¹⁸. It also plays a role in accumulation and immune system regulation of tumor suppressor cells derived from myeloid cells⁷¹⁹. DDIT3 expression can block T cell responses⁷¹⁹, suggesting involvement in tumor-induced tolerance and resistance to immunotherapy. Our data suggest it has a significant, although complicated and unelucidated, role in early postnatal mammary development.

Positive PRC genes also fall into another prominent functional group with interesting roles in cell adhesion, adherens junctions, and cellular migration. In Chapter one, it is described how through basally located integrin molecules and laminins, direct attachment of mammary epithelial cells to the ECM can occur^{138,139}. Laminin-binding integrin molecules commonly are found in focal adhesions (FAs) and HDs, which are two adhesion complexes that form mechanical links between the ECM and the actomyosin cytoskeleton¹⁴⁰. Linking the ECM and the cytoskeleton is critical in regulation of a variety of processes, including proliferation and differentiation, and multiple supporting genes are found in the top 25 positive PRC dataset. With a significant weight, DOCK4, Dedicator of cytokinesis protein 4 is detected in our early mammary development model. This gene encodes a membrane-associated, cytoplasmic protein that activates Rap GTPase in order to regulate formation of adherens junctions between cells and also functions in cell migration^{720–724}. In MDA-MB-

231 TNBC cells, DOCK4 interacts with SH3 and SYLF domain containing protein 1 (SH3YL1) to promote cancer cell migration⁷²⁰. L1CAM, neural cell adhesion molecule L1, also encodes a cell adhesion protein with both a functional role in cell migration and nervous system development through axonin binding^{472,725} and a metastatic role in the progression of human tumors⁷²⁶. L1CAM can bind homophilically to itself or heterophilically to other cell adhesion molecules or integrin molecules⁷²⁷, leading to two disparate functions: static, acting as an adhesion molecule between cells (normally in neural settings), and motile, promoting cell migration^{728,729}. The functional switch occurs when the L1CAM ectodomain is cleaved and free to bind integrin molecules⁷³⁰. This subsequently promotes cell migration^{731,732}, resistance to apoptosis^{733–735}, and angiogenesis⁷³⁶. Recently, studies have shown that L1CAM cleaved by MMPs can heterophilically bind to integrin molecules^{737–740} and activate signaling pathways including PI3K, Rac-1, and ERK^{730,741–744}. Taken together, several recent studies suggest that L1CAM/integrin signaling can converge with growth factor signaling pathways, activate conventional ERK/PI3K signaling, and promote growth and development^{728,729}. L1CAM's presence as a top 25 positive PRC gene demonstrates potential for a similar function in postnatal mammary development. L1CAM is also important in human cancers, appearing overexpressed in ovarian and endometrial carcinoma (EC), pancreatic ductal adenocarcinoma (PDAC), melanoma, and glioblastoma where it is associated with poor prognosis through a highly motile and invasive phenotype (reviewed⁷⁴⁵). As an oncogene, L1CAM expression can induce constitutive activation of NFκB activation through IL-1β^{746,747}. Interestingly, it can also indicate and promote EMT transition at the primary tumor site; mesenchymal tumor cell phenotypes exude high levels of vimentin, SLUG and L1CAM^{733,748}. L1CAM expression in human breast cancer cell lines results in disruption adherens

junctions, increased β -catenin transcription, and high motility at the leading edge of cell monolayers^{748–750} while L1CAM knockdown significantly decreases cell motility⁷⁴⁸, suggesting this gene is an EMT trigger in epithelial cells. In Chapter one, it was presented how the ECM and stroma, including growth factors, cytokines and proteinases within these niches, have long been known for morphogenesis⁷². Not only are ECM-cell interactions, adherens junction regulation, EMT, and cell migration essential for development of a functional mammary gland^{7,73}, these also play important roles in breast cancer development (reviewed by^{74–78}). Current research indicates branching morphogenesis of the mammary gland is controlled by interacting epithelial and mesenchymal cell populations, and cell-cell/cell-matrix adhesions, but that this crosstalk is not comprehensively understood⁷⁹. Our presentation and characterization of positively influencing PRC genes in development, such as L1CAM, suggest previously unrealized potential roles for these molecules in postnatal mammary development. Taken together, the studies discussed above illustrate a relationship between immune cell activation, the ECM, TME, integrin molecules, cell migration, and L1CAM. If L1CAM can induce IL-1B secretion by immune cells, it could promote tumorigenicity and/or inflammation. A review of the literature combined with this novel transcriptomic data confirming that L1CAM is one of the highest weighted genes impacting the PRC curve suggests this gene is functioning in early postnatal mammary development in conjunction with integrin molecules and laminins to promote mammary organogenesis.

FRMD4A, FERM domain-containing 4A (FRMD4A), is a human epidermal stem cell marker⁷⁵¹ encoding a protein that regulates epithelial cell polarity⁷⁵². FRMD4A belongs to the Four-point-one, Ezrin, Radixin, Moesin (FERM) superfamily, whose members all contain a FERM domain^{753–755}. FERM domains have been shown to possess functional roles

in cytoskeleton organization and structure, tissue development, maintenance of cell morphology, cell signaling, and tumor cell migration ^{754,756–760}. Specifically, FRMD4A has been shown to control cytoskeleton protein expression and mediate the assembly of adherens junctions between mammary epithelial cells, essential to tissue morphogenesis, through the activation of a central player in actin cytoskeleton dynamics: ADP ribosylation factor 6 (Arf6) ⁷⁵². FRMD4A has also been found upregulated in squamous cell carcinoma (SCC) cells, especially head and neck squamous cell carcinoma (HNSCC) ⁷⁵¹, and rectal cancer ⁷⁶¹, while FRMD4A silencing *in vitro* stimulates caspase activity and expression of terminal differentiation markers ⁷⁶². This data shows significant impact on mammary development through FRMD4A signaling, suggesting this gene potentially supports organization of the cytoskeleton and maintenance of mammary tissue morphology.

STARD8, START domain protein 8, also known as deleted in liver cancer 3 (DLC-3), encodes a member of the Rho GTPase protein family ⁷⁶³. This protein is widely expressed in normal tissues but downregulated or absent in various cancers including breast cancer ⁷⁶⁴, suggesting it functions as a tumor suppressor. Importantly, STARD8 protein localizes to adherens junctions and is essential for E-cadherin function and maintaining epithelial cell–cell contacts; deletion leads to mislocalization of E-cadherin and disruption of adherens junctions and catenin molecules, impaired cell aggregation, and increased migration ^{763,765,766}. STARD8/DCL3 is best characterized as a novel negative regulator of Rho signaling that can promote carcinogenesis by compromising epithelial cell integrity ⁷⁶³. Taken together with the data showing STARD8 positively contributes to early postnatal mammary development, it may function through potential maintenance/protection of mammary epithelial cell integrity.

LIMS2, LIM Zinc Finger Domain Containing 2, also more commonly known as PINCH2 (Particularly Interesting New Cys-His Protein 2) ⁷⁶⁷, is an important focal adhesion and/or nuclear protein which interacts with ILK (integrin-linked kinase) to mediate extracellular matrix adhesions ⁷⁶⁷. Importantly, LIMS2 can link integrin molecules to the actin cytoskeleton, activate Rac signaling, and create connections to cell surface RTKs and growth factor receptors, and promote mammary epithelial cell differentiation and migration ^{768–771}. Zhang *et al* characterized LIMS2/PINCH-2 in depth and novelly demonstrated that: PINCH-2 is co-expressed with PINCH-1 (another member of the PINCH family) in a variety of human cells; PINCH-2 localizes to both cell-ECM contact sites and the nucleus; PINCH-2 forms a complex with ILK in mammalian cells; the PINCH-2 LIM1 domain, which mediates the association with ILK, is required for the localization of PINCH-2 to cell-ECM contact sites; overexpression of PINCH-2 inhibits the PINCH-1-ILK interaction and reduces cell spreading and migration ⁷⁶⁷. Taken together with the PRC analysis and denotation of this gene as highly positive, a potential role is suggested for PINCH-2/LIMS2 in postnatal mammary epithelial cell differentiation, migration, and adhesion.

Lesser characterized genes that are still significant in development are UBXN1 and SOX13. UBXN1, UBX domain-containing protein 1, is a ubiquitin-binding protein that recognizes autoubiquitinated BRCA1 ⁷⁷². As covered in Chapter 1, one of the most important and well known mutations of TNBC and basal-like breast cancer is that of the BRCA1/BRCA2 genes and associated pathways ^{228–230}. E3 ligase activity of the BRCA1 complex is dramatically reduced in the presence of UBXN1, suggesting that UBXN1 regulates the enzymatic function of BRCA1 ⁷⁷². UBXN1 is also a negative regulator of NFkB activity ⁷⁷³, but most of the mechanism of its function remains to be elucidated. Our data

suggest it highly contributes to normal mammary development, possibly as a tumor suppressor or regulatory element. Only just characterized in 2013, the novel F-box and WD40 containing protein, FBXW4, is part of a ubiquitin ligase complex with SKP1–cullin 1–F-box protein (SCF), highly expressed in mammary gland involution, and may function as a novel tumor suppressor that regulates developmental processes and maintains normal cell growth ⁷⁷⁴. F-Box proteins (FBPs) are responsible for ubiquitination of proteins with varied cellular function and can also regulate the expression and activity of oncogenes and tumor suppressor genes ⁷⁷⁵. However, very little is known regarding the function of FBXW4 in general ⁷⁷⁴, especially in mammary development; a better understanding is needed for how this gene may regulate critical cell signaling pathways. This data suggests it highly contributes to normal mammary development, possibly as a tumor suppressor or regulatory element.

Taken together, PRC analysis (Tables 7-9; Fig. 9) indicated that development and maturation of the mammary gland during the first 39 days after birth heavily relies on the immune system, immune components, integrin molecules, and the ECM, utilizing genes known to function in development and previously uncharacterized, novel genes that are also likely to function as oncogenes or proto-oncogenes. This novel data also confirms the trends seen in phase one vs. phase two genes from DFA; positive PRC genes have similar functions to phase two DFA genes and represent more refined, advanced growth and differentiation rather than a spurt of initial unregulated growth immediately after birth. Additionally, many positive PRC genes have reported links with adherens junctions, cellular migration, the ECM, and apoptosis, indicating the contribution of stromal and extracellular factors in differentiation that were not seen in phase one DFA. As seen in Table 8, the hypothesis has

been confirmed that phase two and positive PRC genes represent refined development and differentiation, while phase one and negative PRC genes function for generic biosynthesis and growth.

Table 7: PRC positive weights

Ranking	PRC Weight	Gene ID	Family/Subfamily
1	0.51559019	SLA-DOA	HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DO A CHAIN (PTHR19944:SF67)
2	0.515449036	PINK1	SERINE/THREONINE-PROTEIN KINASE PINK1, MITOCHONDRIAL (PTHR22972:SF11)
3	0.495199325	EFCC1	EF-HAND AND COILED-COIL DOMAIN-CONTAINING PROTEIN 1-RELATED (PTHR11595:SF36)
4	0.489631223	SNAPIN	SNARE-ASSOCIATED PROTEIN SNAPIN (PTHR31305:SF3)
5	0.461953857	DOCK4	DEDICATOR OF CYTOKINESIS PROTEIN 4 (PTHR23317:SF89)
6	0.454108222	ZNF613	ZINC FINGER PROTEIN 613 (PTHR24404:SF49)
7	0.443115828	GALNT15	POLYPEPTIDE N-ACETYL GALACTOSAMINYLTRANSFERASE 15 (PTHR11675:SF77)
8	0.422306462	L1CAM	NEURAL CELL ADHESION MOLECULE L1 (PTHR10489:SF812)
9	0.416458647	FRMD4A	FERM DOMAIN-CONTAINING PROTEIN 4A (PTHR23281:SF29)
10	0.412167164	STARD8	STAR-RELATED LIPID TRANSFER PROTEIN 8 (PTHR12659:SF11)
11	0.410868587	ZNF174	ZINC FINGER PROTEIN 174 (PTHR10032:SF270)
12	0.404149309	PDE2A	3',5'-CYCLIC PHOSPHODIESTERASE PDE-2-RELATED (PTHR11347:SF164)
13	0.393960511	VLDLR	VERY LOW-DENSITY LIPOPROTEIN RECEPTOR (PTHR10529:SF318)
14	0.391076702	MPND	MPN DOMAIN-CONTAINING PROTEIN (PTHR12802:SF80)
15	0.381996691	SOX13	TRANSCRIPTION FACTOR SOX-13 (PTHR10270:SF273)
16	0.381912433	LIMS2	LIM AND SENESCENT CELL ANTIGEN-LIKE-CONTAINING DOMAIN PROTEIN 2 (PTHR24210:SF13)

17	0.381038717	UBXN1	UBX DOMAIN-CONTAINING PROTEIN 1 (PTHR13020:SF47)
18	0.380960327	DDIT3	DNA DAMAGE-INDUCIBLE TRANSCRIPT 3 PROTEIN (PTHR16833:SF1)
19	0.380488353	TRIM3	TRIPARTITE MOTIF-CONTAINING PROTEIN 3 (PTHR24103:SF458)
20	0.379222784	FBXW4	F-BOX/WD REPEAT-CONTAINING PROTEIN 4 (PTHR14381:SF2)
21	0.379074354	HGF	HEPATOCTE GROWTH FACTOR (PTHR24256:SF291)
22	0.378892532	DMAP1	DNA METHYLTRANSFERASE 1-ASSOCIATED PROTEIN 1 (PTHR12855:SF11)
23	0.377950717	TRAF3	TNF RECEPTOR-ASSOCIATED FACTOR 3 (PTHR10131:SF96)
24	0.367656669	LOC106506316	
25	0.363138248	LHX6	LIM/HOMEODOMAIN PROTEIN LHX6 (PTHR24208:SF148)
26	0.356873254	RGS18	REGULATOR OF G-PROTEIN SIGNALING 18 (PTHR10845:SF226)
27	0.356526442	CTIF	CBP80/20-DEPENDENT TRANSLATION INITIATION FACTOR (PTHR23254:SF18)
28	0.356237482	LOC100739121	
29	0.354164393	C1QTNF9	COMPLEMENT C1Q AND TUMOR NECROSIS FACTOR-RELATED PROTEIN 9A-RELATED (PTHR24023:SF654)
30	0.352729291	GNLY	GRANULYSIN (PTHR15541:SF3)
31	0.352600564	PPFIA1	LIPRIN-A-1 (PTHR12587:SF27)
32	0.350038204	RNF31	E3 UBIQUITIN-PROTEIN LIGASE RNF31 (PTHR16004:SF6)
33	0.349655753	FLI1	
34	0.349297781	A4GALT	LACTOSYLCERAMIDE 4-A-GALACTOSYLTRANSFERASE (PTHR12042:SF27)
35	0.347687199	CCDC28A	COILED-COIL DOMAIN-CONTAINING PROTEIN 28A (PTHR13400:SF5)
36	0.347099383	GPANK1	G PATCH DOMAIN AND ANKYRIN REPEAT-CONTAINING PROTEIN 1 (PTHR20923:SF2)
37	0.346498169	CCDC130	COILED-COIL DOMAIN-CONTAINING PROTEIN 130 (PTHR12111:SF5)
38	0.341624683	CEP95	CENTROSOMAL PROTEIN OF 95 KDA (PTHR22545:SF1)
39	0.341619986	CTSF	CATHEPSIN F (PTHR12411:SF425)
40	0.341569075	LOC1065050	

		43	
41	0.337943464	HMCN2	HEMICENTIN-2 (PTHR10489:SF755)
42	0.337929917	MID2	E3 UBIQUITIN-PROTEIN LIGASE MID2-RELATED (PTHR24103:SF450)
43	0.336405962	ABLIM3	ACTIN-BINDING LIM PROTEIN 3 (PTHR24213:SF24)
44	0.334111451	CUX1	HOMEBOX PROTEIN CUT-LIKE 1 (PTHR14043:SF18)
45	0.333882104	TCTP	TRANSLATIONALLY-CONTROLLED TUMOR PROTEIN (PTHR11991:SF7)
46	0.333209584	KCNB1	POTASSIUM VOLTAGE-GATED CHANNEL SUBFAMILY B MEMBER 1 (PTHR11537:SF177)
47	0.332306307	PTPRM	RECEPTOR-TYPE TYROSINE-PROTEIN PHOSPHATASE MU (PTHR19134:SF391)
48	0.331268784	CUX1	HOMEBOX PROTEIN CUT-LIKE 1 (PTHR14043:SF18)
49	0.33121179	ANKRD24	ANKYRIN REPEAT DOMAIN-CONTAINING PROTEIN 24 (PTHR24173:SF53)
50	0.329721287	PBLD	PHENAZINE BIOSYNTHESIS-LIKE DOMAIN-CONTAINING PROTEIN (PTHR13774:SF23)
51	0.328474016	FADS3	FATTY ACID DESATURASE 3 (PTHR19353:SF39)

Table 8: PRC negative weights

Ranking	PRC Weight	Gene ID	Family/Subfamily
1	-0.755312569	CENPU	CENTROMERE PROTEIN U (PTHR32222:SF2)
2	-0.687764094	TIMM8A	
3	-0.678439898	CCNE2	G1/S-SPECIFIC CYCLIN-E2 (PTHR10177:SF330)
4	-0.650314355	NPM3	NUCLEOPLASMIN-3 (PTHR22747:SF24)
5	-0.646226192	MYBL2	MYB-RELATED PROTEIN B (PTHR10641:SF744)
6	-0.642932862	RNFT1	RING FINGER AND TRANSMEMBRANE DOMAIN-CONTAINING PROTEIN 1 (PTHR15860:SF10)
7	-0.630912134	TTC27	TETRATRICOPEPTIDE REPEAT PROTEIN 27 (PTHR16193:SF1)
8	-0.629593806	GAR1	H/ACA RIBONUCLEOPROTEIN COMPLEX SUBUNIT 1 (PTHR23237:SF9)
9	-0.628766562	DONSON	PROTEIN DOWNSTREAM NEIGHBOR OF SON (PTHR12972:SF1)
10	-0.626469029	RFC4	REPLICATION FACTOR C SUBUNIT 4

			(PTHR11669:SF30)
11	-0.61727193	CCNE1	G1/S-SPECIFIC CYCLIN-E1 (PTHR10177:SF288)
12	-0.612904564	NEDD1	PROTEIN NEDD1 (PTHR22847:SF508)
13	-0.609914388	CIT	
14	-0.608331544	WDR12	RIBOSOME BIOGENESIS PROTEIN WDR12 (PTHR19855:SF13)
15	-0.606545954	DESI1	DESUMOYLATING ISOPEPTIDASE 1 (PTHR12378:SF23)
16	-0.601177154	NUDCD2	NUDC DOMAIN-CONTAINING PROTEIN 2 (PTHR12356:SF22)
17	-0.598877627	FAM64A	PROTEIN FAM64A (PTHR35819:SF2)
18	-0.595805113	GALE	UDP-GLUCOSE 4-EPIMERASE (PTHR43725:SF5)
19	-0.593420751	ABT1	ACTIVATOR OF BASAL TRANSCRIPTION 1 (PTHR12311:SF8)
20	-0.59256649	KCNE5	POTASSIUM VOLTAGE-GATED CHANNEL SUBFAMILY E REGULATORY B SUBUNIT 5 (PTHR15282:SF14)
21	-0.592304891	CDKN2C	CYCLIN-DEPENDENT KINASE 4 INHIBITOR C (PTHR24148:SF54)
22	-0.586226873	LOC102167061	
23	-0.583969995	PRELID3B	PROTEIN SLOWMO HOMOLOG 2 (PTHR11158:SF37)
24	-0.581429846	CTPS1	CTP SYNTHASE 1 (PTHR11550:SF16)
25	-0.578118522	DOHH	DEOXYHYPUSINE HYDROXYLASE (PTHR12697:SF23)
26	-0.576783145	ERCC6L	DNA EXCISION REPAIR PROTEIN ERCC- 6-LIKE (PTHR10799:SF911)
27	-0.572198441	CENPK	CENTROMERE PROTEIN K (PTHR14401:SF7)
28	-0.571350974	DUSP11	
29	-0.56764138	ALG8	DOLICHYL PYROPHOSPHATE GLC1MAN9GLCNAC2 A-1,3- GLUCOSYLTRANSFERASE-RELATED (PTHR12413:SF3)
30	-0.56630587	MTHFD2	BIFUNCTIONAL METHYLENETETRAHYDROFOLATE DEHYDROGENASE/CYCLOHYDROLASE, MITOCHONDRIAL (PTHR10025:SF48)
31	-0.565106896	WDR74	WD REPEAT-CONTAINING PROTEIN 74 (PTHR16038:SF5)
32	-0.562372872	RABGGTB	GERANYLGERANYL TRANSFERASE TYPE-2 SUBUNIT B (PTHR11774:SF13)
33	-0.560353639	MPHOSPH6	M-PHASE PHOSPHOPROTEIN 6 (PTHR13582:SF1)
34	-0.559254552	RNFT1	RING FINGER AND TRANSMEMBRANE

			DOMAIN-CONTAINING PROTEIN 1 (PTHR15860:SF10)
35	-0.557948831	RRP8	RIBOSOMAL RNA-PROCESSING PROTEIN 8 (PTHR12787:SF1)
36	-0.551866159	PRELID3B	PROTEIN SLOWMO HOMOLOG 2 (PTHR11158:SF37)
37	-0.549404483	BUB1	MITOTIC CHECKPOINT SERINE/THREONINE-PROTEIN KINASE BUB1 (PTHR14030:SF10)
38	-0.548547924	CYP26B1	CYTOCHROME P450 26B1 (PTHR24286:SF101)
39	-0.546993466	BCHE	CHOLINESTERASE (PTHR11559:SF337)
40	-0.545884312	MPLKIP	M-PHASE-SPECIFIC PLK1-INTERACTING PROTEIN (PTHR22446:SF4)
41	-0.545783381	LOC106509864	
42	-0.545548898	INCENP	INNER CENTROMERE PROTEIN (PTHR13142:SF2)
43	-0.54113617	MINPP1	MULTIPLE INOSITOL POLYPHOSPHATE PHOSPHATASE 1 (PTHR20963:SF31)
44	-0.540628807	NQO1	NAD(P)H DEHYDROGENASE [QUINONE] 1 (PTHR10204:SF49)
45	-0.537488756	CSRNP1	CYSTEINE/SERINE-RICH NUCLEAR PROTEIN 1 (PTHR13580:SF16)
46	-0.536682609	ANP32A	ACIDIC LEUCINE-RICH NUCLEAR PHOSPHOPROTEIN 32 FAMILY MEMBER A-RELATED (PTHR11375:SF9)
47	-0.536194601	UBE2T	UBIQUITIN-CONJUGATING ENZYME E2 T (PTHR24067:SF205)
48	-0.535012107	ADAMTS4	A DISINTEGRIN AND METALLOPROTEINASE WITH THROMBOSPONDIN MOTIFS 4 (PTHR13723:SF205)
49	-0.533396073	AMD1	S-ADENOSYLMETHIONINE DECARBOXYLASE PROENZYME (PTHR11570:SF8)
50	-0.531299615	C2H1orf31	

Table 9: Summary of PRC trends

	Negative weight (red)	Positive weight (green)
Trend	Decrease with curve abundance	Increase with curve abundance
Top 50	CENPU	SLA-DOA

genes by weight	TIMM8A	PINK1
	CCNE2	EFCC1
	NPM3	SNAPIN
	MYBL2	DOCK4
	RNFT1	ZNF613
	TTC27	GALNT15
	GAR1	L1CAM
	DONSON	FRMD4A
	RFC4	STARD8
	CCNE1	ZNF174
	NEDD1	PDE2A
	CIT	VLDLR
	WDR12	MPND
	DESI1	SOX13
	NUDCD2	LIMS2
	FAM64A	UBXN1
	GALE	DDIT3
	ABT1	TRIM3
	KCNE5	FBXW4
	CDKN2C	HGF
	LOC102167061	DMAP1
	PRELID3B	TRAF3
	CTPS1	LOC106506316
	DOHH	LHX6
	ERCC6L	RGS18
	CENPK	CTIF
	DUSP11	LOC100739121
	ALG8	C1QTNF9
	MTHFD2	GNLY
	WDR74	PPFIA1
	RABGGTB	RNF31
	MPHOSPH6	FLI1
	RNFT1	A4GALT
	RRP8	CCDC28A
	PRELID3B	GPANK1
	BUB1	CCDC130
	CYP26B1	CEP95

	BCHE	CTSF
	MPLKIP	LOC106505043
	LOC106509864	HMCN2
	INCENP	MID2
	MINPP1	ABLIM3
	NQO1	CUX1
	CSRNP1	TCTP
	ANP32A	KCNB1
	UBE2T	PTPRM
	ADAMTS4	CUX1
	AMD1	ANKRD24
	C2H1orf31	PBLD
	CDCA4	FADS3
Hypothesis	These are phase one genes (very early genes)	These are phase two genes; developmental genes
Top pathways in Reactome	Cell Cycle, Mitotic	Immune System
	Cell Cycle	Metabolism of proteins
	Metabolism	Signal Transduction
	Metabolism of proteins	Adaptive Immune System
	Signal Transduction	Cytokine Signaling in Immune system
	RHO GTPase Effectors	Post-translational protein modification
	Signaling by Rho GTPases	Axon guidance
	Gene expression (Transcription)	Hemostasis
	Generic Transcription Pathway	Developmental Biology
		Interleukin-4 and 13 signaling
		Signaling by Interleukins
		Innate immune system
Biological functions manually curated from literature	Organelle biosynthesis	Immune function
	Transcription and gene expression	Stem cell signaling
	Cell control: cycle, division, survival	Antigen presentation, stress protection, inflammation
	RNA binding and	Apoptosis and transcriptional

	processing	repression
	DNA elongation and amplification	Organelle biogenesis, cell fate regulation
	CDK interactions	Integrin signaling- NFkB, Wnt, MAPK, AKT pathways
	Molecular chaperoning	Migration/adhesion/epithelial and mesenchymal cell polarity and morphology
		Focal adhesions, adherens junctions, and ECM
		Cell spreading
Oncogene function	22% are potential oncogenes	65% are potential oncogenes
Top oncogenes	CCNE2	PINK1
	MYBL2	DOCK4
	RFC4	L1CAM
	CCNE1	FRMD4A
	CDKN2C	STARD8
	BUB1	TRIM3
	NQO1	HGF
	ANP32A	TRAF3
	UBE2T	LHX6
	ADAMTS4	PPFIA1
	AMD1	RNF31
		FLI1
		CTSF
		CUX1
		TCTP
		PTPRM
		SLA-DOA
		SNAPIN
		PDE2A
		VLDLR
		SOX13
		LIMS2
		UBXN1
		DDIT3
		FBXW4

	DMAP1
	C1QTNF9
	CCDC28A
	HMCN2
	MID2
	KCNB1
	PBLD

WGCNA demonstrates functional modules in early mammary gland development

Weighted Gene Co-Expression Network Analysis visualized the clustered connectivity of the top 5000 genes, by variance, with a similar expression pattern, resulting in functional modules which will be further explained in a pathway analysis (see Fig. 10). The diagonal line represents the autocorrelation between the gene to itself, resulting in a correlation value of 1 (perfect correlation). The distinct boxes that form around the diagonal line represent the functional modules that contain highly interconnected genes, by correlation, to other genes within the same module. Modules farther away from the diagonal represent inter-module relationships (connected pathways, or one pathway regulating another). Weighting has been used to adjust the correlation matrix such that only the strongest of correlations are depicted. Systems biology approaches can be applied to the RNA-seq data sets to identify groups of genes that are highly correlated (in this case, over time) to form functional modules; WGCNA exists as a popular method to do so since its introduction in 2005. The WGCNA method assumes that genes that are co-expressed with similar variances over time are present in the same subsystem and is appropriate for use here^{776–780}. The gene co-expression network (GCN) shown in Fig. 10 visually expresses these

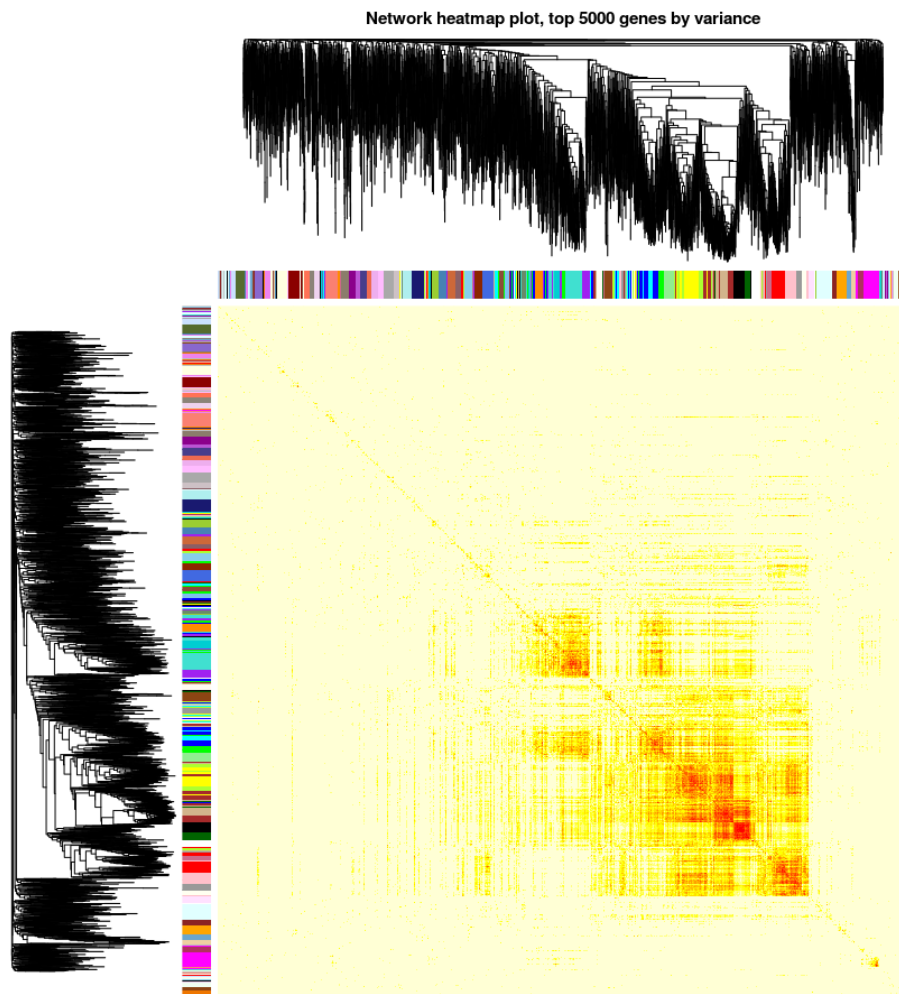


Figure 10: WGCNA demonstrates functional modules in early mammary gland development. Heatmap plot of the top 5000 genes by variance; diagonal line represents the autocorrelation between the gene to itself; distinct boxes around diagonal line represent functional modules that contain highly interconnected genes, by correlation, to other genes within the same module.

subsystems that were clustered based on co-variation within groups of genes (across all 17 samples). GCN can be used to simulate functional significance for groups of genes; tightly correlated gene groups can be thought of as one functional unit and related to

phenotype⁷⁸¹. This pipeline has been used recently to: generate functional conclusions about signaling pathways underlying familial breast cancer susceptibility⁷⁸², identify prognostic biomarkers tumor state based on stromal gene expression⁷⁸⁸, and identify modular heterogeneity of breast cancer subtypes⁷⁸⁹. Based on a review of literature wherein WGCNA methods were applied to RNA-seq datasets, we concluded this method can be used in our study to identify highly correlated groups of genes with functional roles in early postnatal/prepubertal mammary gland development. Fig. 10 confirms the hypothesis that there are genuine functional modules (shown in red/orange) present within the overall RNA-seq dataset generated from samples collected during pre-pubertal mammary gland development. These functional modules can be further dissected using the MCODE algorithm and pathway analysis.

MCODE algorithm identifies highly correlated gene networks within RNA-seq dataset.

MCODE, shown in Fig. 11, shows functional modules that have been isolated from WGCNA and filtered to exclude weak relationships. The ten highest weighted network modules are shown in Fig. 12 and summarized in Table 10. These modules can be further characterized using functional and pathway analysis to determine biological significance in early postnatal mammary development.

After suspected functional modules were identified in Fig. 10, the MCODE algorithm was used as a Cytoscape^{429,790} plug-in to find highly interconnected molecular clusters and remove those with low connectivity from the original plot using the random matrix theory⁷⁹¹. MCODE identifies densely connected nodes to graphically display both nodules and

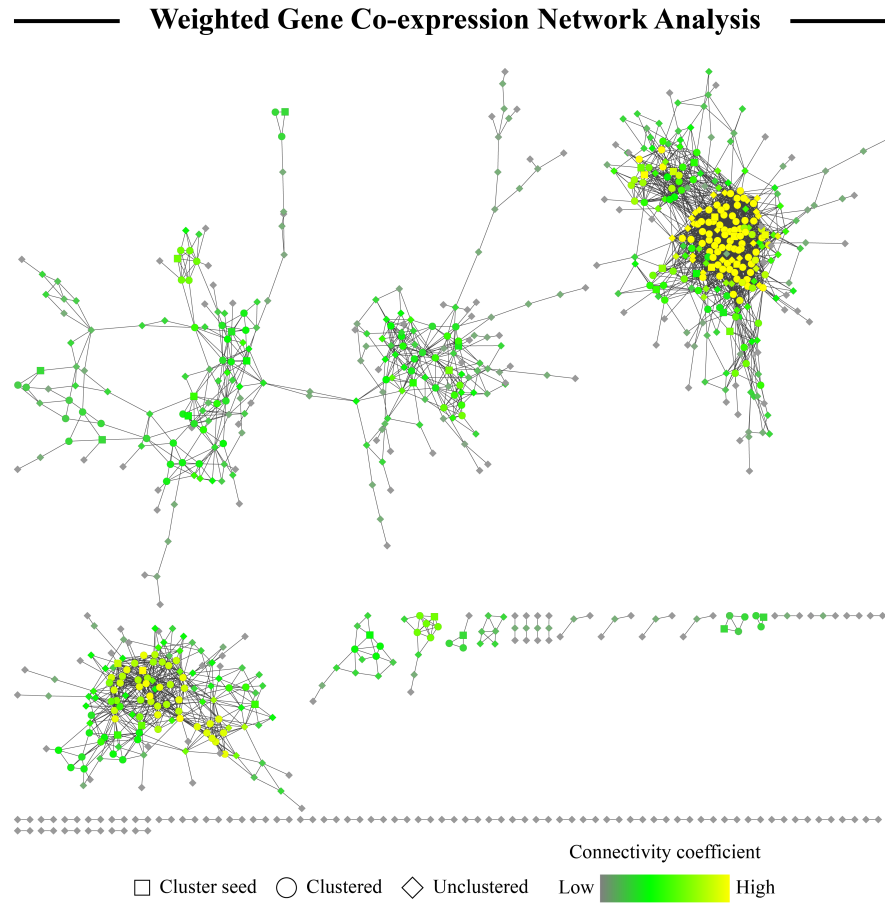


Figure 11: MCODE algorithm identifies highly correlated gene networks within mammary transcriptome dataset.

associated weights, which can be used to sort the relationships⁷⁹², and is commonly used in cancer biology. The genes contributing to each functional module can be extracted as a base for pathway analysis and biological interpretation. Although utilized frequently in the past year to identify biomarkers in colorectal cancer⁷⁹³ and head and neck squamous cell

carcinoma ⁷⁹⁴, as well as attempt to isolate biomarkers and gene networks involved in breast cancer ^{795–797}, to our knowledge, MCODE has not yet been applied to developmental mammary biology.

Functional annotation of MCODE clusters demonstrates biological significance

After MCODE identified densely connected nodes to graphically display both nodules and associated weights (Fig. 11), which can be used to sort the relationships ⁷⁹², the genes contributing to each functional module were extracted for pathway analysis and biological interpretation. The top ten network modules, sorted by weight, are shown below.

Table 10. Functional roles of top ten network modules

Cluster	Score	Function
1	26.25	1. Striated muscle contraction and actin cytoskeleton integrity. Focused around the activity of titin (TTN). Most factors can be found in different domains of muscle. Two oncogenes: TTN and EEF1A. Largest cluster. Can put together good signaling pathway of interactions in muscle development/contraction.
2	7.684	2. Oncogenic cluster. Adopogenesis, adipocyte metabolism, immune function, and metabolic alterations in cancer. Involves AKT/Laminin/FOXO/apoptosis. FAH has a functional link with BRCA. >50% genes in network are oncogenes. Need to design own pathway of normal vs oncogenic interactions. Tissue co-expression could be important in the story here. Interesting interaction with the innate immune system (see Cluster 2 tab).
3	7.417	3. Lipid and protein metabolism + cell cycle control. Common protein interaction: UBC. Definitely an oncogenic cluster.
4	6	4. Keratinocyte differentiation during development. Epithelial markers. Not very oncogenic. Could play a role in melanoma and HPV.
5	5.667	5. Keratinization and differentiation; oncogenic cluster. Contains markers for breast cancer among other types. Specifically contains markers for basal breast cancer and EMT.
6	5.6	6. Muscle contraction. Low oncogenic potential, except a couple genes interact with BRCA1/BRCA2.
7	5.2	7. Muscle contraction/actin binding + calcium signaling. Low oncogenic potential.
8	4.75	8. Protein modification + cell-cell adhesion. Seems to have a neuronal connection. Mixed bag of genes; >50% of them are oncogenes. Potential oncogenic cluster with some laminin/integrin interactions.
9	4.286	9. Muscle contraction and actin binding. Low oncogenic potential; contains several repeats from other clusters.
10	4	10. Metabolism. Genes do not interact clearly with eachother. Some immune factors, muscle factors. Low oncogenic potential.

Table 10 shows functional roles based on review of literature, Panther characterization, and Reactome pathway analysis for the top ten modules. Functional annotation and visualization of network interaction are shown in Fig. 12. Detailed functional analysis is shown in Tables 11-12. The overall network connectivity coefficient indicates the strength of interaction between individual genes in each module, demonstrating novel interactions of genes contributing to early postnatal mammary developmental trends.

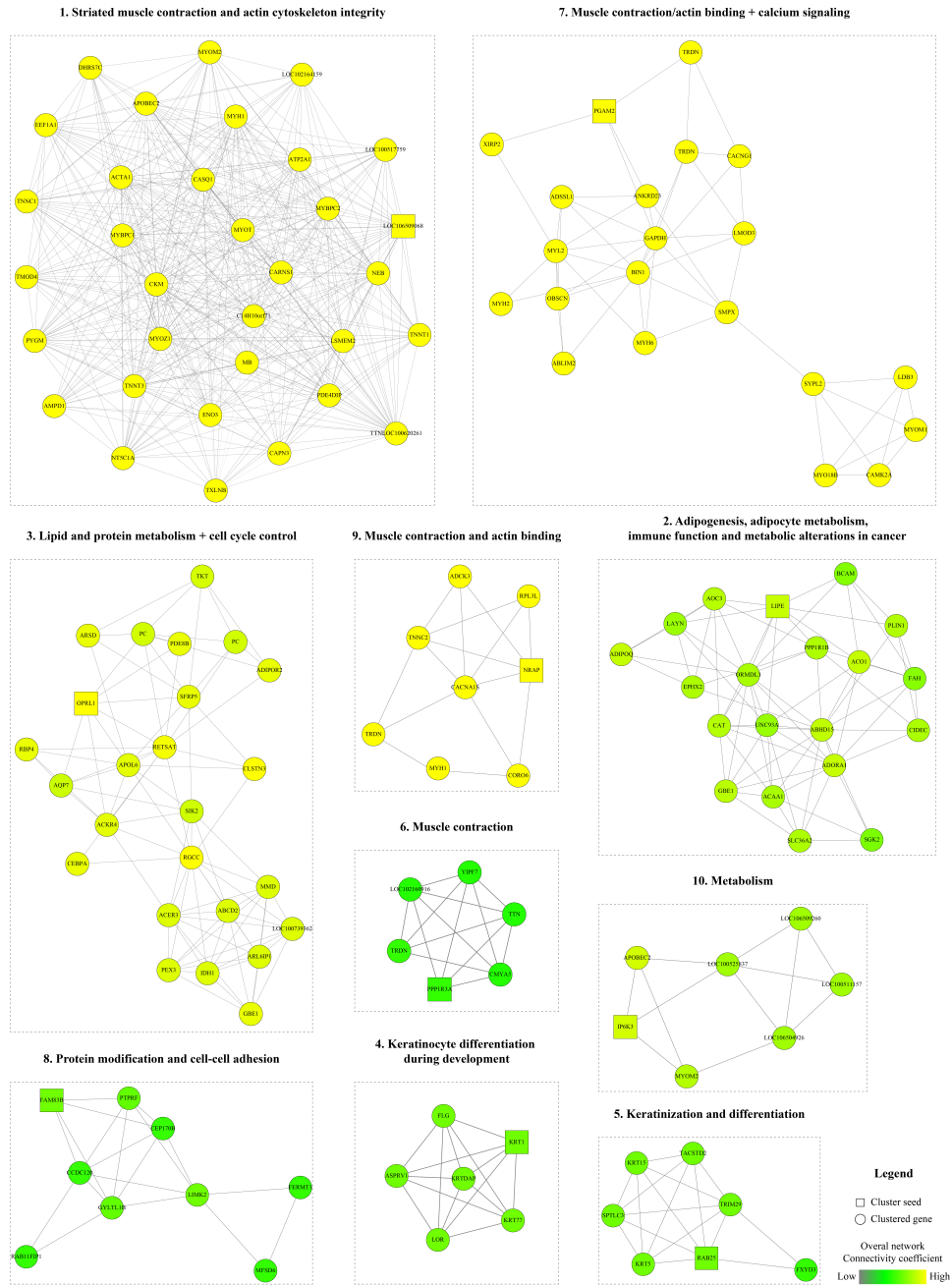


Figure 12. Functional annotation of MCODE clusters. The overall network connectivity coefficient indicates the strength of interaction between individual genes in each module, demonstrating novel interactions of genes contributing to early postnatal mammary developmental trends

Table 11: Reactome pathway summary

Cluster	Weight	Top Pathways	Function
1	26.25	Muscle contraction	Striated muscle contraction and actin cytoskeleton integrity.
		Striated Muscle Contraction	
		Metabolism	
2	7.684	Metabolism	Adopogenesis, adipocyte metabolism, immune function, and metabolic alterations in cancer.
		Immune System	
		Innate Immune System	
		Neutrophil degranulation	
3	7.417	Metabolism	Lipid and protein metabolism + cell cycle control.
		Post-translational protein modification	
		Metabolism of proteins	
		Sphingolipid metabolism	
		Metabolism of lipids	
		Developmental Biology	
		Signal Transduction	
4	6	Developmental Biology	Keratinocyte differentiation during development.
		Formation of the cornified envelope	
		Keratinization	
		FGFR1 mutant receptor activation	
		Signaling by FGFR1 in disease	
		Signaling by FGFR in disease	
5	5.667	Keratinization	Keratinization and differentiation
		Formation of the cornified envelope	
		Developmental Biology	
		Type I hemidesmosome assembly	
		Cell junction organization	
		Cell-Cell communication	
6	5.6	Muscle contraction	Muscle contraction
		Striated Muscle Contraction	
7	5.2	Muscle contraction	Muscle contraction/actin binding + calcium signaling
		Cardiac conduction	
		Metabolism	
		Striated Muscle Contraction	
8	4.75	RHO GTPases Activate ROCKs	Protein modification + cell-cell adhesion

		Receptor protein tyrosine phosphatases interactions	
		Sema4D induced cell migration and growth-cone collapse	
		Sema4D in semaphorin signaling	
		SALM protein interactions at the synapses	
9	4.286	Muscle contraction	Muscle contraction and actin binding
		Ion homeostasis	
		Translocation of GLUT4 to the plasma membrane	
		Cardiac conduction	
		Striated Muscle Contraction	
10	4	Synthesis of pyrophosphates in the cytosol	Metabolism
		mRNA Editing: C to U Conversion	
		Formation of the Editosome	
		mRNA Editing	
		HSF1 activation	

Table 12: Gene families contained in each module (cluster)

Gene ID	Cluster	Gene Name, Gene Symbol, Ortholog	Family/Subfamily
ACTA1	1	Actin, α skeletal muscle;ACTA1;ortholog	ACTIN, A SKELETAL MUSCLE (PTHR11937:SF348)
AMPD1	1	Adenosine monophosphate deaminase 1 isoform M;AMPD1;ortholog	AMP DEAMINASE 1 (PTHR11359:SF10)
APOBEC2	1	Probable C-U-editing enzyme APOBEC-2;APOBEC2;ortholog; Apolipoprotein B mRNA Editing Enzyme Catalytic Subunit 2	C-U-EDITING ENZYME APOBEC-2-RELATED (PTHR13857:SF34)
ATP2A1	1	Calcium-transporting ATPase;ATP2A1;ortholog	SARCOPLASMIC/ENDOPLASMIC RETICULUM CALCIUM ATPASE 1 (PTHR42861:SF17)
C14H10orf71	1	chromosome 14 open reading	

		frame, human C10orf71	
CAPN3	1	Calpain-3;CAPN3;ortholog	CALPAIN-3 (PTHR10183:SF328)
CARNS1	1	Uncharacterized protein (Fragment);CARNS1;ortholog	CARNOSINE SYNTHASE 1 (PTHR10578:SF91)
CASQ1	1	Calsequestrin;CASQ1;ortholog	CALSEQUESTRIN-1 (PTHR10033:SF20)
CKM	1	Creatine kinase M- type;CKM;ortholog	CREATINE KINASE M-TYPE (PTHR11547:SF45)
DHRS7C	1	Uncharacterized protein;DHRS7C;ortholog	DEHYDROGENASE/REDUC TASE SDR FAMILY MEMBER 7C (PTHR43490:SF45)
EEF1A1	1	Elongation factor 1- α ;LOC100514912;ortholog	SUBFAMILY NOT NAMED (PTHR23115:SF198)
ENO3	1	B-enolase;ENO3;ortholog	B-ENOLASE (PTHR11902:SF27)
LOC1005177 59	1	Uncharacterized protein (Fragment);LOC100517759;or tholog; likely myomegalin	SUBFAMILY NOT NAMED (PTHR14199:SF24)
LOC1021641 59	1	LOC102164159	CALSARCIN (PF0556), FATZ RELATED PROTEINCALSARCIN- RELATED (PTHR15941); myozenin-2 like
MB	1	Myoglobin;MB;ortholog	MYOGLOBIN (PTHR11442:SF57)
MYBPC1	1	Uncharacterized protein;MYBPC1;ortholog	MYOSIN-BINDING PROTEIN C, SLOW-TYPE (PTHR13817:SF24)
MYBPC2	1	Uncharacterized protein;MYBPC2;ortholog	MYOSIN-BINDING PROTEIN C, FAST-TYPE (PTHR13817:SF16)
MYH1	1	Myosin-1;MYH1;ortholog	MYOSIN-1 (PTHR13140:SF649)
MYOM2	1	Uncharacterized protein (Fragment);MYOM2;ortholog	MYOMESIN-2 (PTHR13817:SF30)
MYOT	1	Uncharacterized protein;MYOT;ortholog	
MYOZ1	1	Myozenin-1;MYOZ1;ortholog	MYOZENIN-1 (PTHR15941:SF18)
NEB	1	Nebulin;NEB;ortholog	NEBULIN (PTHR11039:SF58)
NT5C1A	1	Uncharacterized protein (Fragment);NT5C1A;ortholog	CYTOSOLIC 5'- NUCLEOTIDASE 1A (PTHR31367:SF7)
PDE4DIP	1	Uncharacterized protein;PDE4DIP;ortholog	MYOMEGALIN (PTHR13140:SF634)

PYGM	1	A-1,4 glucan phosphorylase;PYGM;ortholog	GLYCOGEN PHOSPHORYLASE, MUSCLE FORM (PTHR11468:SF18)
TMOD4	1	Tropomodulin 4;TMOD4;ortholog	TROPOMODULIN-4 (PTHR10901:SF21)
TNNC1	1	Troponin C, slow skeletal and cardiac muscles;TNNC1;ortholog	TROPONIN C, SLOW SKELETAL AND CARDIAC MUSCLES (PTHR23064:SF13)
TNNT1	1	Troponin T, slow skeletal muscle;TNNT1;ortholog	TROPONIN T, SLOW SKELETAL MUSCLE (PTHR11521:SF14)
TNNT3	1	Troponin T, fast skeletal muscle;TNNT3;ortholog	TROPONIN T, FAST SKELETAL MUSCLE (PTHR11521:SF18)
TTN	1	Titin;TTN;ortholog	TITIN (PTHR13817:SF35)
TXLNB	1	Uncharacterized protein;TXLNB;ortholog	B-TAXILIN (PTHR16127:SF23)
ABHD15	2	Uncharacterized protein;ABHD15;ortholog	ABHYDROLASE DOMAIN-CONTAINING PROTEIN 15 (PTHR10794:SF63)
ACAA1	2	Uncharacterized protein;ACAA1;ortholog	3-KETOACYL-COA THIOLASE, PEROXISOMAL (PTHR43853:SF5)
ACO1	2	Aconitate hydratase;ACO1;ortholog	CYTOPLASMIC ACONITATE HYDRATASE (PTHR11670:SF47)
ADORA1	2	Adenosine receptor A1;ADORA1;ortholog	ADENOSINE RECEPTOR A1 (PTHR24246:SF35)
AOC3	2	Amine oxidase;AOC3;ortholog	MEMBRANE PRIMARY AMINE OXIDASE (PTHR10638:SF41)
AOC3	2	Amine oxidase;AOC3;ortholog	MEMBRANE PRIMARY AMINE OXIDASE (PTHR10638:SF41)
BCAM	2	Uncharacterized protein;BCAM;ortholog	BASAL CELL ADHESION MOLECULE (PTHR11973:SF23)
CAT	2	Catalase;CAT;ortholog	CATALASE (PTHR11465:SF29)
EPHX2	2	Bifunctional epoxide hydrolase 2;EPHX2;ortholog	BIFUNCTIONAL EPOXIDE HYDROLASE 2 (PTHR43329:SF1)
FAH	2	Uncharacterized protein;FAH;ortholog	FUMARYLACETOACETASE (PTHR43069:SF2)
GBE1	2	1,4- α -glucan-branching	1,4-A-GLUCAN-

		enzyme;GBE1;ortholog	BRANCHING ENZYME (PTHR43651:SF1)
GBE1	2	Uncharacterized protein;LOC100157421;ortholog	SUBFAMILY NOT NAMED (PTHR10357:SF185)
LAYN	2	Uncharacterized protein;LAYN;ortholog	LAYILIN (PTHR14789:SF5)
LIPE	2	Hormone-sensitive lipase;LIPE;ortholog	HORMONE-SENSITIVE LIPASE (PTHR23025:SF6)
ORM1	2	ORM1-like protein 3;ORMDL3;ortholog	ORM1-LIKE PROTEIN 3 (PTHR12665:SF20)
PLIN1	2	Perilipin-1;PLIN1;ortholog	PERILIPIN-1 (PTHR14024:SF37)
PLIN1	2	Perilipin;PLIN1;ortholog	PERILIPIN-1 (PTHR14024:SF37)
PPP1R1B	2	Protein phosphatase 1 regulatory subunit 1B (Fragment);PPP1R1B;ortholog	PROTEIN PHOSPHATASE 1 REGULATORY SUBUNIT 1B (PTHR15417:SF7)
SGK2	2	Uncharacterized protein;SGK2;ortholog	SERINE/THREONINE-PROTEIN KINASE SGK2 (PTHR24356:SF282)
SLC36A2	2	Proton-coupled amino acid transporter 2;SLC36A2;ortholog	PROTON-COUPLED AMINO ACID TRANSPORTER 2 (PTHR22950:SF395)
UNC93A	2	Protein unc-93 homolog A;UNC93A;ortholog	PROTEIN UNC-93 HOMOLOG A (PTHR19444:SF36)
ACER3	3	Uncharacterized protein;ACER3;ortholog	ALKALINE CERAMIDASE 3 (PTHR12956:SF43)
ACKR4	3	Atypical chemokine receptor 4;ACKR4;ortholog	ATYPICAL CHEMOKINE RECEPTOR 4 (PTHR10489:SF754)
ADIPOR2	3	Adiponectin receptor 2;ADIPOR2;ortholog	ADIPONECTIN RECEPTOR PROTEIN 2 (PTHR20855:SF80)
APOL6	3	Uncharacterized protein (Fragment);APOL6;ortholog	APOLIPOPROTEIN L6 (PTHR14096:SF47)
AQP7	3	Aquaporin 7;AQP7;ortholog	AQUAPORIN-7 (PTHR43829:SF15)
ARL6IP1	3	ADP-ribosylation factor-like protein 6-interacting protein 1;ARL6IP1;ortholog	ADP-RIBOSYLATION FACTOR-LIKE PROTEIN 6-INTERACTING PROTEIN 1 (PTHR20952:SF6)
ARSD	3	Uncharacterized protein (Fragment);ARSD;ortholog	ARYLSULFATASE D (PTHR42693:SF3)
CEBPA	3	Uncharacterized protein;CEBPA;ortholog	CCAAT/ENHANCER-BINDING PROTEIN A

			(PTHR23334:SF30)
CLSTN3	3	Uncharacterized protein;CLSTN3;ortholog	CALSYNTENIN-3 (PTHR14139:SF9)
IDH1	3	Isocitrate dehydrogenase [NADP] cytoplasmic (Fragment);IDH1;ortholog	
MMD	3	Uncharacterized protein;MMD;ortholog	MONOCYTE TO MACROPHAGE DIFFERENTIATION FACTOR (PTHR20855:SF59)
PC	3	Pyruvate carboxylase, mitochondrial;PC;ortholog	PYRUVATE CARBOXYLASE, MITOCHONDRIAL (PTHR43778:SF4)
PDE8B	3	Uncharacterized protein;PDE8B;ortholog	HIGH AFFINITY CAMP-SPECIFIC AND IBMX-INSENSITIVE 3',5'-CYCLIC PHOSPHODIESTERASE 8B (PTHR11347:SF152)
PEX3	3	Uncharacterized protein;PEX3;ortholog	PEROXISOMAL BIOGENESIS FACTOR 3 (PTHR28080:SF2)
RBP4	3	Retinol-binding protein 4;RBP4;ortholog	RETINOL-BINDING PROTEIN 4 (PTHR11873:SF4)
RETSAT	3	All-trans-retinol 13,14-reductase;RETSAT;ortholog	ALL-TRANS-RETINOL 13,14-REDUCTASE (PTHR10668:SF96)
RGCC	3	Regulator of cell cycle RGCC;RGCC;ortholog	REGULATOR OF CELL CYCLE RGCC (PTHR32193:SF5)
SFRP5	3	Uncharacterized protein;SFRP5;ortholog	SECRETED FRIZZLED-RELATED PROTEIN 5 (PTHR11309:SF110)
SIK2	3	Uncharacterized protein;SIK2;ortholog	SERINE/THREONINE-PROTEIN KINASE SIK2 (PTHR24343:SF247)
TKT	3	Transketolase;tkt;ortholog	TRANSKETOLASE (PTHR43195:SF3)
ASPRV1	4	Retroviral-like aspartic protease 1;ASPRV1;ortholog	RETROVIRAL-LIKE ASPARTIC PROTEASE 1 (PTHR37006:SF2)
FLG	4	Uncharacterized protein;FLG;ortholog	
KRT1	4	Uncharacterized protein;KRT1;ortholog	KERATIN, TYPE II CYTOSKELETAL 1 (PTHR23239:SF236)
KRT77	4	Uncharacterized	KERATIN, TYPE II

		protein;KRT77;ortholog	CYTOSKELETAL 1B (PTHR23239:SF322)
KRTDAP	4	Uncharacterized protein;KRTDAP;ortholog	KERATINOCYTE DIFFERENTIATION-ASSOCIATED PROTEIN (PTHR36463:SF2)
LOR	4	Loricrin;LOR;ortholog	LORICRIN (PTHR39228:SF2)
FXYD3	5	FXYD domain-containing ion transport regulator 3;FXYD3;ortholog	FXYD DOMAIN-CONTAINING ION TRANSPORT REGULATOR 3 (PTHR14132:SF19)
KRT15	5	Uncharacterized protein;KRT15;ortholog	KERATIN, TYPE I CYTOSKELETAL 15 (PTHR23239:SF239)
KRT5	5	Uncharacterized protein (Fragment);KRT5;ortholog	KERATIN, TYPE II CYTOSKELETAL 5 (PTHR23239:SF279)
RAB25	5	Uncharacterized protein;RAB25;ortholog	RAS-RELATED PROTEIN RAB-25 (PTHR24073:SF815)
SPTLC3	5	Uncharacterized protein;SPTLC3;ortholog	SERINE PALMITOYLTRANSFERASE 3 (PTHR13693:SF74)
TACSTD2	5	Uncharacterized protein;TACSTD2;ortholog	TUMOR-ASSOCIATED CALCIUM SIGNAL TRANSDUCER 2 (PTHR14168:SF8)
TRIM29	5	Uncharacterized protein;TRIM29;ortholog	TRIPARTITE MOTIF-CONTAINING PROTEIN 29 (PTHR24103:SF530)
CMYA5	6	Uncharacterized protein (Fragment);CMYA5;ortholog	CARDIOMYOPATHY-ASSOCIATED PROTEIN 5 (PTHR24099:SF12)
LOC102160916	6	Myosin light chain 1/3, skeletal muscle isoform;MYL1;ortholog	MYOSIN LIGHT CHAIN 1/3, SKELETAL MUSCLE ISOFORM (PTHR23048:SF21)
PPP1R3A	6	Uncharacterized protein;PPP1R3A;ortholog	PROTEIN PHOSPHATASE 1 REGULATORY SUBUNIT 3A (PTHR12307:SF26)
TRDN	6	Uncharacterized protein (Fragment);TRDN;ortholog	TRIADIN (PTHR14106:SF2)
TTN	6	Titin;TTN;ortholog	TITIN (PTHR13817:SF35)
YIPF7	6	Protein YIPF;YIPF7;ortholog	PROTEIN YIPF7 (PTHR21236:SF11)
ABLIM2	7	Uncharacterized protein;ABLIM2;ortholog	ACTIN-BINDING LIM PROTEIN 2 (PTHR24213:SF21)
ADSSL1	7	Adenylosuccinate synthetase	ADENYLOSUCCINATE

		isozyme 1;ADSSL1;ortholog	SYNTHETASE ISOZYME 1 (PTHR11846:SF5)
ANKRD23	7	Uncharacterized protein;ANKRD23;ortholog	ANKYRIN REPEAT DOMAIN-CONTAINING PROTEIN 23 (PTHR24176:SF22)
BIN1	7	Uncharacterized protein;BIN1;ortholog	MYC BOX-DEPENDENT-INTERACTING PROTEIN 1 (PTHR10663:SF287)
CACNG1	7	Voltage-dependent calcium channel gamma-1 subunit;CACNG1;ortholog	VOLTAGE-DEPENDENT CALCIUM CHANNEL GAMMA-1 SUBUNIT (PTHR15025:SF9)
CAMK2A	7	Uncharacterized protein;CAMK2A;ortholog	CALCIUM/CALMODULIN-DEPENDENT PROTEIN KINASE TYPE II SUBUNIT A (PTHR24347:SF318)
GAPDH	7	Glyceraldehyde-3-phosphate dehydrogenase;GAPDH;ortholog	SUBFAMILY NOT NAMED (PTHR10836:SF69)
MYH6	7	Uncharacterized protein (Fragment);MYH6;ortholog	MYOSIN-6-RELATED (PTHR13140:SF522)
MYL2	7	Myosin regulatory light chain 2, ventricular/cardiac muscle isoform;MYL2;ortholog	MYOSIN REGULATORY LIGHT CHAIN 2, VENTRICULAR/CARDIAC MUSCLE ISOFORM (PTHR23049:SF46)
MYO18B	7	Uncharacterized protein;MYO18B;ortholog	UNCONVENTIONAL MYOSIN-XVIIIB (PTHR13140:SF576)
MYOM1	7	Uncharacterized protein;MYOM1;ortholog	MYOMESIN-1 (PTHR13817:SF28)
OBSCN	7	Obscurin;OBSCN;ortholog	OBSCURIN (PTHR10489:SF903)
PGAM2	7	Phosphoglycerate mutase 2;PGAM2;ortholog	PHOSPHOGLYCERATE MUTASE 2 (PTHR11931:SF24)
SMPX	7	Small muscular protein;SMPX;ortholog	SMALL MUSCULAR PROTEIN (PTHR17416:SF1)
SYPL2	7	Uncharacterized protein;SYPL2;ortholog	SYNAPTOPHYSIN-LIKE PROTEIN 2 (PTHR10306:SF22)
TRDN	7	Uncharacterized protein (Fragment);TRDN;ortholog	TRIADIN (PTHR14106:SF2)
XIRP2	7	Uncharacterized protein;XIRP2;ortholog	SUBFAMILY NOT NAMED (PTHR24206:SF39)
CCDC120	8	Uncharacterized	COILED-COIL DOMAIN-

		protein;CCDC120;ortholog	CONTAINING PROTEIN 120 (PTHR16093:SF7)
CEP170B	8	Uncharacterized protein;CEP170B;ortholog	CENTROSOMAL PROTEIN OF 170 KDA PROTEIN B-RELATED (PTHR15715:SF34)
GYLTL1B	8	Uncharacterized protein;GYLTL1B;ortholog	GLYCOSYLTRANSFERASE-LIKE PROTEIN LARGE2 (PTHR12270:SF40)
LIMK2	8	Uncharacterized protein;LIMK2;ortholog	LIM DOMAIN KINASE 2-RELATED (PTHR23257:SF625)
MFSD6	8	Major facilitator superfamily domain-containing protein 6;MFSD6;ortholog	MAJOR FACILITATOR SUPERFAMILY DOMAIN-CONTAINING PROTEIN 6 (PTHR16172:SF35)
PTPRF	8	Receptor-type tyrosine-protein phosphatase F;PTPRF;ortholog	RECEPTOR-TYPE TYROSINE-PROTEIN PHOSPHATASE F (PTHR19134:SF343)
RAB11FIP1	8	Uncharacterized protein (Fragment);RAB11FIP1;ortholog	RAB11 FAMILY-INTERACTING PROTEIN 1 (PTHR15746:SF26)
ADCK3 (COQ8A)	9	Uncharacterized protein;ADCK3;ortholog	ATYPICAL KINASE ADCK3, MITOCHONDRIAL (PTHR43851:SF1)
CORO6	9	Coronin;CORO6;ortholog	CORONIN-6 (PTHR10856:SF29)
MYH1	9	Myosin-1;MYH1;ortholog	MYOSIN-1 (PTHR13140:SF649)
NRAP	9	Uncharacterized protein (Fragment);NRAP;ortholog	NEBULIN-RELATED-ANCHORING PROTEIN (PTHR11039:SF55)
TNNC2	9	Troponin C, skeletal muscle;TNNC2;ortholog	TROPONIN C, SKELETAL MUSCLE (PTHR23064:SF14)
TRDN	9	Uncharacterized protein (Fragment);TRDN;ortholog	TRIADIN (PTHR14106:SF2)
APOBEC2	10	Probable C-U-editing enzyme APOBEC-2;APOBEC2;ortholog; Apolipoprotein B mRNA Editing Enzyme Catalytic Subunit 2	C-U-EDITING ENZYME APOBEC-2-RELATED (PTHR13857:SF34)
EEF1A1	10	Elongation factor 1- α ;LOC100514912;ortholog	SUBFAMILY NOT NAMED (PTHR23115:SF198)
IP6K3	10	Inositol hexakisphosphate kinase 3;IP6K3;ortholog	INOSITOL HEXAKISPHOSPHATE KINASE 3 (PTHR12400:SF57)

LOC100511157	10	ncRNA	
LOC106504926	10		
LOC106509260	10		
MYOM2	10	Uncharacterized protein (Fragment);MYOM2;ortholog	MYOMESIN-2 (PTHR13817:SF30)

After the hypothesis that there are genuine functional modules present within the overall transcriptomic dataset generated from samples collected during pre-pubertal mammary gland development was confirmed, the functional modules ('clusters') were further dissected using the MCODE algorithm and Reactome pathway analysis/Panther characterization and a review of the literature. Once MCODE identified densely connected nodes to graphically display both nodules and associated weights, which were used to sort the relationships from strongest to weakest⁷⁹², the genes contributing to each functional module were extracted for detailed pathway analysis and biological interpretation based on functional significance in mammary development and cancer if applicable. As shown in Table 10, ten functional modules were extracted from the data, ranging in MCODE weights from 26.25 (greatest connectivity) to 4 (least connectivity). The overall network connectivity coefficient for Cluster 1 indicates the highest strength of interaction between the individual genes in this module, demonstrating a novel interactions of striated muscle contraction and actin cytoskeleton integrity genes that significantly contribute to early postnatal mammary developmental trends. In contrast, the overall network connectivity coefficient for Cluster 10 indicates the lowest strength of interaction between the individual genes in this module, and pathway analysis suggests these genes have weak interaction with each other; these could not

be clearly grouped by their biological function in mammary development. Approximately 50% of the clusters contained genes with oncogenic potential, which will also be discussed.

Reactome pathway analysis confirms validity of the functional interactions (by variance) using the MCODE algorithm, by clustering of distinct pathways utilized into different modules. Cluster 1 genes utilized pathways of mainly striated muscle contraction and actin cytoskeleton integrity, and much of the activity was centered around genes encoding the proteins titin, troponin, actin, myomesin, myozenin, and troponin, with most physically located in different domains of muscle⁴⁷². Cluster 2 genes utilized pathways of adipogenesis, adipocyte metabolism, immune function, and metabolic alterations in cancer. Importantly, there was interaction with the innate immune system and >50% of the genes contained in this module were proto-oncogenes or oncogenes. Cluster 3 genes also were largely proto-oncogenes or oncogenes and utilized lipid and protein metabolism pathways and cell cycle control. Clusters 4 and 5 genes both utilized pathways of keratinocyte differentiation and development, but Cluster 5 contained entirely proto-oncogenes and oncogenes with significant integrin association. Clusters 6, 7, and 9 were weakly interacting modules of genes mainly associated with muscle contraction, actin binding, and calcium signaling. Interestingly, although weakly interacting, Cluster 8 contained genes involved in protein modification and cellular adhesion with a majority acting as proto-oncogenes or oncogenes. Lastly, the weakest interaction is represented by Cluster 10, which is a 'mixed bag' of mainly metabolic genes. Clusters 6, 7, 9, and 10, in addition to being weakly interacting modules, did not contain a majority of genes with interesting functional significance to early pre-pubertal mammary development and will not be discussed at length

here. Relevant clusters to the developmental model (Clusters 1, 2, 3, 4, 5, and 8), their functional significance, and interacting gene members will be subsequently discussed.

Cluster 1, possessing the highest MCODE connectivity weight, contained genes involved in actin cytoskeleton integrity and muscle contraction, which are both important events during postnatal mammary development. This module included the most genes: ACTA1, AMPD1, APOBEC2, ATP2A1, C14H10orf71, CAPN3, CARNS1, CASQ1, CKM, DHRS7C, EEF1A1, ENO3, MB, MYBPC1, MYBPC2, MYH1, MYOM2, MYOT, MYOZ1, NEB, NT5C1A, PDE4DIP, PYGM, TMOD4, TNNC1, TNNT1, TNNT3, TTN, TXLNB, and two uncharacterized LOC IDs. MCODE modeling has been used to isolate biomarkers and gene networks involved in breast cancer^{795–797}, but MCODE has not yet been applied to developmental mammary biology. The applicability of this multivariate technique has been confirmed for developmental biology through recognition of the known interactions of Cluster 1 genes with each other in the literature, reports of Cluster 1 gene presence in mammary epithelial cells, relationships between Cluster 1 genes and developmental DA/PRC genes, and the widely accepted importance of the actin cytoskeleton in epithelial cell adhesion and migration to promote mammary development^{152,377,798–801}. Striated muscle involvement in the mammary gland was reported as early as 1988 by Griffiths *et al*⁸⁰²; it is not surprising that this pathway is utilized by strongly coordinating genes in this instance. As reviewed in Chapter 1, mammary epithelial cells physically associate with each other through adherens junctions; this association allows formation and mediation of stable hemidesmosomes on the basal cell surface that link the cellular cytoskeleton with the BM¹³³. The actin cytoskeleton is directly associated with this process through physical support of hemidesmosomes, cell junctions, epithelial cell assembly, and cellular maintenance and

migration^{803–806}; the prominence of actin complexes within the highest weighted module further suggests significant involvement of cytoskeletal dynamics in early pre-pubertal, postnatal mammary gland development. Additionally, the actin cytoskeleton has been shown to regulate EMT, an important process in early development, in normal and cancerous cells⁸⁰⁷. Pursuant to this research focus, integrin molecules promote assembly of the actin cytoskeleton through recruitment of molecules to promote actin polymerization and relocation of actin to cell adhesion sites⁸⁰⁸; these adhesions are predominantly governed through Rac signaling^{269,809–811}. ACTA1 (skeletal muscle actin α), presented in Cluster 1, is a marker of basal epithelial cell differentiation^{812,813}, elucidating affiliation between actin dynamics and mammary epithelial cell growth and maturation. ACTA1 is also a member of the highly conserved family of actin isoforms that exhibit functional redundancy; this family includes smooth muscle α -actin (ACTA2) that has recently been proven required specifically for lactation in mammals⁸⁰⁰. Additionally, it has been shown that during postnatal development, mammary myoepithelial cells acquire a differentiated phenotype while gaining expression of smooth muscle markers, associating muscular genes with epithelial cell differentiation⁸⁰¹. It is hypothesized that a similar functional relationship is depicted with the Cluster 1 module of genes. Discriminate analysis in Fig X. showed that during phase two of development, Type VIII collagen was upregulated. Hou *et al* investigated smooth muscle cell (SMC) migration via type VIII collagen and concluded its effects are mediated by $\alpha 2\beta 1$ and $\alpha 1\beta 1$ integrin receptors as well as MMP-2 and MMP-9 expression⁴⁸⁸, further suggesting validity of the novel cluster results. A functional link between the network module in Cluster 1 and the positive PRC curve (consisting of genes supporting normal postnatal mammary development) is demonstrated by the role of PINCH-2 in both the top 25 positive PRC genes

and in the literature in actin dynamics; actin polymerization is stimulated at integrin clustering sites through via recruitment of a complex of proteins, commonly ILK–PINCH-2⁸¹⁴. This increases confidence and usability in the novel application of these multivariate techniques, frequently only used in ecological studies, to mammary biology and developmental biology. In adaptive immune system development, the actin cytoskeleton has critical roles in T-cell activation, development, and motility, helping to form immunological synapses⁸¹⁵. Furthermore, the Cluster 1 gene Elongation factor 1- α (EEF1A1) behaves as a T helper 1 (Th1) cell-specific translation factor and is involved in Th1 cytokine production⁸¹⁶. EEF1A is also commonly applied as an internal control gene when normalizing transcriptional data of mammary epithelial cells, confirming its widespread presence in the mammary gland⁸¹⁷. However, with roles in proliferation, invasion and migration of cancer cells, its overexpression has been seen in cancers of the head and neck, breast, as well as leukemia and hepatocarcinoma⁸¹⁸. Taken together, Cluster 1 module networks suggest that the muscle-related genes are, in a coordinated fashion, supporting correct actin dynamics for mammary epithelial cell migration, differentiation, and signaling; the links with published literature and developmental genes discovered using other methods (DA/PRC) confirm validity of the novel approach to characterize early pre-pubertal mammary development.

Cluster 2 is a highly interconnected module of genes utilizing metabolic and immune system pathways to execute functional roles in adipogenesis, adipocyte metabolism, and innate immune function; metabolic alterations present in cancers also appear in this module^{430,431,472}. Greater than 50% of the genes in this module are proto-oncogenes or oncogenes, leading to our characterization of this module as an ‘oncogenic cluster’. Genes included in this module are: ABHD15, ACAA1, ACO1, ADORA1, AOC3, BCAM, CAT, EPHX2, FAH,

LAYN, LIPE, ORM1, PLIN1, PPP1R1, SGK2, SLC36A2, and UNC93A; top Reactome pathways utilized were metabolism, immune system, innate immune system, and neutrophil degranulation. The significant involvement of adipogenesis in this postnatal mammary development module can be explained by the concept that shortly after birth, adipocytes densely packed in the fat pad begin to mature and become capable of supporting further mammary epithelial cell morphogenesis^{46,47}. Adipocytes were formerly viewed as a storage depot for energy; it is accepted currently that they also have endocrine functions and metabolic roles in regulating whole body homeostasis, promoting inflammation, and secretion of a range of factors⁸¹⁹. Cluster 2 suggests that as the mammary epithelial ducts penetrate the surrounding mesenchyme during development, utilizing invasive cell signaling detected with our PRC and DA analysis, surrounding fibroblasts and adipocytes play a large role in the processes^{9,10}. It has been demonstrated that the absence of adipocytes in the mammary gland results in disruption of ECM signaling events and normal mammary development, suggesting these cells do regulate and sustain epithelial cell growth^{820,821}. Furthermore, adipose-derived stem cells (ASCs) are capable of differentiating into both adipocytes and epithelial cells in the mammary gland⁸²². Literature suggests that adipocytes are critical for proper mammary gland development during puberty and for the maintenance of the ductal architecture in the adult mammary gland⁸²³; Cluster 2 supports this notion while illustrating similar functions that could be essential for early postnatal mammary gland development. Additionally, in tumors, there exists a complex metabolic collaboration between TME adipocytes and cancer cells that promotes invasiveness and metastasis⁸²⁴, suggesting potential interaction between mammary epithelial cells and adipocytes during normal mammary organogenesis. Knowledge of adipocyte metabolism and mammary

development is simultaneously improving, but many factors (and genes generated by our functional modules) remain to be studied and characterized. The Cluster 2 gene α/β -hydrolase domain containing protein 15 (ABHD15) is a direct target gene of peroxisome proliferator-activated receptor gamma (PPAR γ), the master regulator of adipogenesis^{825–827} and is upregulated during adipocyte development. Knockdown of ABHD15 *in vitro* resulted in apoptosis and a reduced proliferation capability of the cells⁸²⁸. Prokesch *et al* in 2014 presented the idea of reciprocal transdifferentiation between mammary adipocytes and epithelial cells; the researchers hypothesize it is one way in which the mammary gland accomplishes cyclical remodeling throughout development, pregnancy, and lactation⁸²⁹. Importantly, they determined Perilipin1 (PLIN1; a Cluster 2 gene) is a critical regulator of mammary transdifferentiation and detected cells with an intermediate phenotype – termed ‘adipoepithelial’ cells – in tissue during pregnancy that may be regulated by integrin complexes⁸²⁹. The presence of PLIN1 in a module aside adipogenesis genes suggests possible transdifferentiation between adipocytes and mammary epithelial cells during postnatal pre-pubertal mammary development.

Several Cluster 2 genes are overexpressed in breast cancers: adenosine A1 receptor (ADORA1) and Protein Phosphatase 1 Regulatory Inhibitor Subunit 1B (PP1R1B; also known as DARPP-32). ADORA1 is a member of the G protein-coupled receptor superfamily that may regulate both normal and tumor cell development through promotion of survival and inhibition of apoptosis. This gene acts as an E2/ER α target and a regulator of ER α transcriptional activity, promoting breast cancer growth⁸³⁰ and is overexpressed in breast cancer cell lines⁸³¹, suggesting it is a mediator of estrogen activity in mammary epithelial cell growth. PP1R1B/DARPP-32 is a signal transduction molecule, possibly linked to Wnt5a

signaling⁸³², expressed in both normal and malignant breast tissue⁸³³, with knockout *in vivo* demonstrating decreased tumor growth⁸³³. However, phosphorylation sites of DARPP-32 and their downstream effects indicate a complicated network of positive and negative feedback/feedforward loops; several studies suggest its presence actually inhibits cell growth^{833,834}, and thus its function in development remains largely unknown⁸³⁵. Detection of DARPP-32 in Cluster 2 suggests a relationship with genes regulating adipogenesis and adipocyte metabolism. The Layilin (LAYN) gene encodes a cell surface glycoprotein thought to function in cell migration by membrane anchorage and promotion of F-actin binding and integrin/focal adhesion complexes⁸³⁶; it is also a Treg cell-signature gene⁸³⁷. LAYN is present in Cluster 2, suggesting it may interact with adipocytes for cell migration during postnatal mammary development and branching of the gland. Pursuant to our focus is the activity of Basal Cell Adhesion Molecule (B-CAM; also known as Lutheran), a Cluster 2 gene proven to compete with integrin molecules for laminin $\alpha 5$ association in the BM in order to maintain balance between static and migratory cell behaviors⁸³⁸. Studies also suggest BCAM may act as a suppressive oncogene⁸³⁹ and/or join with AKT to form a BCAM/AKT fusion kinase that is constitutively activated in ovarian carcinoma⁸⁴⁰. Taken together, the results of Cluster 2 module suggest a unique collaborative relationship between adipocytes, the ECM, and mammary epithelial cells with some integrin involvement in promotion of growth of the mammary gland.

Cluster 3 contains genes mainly involved in lipid metabolism, protein metabolism, and cell cycle control. Reactome pathway analysis indicated the most common signaling pathways utilized were: Metabolism, Post-translational protein modification, Metabolism of proteins, Sphingolipid metabolism, Metabolism of lipids, Developmental Biology, and Signal

Transduction. There are also several functional immune genes present in this cluster. Genes included in this module were: ACER3, the uncharacterized LOCID LOC100739362, ACKR4, ADIPOR2, APOL6, AQP7, ARL6IP1, ARSD, CEBPA, CLSTN3, PC, IDH1, MMD, PC, PDE8B, PEX3, RBP4, RETSAT, RGCC, SFRP5, SIK2, and TKT. Immune-related genes in Cluster 3 include ACKR4 (Atypical chemokine receptor 4), CEBPA (CCAAT/enhancer binding protein A), and MMD (Monocyte-Macrophage Differentiation). Not immediately obvious from pathway analysis results was the prevalence of several important immune system relationships in this cluster, specifically those of CEBPA and MMD, which will be discussed. ACKR4 is an atypical chemokine receptor that plays a role in the migration of immune cells expressing chemokine receptors CCR7 and CCR9 by reducing the availability of CCL19, CCL21, and CCL25^{841–844} and in cancer cells, may promote metastasis via an effect on EMT signaling⁸⁴³. CEBPA, encoding the CCAAT enhancer-binding protein- α (C/EBP α) is a member of a large family of C/EBP proteins that regulate energy metabolism, immunity, inflammation, hematopoiesis, and adipogenesis^{845–849}. C/EBP α is expressed by early myeloid progenitors and may regulate monocyte and macrophage development^{850–852} and has a lesser understood, but important, impact on energy metabolism of surrounding cells⁸⁵³. Studies have shown that this gene regulates other non-immune cell types and controls genes required for lipid accumulation and glycogen storage^{854,855}, illustrating another potential association between immune development and adipogenesis in the mammary gland. However, there is conflicting evidence for the presence and role of C/EBP α in the mammary gland; some data suggest C/EBP α levels are low and static throughout mouse mammary gland development^{856,857} while others suggest dynamic upregulation during different developmental phases⁸⁵⁸. This data, indicating the presence of

this gene in a highly interconnected cluster of other adipogenesis and mammary development genes suggest that it is indeed active and sustains a functional role in development of the gland. Interestingly, a separate Cluster 3 gene is also expressed by differentiated macrophages; as its name indicates, MMD (monocyte to macrophage differentiation) supports differentiation of monocytes to macrophages after they migrate through the endothelial membrane ⁸⁵⁹. Differentiated macrophages can then subsequently secrete growth factors, cytokines, and chemoattractants to recruit additional monocytes, support smooth muscle cell development ⁸⁶⁰, and assist in lipoprotein binding through proteoglycan secretion ⁸⁶¹. A significant matrix effect, including integrin involvement, on monocyte to macrophage differentiation has been recently demonstrated ⁸⁶². In breast cancer models, initiation of mammary tumors by tumor-associated macrophages results mainly from monocyte to macrophage differentiation (reviewed by ⁸⁶³), further suggesting a role for MMD in the growth of epithelial cells and ECM integrity and function. ⁸⁶⁴ In Chapter 1, the contribution of macrophages in mammary development was reviewed ⁹¹. Taken together with reports that in the developing mammary gland, regulation of genes contributing to macrophage function (immunity and defense, lipid, fatty acid and steroid metabolism, cell adhesion, carbohydrate metabolism, amino acid metabolism, and endocytosis) are found ⁸⁶⁵, it is hypothesized that MMD has a significant role within Cluster 3. The unique network of genes presented in the Cluster 3 module further confirms the collaboration between immune cells, adipocytes, and mammary epithelial cells during early mammary organogenesis.

Cluster 4 mainly contains genes utilizing pathways of keratinization during development; these are predominantly epithelial cell markers that support keratinocyte differentiation and structural integrity of epithelial cells ^{472,603} with low oncogenic potential.

Interestingly, Cluster 5 utilizes the same pathways as Cluster 4, but is a small and highly oncogenic cluster with multiple genetic markers of basal breast cancer and EMT. The significance of the ECM, adherens junctions, and epithelial cell interactions in mammary development has already been thoroughly covered in the discussion above; the same significance applies here. The genes included in Cluster 5 are FXYD3, KRT15, KRT5, RAB25, SPTLC3, TACSTD2, and TRIM29. Importantly, 100% of the genes in this cluster have been reported to function as oncogenes or proto-oncogenes. FXYD3, also known as mammary tumor protein 8, is an enzyme and chloride ion channel or chloride channel regulator⁸⁶⁶, but also a biomarker for breast^{867,868}, endometrial⁸⁶⁹, colon and thyroid cancer⁸⁷⁰ and is suspected to support extensive cellular proliferation. KRT15, keratin 15, has been identified as a biomarker in TNBC and mammary stem cells³⁶⁴. KRT5, keratin 5, is a marker of basal epithelial differentiation that mediates cell-cell adhesion and hemidesmosome formation in conjunction with integrin molecules⁸⁷¹. As a desmosome anchor for basal cells, KRT5 has been implicated in the ECM-dependent transition of tumor cells in basal breast cancer⁸⁷². RAB25, a member of the rat sarcoma (RAS) family of small GTPases, encodes a protein whose functional role is recycling of cell surface receptors and signaling proteins and regulation of cellular processes including proliferation, signal transduction, apoptosis, microtubule organization, and integrin trafficking^{873–881}. Importantly, RAB-mediated trafficking that contributes to invasion targets and upregulates integrin $\alpha 6\beta 4$ while not affecting levels of $\alpha 3\beta 1$ integrin⁸⁷³. This gene has also been reported to act as both a tumor suppressor and tumor initiator, acting through multiple pathways to promote tumorigenesis in TNBC⁸⁸¹, suggesting a functional relationship between RAB25 and $\alpha 6\beta 4$ integrin could influence TNBC development. SPTLC3 (Serine Palmitoyltransferase Long Chain Base

Subunit 3), a sphingolipid biosynthesis gene, has been implicated in hepatocellular carcinoma⁸⁸² while TACSTD2 (Tumor-associated calcium signal transducer 2; also known as TROP2), a calcium signal transducer with roles in maintenance of the epithelial barrier through regulation of proper localization of tight junction proteins^{883,884}, has been seen overexpressed in epithelial tumors⁸⁸⁵ and TNBC^{886,887} where it promotes stem cell renewal⁸⁸⁵, EMT signaling^{888,889}, PI3K activation⁸⁹⁰, and β catenin signaling⁸⁸⁵ – all processes leading to cellular proliferation, migration, and metastasis. However, contradictory data by Wang *et al* uncover a role for TACSTD2 loss in tumorigenesis in SCC. Despite its expression in a wide range of cancers, little is known about normal TACSTD2 signaling or downstream targets and its role in cell adhesion to the ECM is unclear⁸⁸⁵. Interestingly, it was demonstrated recently that TACSTD2 inhibits prostate cancer cell adhesion to fibronectin and promotes β 1 integrin activation of the Src/FAK signaling pathway⁸⁹¹. Another related molecule also found in Cluster 5 that has an important role in EMT and regulation of ECM components is Tripartite motif-containing 29 (TRIM29)^{892–897}. It has been demonstrated that TRIM29 suppresses the activity of integrin β 1 and regulates the p63-mediated pathway in cervical cancer p63⁸⁹², reduces PTEN expression and increases phosphorylated AKT levels which induces proliferation through the PTEN/AKT/mTOR pathway in nasopharyngeal carcinoma (NPC)⁸⁹⁸, and directly binds p53 to increase proliferation by sequestering it outside the nucleus⁸⁹³. Clearly, the roles of TRIM29 across different cell types is still not completely understood. Cluster 5 is an important functional module of genes present in our transcriptomic data set – not only are all of the included genes oncogenes or proto-oncogenes, they all function mainly to maintain epithelial integrity, mediate cell-cell adhesion, and have direct links with integrin molecules.

Clusters 6-10 have weaker interactions; Clusters 6, 7, and 9 all contain genes mainly related to muscle contraction while Cluster 10 does not have strong organization. Cluster 8 is a group of genes that do have some function in cell adhesion, but the most significant functional modules from our transcriptomic dataset were Clusters 1-3 and 5. Taken together, application of the MCODE algorithm to transcriptomic data collected over time in early postnatal mammary development generated distinct modules, or clusters, containing highly interconnected genes. These clusters each had a unique biological function and strong implications for the novel mammary development model.

A model has been provided to achieve what was previously lacking in mammary biology: continuous developmental data collected from the same live animal over time. In contrast with previous attempts using necropsy samples from different animals at different ages¹⁴⁴, which does not completely control for variability between animals, mammary tissue has been obtained from the same animal throughout a period of early postnatal development without adverse effect to the animal or tissue. Anatomically and physiologically, the swine model is more closely related to that of the human; this data supports this concept and the swine model proved a good tool for histologic and transcriptomic evaluation of early mammary development¹⁵⁰. Availability of the swine genome sequence that shows high homology with humans¹⁵⁰ was especially helpful in interpretation of the data. Since the immune system of humans varies considerably from that of the mouse, the previously accepted model, this data suggests the use of swine to study immune system development in mammals. Along with the assessment of key immune regulators and markers in swine, this data supports use of the swine model to study human innate immunity and disease instead of rodents¹⁵¹. Additionally, multiple advanced multivariate techniques frequently applied to

microbial ecology and less frequently, mammary biology, were utilized to generate several models of postnatal, pre-pubertal mammary development. With the application of statistical techniques to a new area of biology, there exists concerns of reproducibility; it is important to note that these methods do not necessarily build on each other, meaning that the output of one method does not always generate the input for the next method. The whole data set, or a subset of variables sorted by variance, is used as the input for each method, so an incorrect output of an earlier method (for example, DFA) would not lead to an inherent inaccuracy in a latter method (for example, WGCNA).

As discussed, studying early mammary development is critical both due to the suspected links between development and invasive cancer and the conflicting views of researchers on the presence of developmental activity between birth and puberty. In contrast with the popular perspective that postnatally, the mammary gland exists as a rudimentary organ that remains quiescent and dormant until puberty⁶⁻⁹, this data suggests that there is significant growth occurring immediately after birth, including events previously suspected such as epithelial cell proliferation, branching, and expansion¹¹⁻¹⁴ as well as significant immune system and ECM involvement. It is also accepted that in humans, the formation of the breast in the embryo and early postnatal development contains some events akin to early breast carcinogenesis¹⁶. This data supports this concept and the main signaling pathways responsible for early postnatal mammary development and patterning over time have been identified and characterized. Awareness of their downstream effects and targets has both advanced the understanding of developmental biology and could provide future areas of study for carcinogenesis and breast cancer treatment.

Chapter two results and discussion

Given the disclosure of many proto-oncogenes or oncogenes involved in novel interactions contributing to early postnatal mammary developmental trends in Chapter 1, the hypothesis was strengthened that there are transcriptional similarities in early postnatal mammary development and TNBC that can be modulated through the $\alpha 6 \beta 4$ integrin signaling pathway. Therefore, it is expected that the MDA-MB-231 TNBC cell line should utilize growth pathways similar to those uncovered in early postnatal mammary development which are potentially affected by treatment with the recombinant rG3 protein. In this Chapter, the effect of rG3 on viability is evaluated in multiple cell lines across species, while RNA sequencing data reveal the prominent cell signaling pathways influenced by rG3 treatment, suggesting this recombinant protein can be used as a novel tool to study apoptosis signaling in TNBC.

Production of purest form of rG3

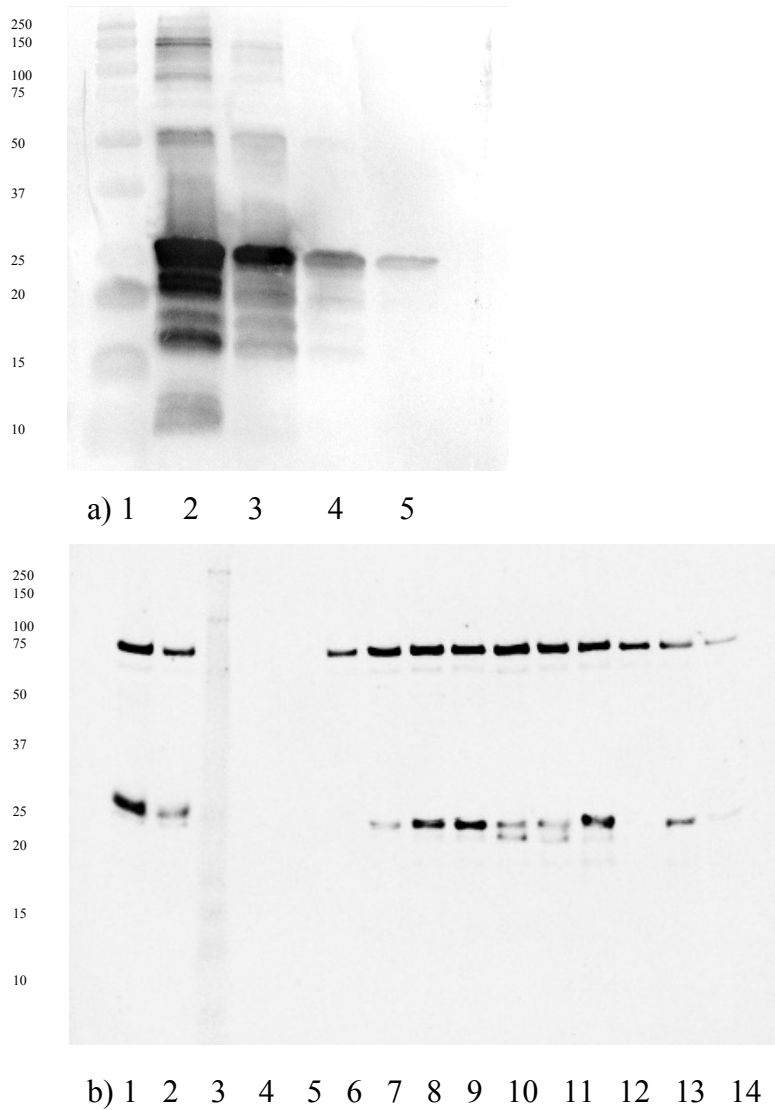


Figure 13. a) rG3 Western Blot using anti-polyhistidine monoclonal antibody after in-house production. Lane 1: Marker. Lane 2: First isolation elution. Lane 3: Second isolation elution. Lane 4: Third isolation elution. Lane 5: Fourth isolation elution. kDa size marker shown on left. b) rG3 Western Blot using anti-polyhistidine monoclonal antibody after Vanderbilt production and purification. Purest fractions were chosen for further analysis.

Fig 13. shows the purity of rG3 after production in-house or after further purification by MSRC Proteomics Laboratory at Vanderbilt University School of Medicine. In the Qiagen Fast Start Ni-NTA Kit used for in-house production, rG3 elution is repeated four times to maximize protein yield. In the first elution, the majority of rG3 is captured, and the highest concentration of protein is present at the ~27 kDa (rG3) band. The newest purified rG3 produced using Vanderbilt's technique can be seen in the Western Immunoblot in Figure 13B. Lanes 4-15 contain the fractions from purification. The cleanest fractions were chosen to be used in subsequent assays and can be seen in Lanes 8, 9, and 12. For the final experiments, production and isolation of rG3 were modified to obtain the purest form of rG3 since purification of rG3 in-house using the Ni-NTA Fast Start purification yielded a protein with two main components (as seen in Figure 13B at approximately 27 kDa and approximately 57 kDa). Since the active portion of rG3 is at approximately 27 kDa, any non-specific proteins are a concern.

rG3 isolation was confirmed by both gel electrophoresis and Western Blotting using an anti-polyhistidine monoclonal antibody (As previously demonstrated by Turner in 2005, the molecular weight of rG3 including the histidine tag is ~27 kDa). Figure X confirms the presence of rG3 with the histidine tag, but the inclusion of larger MW bands contain the histidine tag suggest the presence of larger protein fragments (~70 kDa, band 2) that are an rG3 protein complex. As shown in Figure XB, Vanderbilt's purification procedure did yield cleaner fractions. The cleanest fractions, in Lanes 8, 9, and 12, were chosen for subsequent assays. Mass spectrometry results confirmed that the upper MW band contains the chaperone heat shock Protein 60 complexed with the rG3 protein (by sequence) and the majority of the lower band is the rG3 protein sequence. Mass spectrometry, also performed at Vanderbilt,

confirmed that the upper MW band is the chaperone heat shock Protein 60 (accession number CH60_ECOLI).

Evaluation of hsp-60 effect on viability

MTT viability assay results using representative concentrations of pure heat shock protein 60 on the MDA-MB-231 cancer cells can be seen in Fig. 14.

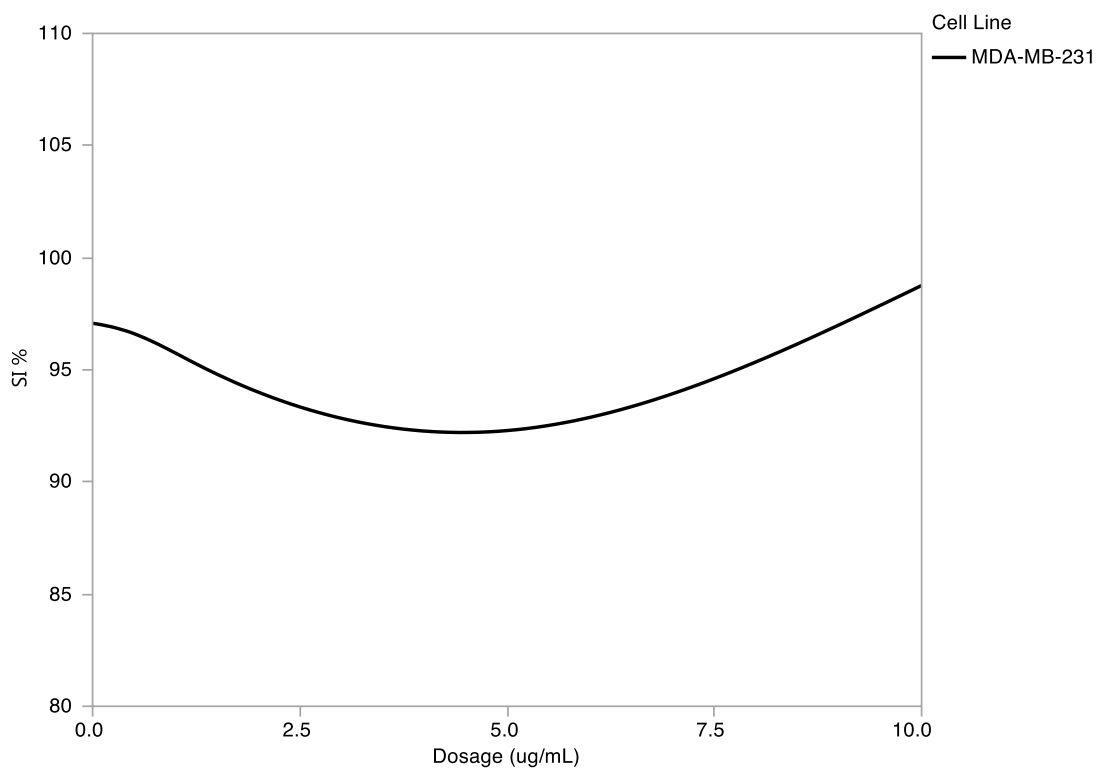


Figure 14. MTT results for MDA-M-231 cell line after 24 hr treatment in vitro with 0 – 10µg/mL pure hsp-60.

The MTT results suggest that pure hsp-60, up to 10µg/mL, does not have a significant effect on MDA-MB-231 cell viability. Therefore, it is likely that the biological

effect of rG3 is specifically due to the recombinant portion of Laminin-5 G3 domain and not the inclusion of a hsp-60/rG3 protein complex.

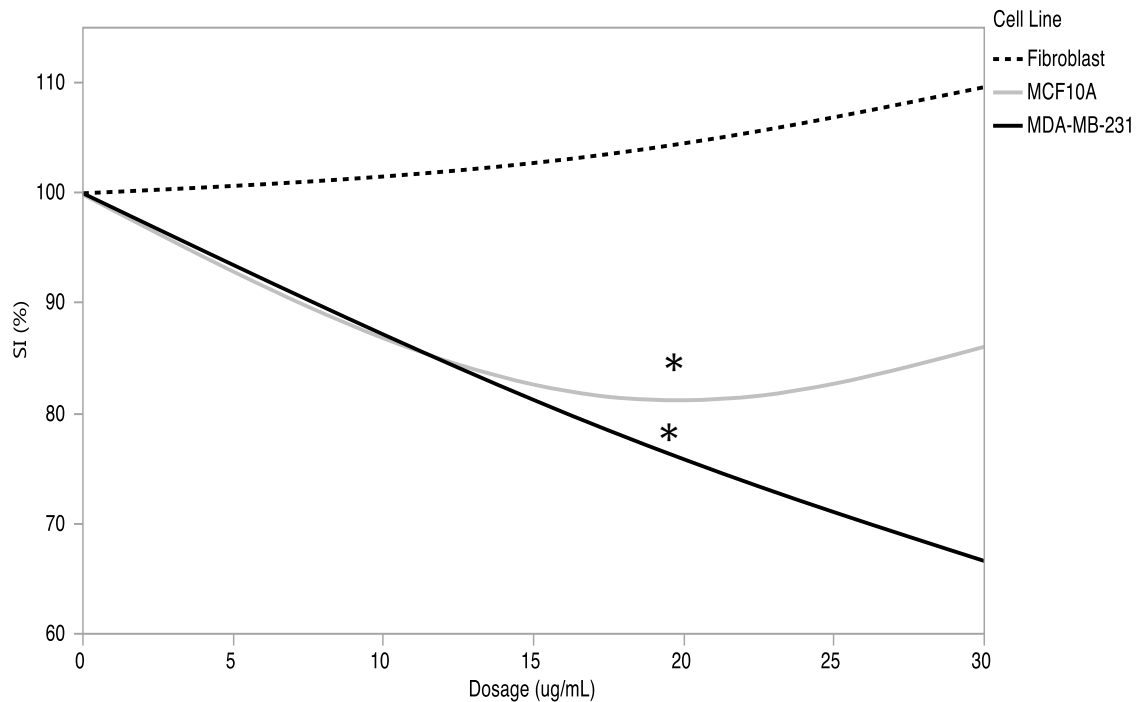


Figure 15. MTT results for MDA-M-231, MCF-10A, and primary swine fibroblast cell line after 24 hr treatment in vitro with 0 – 30 µg/mL rG3.

Evaluation of rG3 effect on TNBC viability in vitro

MTT viability assays were performed with 0, 20, 30, 40, and 50 µg/ml concentrations of rG3. As shown in Fig. 15., all concentrations of rG3 induced % SI values that were significantly different from each other. At 20 µg/mL, the 231s have the most significant reduction in viability (% SI) compared to 0 µg/mL. At 20 µg/mL, primary swine fibroblasts do not show a significant reduction in viability compared to 0 µg/mL and have a significantly

higher viability in comparison with the 231s ($p \leq 0.0001$). Although the remaining cell line (MCF-10A) has a higher viability than the 231s at 20 $\mu\text{g/mL}$, their SI values are not significantly greater than the 231 cells at this concentration.

At 20 $\mu\text{g/mL}$, the lowest concentration where significant reduction in proliferation was seen (Fig 15), MDA-MB 231 exhibit the highest inhibition of proliferation. MCF10A cells, the “normal” mammary epithelial cell line, exhibit a significant decrease in viability at 20 $\mu\text{g/mL}$, which raises questions regarding the suitability of these cells as a control as previously discussed^{146,899}. Any response of MCF10A cells to rG3 treatment may be due to the level of integrin $\beta 4$ expression on these cells (Fig. 17) being comparable to the level of integrin $\beta 4$ expression in MDA-MB-231 cancer cells.

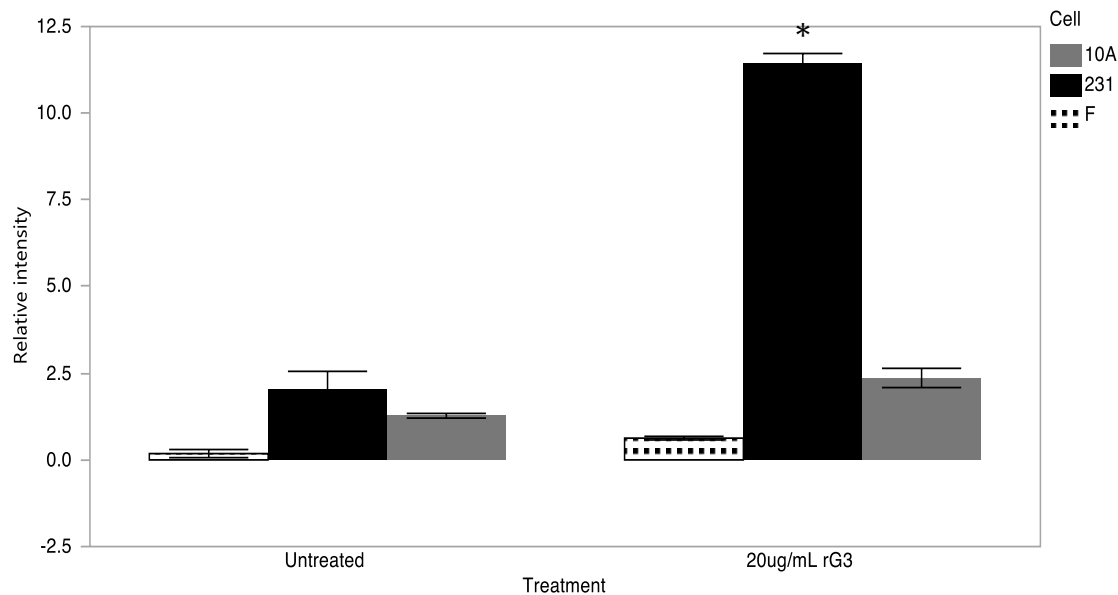


Figure 16. Western Blotting for p53 after treatment with 20ug/mL rG3 in vitro using MCF-10A, MDA-MB-231, and primary swine fibroblast lysate.

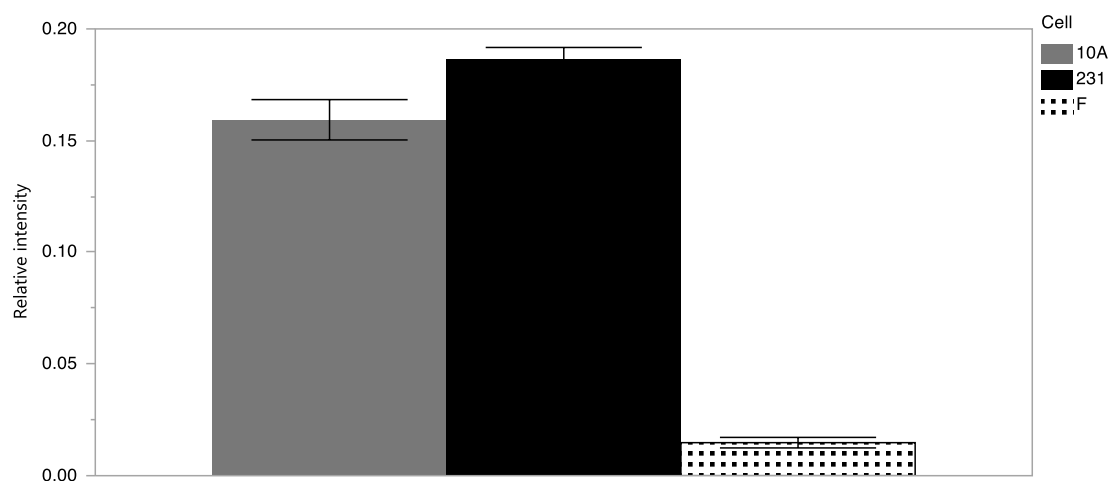


Figure 17. Western Blotting for integrin $\beta 4$ *in vitro* using MCF-10A, MDA-MB-231, and primary swine fibroblast lysate. MCF10 and MDA-MB-231 lysate showed significantly higher integrin $\beta 4$ expression in comparison with the primary swine fibroblasts ($p = 0.0016$)

Analysis of integrin $\beta 4$ and p53 protein expression after rG3 treatment in multiple cell lines

To understand at what level integrin $\beta 4$, the indirect target of rG3, is expressed in the cell lines used for *in vitro* assays, Western Immunoblotting was performed using an anti-integrin $\beta 4$ antibody on cell lysates collected from MDA-MB-231 cells, MCF-10A cells, and primary swine Fibroblasts. All images were analyzed at $n=3$ and integrin $\beta 4$ expression was normalized to β actin expression in respective samples. ANOVA indicated significant differences in means among the three different cell types, so Tukey HSD analysis was chosen to determine if tumorigenic cell types (MDA-MB-231) express significantly more integrin $\beta 4$. To understand how rG3 treatment is affecting downstream protein expression of a common apoptosis marker, p53, Western Immunoblotting was performed on cell lysates collected from MDA-MB-231 cells, MCF-10A cells, and primary swine Fibroblasts with and without 24 hr of treatment with 20 $\mu\text{g/mL}$ rG3. All images were analyzed at $n=3$ and integrin

$\beta 4$ expression was normalized to β actin expression in respective samples. ANOVA indicated differences in means among the three cell types and between treatments, so Tukey HSD analysis was chosen to determine if MDA-MB-231 cells expressed significantly more p53 after treatment than the other three non-tumorigenic cell types. In Figure 16, these differences are shown between cell lines and with or without rG3 treatment.

In Figure 17, integrin $\beta 4$ protein expression is significantly higher ($p = 0.0016$) in the MDA-MB-231 and MCF-10A cell lines in comparison to the primary swine fibroblasts, which have low integrin $\beta 4$ protein expression. MDA-MB-231 cells are the only cell line that has a significant increase in p53 expression ($p \leq 0.001$) after 24 hours of 20 $\mu\text{g/mL}$ rG3 treatment. MCF-10A cells do not indicate a significant difference in p53 protein expression after 24 hours of 20 $\mu\text{g/mL}$ rG3 treatment. Additionally, p53 expression after rG3 treatment in the MCF-10A cells is significantly lower than MDA-MB-231 cells after rG3 treatment ($p \leq 0.001$). Primary swine fibroblasts do not show significantly different p53 expression levels with or without 24 hours of 20 $\mu\text{g/mL}$ rG3 treatment. Cell death due to p53-dependent apoptosis typically follows the mitochondrial pathway, however p53 can also modulate cell death through death receptors⁹⁰⁰; it is suspected that one or both of these p53 associated pathways are utilized upon binding of rG3 to the integrin $\beta 4$ ligand. As shown in Figure 16, significantly increased p53 expression after rG3 treatment was only detected in the MDA-MB-231 cell line and not MCF10A or fibroblast cells, indicating induction of the p53-specific apoptotic pathway in MDA-MB-231 cells only. It is important to note that although MCF10A cells show expression of integrin $\beta 4$ (Fig 17) and a decrease in viability after rG3 treatment (Fig 15), a significant increase of apoptotic protein p53 is not seen after treatment. This could indicate that although rG3 affects viability of these cells, it is not through a p53-

dependent manner and therefore they are not an appropriate control or tool to study p53-mediated apoptosis induction and were not included in subsequent assays.

RNA sequencing confirms primary swine fibroblasts utilized generic growth pathways and are appropriate control

RNA sequencing provided data for approximately 25,000 gene transcripts for the primary swine fibroblast line. After 24 hours of growth *in vitro*, the most commonly utilized Reactome pathways by fibroblasts were metabolism of proteins, immune system, signal transduction, metabolism, Post-translational protein modification, Gene expression (Transcription), Innate Immune System, RNA Polymerase II Transcription, Metabolism of RNA, Cytokine Signaling in Immune system, Generic Transcription Pathway, Developmental Biology, and Adaptive Immune System.

As mentioned and seen in MTT results (Fig. 15), MCF-10A cells can exhibit stem cell-like features and possess a variable phenotype dependent on culture conditions¹⁴⁶. Due to the discrepancy in results using MCF10As, and the anatomical and physiological resemblance to human tissue^{147,150}, swine mammary tissue was collected *in vivo* and a primary cell line was established *in vitro*. Fibroblasts are the most common cells of connective tissue in mammals⁹⁰¹; these cells can direct cell-cell interactions, release various factors, and modulate ECM interactions⁹⁰². These properties led to the choice of primary swine fibroblasts as a control cell line for rG3 studies because of their close association with mammary tissue and low $\beta 4$ integrin expression, which is the indirect target of rG3 (see Fig. 17). Additionally, a primary cell line serves as a better control for *in vitro* assays than an immortalized cell line due to the loss of the p53 tumor suppressor pathway function in many

immortalized lines^{903–905} which is a major pathway focus of this dissertation. Interestingly, Reactome analysis of transcripts after RNA sequencing indicates pathways utilized by primary swine fibroblasts are very similar to those identified by K-means clustering in early mammary development, including immune system, metabolism, signal transduction, cell cycle regulation, gene transcription, and protein translation, as shown in Table 3. Importantly, TP53 and CASP genes, critical inducers of apoptosis⁹⁰⁶, are not found expressed in the top 5% of genes expressed while integrin β 4, integrin β 1, and integrin α 6 are among the top 5% genes expressed (see Appendix C). As seen in Fig. 17, integrin β 4 protein expression is significantly higher ($p = 0.0016$) in the MDA-MB-231 and MCF-10A cell lines in comparison to the primary swine fibroblasts, but RNA sequencing results confirm that integrin β 4 is still expressed to some level in this line. Panther analysis for biological process of the top 5% of expressed transcripts indicates generic cellular and metabolic processes are being executed for growth, suggesting this line is a good choice for comparison against the MDA-MB-231 cell line (Fig. 18).

Table 13: Reactome analysis of top 15 pathways utilized by fibroblasts *in vitro* after 24 hr growth

Pathway identifier	Pathway name	#Entities found	#Entities total
R-HSA-168256	Immune System	187	2765
R-HSA-1430728	Metabolism	172	3461
R-HSA-162582	Signal Transduction	169	3002
R-HSA-392499	Metabolism of proteins	119	2373
R-HSA-1266738	Developmental Biology	104	1175
R-HSA-1280215	Cytokine Signaling in Immune system	103	1222
R-HSA-74160	Gene expression (Transcription)	97	1619
R-HSA-168249	Innate Immune System	92	1632
R-HSA-73857	RNA Polymerase II Transcription	90	1440

R-HSA-597592	Post-translational protein modification	88	1530
R-HSA-556833	Metabolism of lipids	87	1316
R-HSA-449147	Signaling by Interleukins	85	910
R-HSA-212436	Generic Transcription Pathway	84	1300
R-HSA-372790	Signaling by GPCR	65	1509
R-HSA-1280218	Adaptive Immune System	57	1189

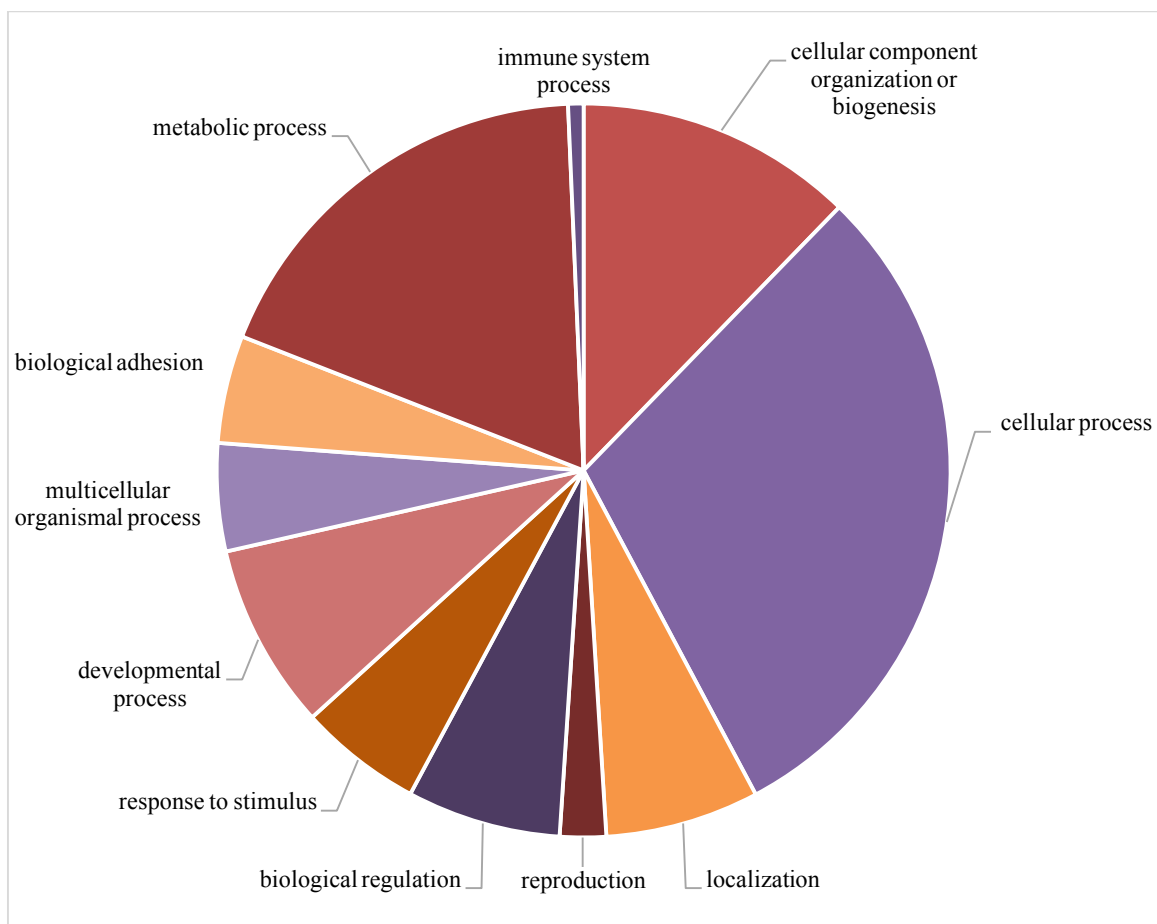


Figure 18. Panther biological Process analysis; primary fibroblasts after 24 hr growth in vitro

Swine fibroblasts are unaffected by rG3 treatment at 20ug/mL

Differential gene expression analysis performed on primary swine fibroblasts treated with 20ug/mL rG3 or with media further confirms the usage of this cell line as an *in vitro* control for rG3 studies. After normalization, the maximum shift change between treated and untreated fibroblasts was <31 with only 0.33% of the whole transcriptome resulting in a shift change >5 with rG3 treatment. For the differentially expressed genes that increased with treatment by a shift change of ≥ 5 , Reactome analysis indicated the pathways differentially utilized were: Immune System, Signal Transduction, Metabolism, Cytokine Signaling in Immune system, Metabolism of proteins, Innate Immune System, Gene expression (Transcription), RNA Polymerase II Transcription, Generic Transcription Pathway, Signaling by Interleukins, Post-translational protein modification, and Developmental Biology. A summary of rG3 effect on primary swine fibroblasts can be seen in Table 14. Taken together with the MTT results (Fig. 15), this transcriptome data indicates the primary swine fibroblasts are largely unaffected by rG3 treatment, although they express the receptor for rG3. This confirms their suitability as a control for subsequent *in vitro* assays.

Table 14. Top 75 Reactome pathways utilized by differentially expressed genes – rG3 treated fibroblasts

Pathway identifier	Pathway name	#Entities found	#Entities total
R-HSA-168256	Immune System	187	2765
R-HSA-1430728	Metabolism	172	3461
R-HSA-162582	Signal Transduction	169	3002
R-HSA-392499	Metabolism of proteins	119	2373
R-HSA-1266738	Developmental Biology	104	1175
R-HSA-1280215	Cytokine Signaling in Immune system	103	1222

R-HSA-74160	Gene expression (Transcription)	97	1619
R-HSA-168249	Innate Immune System	92	1632
R-HSA-73857	RNA Polymerase II Transcription	90	1440
R-HSA-597592	Post-translational protein modification	88	1530
R-HSA-556833	Metabolism of lipids	87	1316
R-HSA-449147	Signaling by Interleukins	85	910
R-HSA-212436	Generic Transcription Pathway	84	1300
R-HSA-372790	Signaling by GPCR	65	1509
R-HSA-1280218	Adaptive Immune System	57	1189
R-HSA-382551	Transport of small molecules	56	987
R-HSA-109582	Hemostasis	56	798
R-HSA-1643685	Disease	55	1496
R-HSA-422475	Axon guidance	51	611
R-HSA-166520	Signalling by NGF	45	615
R-HSA-5653656	Vesicle-mediated transport	43	828
R-HSA-199991	Membrane Trafficking	40	667
R-HSA-388396	GPCR downstream signaling	38	1140
R-HSA-6785807	Interleukin-4 and 13 signaling	37	212
R-HSA-1474244	Extracellular matrix organization	36	329
R-HSA-1433557	Signaling by SCF-KIT	34	468
R-HSA-186797	Signaling by PDGF	33	520
R-HSA-187037	NGF signalling via TRKA from the plasma membrane	33	527
R-HSA-194315	Signaling by Rho GTPases	32	445
R-HSA-186763	Downstream signal transduction	31	489
R-HSA-2454202	Fc epsilon receptor (FCERI) signaling	31	594
R-HSA-177929	Signaling by EGFR	31	502
R-HSA-2424491	DAP12 signaling	30	492
R-HSA-2172127	DAP12 interactions	30	509
R-HSA-112316	Neuronal System	29	425
R-HSA-6809371	Formation of the cornified envelope	28	137
R-HSA-6805567	Keratinization	28	225
R-HSA-881907	Gastrin-CREB signalling pathway via PKC and MAPK	28	504
R-HSA-194138	Signaling by VEGF	28	381
R-HSA-1989781	PPARA activates gene expression	27	177
R-HSA-400206	Regulation of lipid metabolism by Peroxisome proliferator-activated	27	180

	receptor α (PPAR α)		
R-HSA-6798695	Neutrophil degranulation	27	480
R-HSA-4420097	VEGFA-VEGFR2 Pathway	27	372
R-HSA-1852241	Organelle biogenesis and maintenance	26	334
R-HSA-8957322	Metabolism of steroids	25	316
R-HSA-983169	Class I MHC mediated antigen processing & presentation	25	464
R-HSA-5683057	MAPK family signaling cascades	25	324
R-HSA-8953897	Cellular responses to external stimuli	25	598
R-HSA-375165	NCAM signaling for neurite out-growth	24	306
R-HSA-74752	Signaling by Insulin receptor	24	349
R-HSA-76002	Platelet activation, signaling and aggregation	24	305
R-HSA-5663202	Diseases of signal transduction	24	463
R-HSA-451927	Interleukin-2 family signaling	23	288
R-HSA-512988	Interleukin-3, 5 and GM-CSF signaling	23	293
R-HSA-448424	Interleukin-17 signaling	23	332
R-HSA-5688426	Deubiquitination	23	289
R-HSA-2262752	Cellular responses to stress	23	511
R-HSA-194840	Rho GTPase cycle	22	144
R-HSA-912526	Interleukin receptor SHC signaling	22	277
R-HSA-8853659	RET signaling	22	292
R-HSA-2428928	IRS-related events triggered by IGF1R	22	325
R-HSA-2428924	IGF1R signaling cascade	22	326
R-HSA-2404192	Signaling by Type 1 Insulin-like Growth Factor 1 Receptor (IGF1R)	22	327
R-HSA-74751	Insulin receptor signalling cascade	22	324
R-HSA-425407	SLC-mediated transmembrane transport	22	413
R-HSA-500792	GPCR ligand binding	22	592
R-HSA-112315	Transmission across Chemical Synapses	22	278
R-HSA-2871796	FCERI mediated MAPK activation	21	377
R-HSA-112399	IRS-mediated signalling	21	321
R-HSA-913531	Interferon Signaling	21	292
R-HSA-3700989	Transcriptional Regulation by TP53	21	485
R-HSA-5673001	RAF/MAP kinase cascade	20	267
R-HSA-112412	SOS-mediated signalling	20	267
R-HSA-187706	Signalling to p38 via RIT and RIN	20	272
R-HSA-167044	Signalling to RAS	20	278

Table 15. Shift change values for rG3 treated fibroblasts

Condition	Shift change	GeneID	Gene Name
rG3 treated fibroblasts at 24 hours. Formula = (Fibro treated 24 hr - Fibro control 24 hr)/(Fibro control 24 hr)	31.00	RND1	Rho-related GTP-binding protein Rho6;RND1;ortholog
	13.33	EGR3	Early growth response protein 3;EGR3;ortholog
	13.25	CHI3L2	Chitinase-3-like protein 2;CHI3L2;ortholog
	13.00	KLHL9	Kelch-like protein 9;KLHL9;ortholog
	13.00	RORC	Nuclear receptor ROR-gamma;RORC;ortholog
	12.00	ALOX5AP	Arachidonate 5-lipoxygenase-activating protein;ALOX5AP;ortholog
	10.00	RASIP1	Ras-interacting protein 1;RASIP1;ortholog
	10.00	MLXIPL	Carbohydrate-responsive element-binding protein;MLXIPL;ortholog
	9.14	TRIB1	Tribbles homolog 1;TRIB1;ortholog
	9.00	NR4A3	Nuclear receptor subfamily 4 group A member 3;NR4A3;ortholog
	8.33	SLC16A1	Monocarboxylate transporter 1;SLC16A1;ortholog
	8.00	WDFY4	WD repeat- and FYVE domain-containing protein 4;WDFY4;ortholog
	8.00	TTC21A	Tetratricopeptide repeat protein 21A;TTC21A;ortholog
	8.00	GATM	Glycine amidinotransferase, mitochondrial;GATM;ortholog
	8.00	SOST	Sclerostin;SOST;ortholog
	8.00	NLRP3	NACHT, LRR and PYD domains-containing protein 3;NLRP3;ortholog
	8.00	SRCIN1	SRC kinase signaling inhibitor 1;SRCIN1;ortholog
	8.00	IQSEC3	IQ motif and SEC7 domain-containing protein 3;IQSEC3;ortholog
	7.32	DUSP5	Dual specificity protein phosphatase 5;DUSP5;ortholog
	7.00	TCF7L1	Transcription factor 7-like 1;TCF7L1;ortholog
	7.00	RASGEF1A	Ras-GEF domain-containing family member 1A;RASGEF1A;ortholog
	7.00	LTB	Lymphotoxin- β ;LTB;ortholog
	6.95	HES1	Transcription factor HES-1;HES1;ortholog
	6.60	OLFM1	Noelin;OLFM1;ortholog
	6.10	HCLS1	Hematopoietic lineage cell-specific

			protein;HCLS1;ortholog
	6.00	PIANP	PILR α -associated neural protein;PIANP;ortholog
	6.00	TMC4	Transmembrane channel-like protein 4;TMC4;ortholog
	6.00	TMEM198	Transmembrane protein 198;TMEM198;ortholog
	6.00	FAM189A2	Protein FAM189A2;FAM189A2;ortholog
	6.00	MYH7B	Myosin-7B;MYH7B;ortholog
	5.64	EGR2	E3 SUMO-protein ligase EGR2;EGR2;ortholog
	5.50	USP2	Ubiquitin carboxyl-terminal hydrolase 2;USP2;ortholog
	5.50	LRRK2	Leucine-rich repeat serine/threonine-protein kinase 2;LRRK2;ortholog
	5.33	SLC7A8	Large neutral amino acids transporter small subunit 2;SLC7A8;ortholog
	5.25	ZNF404	Zinc finger protein 404;ZNF404;ortholog
	5.25	ZNF704	Zinc finger protein 704;ZNF704;ortholog
	5.17	DOCK8	Dedicator of cytokinesis protein 8;DOCK8;ortholog
	5.00	NRARP	Notch-regulated ankyrin repeat-containing protein;NRARP;ortholog
	5.00	ARHGAP30	Rho GTPase-activating protein 30;ARHGAP30;ortholog
	5.00	KCNK1	Potassium channel subfamily K member 1;KCNK1;ortholog
	5.00	RAB37	Ras-related protein Rab-37;RAB37;ortholog
	5.00	RGL3	Ral guanine nucleotide dissociation stimulator-like 3;RGL3;ortholog
	5.00	TTC22	Tetratricopeptide repeat protein 22;TTC22;ortholog
	5.00	NCF4	Neutrophil cytosol factor 4;NCF4;ortholog
	5.00	AGER	Advanced glycosylation end product-specific receptor;AGER;ortholog

MDA-MB-231 TNBC transcriptome from in vitro growth suggests signaling similarities with early postnatal mammary development in vivo

RNA sequencing provided data for approximately 58,000 gene transcripts for the MDA-MB-231 TNBC cell line, 30,000 of which were actively expressed (raw transcript

count ≥ 1). As seen in Table 16, after 24 hours of growth *in vitro*, the most frequently utilized Reactome pathways by the TNBC cells were: Metabolism of proteins, Immune System, Signal Transduction, Metabolism, Post-translational protein modification, Innate Immune System, Disease, Gene expression (Transcription), Metabolism of RNA, Cytokine Signaling in Immune system, Developmental Biology, RNA Polymerase II Transcription, Adaptive Immune System, Signaling by Interleukins, and Generic Transcription Pathway. Importantly, the top pathways utilized by the MDA-MB-231 during normal growth and proliferation cycles are almost identical to the signaling pathways identified by K-means clustering in early mammary development AND the pathways utilized by primary swine fibroblasts during growth.

The MDA-MB-231 line is an invasive breast cancer cell line that lacks expression of estrogen (ER) α , progesterone (PR) and human epidermal growth factor receptor 2 (HER2), and is classified as a TNBC line³⁷⁸. As previously discussed, MDA-MB-231 cells express some level of integrin $\alpha 6$ and integrin $\beta 4$ ^{284,312,373–377}, and the $\alpha 6\beta 4$ complex is necessary for the tumorigenic properties of these cells³⁰³. In addition to the Western Blotting results in Fig. 17, the expression of these integrin molecules is confirmed through RNA sequencing. Importantly, TP53 and CASP genes, critical inducers of apoptosis⁹⁰⁶, are not found expressed in the top 5% of active genes (see Appendix C) expressed in untreated, normally proliferating MDA-MB-231 cells *in vitro*. Panther analysis for biological process indicates common cellular and metabolic processes are being executed for growth of this TNBC cell line *in vitro* (Fig. 19).

Table 16: Reactome pathways utilized by MDA-MB-231 cells after 24 hrs growth *in vitro*

Pathway identifier	Pathway name	#Entities found	#Entities total
R-HSA-392499	Metabolism of proteins	511	2373
R-HSA-168256	Immune System	464	2765
R-HSA-162582	Signal Transduction	427	3002
R-HSA-1430728	Metabolism	344	3461
R-HSA-597592	Post-translational protein modification	289	1530
R-HSA-168249	Innate Immune System	279	1632
R-HSA-1643685	Disease	271	1496
R-HSA-74160	Gene expression (Transcription)	255	1619
R-HSA-8953854	Metabolism of RNA	249	782
R-HSA-1280215	Cytokine Signaling in Immune system	245	1222
R-HSA-1266738	Developmental Biology	229	1175
R-HSA-73857	RNA Polymerase II Transcription	225	1440
R-HSA-1280218	Adaptive Immune System	223	1189
R-HSA-449147	Signaling by Interleukins	215	910
R-HSA-212436	Generic Transcription Pathway	193	1300
R-HSA-5653656	Vesicle-mediated transport	184	828
R-HSA-422475	Axon guidance	172	611
R-HSA-199991	Membrane Trafficking	171	667
R-HSA-5663205	Infectious disease	163	531
R-HSA-1640170	Cell Cycle	154	680
R-HSA-166520	Signalling by NGF	150	615
R-HSA-8953897	Cellular responses to external stimuli	141	598
R-HSA-69278	Cell Cycle, Mitotic	140	566
R-HSA-186797	Signaling by PDGF	135	520
R-HSA-71291	Metabolism of amino acids and derivatives	133	656
R-HSA-2262752	Cellular responses to stress	133	511
R-HSA-187037	NGF signalling via TRKA from the plasma membrane	133	527
R-HSA-382551	Transport of small molecules	133	987
R-HSA-72766	Translation	131	381

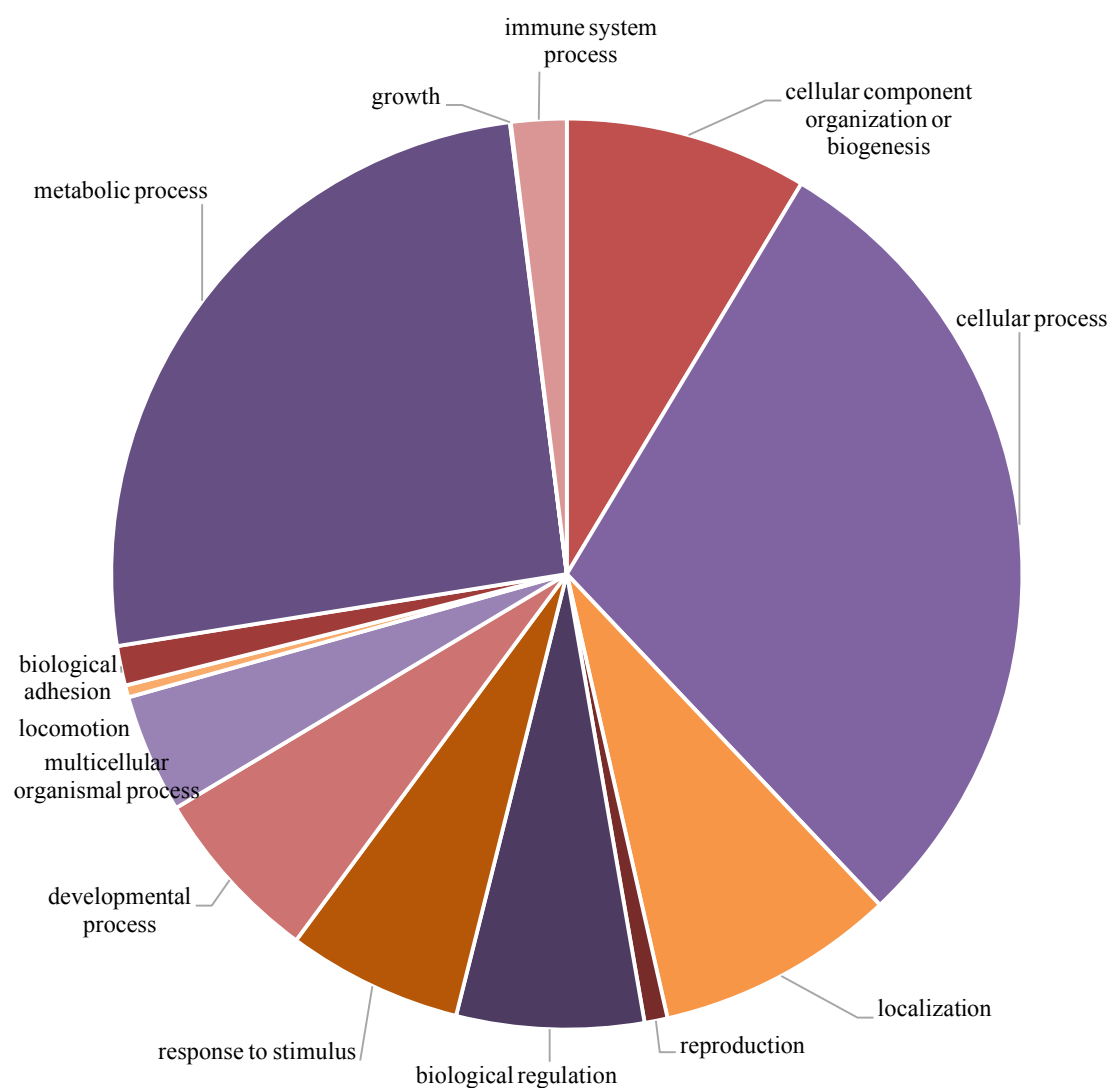


Figure 19. Panther Biological Processes utilized by MDA-MB-231 transcriptome after 24 hrs growth in vitro

Differential expression analysis of MDA-MB-231 cells with and without 20ug/mL rG3 treatment for 24 hours indicates widespread apoptotic induction and downregulation of extracellular matrix organization

In contrast with the primary swine fibroblasts (Table 14), when MDA-MB-231 cells were treated with rG3, 18% of the active transcriptome showed an upregulation in gene expression (shift change >5; maximum 14,000) and 5.8% of the active transcriptome was downregulated (shift change < -5). In contrast with fibroblast controls and untreated 231s, Reactome analysis indicates the top 75 pathways utilized by upregulated genes are indicative of apoptosis and cellular damage, as shown in Table 17. Seventy-five was chosen as a cutoff due to the decrease in functional significance of pathways below this. The top 75 pathways utilized by upregulated genes with a significant z-score (> 2), shown in Table 18, are similarly indicative of apoptosis and IFN (link to TRAIL and p53) signaling. The top 75 pathways utilized by downregulated genes with significant z-score (Table 19) are:

Extracellular matrix organization, ECM proteoglycans, Integrin cell surface interactions, Degradation of the extracellular matrix, Immune System, Signal Transduction, Non-integrin membrane-ECM interactions, Cytokine Signaling in Immune system, Collagen chain trimerization, Assembly of collagen fibrils and other multimeric structures, Collagen degradation, Collagen biosynthesis and modifying enzymes, Collagen formation, MET activates PTK2 signaling, MET promotes cell motility, and Signaling by MET. Importantly, Transcriptional regulation by TP53 is not present in this Reactome list and apoptosis and programmed cell death are not present in the top 100 pathways generated.

This indicates a much greater effect of rG3 on the transcriptome of MDA-MB-231 cells than primary swine fibroblast control cells. A literature review indicates that 56% of the

genes in the **top 0.15%** of differentially expressed genes after rG3 treatment (normalized by control) are positive regulators of apoptosis (with 10% both positive and negative), 14% are negative, and 18% are unknown or unconfirmed. When z-scores are applied, 40% of upregulated genes with a z-score >2 could function as positive apoptosis regulators.

Interestingly, genes downregulated with rG3 treatment (Table 19) predominantly fall into ECM organization and stability categories and are similar to those genes that contribute positively to early postnatal mammary gland development (as seen in Fig. 9). Since direct attachment of mammary epithelial cells to the ECM can occur through focal adhesions (FAs) and HDs, which are two adhesion complexes that form mechanical links between the ECM and the actomyosin cytoskeleton¹⁴⁰, linking the ECM and the cytoskeleton is critical in regulation of a variety of processes, including proliferation and differentiation. Additionally, apoptosis is frequently characterized largely by a large loss of cellular contact with the matrix⁹⁰⁷, so the downregulation of ECM genes after rG3 treatment of TNBC cells *in vitro* suggest that rG3 leads to decreased stability and integrity of the ECM and cell cytoskeleton, supporting apoptosis. Taken together, functional analysis of differentially expressed genes after rG3 treatment indicates widespread apoptotic induction in MDA-MB-231 cells in comparison to treated fibroblasts and suggests TP53- dependent signaling is partially responsible.

Table 17. Top 75 Reactome pathways utilized by **differentially expressed genes** in rG3 treated MDA-MB-231s

Pathway identifier	Pathway name	#Entities found	#Entities total
R-HSA-168256	Immune System	364	2765

R-HSA-1430728	Metabolism	335	3461
R-HSA-392499	Metabolism of proteins	217	2373
R-HSA-1280215	Cytokine Signaling in Immune system	208	1222
R-HSA-162582	Signal Transduction	189	3002
R-HSA-168249	Innate Immune System	177	1632
R-HSA-74160	Gene expression (Transcription)	169	1619
R-HSA-1280218	Adaptive Immune System	156	1189
R-HSA-73857	RNA Polymerase II Transcription	153	1440
R-HSA-212436	Generic Transcription Pathway	147	1300
R-HSA-597592	Post-translational protein modification	135	1530
R-HSA-556833	Metabolism of lipids	132	1316
R-HSA-913531	Interferon Signaling	117	292
R-HSA-983169	Class I MHC mediated antigen processing & presentation	109	464
R-HSA-1643685	Disease	98	1496
R-HSA-877300	Interferon gamma signaling	97	176
R-HSA-6798695	Neutrophil degranulation	95	480
R-HSA-449147	Signaling by Interleukins	90	910
R-HSA-909733	Interferon α/β signaling	90	141
R-HSA-198933	Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	87	316
R-HSA-1266738	Developmental Biology	85	1175
R-HSA-1236975	Antigen processing-Cross presentation	83	186
R-HSA-1236974	ER-Phagosome pathway	80	164
R-HSA-8953897	Cellular responses to external stimuli	79	598
R-HSA-382551	Transport of small molecules	78	987
R-HSA-2262752	Cellular responses to stress	76	511
R-HSA-983170	Antigen Presentation: Folding, assembly and peptide loading of class I MHC	76	102
R-HSA-1236977	Endosomal/Vacuolar pathway	74	82
R-HSA-1640170	Cell Cycle	71	680
R-HSA-109582	Hemostasis	68	798
R-HSA-3700989	Transcriptional Regulation by TP53	67	485
R-HSA-8953854	Metabolism of RNA	60	782
R-HSA-69278	Cell Cycle, Mitotic	57	566
R-HSA-372790	Signaling by GPCR	56	1509
R-HSA-5653656	Vesicle-mediated transport	55	828

R-HSA-199991	Membrane Trafficking	52	667
R-HSA-422475	Axon guidance	44	611
R-HSA-1428517	The citric acid (TCA) cycle and respiratory electron transport	44	219
R-HSA-72766	Translation	43	381
R-HSA-166520	Signalling by NGF	42	615
R-HSA-8957322	Metabolism of steroids	42	316
R-HSA-425407	SLC-mediated transmembrane transport	39	413
R-HSA-1852241	Organelle biogenesis and maintenance	39	334
R-HSA-68886	M Phase	39	390
R-HSA-163200	Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	39	146
R-HSA-73894	DNA Repair	38	348
R-HSA-71291	Metabolism of amino acids and derivatives	37	656
R-HSA-2172127	DAP12 interactions	37	509
R-HSA-71387	Metabolism of carbohydrates	36	414
R-HSA-211859	Biological oxidations	36	548
R-HSA-196854	Metabolism of vitamins and cofactors	36	356
R-HSA-187037	NGF signalling via TRKA from the plasma membrane	35	527
R-HSA-1474244	Extracellular matrix organization	34	329
R-HSA-186797	Signaling by PDGF	34	520
R-HSA-2424491	DAP12 signaling	34	492
R-HSA-194315	Signaling by Rho GTPases	34	445
R-HSA-388396	GPCR downstream signaling	34	1140
R-HSA-611105	Respiratory electron transport	34	115
R-HSA-177929	Signaling by EGFR	33	502
R-HSA-6785807	Interleukin-4 and 13 signaling	33	212
R-HSA-2454202	Fc epsilon receptor (FCER1) signaling	32	594
R-HSA-186763	Downstream signal transduction	32	489
R-HSA-72312	rRNA processing	32	239
R-HSA-1433557	Signaling by SCF-KIT	31	468
R-HSA-8978868	Fatty acid metabolism	30	435
R-HSA-69620	Cell Cycle Checkpoints	30	279
R-HSA-72306	tRNA processing	30	205
R-HSA-5663205	Infectious disease	29	531
R-HSA-3371556	Cellular response to heat stress	29	133

R-HSA-5368287	Mitochondrial translation	29	143
R-HSA-1483257	Phospholipid metabolism	28	315
R-HSA-446203	Asparagine N-linked glycosylation	28	420
R-HSA-5357801	Programmed Cell Death	28	187
R-HSA-109581	Apoptosis	28	179
R-HSA-76002	Platelet activation, signaling and aggregation	28	305

Table 18: Top 75 pathways utilized by upregulated genes in MDA-MB-231 cells after rG3 treatment (with significant z-score)

Pathway identifier	Pathway name	#Entities found	#Entities total
R-HSA-168256	Immune System	107	2765
R-HSA-1280215	Cytokine Signaling in Immune system	91	1222
R-HSA-1280218	Adaptive Immune System	69	1189
R-HSA-983169	Class I MHC mediated antigen processing & presentation	63	464
R-HSA-913531	Interferon Signaling	63	292
R-HSA-983170	Antigen Presentation: Folding, assembly and peptide loading of class I MHC	61	102
R-HSA-1236974	ER-Phagosome pathway	61	164
R-HSA-1236975	Antigen processing-Cross presentation	61	186
R-HSA-877300	Interferon gamma signaling	60	176
R-HSA-198933	Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	59	316
R-HSA-1236977	Endosomal/Vacuolar pathway	58	82
R-HSA-909733	Interferon α/β signaling	57	141
R-HSA-1430728	Metabolism	50	3461
R-HSA-392499	Metabolism of proteins	39	2373
R-HSA-168249	Innate Immune System	32	1632
R-HSA-449147	Signaling by Interleukins	31	910
R-HSA-1643685	Disease	30	1496
R-HSA-8953854	Metabolism of RNA	28	782
R-HSA-162582	Signal Transduction	26	3002
R-HSA-109582	Hemostasis	25	798

R-HSA-6798695	Neutrophil degranulation	20	480
R-HSA-5663205	Infectious disease	19	531
R-HSA-72312	rRNA processing	18	239
R-HSA-597592	Post-translational protein modification	18	1530
R-HSA-382551	Transport of small molecules	16	987
R-HSA-5653656	Vesicle-mediated transport	16	828
R-HSA-71387	Metabolism of carbohydrates	15	414
R-HSA-2262752	Cellular responses to stress	15	511
R-HSA-8953897	Cellular responses to external stimuli	15	598
R-HSA-1428517	The citric acid (TCA) cycle and respiratory electron transport	14	219
R-HSA-76002	Platelet activation, signaling and aggregation	14	305
R-HSA-199991	Membrane Trafficking	14	667
R-HSA-74160	Gene expression (Transcription)	14	1619
R-HSA-114608	Platelet degranulation	13	137
R-HSA-76005	Response to elevated platelet cytosolic Ca ²⁺	13	144
R-HSA-1266738	Developmental Biology	13	1175
R-HSA-447115	Interleukin-12 family signaling	12	95
R-HSA-422475	Axon guidance	12	611
R-HSA-212436	Generic Transcription Pathway	12	1300
R-HSA-73857	RNA Polymerase II Transcription	12	1440
R-HSA-6785470	tRNA processing in the mitochondrion	11	47
R-HSA-8868766	rRNA processing in the mitochondrion	11	41
R-HSA-72306	tRNA processing	11	205
R-HSA-1474244	Extracellular matrix organization	11	329
R-HSA-168255	Influenza Life Cycle	10	183
R-HSA-8950505	Gene and protein expression by JAK-STAT signaling after Interleukin-12 stimulation	10	74
R-HSA-168254	Influenza Infection	10	194
R-HSA-72766	Translation	10	381
R-HSA-3700989	Transcriptional Regulation by TP53	10	485
R-HSA-1640170	Cell Cycle	10	680
R-HSA-556833	Metabolism of lipids	10	1316
R-HSA-163200	Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling	9	146

	proteins.		
R-HSA-2172127	DAP12 interactions	9	509
R-HSA-372790	Signaling by GPCR	9	1509
R-HSA-156902	Peptide chain elongation	8	99
R-HSA-70171	Glycolysis	8	108
R-HSA-156842	Eukaryotic Translation Elongation	8	104
R-HSA-611105	Respiratory electron transport	8	115
R-HSA-70326	Glucose metabolism	8	139
R-HSA-2408522	Selenoamino acid metabolism	8	189
R-HSA-168273	Influenza Viral RNA Transcription and Replication	8	171
R-HSA-917937	Iron uptake and transport	8	83
R-HSA-6785807	Interleukin-4 and 13 signaling	8	212
R-HSA-69275	G2/M Transition	8	210
R-HSA-453274	Mitotic G2-G2/M phases	8	212
R-HSA-381119	Unfolded Protein Response (UPR)	8	160
R-HSA-71291	Metabolism of amino acids and derivatives	8	656
R-HSA-69278	Cell Cycle, Mitotic	8	566
R-HSA-2424491	DAP12 signaling	8	492
R-HSA-72689	Formation of a pool of free 40S subunits	7	106
R-HSA-72764	Eukaryotic Translation Termination	7	108
R-HSA-975956	Nonsense Mediated Decay (NMD) independent of the Exon Junction Complex (EJC)	7	102
R-HSA-156827	L13a-mediated translational silencing of Ceruloplasmin expression	7	118
R-HSA-1799339	SRP-dependent cotranslational protein targeting to membrane	7	120

Table 19: Top 75 pathways utilized by downregulated genes with significant z-score

Pathway identifier	Pathway name	#Entities found	#Entities total
R-HSA-1474244	Extracellular matrix organization	14	329
R-HSA-3000178	ECM proteoglycans	12	78
R-HSA-216083	Integrin cell surface interactions	10	87
R-HSA-	Degradation of the extracellular matrix	9	148

1474228			
R-HSA-168256	Immune System	9	2765
R-HSA-162582	Signal Transduction	9	3002
R-HSA-3000170	Syndecan interactions	7	29
R-HSA-3000171	Non-integrin membrane-ECM interactions	7	61
R-HSA-1280215	Cytokine Signaling in Immune system	7	1222
R-HSA-8948216	Collagen chain trimerization	6	44
R-HSA-2022090	Assembly of collagen fibrils and other multimeric structures	6	67
R-HSA-1442490	Collagen degradation	6	69
R-HSA-1650814	Collagen biosynthesis and modifying enzymes	6	76
R-HSA-1474290	Collagen formation	6	104
R-HSA-6785807	Interleukin-4 and 13 signaling	6	212
R-HSA-186797	Signaling by PDGF	6	520
R-HSA-449147	Signaling by Interleukins	6	910
R-HSA-8874081	MET activates PTK2 signaling	5	32
R-HSA-8875878	MET promotes cell motility	5	45
R-HSA-6806834	Signaling by MET	5	87
R-HSA-2173782	Binding and Uptake of Ligands by Scavenger Receptors	5	169
R-HSA-397014	Muscle contraction	5	216
R-HSA-76002	Platelet activation, signaling and aggregation	5	305
R-HSA-375165	NCAM signaling for neurite out-growth	5	306
R-HSA-422475	Axon guidance	5	611
R-HSA-109582	Hemostasis	5	798
R-HSA-5653656	Vesicle-mediated transport	5	828
R-HSA-1266738	Developmental Biology	5	1175
R-HSA-1643685	Disease	5	1496
R-HSA-390522	Striated Muscle Contraction	4	40

R-HSA-419037	NCAM1 interactions	4	44
R-HSA-3781865	Diseases of glycosylation	4	190
R-HSA-597592	Post-translational protein modification	4	1530
R-HSA-392499	Metabolism of proteins	4	2373
R-HSA-1430728	Metabolism	4	3461
R-HSA-445355	Smooth Muscle Contraction	3	39
R-HSA-3560782	Diseases associated with glycosaminoglycan metabolism	3	49
R-HSA-3000480	Scavenging by Class A Receptors	3	50
R-HSA-76009	Platelet Aggregation (Plug Formation)	3	52
R-HSA-8957275	Post-translational protein phosphorylation	3	109
R-HSA-381426	Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)	3	127
R-HSA-114608	Platelet degranulation	3	137
R-HSA-76005	Response to elevated platelet cytosolic Ca ²⁺	3	144
R-HSA-1630316	Glycosaminoglycan metabolism	3	183
R-HSA-202733	Cell surface interactions at the vascular wall	3	256
R-HSA-198933	Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	3	316
R-HSA-71387	Metabolism of carbohydrates	3	414
R-HSA-1280218	Adaptive Immune System	3	1189
R-HSA-3595174	Defective CHST14 causes EDS, musculocontractural type	2	9
R-HSA-3595172	Defective CHST3 causes SEDCJD	2	9
R-HSA-3595177	Defective CHSY1 causes TPBS	2	10
R-HSA-430116	GP1b-IX-V activation signalling	2	12
R-HSA-2022923	Dermatan sulfate biosynthesis	2	13
R-HSA-2214320	Anchoring fibril formation	2	15
R-HSA-75892	Platelet Adhesion to exposed collagen	2	16
R-HSA-3560801	Defective B3GAT3 causes JDSSDHD	2	21

R-HSA-4420332	Defective B3GALT6 causes EDSP2 and SEMDJL1	2	21
R-HSA-3560783	Defective B4GALT7 causes EDS, progeroid type	2	21
R-HSA-2243919	Crosslinking of collagen fibrils	2	24
R-HSA-2022870	Chondroitin sulfate biosynthesis	2	25
R-HSA-2024101	CS/DS degradation	2	29
R-HSA-1971475	A tetrasaccharide linker sequence is required for GAG synthesis	2	31
R-HSA-2129379	Molecules associated with elastic fibres	2	38
R-HSA-1566948	Elastic fibre formation	2	46
R-HSA-114604	GPVI-mediated activation cascade	2	64
R-HSA-1793185	Chondroitin sulfate/dermatan sulfate metabolism	2	73
R-HSA-8936459	RUNX1 regulates genes involved in megakaryocyte differentiation and platelet function	2	78
R-HSA-1638091	Heparan sulfate/heparin (HS-GAG) metabolism	2	88
R-HSA-8878171	Transcriptional regulation by RUNX1	2	260
R-HSA-212436	Generic Transcription Pathway	2	1300
R-HSA-73857	RNA Polymerase II Transcription	2	1440
R-HSA-74160	Gene expression (Transcription)	2	1619
R-HSA-1566977	Fibronectin matrix formation	1	7
R-HSA-3656244	Defective B4GALT1 causes B4GALT1-CDG (CDG-2d)	1	9
R-HSA-3656225	Defective CHST6 causes MCDC1	1	9

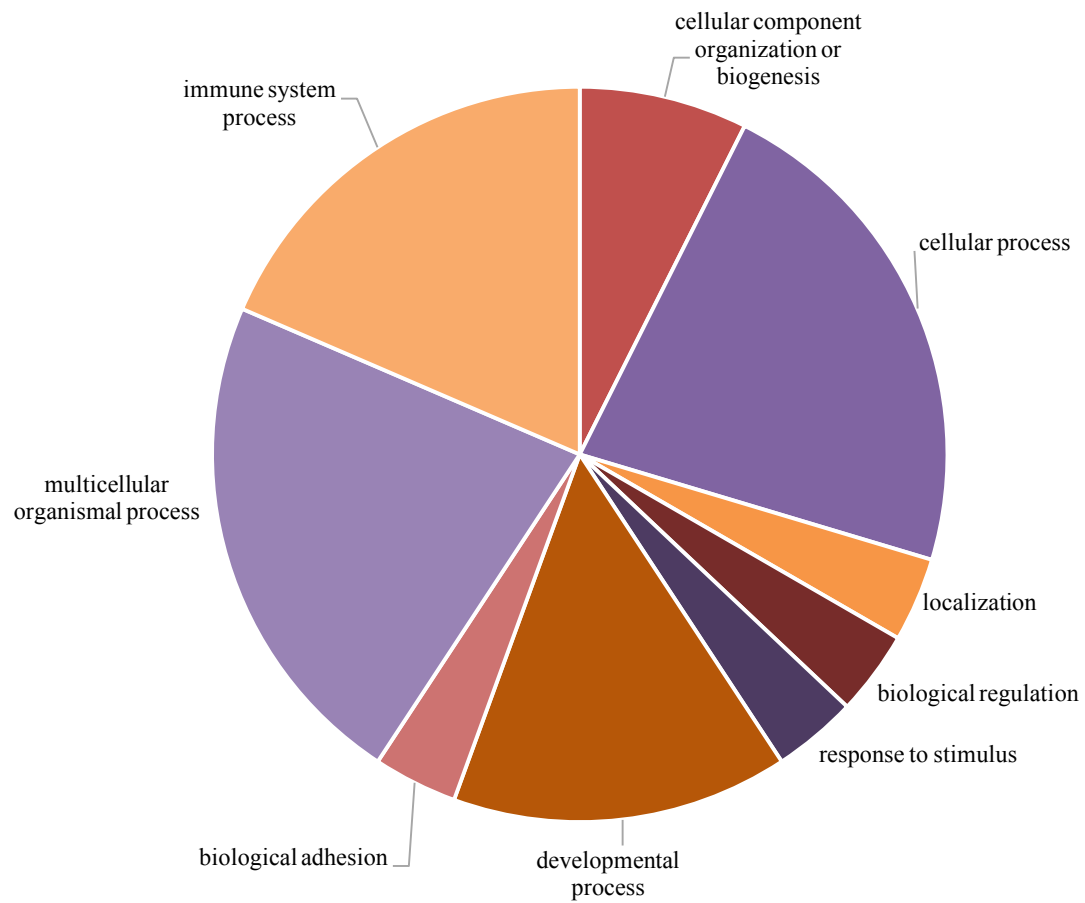


Figure 20. Panther Biological Process utilized by top genes expressed with negative z-score (downregulated) in MDA-MB-231 cells in vitro

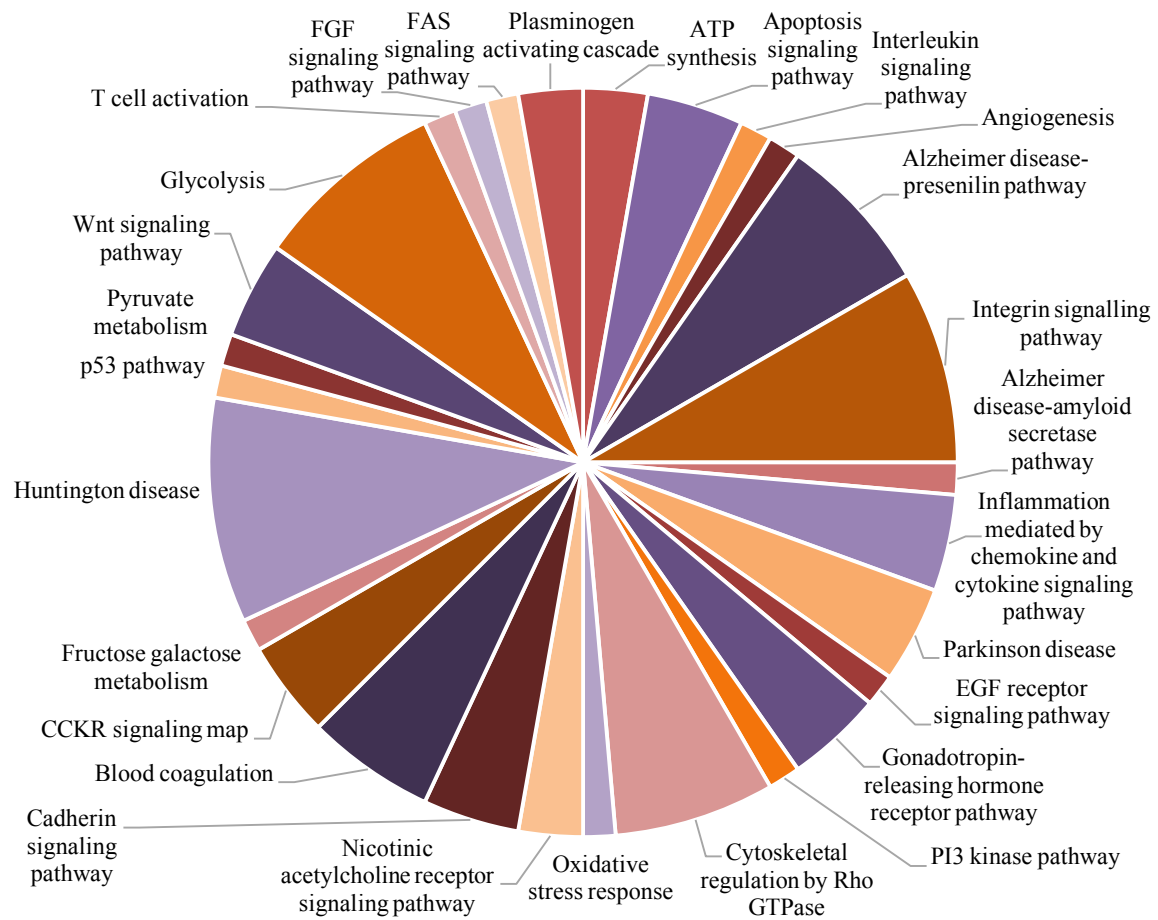


Figure 21. Panther Pathway utilized by top genes expressed with positive z-score (upregulated) in MDA-MB-231 cells in vitro

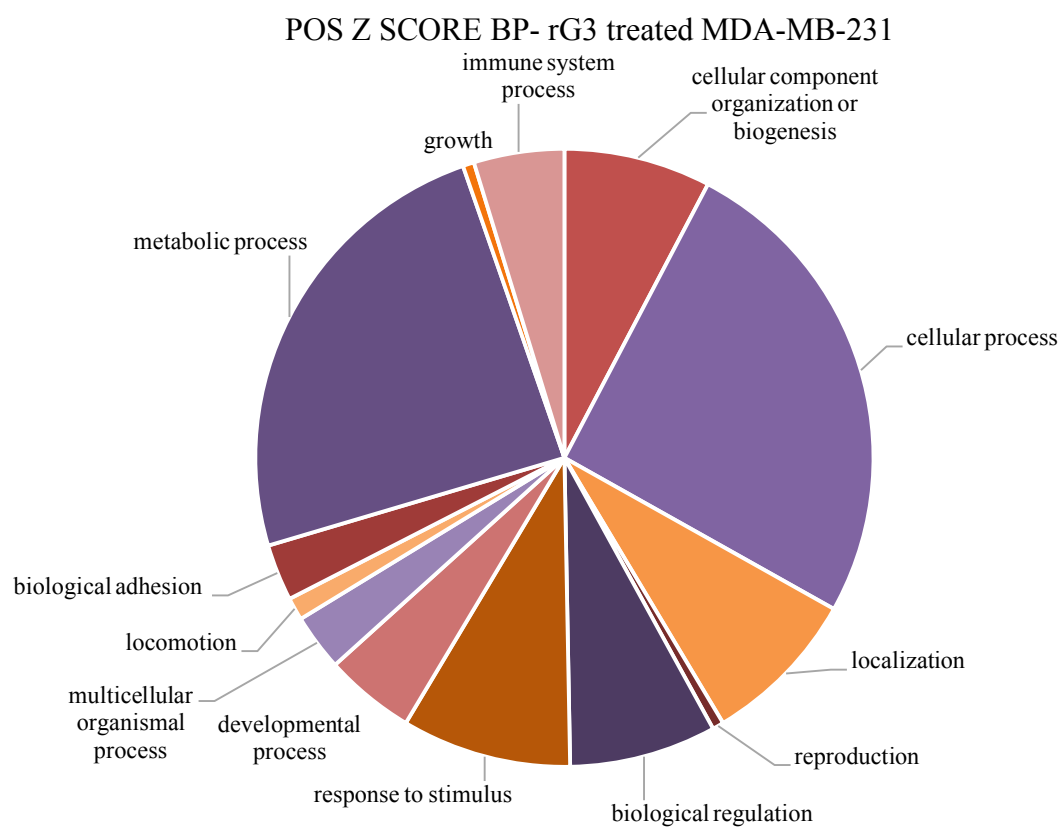


Figure 22. Panther Biological Process utilized by top genes expressed with positive z-score (upregulated) in MDA-MB-231 cells in vitro

rG3 binds to integrin α -6 β -4 to modulate downstream signaling

Integrin transcript expression, shown in Figure 23, indicates that rG3 highly upregulates the expression of ITGA2, ITGA3, ITGA6, ITGB1, ITGB4, and ITGB5 on the MDA-MB-231 cell surface after 24 hours of *in vitro* treatment. Additionally, the α 6 β 4 signaling pathway specific to rG3 originally proposed is shown in Fig. 24 with RNA sequencing results applied. Taken together with the heat map in Figure 24, these results show

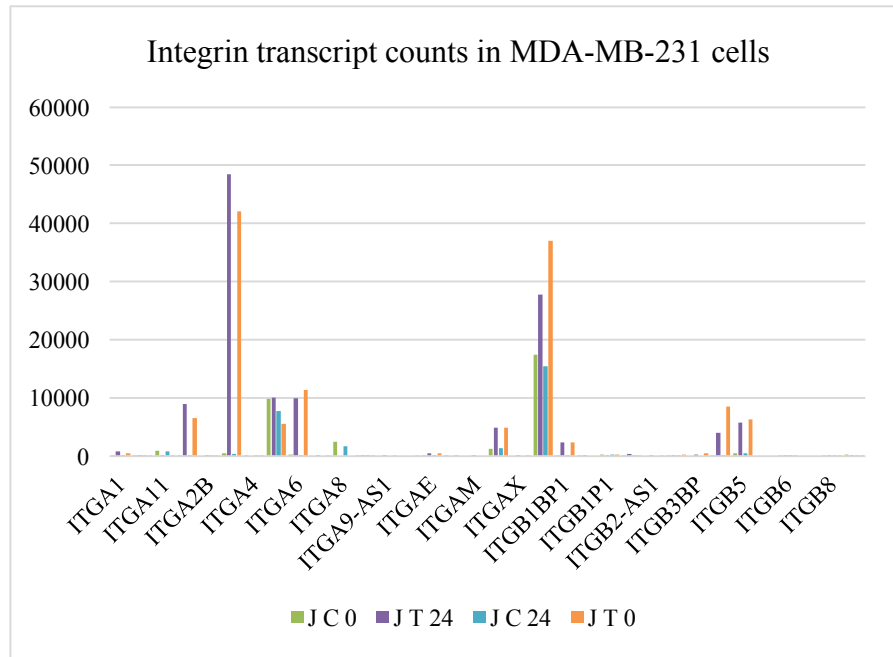


Figure 23. Integrin transcript expression in MDA-MB-231 cells at 0 and after 24 hours of *in vitro* treatment. JC = control; JT = rG3 treated.

modulation of the PI3K/AKT pathway through integrin expression by rG3 treatment.

As previously discussed, any normal cellular signaling and ECM pathways containing integrin receptors are deregulated in cancer to allow cellular migration, as evidenced by differential integrin expression among normal breast, simple hyperplasia,

atypical hyperplasia, ductal carcinoma *in situ* (DCIS) and invasive tumors³⁰⁷. $\alpha 6\beta 4$ integrin is thought to remodel the ECM through interaction with the PI3K and RhoA cell signaling pathways^{308,309}, supporting PI3K activation through different receptor tyrosine kinases (RTKs)³¹². targeted therapies¹³⁷. Notably, $\alpha 6\beta 4$ integrin may modulate oncogenic signaling through association with Laminin molecules in the ECM. Modulation of integrin transcript expression after rG3 treatment indicates that this recombinant form of Laminin-5 is binding to or influencing integrin expression in MDA-MB-231 cells.

Title: Alpha 6 Beta 4 signaling - rG3 M
Organism: Homo sapiens

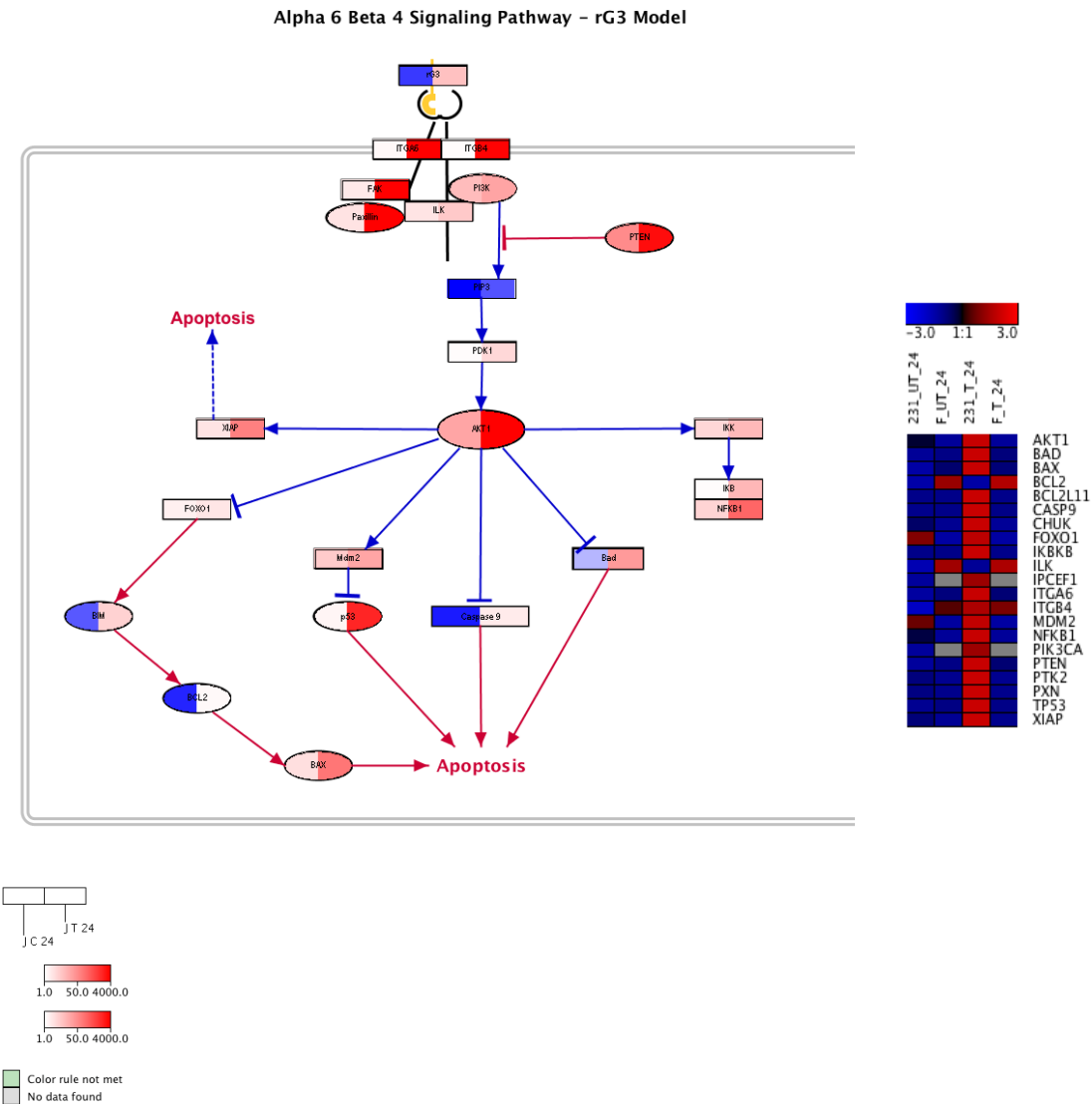


Figure 24: $\alpha 6\beta 4$ integrin modulation and signaling in MDA-MB-231 cells with and without rG3 treatment. Left half of node represents untreated cells; right half of node represents treated cells. Blue represents expression levels of 0-1. Shades of red represent high expression levels.

Modulation of the $\alpha 6\beta 4$ integrin signaling pathway through rG3

RNA seq transcript count values between rG3 treated and untreated MDA-MB-231 cells were used to curate pathways using PathVisio and WikiPathways^{435–438}; raw counts for rG3 treated and untreated MDA-MB-231 cells and primary swine fibroblasts were normalized and prepared into heat maps with Genesis⁴²⁸. Blue = downregulation/low expression; red = upregulation /high expression; Pathvisio gray = no change; heatmap gray = gene not found. For comparison of human and swine genes, swine gene orthologs were included if possible using OrthoDB⁴³⁹. After rG3 treatment, upregulation of multiple members of the PI3K/AKT pathway is shown in Fig 24. Upregulated gene transcripts include ITGA6 and ITGB4; especially important for apoptosis are the upregulated gene transcripts of BIM, BCL2, BAX, P53, BAD, and CASP9.

Upon binding to its ligand, ITGA6 and ITGB4 are likely upregulated due to “inside-out” signaling characteristic of integrin molecules⁹⁰⁸ that allow intracellular signals to increase the affinity of integrin molecules to their extracellular ligands. Downstream, our data suggests slight upregulation in PTEN, a multifunctional tumor suppressor previously discussed^{337,909–911}. However, pro-survival signaling molecules such as Paxillin and AKT1 are not downregulated, but pro-apoptosis molecules downstream of AKT1 including BIM, BCL2, BAX, P53, BAD, and CASP9 are upregulated with rG3 treatment. This suggests that the signaling mechanisms between ITGA6 and ITGB4 induction and apoptosis induction are more complicated than the original hypothesis, and will be elucidated using a more comprehensive apoptosis modulation pathway.

Differential expression analysis of MDA-MB-231 cells with and without 20ug/mL rG3 treatment for 24 hours indicates widespread apoptotic signaling and modulation of cell death pathways

RNA seq transcript count values between rG3 treated and untreated MDA-MB-231 cells were used to curate pathways using PathVisio and WikiPathways^{435–438}; raw counts for rG3 treated and untreated MDA-MB-231 cells and primary swine fibroblasts were normalized and prepared into heat maps in Genesis⁴²⁸. Blue = downregulation/low expression; red = upregulation /high expression; Pathvisio gray = no change; heatmap gray = gene not found. For comparison of human and swine genes, swine gene orthologs were included if possible using OrthoDB⁴³⁹. Pathway analysis demonstrates the widespread apoptotic signaling that occurs with rG3 treatment after 24 hours (Figure 25).

A better understanding of the diverse modes of cell death in TNBC cells could provide a molecular basis for new strategies targeting apoptotic molecules and programmed cell death pathways. Our results elucidate further the cell signaling pathways and factors that may modulate apoptosis; widespread apoptotic signaling is seen after rG3 treatment in MDA-MB-231 cells, confirming the results of rG3 effect on cell viability in Fig 25. Apoptosis, or programmed cell death, is tightly regulated process possessing a critical role in cellular functions such as cell turnover, homeostasis, growth regulation, early development, immune system function, and disease^{906,912–916}. This defensive mechanism acts to remove infected, mutated, or damaged cells in response to a variety of internal and external stimuli, and is characterized primarily by a large loss of cellular contact with the matrix⁹⁰⁷. Apoptosis can occur extrinsically, through activation of death receptors, or intrinsically, through the mitochondria^{913,914,917}; evidence supports that in some cases these two pathways may be

converging rather than distinct entities ⁹⁰⁷. As shown in Fig 25 and 26, rG3 treatment of TNBC cells *in vitro* leads to upregulation of multiple apoptotic molecules and modulation of both intrinsic and extrinsic apoptosis, suggesting crosstalk between the two cell signaling pathways is induced by rG3 treatment. A series of steps in either the extrinsic or intrinsic pathway leads to activation of effector cysteine protease molecules known as caspases. In a cascade of apoptotic activity, caspase molecules cleave both each other and nuclear and cytoplasmic substrates responsible for the maintenance of nuclear integrity, cell cycle progression, and DNA repair ⁹¹⁸. Although less frequently, apoptosis can occur independently of caspase molecules ⁹¹⁹.

Caspases can be initiated extrinsically in several different ways, including ligation of transmembrane members of the TNF receptor superfamily, specifically CD95/Fas or TNF-related apoptosis-inducing ligand (TRAIL) ⁹²⁰⁻⁹²². Upregulated apoptotic molecules after rG3 treatment include the transmembrane members of the TNF receptor superfamily capable of downstream apoptosis induction (TRAIL, TRAIL-R4, TRAIL-R1, TRAIL-R2, TNFR-1, TNFR-2) and FAS. These data suggest that rG3 treatment is initiating caspases extrinsically, which leads to downstream apoptotic signaling modulation. Extrinsic initiation results in activation of the initiator caspase-8, which cleaves downstream effector caspases including caspase-3 ⁹²⁰. Results show that after rG3 treatment, through activation of the TRAIL receptors, FADD, CASP8, and CASP10 are upregulated in comparison to untreated TNBC cells.

A wide range of signals in the extrinsic apoptosis pathway are stimulated by p53, the most extensively studied and tightly regulated tumor suppressor that can be activated by DNA damage, hypoxia, or oncogene expression ⁹²³. Our results also show an upregulation of

TP53 and TP53 targets (Fig. 26), after rG3 treatment, indicating this tumor suppressor is highly involved in apoptosis induced by rG3. When triggered by stress, p53 selectively regulates cell cycle arrest, DNA repair, and apoptosis to protect genomic stability and prevent tumor formation²⁶⁴; p53 is often mutated in metastatic tumors; somatic p53 mutations occur in greater than 50% of tumors and almost every cancer type^{235,265,266}. p53 is important in therapeutic development as p53-dependent apoptosis contributes to cell death induced by chemotherapy⁹²⁴ and loss of p53 function has been associated with chemoresistance in certain tumor types^{925,926}. Research suggests that p53-mediated apoptosis proceeds primarily through the intrinsic apoptotic pathway⁹²⁷, but the extrinsic pathway is also regulated by p53 though poorly understood⁹²³. A greater understanding of the p53 modulated apoptosis pathways and their crosstalk in response to rG3 treatment suggests rG3 is a valuable tool for understanding the molecular mechanisms of cancer development, diagnosis, and therapy.

The most significant link between p53 activation and apoptosis comes from its ability to control pro-apoptotic members of the Bcl-2 family, including Bax, Noxa, Puma, and Bid⁹²⁸⁻⁹³². The intrinsic apoptosis pathway relies upon the the Bcl-2 protein family for release of cytochrome c from the mitochondria⁹³³⁻⁹³⁵, further linking p53 activation to intrinsic apoptosis. Much remains to be uncovered regarding the mechanism of apoptotic control by p53 targets, but it is accepted that the net effect is increase of the ratio of pro- to anti-apoptotic proteins, shifting the balance to favor caspase activation and cell death⁹²³. Bax-induced release of cytochrome c from the mitochondria subsequently activates caspase-9^{929,936}. Interestingly, Bax has been recently shown to participate in apoptosis as an indirect p53 target with the primary target being PUMA/BBC3⁹³⁷. Noxa is thought to be involved in p53 mediated apoptosis in a similar fashion to PUMA/BBC3 and Bax⁹³⁸. The pro-apoptotic

protein Bid has a unique role in uniting the extrinsic death pathway to the intrinsic pathway through activation of mitochondrial disruption⁹²⁹; impact of both pathways may be enhanced when these converge through Bid⁹²⁹. After rG3 treatment of MDA-MB-231 cells *in vitro*, intrinsic apoptotic factors are highly upregulated, including BID, BAK1, BOK, BNIP3, NOXA, PUMA, CYTOC, and CASP9 (seen in Fig 25 and 26 in red). Interestingly, although anti-apoptotic BCL-XL⁹³⁹ is upregulated after rG3 treatment, expression of BAK and BOK is not inhibited. Modulation of apoptosis through cell cycle regulation is an important cell signaling event, and the cell cycle regulator p16(INK4A) (p16) has been reported to be upregulated in cervical cancer tissues where it promotes cell survival⁹⁴⁰. After rG3 treatment, expression of p16 is highly downregulated, suggesting it no longer can positively regulate growth. Importantly, we do not see the same upregulation of pro-apoptotic factors in rG3 treated fibroblasts, although these genes are present; they remain mostly unchanged, as seen in Fig. 25.

Extrinsically, p53 can induce mRNA expression of Fas receptor (CD95/Apo-1), a member of the TNF-R family of receptors and a key component of the extrinsic death pathway that is activated by binding of its ligand expressed predominantly on T-cells, FasL/TNFSF6⁹⁴¹⁻⁹⁴³. When overexpressed, p53 can also stimulate Fas levels on the cell surface, rapidly sensitizing cells to Fas-induced apoptosis before transcription-dependent effects are seen⁹⁴⁴. In human tumors, there is often a decrease in or disappearance of the expression of Fas receptor (Apo-1; CD95) on the cell surface, inhibiting or decreasing Fas-mediated death signaling that normally would recruit FADD and caspase-8 to the Fas receptor to form the death-inducing signaling complex^{945,946}. In some cases, loss of Fas expression can be a result of silencing by STAT-3 and CJUN, which is regulated by PI3K

signaling^{947,948}. FAP-1 has been shown to suppress Fas-mediated apoptosis, but the specific mechanisms of inhibition remain unknown⁹⁴⁹. FAP-1, in association with either DPP4, PLAUR or integrin molecules, is a negative regulator of apoptosis that is involved in proteolysis of the extracellular matrix to promote cell adhesion, migration, and invasion; it can also enhance tumor growth progression and reducing antitumor response of the immune system. During autophagy, FAP-1 protein is degraded, promoting apoptosis⁹⁵⁰. Our results show that FAP-1 is highly downregulated during rG3 induced apoptosis where it is not able to efficiently inhibit FADD/CASP8 signaling. PRKD1, another molecule capable of inhibiting FADD/CASP8 signaling, is linked to increased proliferation and survival when activated or expressed in various cell lines⁹⁵¹⁻⁹⁵⁷. After rG3 treatment in vitro, PRKD1 expression is highly downregulated, indicating its contribution to the decrease in cell survival.

p53 can also induce TRAIL-R2, as confirmed by our results, which is the death-domain-containing receptor for TNF-related apoptosis-inducing ligand (TRAIL)⁹⁵⁸. PERP, p53 apoptosis effector related to PMP-22, is another gene that is a direct p53 transcriptional target and novel effector of apoptosis⁹⁵⁹, but it remains to be defined how specifically this gene contributes to p53-mediated apoptosis⁹⁶⁰. However, it is known that PERP is induced during apoptosis only and not during cell cycle arrest⁹⁶¹. After rG3 treatment in vitro, the PERP gene is highly upregulated (see Appendix C). Elevation of the PERP protein positively influences p53 through enhancement of nuclear localization and stabilization, revealing a potential target for exploitation in enhancing p53 activity⁹⁶¹. Normally, caspase-8 and -10 respond to external death ligands while caspase-2 and -9 sense changes in mitochondrial potential from the intrinsic pathway⁹²⁹. After treatment, upregulation of

CASP8, CASP10, CASP2, and CASP9 is seen, in addition to upregulation of CASP7, CASP3, CASP4, CASP6, and DFFA indicating rG3 activates both the extrinsic and intrinsic apoptotic pathway and modulates caspase signaling (Fig. 25/26).

P53, in addition to its ability to activate genes directly supporting apoptosis, can also regulate genes that elude typical anti-apoptotic pathways. Pursuant to our focus, p53 has the ability to regulate PTEN, the negative regulator of the PI3K/AKT pathway⁹⁶² Normally, the PI3K pathway phosphorylates the major inhibitor of p53, Mdm2, in order to promote cell survival under normal conditions⁹²⁹. During stress-induced activation of p53, however, caspase-mediated cleavage and degradation of the AKT protein^{929,963} and expression of the PTEN tumor suppressor gene (which dephosphorylates PI3K), are induced, impairing PI3K/AKT signaling function⁹⁶⁴. Our data show upregulation of PTEN and modulation of the PI3K cell signaling pathway after rG3 treatment (Fig. 24). Taken together, this indicates that p53-dependent apoptosis induced by rG3 treatment allows cells to counteract survival signals - although still present - possibly reducing the threshold required for pro-apoptotic factors to activate apoptotic signaling and again shifting the balance from survival to apoptosis. This data, demonstrating widespread apoptotic signaling, suggests that by simultaneously targeting multiple pathways of the apoptotic process, p53 increases the probability of successful programmed cell death. The ability of rG3 to induce such a widespread p53-dependent effect makes it an incredibly valuable research tool. Additionally, since different cell types have different induction programs after p53 activation⁹²³, rG3 could be used as a targeted tool to study apoptosis across multiple tissue types and species.

Interestingly, some upregulation of pro-survival factors can be seen with rG3 treatment, though the effect is not extreme in comparison to the upregulation of pro-apoptotic factors (Fig. 24), suggesting that apoptosis is about shifting a balance between two programs, rather than specifically silencing all growth signals. An important PI3K/AKT factor that is downregulated with rG3 treatment is PIK3R2, which normally promotes cell growth and survival ⁹⁶⁵. Interestingly, GSK3 is also downregulated with rG3 treatment. GSK3 hyperactivation has been implicated in inflammation, neurological disorders, and tumorigenesis ⁹⁶⁶. It is also implicated in the Wnt- β -catenin signaling pathway with a positive role in Wnt signal transduction by phosphorylating the Wnt receptors low density lipoprotein receptor-related protein (LRP5/6) ⁹⁶⁶. As seen in Fig. 24/25, MDM2 is slightly regulated with treatment, but possibly cannot inhibit p53 efficiently due to highly upregulated p53 activator molecules such as BAX.

Taken together, the modulation of the apoptotic pathway seen with rG3 treatment, especially the induction of p53-dependent signaling, suggests this recombinant protein has significant potential for use as a research tool to further study the complicated interaction of apoptotic signaling pathways.

Title: TP53 Network
 Availability: CC BY 2.0
 Organism: Homo sapiens

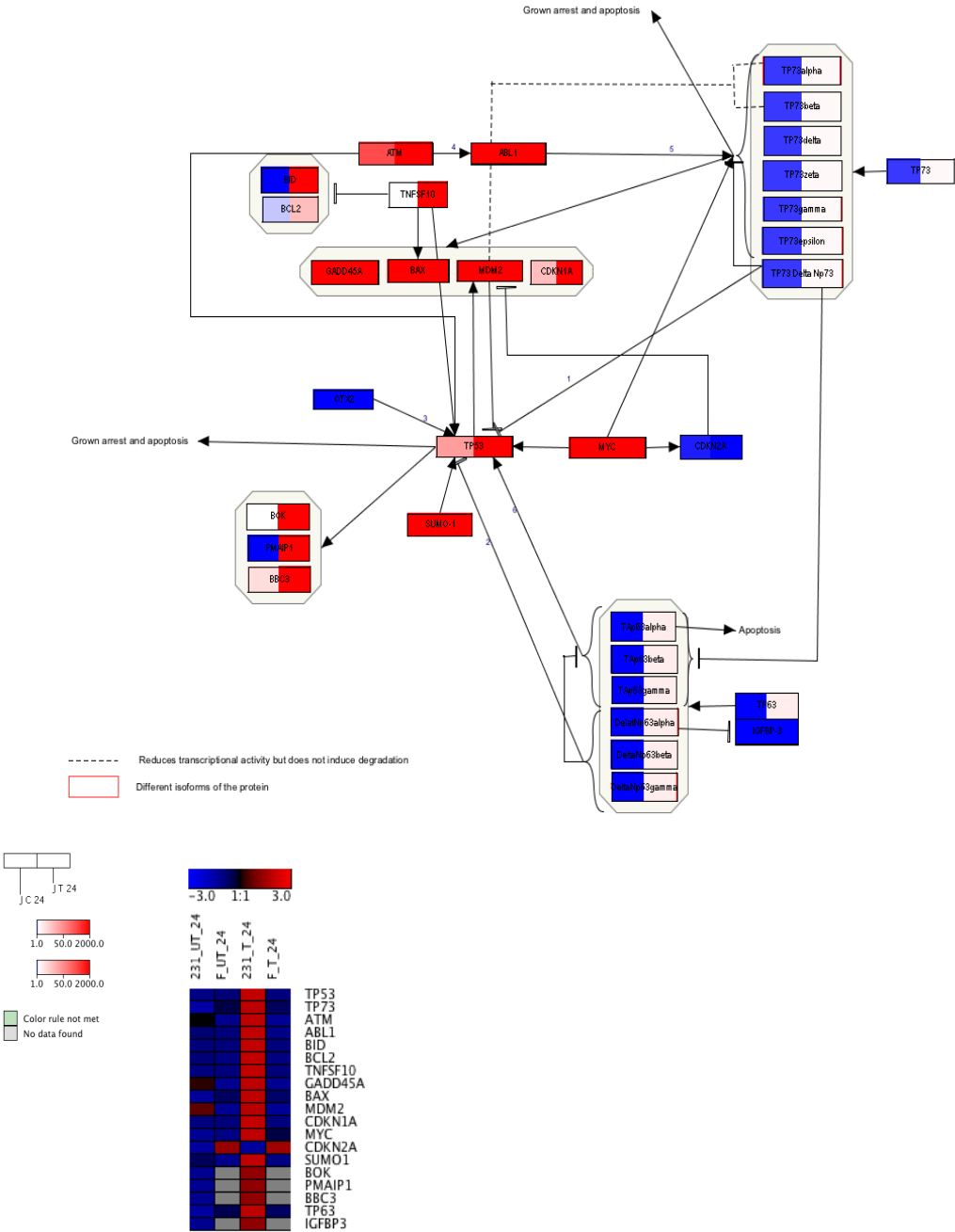


Figure 26: TP53 signaling modulation and signaling in MDA-MB-231 cells with and without rG3 treatment. Left half of node represents untreated cells; right half of node represents treated cells. Blue represents expression levels of 0-1. Shades of red represent high expression levels.

CHAPTER FIVE: SUMMARY AND CONCLUSIONS

Using a global transcriptomic approach, biomarkers, signaling events, and novel pathways of early postnatal mammary gland development were elucidated. Multivariate techniques that are not frequently applied to developmental biology were used to analyze the developing mammary transcriptome with success. PCA confirmed the hypothesis that there are unique signaling events occurring over time and confirmed our transcriptomic data set allows for evaluation of mammary development throughout time without interference from variability between animals used. RDA confirmed that although variability in animals sampled does exist, it does not interfere with the variable of interest: that of over time (age). CLA confirmed trends are present in the transcriptomic data set correlating to time; K-means clustering of our RNA-seq data set in Genesis generated unique gene sets demonstrating similar gene expression over time which were further elucidated using discriminant function analysis. Discriminate analysis of gene expression during the two phases chosen from histological analysis revealed temporal regulation of important signaling events and importantly, confirmed significant involvement of the immune system in early mammary development, specifically during the first two months after birth. Pathway analysis on the differentially expressed genes during early mammary development also indicated that the adaptive immune system is highly utilized beginning in the period of 18-25 days after birth and remains active until at least 39 days after birth, during which it contributes to developmental trends. PRC analysis further confirmed these results, indicating that these same immune pathways are utilized by genes with positive PRC values (increasing in abundance with the curve). Uniquely, PRC allowed identification of specific genes contributing functionally, and possibly interacting with each other, toward development over

time. We hypothesized that the genes decreasing in abundance with the PRC curve are very early genes contributing to the phenotype at birth or prior; these could also be similar to phase one genes (upregulated during 4-18 DOA) discussed in the discriminant analysis section, which was successfully confirmed. Conversely, we hypothesized that the positive PRC genes, those that increase in abundance with the PRC curve, are developmental genes, possibly similar to phase two DA genes discussed previously. Our data suggest that these positive PRC genes function synergistically to significantly contribute to early postnatal mammary development over time. WGCNA was used to confirm the hypothesis that there are genuine functional modules present within the overall RNA-seq dataset generated from samples collected during pre-pubertal mammary gland development, which were dissected using the MCODE algorithm and pathway analysis. Several clusters were significant to our research, especially those demonstrating a collaborative relationship between adipocytes, immune cells, the ECM, and mammary epithelial cells with some integrin involvement in promotion of growth of the mammary gland.

To our knowledge, swine models have not been used to investigate functions of any of these genes in early postnatal development, adding further novelty to our approach. Taken together, our data and the assessment of key immune regulators and markers in swine by other researchers suggests this is a much better model of human innate immunity and disease than rodents. Accordingly, understanding the function of immune molecules in normal development and their concert with other signaling mechanisms can provide insight into these roles during cancer development. In Chapter 1, we have provided a model to achieve what was previously lacking in mammary biology: continuous developmental data collected from the same live animal over time.

Our data suggested that there is significant growth occurring immediately after birth, including events previously suspected such as epithelial cell proliferation, branching, and expansion, as well as significant immune system interaction. The presence of oncogenes and proto-oncogenes in our model suggest that a better understanding of developmental mammary biology and could provide future areas of study for carcinogenesis and breast cancer treatment, leading us to investigate cellular signaling at the same global transcriptomic level in a TNBC model *in vitro*.

In Chapter 2, a recombinant laminin protein (rG3) with significant potential to be used as a research tool was evaluated to study apoptosis in both normal and tumorigenic models. By using primary swine fibroblasts as a control in comparison to the frequently debated MCF-10A line, we were able to confirm that signaling pathways utilized by this normal cell line were very similar to those identified by K-means clustering in early mammary development, including immune system, metabolism, signal transduction, cell cycle regulation, gene transcription, and protein translation. The fibroblasts lacked TP53 and CASP in the top 5% of genes expressed while integrin β 4, integrin β 1, and integrin α 6 were still present. Combined with the absence of an rG3 effect at 20ug/mL, these cells demonstrated appropriateness as a control. Interestingly, the top pathways utilized by the MDA-MB-231 cells during normal growth and proliferation cycles were almost identical to the signaling pathways identified by K-means clustering in early mammary development and the pathways utilized by primary swine fibroblasts during growth.

Possibly the most significant finding in Chapter 2 is the widespread apoptotic induction and downregulation of extracellular matrix organization that occurs when rG3 is applied *in vitro*. Further functional analysis of differentially expressed genes after rG3

treatment indicated TP53- dependent signaling is partially responsible, and suggest crosstalk between intrinsic and extrinsic apoptotic pathways. A greater understanding of a range of apoptosis pathways and their crosstalk can be gained by using rG3 as a tool.

The original hypothesis of this dissertation has been confirmed; there are specific transcriptional differences and a unique interaction of a gene set over time that contributes to postnatal mammary gland development, and this model clearly shares similarities and signaling pathways with oncogenic development. Now that we have a better understanding of normal development using high throughput analysis and a clinically relevant animal model, it is obvious which pathways should be studied in depth to better understand oncogenic development- those of the adaptive and innate immune system, ECM remodeling and integrin interactions, and extrinsic and intrinsic TP53-mediated apoptosis, which could lead to early detection of potential tumorigenic growth and identification of potential treatment avenues. Presented is a novel, comprehensive, transcriptomic model of early mammary development along with several panels of biomarkers that possess a role in normal mammary development, are involved in aggressive cancers, and are affected by apoptosis induced by rG3 treatment. These biomarkers and the cell signaling modules discussed in this dissertation are highly clinically relevant and should be further dissected using more targeted and specific techniques.

APPENDICES

APPENDIX A. PRC Top 50 positive weighted genes

PRC Weight	Gene ID	Gene ID	Mapped ID2	Gene Name, Gene Symbol, Ortholog
0.75531 2569	CENPU	HUMAN HGNC=21348 UniProtKB=Q71F23	CENPU	Centromere protein U;CENPU;ortholog
0.68776 4094	TIMM8A	PIG Ensembl=ENSSSCG00000012494 UniProtKB=K7GKR4	TIMM8A	Uncharacterized protein;TIMM8A;ortholog
0.67843 9898	CCNE2	PIG Ensembl=ENSSSCG00000006095 UniProtKB=F1RY56	CCNE2	Uncharacterized protein;CCNE2;ortholog
0.65031 4355	NPM3	PIG Ensembl=ENSSSCG00000010568 UniProtKB=F1S8T7	NPM3	Uncharacterized protein;NPM3;ortholog
0.64622 6192	MYBL2	PIG Ensembl=ENSSSCG00000007366 UniProtKB=F1SDT3	MYBL2	Uncharacterized protein;MYBL2;ortholog
0.64293 2862	RNFT1	HUMAN HGNC=30206 UniProtKB=Q5M7Z0	RNFT1	RING finger and transmembrane domain-containing protein 1;RNFT1;ortholog
0.63091 2134	TTC27	HUMAN HGNC=25986 UniProtKB=Q6P3X3	TTC27	Tetratricopeptide repeat protein 27;TTC27;ortholog
0.62959 3806	GAR1	HUMAN HGNC=14264 UniProtKB=Q9NY12	GAR1	H/ACA ribonucleoprotein complex subunit 1;GAR1;ortholog
0.62876 6562	DONSON	PIG Ensembl=ENSSSCG000000027089 UniProtKB=I3LJ77	DONSON	Uncharacterized protein;DONSON;ortholog
0.62646 9029	RFC4	PIG Ensembl=ENSSSCG00000011804 UniProtKB=F1SF13	RFC4	Uncharacterized protein;RFC4;ortholog
0.61727 193	CCNE1	PIG Ensembl=ENSSSCG000000002855 UniProtKB=F1RNY7	CCNE1	Uncharacterized protein;CCNE1;ortholog
0.61290 4564	NEDD1	PIG Ensembl=ENSSSCG000000000890 UniProtKB=F1SQR9	NEDD1	Uncharacterized protein;NEDD1;ortholog
0.60991 4388	CIT	PIG Ensembl=ENSSSCG000000009839 UniProtKB=F1RL46	CIT	Uncharacterized protein (Fragment);CIT;ortholog
0.60833 1544	WDR12	PIG Ensembl=ENSSSCG00000016116 UniProtKB=F1SHE8	WDR12	Ribosome biogenesis protein WDR12;WDR12;ortholog
0.60654 5954	DESI1	PIG Ensembl=ENSSSCG000000000060 UniProtKB=F1SRD9	DESI1	Uncharacterized protein;DESI1;ortholog

0.60117 7154	NUDCD2	PIG Ensembl=ENSSSCG00000027931 UniProtKB=F1RR73	NUDCD2	Uncharacterized protein;NUDCD2;ortholog
0.59887 7627	FAM64A	PIG Ensembl=ENSSSCG00000017893 UniProtKB=F1RGM1	FAM64A	Uncharacterized protein;FAM64A;ortholog
0.59580 5113	GALE	PIG Ensembl=ENSSSCG00000027827 UniProtKB=I3LL84	GALE	Uncharacterized protein;GALE;ortholog
0.59342 0751	ABT1	HUMAN HGNC=17369 UniProtKB=Q9ULW3	ABT1	Activator of basal transcription 1;ABT1;ortholog
0.59256 649	KCNE5	PIG Ensembl=ENSSSCG00000012582 UniProtKB=F1RWU6	KCNE5	Uncharacterized protein;KCNE5;ortholog
0.59230 4891	CDKN2C	PIG Ensembl=ENSSSCG000000003876 UniProtKB=F1S6F7	CDKN2C	Uncharacterized protein;CDKN2C;ortholog
0.58622 6873	LOC102167061	Truly uncharacterized		
0.58396 9995	PRELID3B	HUMAN HGNC=15892 UniProtKB=Q9Y3B1	PRELID3B	Protein slowmo homolog 2;SLMO2;ortholog
0.58142 9846	CTPS1	PIG Ensembl=ENSSSCG000000003973 UniProtKB=F1SF78	CTPS1	CTP synthase;CTPS1;ortholog
0.57811 8522	DOHH	PIG Ensembl=ENSSSCG00000013478 UniProtKB=F1S8E6	DOHH	Deoxyhypusine hydroxylase;DOHH;ortholog
0.57678 3145	ERCC6L	PIG Ensembl=ENSSSCG000000021834 UniProtKB=I3LFY4	ERCC6L	Uncharacterized protein (Fragment);ERCC6L;ortholog
0.57219 8441	CENPK	PIG Ensembl=ENSSSCG000000016946 UniProtKB=F1SKT7	CENPK	Uncharacterized protein;CENPK;ortholog
0.57135 0974	DUSP11	PIG Ensembl=ENSSSCG000000008298 UniProtKB=F1SLG2	DUSP11	Uncharacterized protein;DUSP11;ortholog
0.56764 138	ALG8	PIG Ensembl=ENSSSCG000000014887 UniProtKB=F1STY3	ALG8	Uncharacterized protein;ALG8;ortholog
0.56630 587	MTHFD2	PIG Ensembl=ENSSSCG000000023444 UniProtKB=F1SLH1	MTHFD2	Uncharacterized protein (Fragment);MTHFD2;ortholog
0.56510 6896	WDR74	PIG Ensembl=ENSSSCG000000028025 UniProtKB=I3LD53	WDR74	Uncharacterized protein;WDR74;ortholog
0.56237 2872	RABGGTB	PIG Ensembl=ENSSSCG000000003775 UniProtKB=F1S9R0	RABGGTB	Uncharacterized protein;RABGGTB;ortholog
0.56035 3639	MPHOSH6	PIG Ensembl=ENSSSCG000000010818 UniProtKB=F1S9K8	MPHOSH6	Uncharacterized protein (Fragment);MPHOSH6;ortholog
0.55925 4552	RNFT1	PIG Ensembl=ENSSSCG000000017671 UniProtKB=F1S223	RNFT1	Uncharacterized protein (Fragment);RNFT1;ortholog

				og
0.55794 8831	RRP8	PIG Ensembl=ENSSSCG00000023880 UniProtKB=I3LPL4	RRP8	Uncharacterized protein;RRP8;ortholog
0.55186 6159	PRELID3 B	PIG Ensembl=ENSSSCG00000007525 UniProtKB=A5GFX0	PRELID3 B	Protein slowmo homolog 2;SLMO2;ortholog
0.54940 4483	BUB1	PIG Ensembl=ENSSSCG000000030469 UniProtKB=I3LRF1	BUB1	Uncharacterized protein;BUB1;ortholog
0.54854 7924	CYP26B1	PIG Ensembl=ENSSSCG000000008311 UniProtKB=F1SLE8	CYP26B1	Uncharacterized protein;CYP26B1;ortholog
0.54699 3466	BCHE	PIG Gene=BCHE UniProtKB=P32752	BCHE	Cholinesterase (Fragment);BCHE;ortholog
0.54588 4312	MPLKIP	PIG Ensembl=ENSSSCG000000016768 UniProtKB=F1SSC4	MPLKIP	Uncharacterized protein;MPLKIP;ortholog
0.54578 3381	LOC10650 9864	truly uncharacterized		
0.54554 8898	INCENP	PIG Ensembl=ENSSSCG000000013066 UniProtKB=F1RPW7	INCENP	Uncharacterized protein;INCENP;ortholog
0.54113 617	MINPP1	PIG Ensembl=ENSSSCG000000010435 UniProtKB=F1SCZ6	MINPP1	Uncharacterized protein (Fragment);MINPP1;ortholog
0.54062 8807	NQO1	PIG Ensembl=ENSSSCG000000002754 UniProtKB=F1S395	NQO1	Uncharacterized protein;NQO1;ortholog
0.53748 8756	CSRNP1	PIG Ensembl=ENSSSCG000000011264 UniProtKB=F1SJR3	CSRNP1	Uncharacterized protein;CSRNP1;ortholog
0.53668 2609	ANP32A	PIG Ensembl=ENSSSCG000000004963 UniProtKB=F1SIU2	ANP32A	Uncharacterized protein (Fragment);ANP32A;ortholog
0.53619 4601	UBE2T	PIG Ensembl=ENSSSCG000000010924 UniProtKB=F1S5B2	UBE2T	Uncharacterized protein;UBE2T;ortholog
0.53501 2107	ADAMTS 4	PIG Ensembl=ENSSSCG000000006359 UniProtKB=F1S1A7	ADAMTS 4	Uncharacterized protein;ADAMTS4;ortholog
0.53339 6073	AMD1	PIG Ensembl=ENSSSCG000000004392 UniProtKB=F1RT08	AMD1	S-adenosylmethionine decarboxylase proenzyme;AMD1;ortholog
0.53129 9615	C2H1lorf 31	truly uncharacterized		
0.53078 8325	CDCA4	PIG Ensembl=ENSSSCG000000002559 UniProtKB=F1S970	CDCA4	Uncharacterized protein;CDCA4;ortholog

APPENDIX B. PRC Top 50 negative genes

PRC Weight	Gene ID	Gene ID2	Mapped ID2	Gene Name, Gene Symbol, Ortholog
- 0.51559 019	SLA-DOA	HUMAN HGNC=4936 UniProtKB=P06340	HLA-DOA	HLA class II histocompatibility antigen, DO alpha chain;HLA-DOA;ortholog
- 0.51544 9036	PINK1	PIG Ensembl=ENSSSCG00000003506 UniProtKB=F1SU10	PINK1	Uncharacterized protein (Fragment);PINK1;ortholog
- 0.49519 9325	EFCC1	PIG Ensembl=ENSSSCG00000024000 UniProtKB=I3L9P1	EFCC1	Uncharacterized protein;EFCC1;ortholog
- 0.48963 1223	SNAPIN	PIG Ensembl=ENSSSCG00000027049 UniProtKB=F1SFV0	SNAPIN	Uncharacterized protein;SNAPIN;ortholog
- 0.46195 3857	DOCK4	HUMAN HGNC=19192 UniProtKB=Q8N1I0	DOCK4	Dedicator of cytokinesis protein 4;DOCK4;ortholog
- 0.45410 8222	ZNF613	HUMAN HGNC=25827 UniProtKB=Q6PF04	ZNF613	Zinc finger protein 613;ZNF613;ortholog
- 0.44311 5828	GALNT15	HUMAN HGNC=21531 UniProtKB=Q8N3T1	GALNT15	Polypeptide N-acetylgalactosaminyltransferase 15;GALNT15;ortholog
- 0.42230 6462	L1CAM	PIG Ensembl=ENSSSCG00000012787 UniProtKB=F1S293	L1CAM	Uncharacterized protein;L1CAM;ortholog
- 0.41645 8647	FRMD4A	PIG Ensembl=ENSSSCG00000011056 UniProtKB=F1RTX2	FRMD4A	Uncharacterized protein;FRMD4A;ortholog
- 0.41216 7164	STARD8	PIG Ensembl=ENSSSCG00000012828 UniProtKB=F1RZ20	STARD8	Uncharacterized protein (Fragment);STARD8;ortholog
- 0.41086 8587	ZNF174	PIG Ensembl=ENSSSCG00000007958 UniProtKB=F1RK39	ZNF174	Uncharacterized protein;ZNF174;ortholog
- 0.40414 9309	PDE2A	PIG Ensembl=ENSSSCG00000029388 UniProtKB=I3LNQ5	PDE2A	Uncharacterized protein (Fragment);PDE2A;ortholog
- 0.39396 0511	VLDLR	PIG Ensembl=ENSSSCG00000005229 UniProtKB=E7CXS1	VLDLR	Uncharacterized protein;Vldlr;ortholog
- 0.39107 6702	MPND	PIG Ensembl=ENSSSCG00000013507 UniProtKB=F1S7L9	MPND	Uncharacterized protein;MPND;ortholog
- 0.38199 6691	SOX13	PIG Ensembl=ENSSSCG00000015277 UniProtKB=F1S3U8	SOX13	Uncharacterized protein (Fragment);SOX13;ortholog

- 0.38191 2433	LIMS2	PIG Ensembl=ENSSSCG00000015862 UniProtKB=F1RQH7	LIMS2	Uncharacterized protein (Fragment);LIMS2;ortholog
- 0.38103 8717	UBXN1	PIG Ensembl=ENSSSCG00000023110 UniProtKB=I3LUD5	UBXN1	Uncharacterized protein;UBXN1;ortholog
- 0.38096 0327	DDIT3	PIG Ensembl=ENSSSCG00000022482 UniProtKB=B8Q504	DDIT3	DNA-damage-inducible transcript 3;Ddit3;ortholog
- 0.38048 8353	TRIM3	PIG Ensembl=ENSSSCG00000014625 UniProtKB=F1RMN8	TRIM3	Uncharacterized protein;TRIM3;ortholog
- 0.37922 2784	FBXW4	PIG Ensembl=ENSSSCG00000010566 UniProtKB=F1S8T9	FBXW4	Uncharacterized protein;FBXW4;ortholog
- 0.37907 4354	HGF	PIG Ensembl=ENSSSCG00000015403 UniProtKB=F1SB93	HGF	Uncharacterized protein (Fragment);HGF;ortholog
- 0.37889 2532	DMAP1	PIG Ensembl=ENSSSCG00000028312 UniProtKB=I3LNQ2	DMAP1	Uncharacterized protein;DMAP1;ortholog
- 0.37795 0717	TRAF3	PIG Ensembl=ENSSSCG00000002525 UniProtKB=F1SA11	TRAF3	Uncharacterized protein;TRAF3;ortholog
- 0.36765 6669	LOC106506316	truly uncharacterized		
- 0.36313 8248	LHX6	PIG Ensembl=ENSSSCG00000005530 UniProtKB=F1SLQ9	LHX6	Uncharacterized protein;LHX6;ortholog
- 0.35687 3254	RGS18	PIG Ensembl=ENSSSCG00000010808 UniProtKB=F1SA94	RGS18	Uncharacterized protein;RGS18;ortholog
- 0.35652 6442	CTIF	PIG Ensembl=ENSSSCG00000004506 UniProtKB=F1RPR2	CTIF	Uncharacterized protein;CTIF;ortholog
- 0.35623 7482	LOC100739121	truly uncharacterized		
- 0.35416 4393	C1QTNF9	HUMAN HGNC=28732 UniProtKB=P0C862	C1QTNF9	Complement C1q and tumor necrosis factor-related protein 9A;C1QTNF9;ortholog
- 0.35272 9291	GNLY	PIG Ensembl=ENSSSCG00000008228 UniProtKB=F1SVC1	GNLY	Uncharacterized protein;GNLY;ortholog
- 0.35260 0564	PPFIA1	HUMAN HGNC=9245 UniProtKB=Q13136	PPFIA1	Liprin-alpha-1;PPFIA1;ortholog
- 0.35003 8204	RNF31	PIG Ensembl=ENSSSCG00000002003 UniProtKB=F1SGN1	RNF31	Uncharacterized protein;RNF31;ortholog

- 0.34965 5753	FLI1	PIG Ensembl=ENSSSCG00000015237 UniProtKB=F1S6G7	FLI1	Uncharacterized protein;FLI1;ortholog
- 0.34929 7781	A4GALT	PIG Ensembl=ENSSSCG00000026962 UniProtKB=I3L755	A4GALT	Uncharacterized protein;A4GALT;ortholog
- 0.34768 7199	CCDC28A	PIG Ensembl=ENSSSCG00000004148 UniProtKB=F1S6Y0	CCDC28A	Uncharacterized protein;CCDC28A;ortholog
- 0.34709 9383	GPANK1	HUMAN HGNC=13920 UniProtKB=O95872	GPANK1	G patch domain and ankyrin repeat-containing protein 1;GPANK1;ortholog
- 0.34649 8169	CCDC130	PIG Ensembl=ENSSSCG00000013755 UniProtKB=F1SD79	CCDC130	Uncharacterized protein;CCDC130;ortholog
- 0.34162 4683	CEP95	PIG Ensembl=ENSSSCG00000028925 UniProtKB=I3LTR6	CEP95	Uncharacterized protein;CEP95;ortholog
- 0.34161 9986	CTSF	PIG Ensembl=ENSSSCG00000027368 UniProtKB=F1RU48	CTSF	Uncharacterized protein;CTSF;ortholog
- 0.34156 9075	LOC106505043	truly uncharacterized		
- 0.33794 3464	HMCN2	PIG Ensembl=ENSSSCG00000005699 UniProtKB=F1S0Y1	HMCN2	Uncharacterized protein (Fragment);HMCN2;ortholog
- 0.33792 9917	MID2	PIG Ensembl=ENSSSCG00000012564 UniProtKB=F1RXJ0	MID2	Uncharacterized protein;MID2;ortholog
- 0.33640 5962	ABLIM3	PIG Ensembl=ENSSSCG00000014430 UniProtKB=F1RLA7	ABLIM3	Uncharacterized protein;ABLIM3;ortholog
- 0.33411 1451	CUX1	PIG Ensembl=ENSSSCG00000007681 UniProtKB=F1RKF9	CUX1	Uncharacterized protein;CUX1;ortholog
- 0.33388 2104	TCTP	PIG Ensembl=ENSSSCG00000000622 UniProtKB=F1SQ68	TCTP	Uncharacterized protein;TCTP;ortholog
- 0.33320 9584	KCNB1	PIG Gene=KCNB1 UniProtKB=O18868	KCNB1	Potassium voltage-gated channel subfamily B member 1;KCNB1;ortholog
- 0.33230 6307	PTPRM	PIG Ensembl=ENSSSCG00000028805 UniProtKB=I3LC13	PTPRM	Uncharacterized protein (Fragment);PTPRM;ortholog
- 0.33126 8784	CUX1	PIG Ensembl=ENSSSCG00000007681 UniProtKB=F1RKF9	CUX1	Uncharacterized protein;CUX1;ortholog
- 0.33121	ANKRD24	PIG Ensembl=ENSSSCG00000013497 UniProtKB=F1S7P0	ANKRD24	Uncharacterized protein;ANKRD24;ortholog

179				log
- 0.32972 1287	PBLD	PIG Ensembl=ENSSSCG00000025876 UniProtKB=I3LTZ8	PBLD	Uncharacterized protein;PBLD;ortholog

APPENDIX C. rG3 treated MDA-MB-231 transcriptome shift changes

ENSID	GENEID	Shift change
ENSG00000169429	CXCL8	14400
ENSG00000175061	LRRC75A-AS1	7937
ENSG00000151651	ADAM8	7685
ENSG00000090339	ICAM1	6827
ENSG00000162645	GBP2	6697
ENSG00000198840	MT-ND3	6521
ENSG00000183696	UPP1	6440
ENSG00000186480	INSIG1	5968
ENSG00000214049	UCA1	5799.5
ENSG00000196352	CD55	5793.5
ENSG00000073008	PVR	5616
ENSG00000021355	SERPINB1	5519
ENSG00000117335	CD46	5257
ENSG00000197249	SERPINA1	4855
ENSG00000160932	LY6E	4722
ENSG00000049245	VAMP3	4634
ENSG00000204592	HLA-E	4584
ENSG00000128340	RAC2	4534
ENSG00000162433	AK4	4517
ENSG00000163739	CXCL1	4480
ENSG00000138771	SHROOM3	4386
ENSG00000107959	PITRM1	3633
ENSG00000126860	EVI2A	3608
ENSG00000141682	PMAIP1	3559
ENSG00000197971	MBP	3464
ENSG00000275216	RP11-54H7.4	3463.8
ENSG00000198695	MT-ND6	3459

ENSG00000204388	HSPA1B	3355
ENSG00000135373	EHF	3347
ENSG00000137210	TMEM14B	3235
ENSG00000091317	CMTM6	3184
ENSG00000019582	CD74	3025
ENSG00000176171	BNIP3	3020.25
ENSG00000162892	IL24	2958
ENSG00000166710	B2M	2946.5
ENSG00000177410	ZFAS1	2908
ENSG00000160213	CSTB	2890
ENSG00000173432	SAA1	2858
ENSG00000189143	CLDN4	2856
ENSG00000234741	GAS5	2849
ENSG00000183255	PTTG1IP	2828
ENSG00000119917	IFIT3	2816
ENSG00000130303	BST2	2773
ENSG00000110218	PANX1	2739
ENSG00000182952	HMGN4	2721
ENSG00000158825	CDA	2669
ENSG00000011422	PLAUR	2652
ENSG00000099204	ABLIM1	2644.5
ENSG00000204520	MICA	2630
ENSG00000172137	CALB2	2571
ENSG00000121797	CCRL2	2544
ENSG00000087074	PPP1R15A	2418.8
ENSG00000138772	ANXA3	2369.666667
ENSG00000133142	TCEAL4	2345
ENSG00000002586	CD99	2335.5
ENSG00000204381	LAYN	2291
ENSG00000108679	LGALS3BP	2201.111111
ENSG00000160326	SLC2A6	2177
ENSG00000134802	SLC43A3	2133.142857
ENSG00000211459	MT-RNR1	2116.8
ENSG00000085063	CD59	2105.333333
ENSG00000204387	C6orf48	2085.5
ENSG00000111846	GCNT2	2064
ENSG00000146678	IGFBP1	2050
ENSG00000101846	STS	2048

ENSG00000212907	MT-ND4L	2032.5
ENSG00000186866	POFUT2	2030
ENSG00000184584	TMEM173	1995
ENSG00000185499	MUC1	1977
ENSG00000165949	IFI27	1933
ENSG00000112379	ARFGEF3	1912
ENSG00000107984	DKK1	1910.166667
ENSG00000166920	C15orf48	1884
ENSG00000198712	MT-CO2	1867.666667
ENSG00000132334	PTPRE	1853
ENSG00000163235	TGFA	1843
ENSG00000137275	RIPK1	1816
ENSG00000132561	MATN2	1788.5
ENSG00000130513	GDF15	1745.5
ENSG00000163694	RBM47	1735
ENSG00000101439	CST3	1675.2
ENSG00000148346	LCN2	1675
ENSG00000166145	SPINT1	1661
ENSG00000108064	TFAM	1660
ENSG00000222041	LINC00152	1652
ENSG00000198886	MT-ND4	1644.311475
ENSG00000128487	SPECC1	1640
ENSG00000198804	MT-CO1	1630.527027
ENSG00000132821	VSTM2L	1629
ENSG00000198113	TOR4A	1627
ENSG00000198899	MT-ATP6	1623.428571
ENSG00000164818	DNAAF5	1616
ENSG00000136205	TNS3	1612.2
ENSG00000185567	AHNAK2	1604.25
ENSG00000117151	CTBS	1594
ENSG00000198768	APCDD1L	1591
ENSG00000177106	EPS8L2	1579
ENSG00000169129	AFAP1L2	1575
ENSG00000125538	IL1B	1569
ENSG00000248527	MTATP6P1	1560.666667
ENSG00000003436	TFPI	1559.666667
ENSG00000149564	ESAM	1554

ENSG00000136244	IL6	1548.4
ENSG00000127399	LRRC61	1532.666667
ENSG00000198938	MT-CO3	1529.179487
ENSG00000133321	RARRES3	1519
ENSG00000013588	GPRC5A	1481.818182
ENSG00000198888	MT-ND1	1479.6
ENSG00000069122	ADGRF5	1462.4
ENSG00000140950	TLDC1	1454
ENSG00000228253	MT-ATP8	1439.5
ENSG00000167642	SPINT2	1410
ENSG00000171606	ZNF274	1401
ENSG00000027697	IFNGR1	1388
ENSG00000150403	TMCO3	1388
ENSG00000164251	F2RL1	1383
ENSG00000146833	TRIM4	1381
ENSG00000116701	NCF2	1373
ENSG00000198763	MT-ND2	1364.615385
ENSG00000162511	LAPTM5	1351
ENSG00000104611	SH2D4A	1340
ENSG00000163814	CDCP1	1329.08
ENSG00000128567	PODXL	1321.25
ENSG00000141524	TMC6	1321
ENSG00000122862	SRGN	1311
ENSG00000050344	NFE2L3	1307
ENSG00000206053	HN1L	1302
ENSG00000142166	IFNAR1	1286.5
ENSG00000112096	SOD2	1286.018868
ENSG00000181381	DDX60L	1267
ENSG00000099940	SNAP29	1260
ENSG00000065923	SLC9A7	1248
ENSG00000214113	LYRM4	1238
ENSG00000089327	FXVD5	1227.333333
ENSG00000099998	GGT5	1223
ENSG00000137699	TRIM29	1213.666667
ENSG00000169174	PCSK9	1195
ENSG00000138134	STAMBPL1	1187
ENSG00000204604	ZNF468	1185
ENSG00000105612	DNASE2	1158

ENSG00000132182	NUP210	1153
ENSG00000101191	DIDO1	1122
ENSG00000109519	GRPEL1	1119
ENSG00000112763	BTN2A1	1111
ENSG00000198142	SOWAHC	1075
ENSG00000160285	LSS	1072.888889
ENSG00000196743	GM2A	1056.5
ENSG00000117984	CTSD	1056.333333
ENSG00000115648	MLPH	1053
ENSG00000188112	C6orf132	1046.2
ENSG00000104368	PLAT	1046
ENSG00000120889	TNFRSF10B	1043.75
ENSG00000163565	IFI16	1037.333333
ENSG00000172270	BSG	1035.136364
ENSG00000164096	C4orf3	1031.75
ENSG00000188707	ZBED6CL	1026.5
ENSG00000111057	KRT18	1026.5
ENSG00000130707	ASS1	1022
ENSG00000171067	C11orf24	1011
ENSG00000116815	CD58	1001
ENSG00000214944	ARHGEF28	996.5
ENSG00000130312	MRPL34	991
ENSG00000133216	EPHB2	991
ENSG00000228300	C19orf24	987
ENSG00000125657	TNFSF9	970
ENSG00000240694	PNMA2	946
ENSG00000198727	MT-CYB	945.8636364
ENSG00000059804	SLC2A3	942.25
ENSG00000148300	REXO4	939
ENSG00000170949	ZNF160	937
ENSG00000163597	SNHG16	895.5
ENSG00000106305	AIMP2	889
ENSG00000203875	SNHG5	886.5
ENSG00000135114	OASL	875
ENSG00000219481	NBPF1	874
ENSG00000181104	F2R	866
ENSG00000123843	C4BPB	862
ENSG00000170542	SERPINB9	859.5

ENSG00000197077	KIAA1671	859
ENSG00000181634	TNFSF15	858.3333333
ENSG00000181885	CLDN7	855
ENSG00000231925	TAPBP	853.8823529
ENSG00000008394	MGST1	853.2
ENSG00000210082	MT-RNR2	852.990099
ENSG00000227507	LTB	836
ENSG00000176974	SHMT1	834
ENSG00000182307	C8orf33	824
ENSG00000174652	ZNF266	822
ENSG00000176340	COX8A	819.6
ENSG00000130023	ERMARD	818
ENSG00000107738	C10orf54	813
ENSG00000041353	RAB27B	803
ENSG00000196878	LAMB3	802.7083333
ENSG00000198910	L1CAM	802.3
ENSG00000146670	CDCA5	801.5
ENSG00000198585	NUDT16	789
ENSG00000151388	ADAMTS12	788.75
ENSG00000132846	ZBED3	785
ENSG00000184292	TACSTD2	778.5
ENSG00000184281	TSSC4	773
ENSG00000189223	PAX8-AS1	772.6666667
ENSG00000026751	SLAMF7	772.5
ENSG00000145623	OSMR	767.4761905
ENSG00000127666	TICAM1	765
ENSG00000075218	GTSE1	762
ENSG00000185803	SLC52A2	758.6666667
ENSG00000141504	SAT2	755.5
ENSG00000221963	APOL6	754.5
ENSG00000134955	SLC37A2	747.7692308
ENSG00000105355	PLIN3	747
ENSG00000172927	MYEOV	743
ENSG00000214026	MRPL23	737
ENSG00000181192	DHTKD1	733
ENSG00000166801	FAM111A	732
ENSG00000169689	STRA13	730
ENSG00000204525	HLA-C	727.95

ENSG00000165512	ZNF22	724
ENSG00000198546	ZNF511	719
ENSG00000063241	ISOC2	717
ENSG00000197608	ZNF841	712
ENSG00000242485	MRPL20	711
ENSG00000145365	TIFA	704
ENSG00000166012	TAF1D	703.6
ENSG00000065809	FAM107B	702.4285714
ENSG00000117602	RCAN3	701
ENSG00000148677	ANKRD1	698.2666667
ENSG00000137274	BPHL	683
ENSG00000163257	DCAF16	680
ENSG00000181061	HIGD1A	678.5
ENSG00000146859	TMEM140	672
ENSG00000204054	LINC00963	671
ENSG00000247271	ZBED5-AS1	669
ENSG00000159212	CLIC6	662
ENSG00000198931	APRT	661.5
ENSG00000135929	CYP27A1	659
ENSG00000121858	TNFSF10	657
ENSG00000232774	FLJ22447	654
ENSG00000204941	PSG5	648
ENSG00000006453	BAIAP2L1	647.5
ENSG00000186283	TOR3A	644
ENSG00000119673	ACOT2	642
ENSG00000198959	TGM2	641.893617
ENSG00000163191	S100A11	639.7916667
ENSG00000108342	CSF3	639
ENSG00000119922	IFIT2	636.6666667
ENSG00000167034	NKX3-1	634
ENSG00000163975	MELTF	628.8333333
ENSG00000127870	RNF6	624.6666667
ENSG00000065518	NDUFB4	624.5
ENSG00000159128	IFNGR2	623
ENSG00000142252	GEMIN7	622
ENSG00000111335	OAS2	613
ENSG00000136193	SCRN1	612.8571429
ENSG00000091592	NLRP1	612

ENSG00000173272	MZT2A	603
ENSG00000059378	PARP12	595
ENSG00000103888	CEMIP	592.5151515
ENSG00000148773	MKI67	586.2
ENSG00000046604	DSG2	581
ENSG00000174718	KIAA1551	580
ENSG00000130304	SLC27A1	580
ENSG00000103381	CPPED1	580
ENSG00000163013	FBXO41	579
ENSG00000134070	IRAK2	578.8
ENSG00000130193	THEM6	578
ENSG00000213654	GPSM3	576
ENSG00000060558	GNA15	575
ENSG00000272398	CD24	575
ENSG00000160336	ZNF761	574
ENSG00000101187	SLCO4A1	574
ENSG00000109452	INPP4B	570.75
ENSG00000119888	EPCAM	564
ENSG00000132256	TRIM5	558
ENSG00000067066	SP100	558
ENSG00000248092	NNT-AS1	557
ENSG00000162777	DENND2D	553
ENSG00000106538	RARRES2	553
ENSG00000166900	STX3	552.5714286
ENSG00000006744	ELAC2	552.5
ENSG00000197586	ENTPD6	547.3333333
ENSG00000185862	EVI2B	545
ENSG00000163435	ELF3	542
ENSG00000196422	PPP1R26	542
ENSG00000175567	UCP2	541.25
ENSG00000147689	FAM83A	540
ENSG00000123609	NMI	533
ENSG00000164744	SUN3	532
ENSG00000171425	ZNF581	532
ENSG00000104689	TNFRSF10A	531.5
ENSG00000147676	MAL2	531
ENSG00000128335	APOL2	527
ENSG00000029534	ANK1	525

ENSG00000151690	MFSD6	524.75
ENSG00000090661	CERS4	522
ENSG00000130775	THEMIS2	521
ENSG00000125798	FOXA2	521
ENSG00000134905	CARS2	519.5
ENSG00000122644	ARL4A	518
ENSG00000128510	CPA4	515
ENSG00000186814	ZSCAN30	510
ENSG00000189339	SLC35E2B	510
ENSG00000100426	ZBED4	506.5
ENSG00000157870	FAM213B	505.5
ENSG00000171051	FPR1	500
ENSG00000148426	PROSER2	499
ENSG00000085662	AKR1B1	497.5935484
ENSG00000074416	MGLL	496.7142857
ENSG00000156711	MAPK13	496.6666667
ENSG00000069998	CECR5	495
ENSG00000244045	TMEM199	489
ENSG00000197467	COL13A1	488.5
ENSG00000170421	KRT8	485.8461538
ENSG00000227036	LINC00511	483
ENSG00000160953	MUM1	482
ENSG00000109390	NDUFC1	480
ENSG00000168393	DTYMK	479.5
ENSG00000260032	NORAD	478.9545455
ENSG00000185608	MRPL40	477
ENSG00000198169	ZNF251	475
ENSG00000149792	MRPL49	474
ENSG00000183828	NUDT14	473
ENSG00000250462	LRRC37BP1	472
ENSG00000172031	EPHX4	470
ENSG00000167767	KRT80	469.5714286
ENSG00000142794	NBPF3	465
ENSG00000168653	NDUFS5	465
ENSG00000168528	SERINC2	463.0625
ENSG00000106610	STAG3L4	462

ENSG00000106853	PTGR1	460.6666667
ENSG00000183421	RIPK4	460
ENSG00000164509	IL31RA	459
ENSG00000163956	LRPAP1	456.1666667
ENSG00000122952	ZWINT	455.5
ENSG00000046651	OFD1	454
ENSG00000226137	BAIAP2-AS1	454
ENSG00000146674	IGFBP3	452.125
ENSG00000251003	ZFPM2-AS1	452
ENSG00000115041	KCNIP3	452
ENSG00000186281	GPAT2	451
ENSG00000198756	COLGALT2	445
ENSG00000223547	ZNF844	444
ENSG00000117525	F3	443.5
ENSG00000048052	HDAC9	441.7777778
ENSG00000149798	CDC42EP2	441
ENSG00000001460	STPG1	440.5
ENSG00000103018	CYB5B	438
ENSG00000186918	ZNF395	437.3333333
ENSG00000152078	TMEM56	437
ENSG00000167595	PROSER3	435
ENSG00000154319	FAM167A	433
ENSG00000090263	MRPS33	432
ENSG00000008283	CYB561	430.1666667
ENSG00000156411	C14orf2	429.3333333
ENSG00000112186	CAP2	429
ENSG00000100422	CERK	425.5
ENSG00000204370	SDHD	424
ENSG00000166452	AKIP1	421
ENSG00000113070	HBEGF	420.6
ENSG00000153558	FBXL2	419
ENSG00000167378	IRGQ	418
ENSG00000162976	PQLC3	417
ENSG00000141179	PCTP	416
ENSG00000134463	ECHDC3	415.5
ENSG00000223891	OSER1-AS1	415
ENSG00000203805	PLPP4	415
ENSG00000155850	SLC26A2	414

ENSG00000182325	FBXL6	412
ENSG00000105053	VRK3	411.6666667
ENSG00000186532	SMYD4	408
ENSG00000164171	ITGA2	405.5
ENSG00000198743	SLC5A3	403.3214286
ENSG00000100814	CCNB1IP1	403.25
ENSG00000169252	ADRB2	401
ENSG00000173193	PARP14	400
ENSG00000197620	CXorf40A	399
ENSG00000145536	ADAMTS16	395
ENSG00000137965	IFI44	393
ENSG00000122711	SPINK4	393
ENSG00000170647	TMEM133	391
ENSG00000150433	TMEM218	390
ENSG00000181924	COA4	387.5
ENSG00000166689	PLEKHA7	386.25
ENSG00000196139	AKR1C3	384.1111111
ENSG00000231074	HCG18	383.2
ENSG00000169093	ASMTL	383
ENSG00000130589	HELZ2	382.6666667
ENSG00000081692	JMJD4	382
ENSG00000106078	COBL	382
ENSG00000125148	MT2A	381.4545455
ENSG00000168116	KIAA1586	381
ENSG00000173818	ENDOV	379
ENSG00000156675	RAB11FIP1	378.4
ENSG00000188015	S100A3	377.2
ENSG00000131115	ZNF227	377
ENSG00000183291	42993	375.8
ENSG00000184985	SORCS2	375.75
ENSG00000167700	MFSD3	373
ENSG00000204287	HLA-DRA	370.5
ENSG00000110628	SLC22A18	370
ENSG00000185917	SETD4	369
ENSG00000255737	AGAP2-AS1	369
ENSG00000176396	EID2	369
ENSG00000124275	MTRR	367.5
ENSG00000242125	SNHG3	362.5

ENSG00000137331	IER3	360.8582677
ENSG00000158828	PINK1	360.3333333
ENSG00000189067	LITAF	358.6666667
ENSG00000163283	ALPP	356.6666667
ENSG00000167173	C15orf39	355.8333333
ENSG00000131943	C19orf12	354.5
ENSG00000187609	EXD3	347
ENSG00000204267	TAP2	347
ENSG00000198624	CCDC69	346.8571429
ENSG00000081041	CXCL2	346.5833333
ENSG00000174749	C4orf32	346
ENSG00000173530	TNFRSF10D	346
ENSG00000133317	LGALS12	346
ENSG00000112679	DUSP22	344
ENSG00000180921	FAM83H	344
ENSG00000189114	BLOC1S3	342.5
ENSG00000121966	CXCR4	341
ENSG00000105928	DFNA5	340.6666667
ENSG00000140545	MFGE8	338.8
ENSG00000138166	DUSP5	338.6428571
ENSG00000219200	RNASEK	338
ENSG00000148399	DPH7	338
ENSG00000100014	SPECC1L	337.5
ENSG00000106479	ZNF862	337
ENSG00000119986	AVPI1	336.8
ENSG00000062582	MRPS24	335
ENSG00000188321	ZNF559	332
ENSG00000105048	TNNT1	331.1666667
ENSG00000215012	C22orf29	331
ENSG00000179958	DCTPP1	331
ENSG00000072135	PTPN18	330
ENSG00000227706	RP11-301G19.1	327
ENSG00000164054	SHISA5	326.1428571
ENSG00000112977	DAP	325.3714286
ENSG00000180448	ARHGAP45	324.2222222
ENSG00000196917	HCAR1	324
ENSG00000110031	LPXN	323.5
ENSG00000180573	HIST1H2AC	323

ENSG00000151503	NCAPD3	322.5
ENSG00000103160	HSDL1	320.3333333
ENSG00000164379	FOXQ1	316
ENSG00000119900	OGFRL1	315.4358974
ENSG00000272391	POM121C	315
ENSG00000217555	CKLF	315
ENSG00000143845	ETNK2	311.5
ENSG00000232119	MCTS1	310.2
ENSG00000189180	ZNF33A	309.75
ENSG00000023191	RNH1	308.4736842
ENSG00000206503	HLA-A	306.7357143
ENSG00000135925	WNT10A	305
ENSG00000198168	SVIP	305
ENSG00000144645	OSBPL10	304.8
ENSG00000078237	TIGAR	304
ENSG00000270231	NBPF8	303.5
ENSG00000100058	CRYBB2P1	303
ENSG00000242259	C22orf39	301
ENSG00000130202	NECTIN2	299.8947368
ENSG00000105877	DNAH11	299
ENSG00000110013	SIAE	297.3333333
ENSG00000146066	HIGD2A	297.2
ENSG00000155275	TRMT44	297
ENSG00000112877	CEP72	297
ENSG00000067064	IDI1	296.875
ENSG00000104979	C19orf53	295.1666667
ENSG00000253161	LINC01605	295
ENSG00000186951	PPARA	294.8333333
ENSG00000102699	PARP4	293.625
ENSG00000096696	DSP	292.5384615
ENSG00000114544	SLC41A3	291.6666667
ENSG00000171943	SRGAP2C	290
ENSG00000255248	MIR100HG	288.75
ENSG00000111348	ARHGDIB	288
ENSG00000168394	TAP1	286.7692308
ENSG00000253910	PCDHGB2	283

ENSG00000133101	CCNA1	283
ENSG00000112343	TRIM38	282.2
ENSG00000188827	SLX4	281
ENSG00000173898	SPTBN2	281
ENSG00000166592	RRAD	280.5
ENSG00000198786	MT-ND5	279.7851852
ENSG00000131941	RHPN2	278.6666667
ENSG00000093010	COMT	278
ENSG00000141562	NARF	278
ENSG00000099194	SCD	277.3877551
ENSG00000142920	AZIN2	277
ENSG00000156510	HKDC1	276.8461538
ENSG00000164535	DAGLB	276.2
ENSG00000163945	UVSSA	275
ENSG00000239887	C1orf226	274
ENSG00000171490	RSL1D1	273.4375
ENSG00000141030	COPS3	272.625
ENSG00000179532	DNHD1	272
ENSG00000196843	ARID5A	270
ENSG00000253729	PRKDC	269.48
ENSG00000234608	MAPKAPK5-AS1	269
ENSG00000132740	IGHMBP2	268.6666667
ENSG00000088888	MAVS	268.5
ENSG00000122778	KIAA1549	268.2
ENSG00000168884	TNIP2	268
ENSG00000176485	PLA2G16	266.4210526
ENSG00000091622	PITPNM3	265.5
ENSG00000229833	PET100	265
ENSG00000111843	TMEM14C	263.4375
ENSG00000225630	MTND2P28	263.3333333
ENSG00000161277	THAP8	263
ENSG00000108021	FAM208B	261.7272727
ENSG00000065609	SNAP91	260
ENSG00000164919	COX6C	258
ENSG00000137337	MDC1	256.9
ENSG00000100034	PPM1F	256.25
ENSG00000108179	PPIF	256.2083333

ENSG00000167705	RILP	255.5
ENSG00000153363	LINC00467	255
ENSG00000131591	C1orf159	255
ENSG00000187266	EPOR	252.5
ENSG00000085117	CD82	250.8125
ENSG00000183161	FANCF	248
ENSG00000103152	MPG	248
ENSG00000067182	TNFRSF1A	245.8571429
ENSG00000132109	TRIM21	245.5
ENSG00000130270	ATP8B3	245
ENSG00000128578	STRIP2	244.5
ENSG00000179526	SHARPIN	244
ENSG00000152443	ZNF776	243.5
ENSG00000167397	VKORC1	242.5
ENSG00000173821	RNF213	242.3913043
ENSG00000247516	MIR4458HG	242
ENSG00000028277	POU2F2	241.75
ENSG00000058335	RASGRF1	241
ENSG00000124588	NQO2	240.875
ENSG00000118197	DDX59	240
ENSG00000120256	LRP11	239.8
ENSG00000242498	ARPIN	238.5
ENSG00000106789	CORO2A	238.25
ENSG00000141556	TBCD	237.4285714
ENSG00000170006	TMEM154	237
ENSG00000139410	SDSL	237
ENSG00000003402	CFLAR	236.9230769
ENSG00000196449	YRDC	236.75
ENSG00000056558	TRAF1	236.2222222
ENSG00000205763	RP9P	236
ENSG00000170791	CHCHD7	236
ENSG00000135953	MFSD9	235
ENSG00000160072	ATAD3B	234
ENSG00000275342	SGK223	233.3333333
ENSG00000246273	SBF2-AS1	233
ENSG00000130489	SCO2	233
ENSG00000137936	BCAR3	232.7272727

ENSG00000197208	SLC22A4	232
ENSG00000138744	NAAA	232
ENSG00000139178	C1RL	231
ENSG00000116285	ERRFI1	230.3252033
ENSG00000119655	NPC2	229.3030303
ENSG00000204922	UQCC3	229
ENSG00000120875	DUSP4	228.9607843
ENSG00000214194	LINC00998	228.6666667
ENSG00000188938	FAM120AOS	227
ENSG00000113845	TIMMDC1	227
ENSG00000127124	HIVEP3	226.75
ENSG00000149948	HMGA2	226.4285714
ENSG00000142556	ZNF614	226
ENSG00000275993	CH507-42P11.8	225.3333333
ENSG00000116260	QSOX1	225.3088235
ENSG00000165609	NUDT5	224.3
ENSG00000104823	ECH1	221.6666667
ENSG00000122687	MRM2	221.25
ENSG00000088876	ZNF343	221
ENSG00000163624	CDS1	220
ENSG00000002587	HS3ST1	220
ENSG00000181274	FRAT2	219
ENSG00000159733	ZFYVE28	218.5
ENSG00000129355	CDKN2D	218.3333333
ENSG00000183087	GAS6	217.8695652
ENSG00000141526	SLC16A3	217.6190476
ENSG00000180190	TDRP	217.5
ENSG00000103253	HAGHL	217
ENSG00000076351	SLC46A1	217
ENSG00000154328	NEIL2	217
ENSG00000154134	ROBO3	216.5
ENSG00000109084	TMEM97	216.4285714
ENSG00000111321	LTBR	216.1666667
ENSG00000182362	YBEY	216
ENSG00000224531	SMIM13	215.3333333
ENSG00000129422	MTUS1	215.3
ENSG00000100099	HPS4	213.5
ENSG00000188785	ZNF548	213

ENSG00000185442	FAM174B	212.5
ENSG00000160062	ZBTB8A	212
ENSG00000104679	R3HCC1	212
ENSG00000123146	ADGRE5	211.7777778
ENSG00000212916	MAP10	211
ENSG00000154710	RABGEF1	211
ENSG00000196976	LAGE3	211
ENSG00000008513	ST3GAL1	210.1111111
ENSG00000181458	TMEM45A	209.5
ENSG00000188157	AGRN	209.3829787
ENSG00000124614	RPS10	208.4814815
ENSG00000138759	FRAS1	208
ENSG00000196372	ASB13	207.5
ENSG00000039068	CDH1	207.3333333
ENSG00000143740	SNAP47	206.8
ENSG00000177685	CRACR2B	206
ENSG00000109062	SLC9A3R1	206
ENSG00000078401	EDN1	206
ENSG00000146410	MTFR2	205.5
ENSG00000115107	STEAP3	204.8421053
ENSG00000146232	NFKBIE	204.5
ENSG00000167536	DHRS13	203
ENSG00000113645	WWC1	202.2083333
ENSG00000205476	CCDC85C	201.75
ENSG00000179271	GADD45GIP1	200.7142857
ENSG00000133678	TMEM254	200
ENSG00000013810	TACC3	200
ENSG00000078804	TP53INP2	199.6666667
ENSG00000170271	FAXDC2	199
ENSG00000165684	SNAPC4	199
ENSG00000099330	OCEL1	198.6666667
ENSG00000198286	CARD11	198.6666667
ENSG00000165480	SKA3	198.3333333
ENSG00000105341	ATP5SL	198.25
ENSG00000253352	TUG1	197.8235294
ENSG00000143786	CNIH3	197.75
ENSG00000010219	DYRK4	197.5
ENSG00000028137	TNFRSF1B	196.25

ENSG00000185022	MAFF	195.9285714
ENSG00000075643	MOCOS	194.7
ENSG00000168765	GSTM4	193.8571429
ENSG00000182957	SPATA13	193
ENSG00000172888	ZNF621	192.8333333
ENSG00000147804	SLC39A4	192.5
ENSG00000122376	FAM35A	192.3333333
ENSG00000198721	ECI2	191.2666667
ENSG00000026103	FAS	191
ENSG00000157510	AFAP1L1	190.3333333
ENSG00000185298	CCDC137	189.3333333
ENSG00000152558	TMEM123	189.2631579
ENSG00000120217	CD274	189
ENSG00000160961	ZNF333	188
ENSG00000185127	C6orf120	187.125
ENSG00000168672	FAM84B	186
ENSG00000173706	HEG1	185.3739837
ENSG00000157601	MX1	185.25
ENSG00000163131	CTSS	185.0833333
ENSG00000171813	PWWP2B	184.3333333
ENSG00000070814	TCOF1	182.5833333
ENSG00000256683	ZNF350	182
ENSG00000173210	ABLIM3	181
ENSG00000137185	ZSCAN9	181
ENSG00000054277	OPN3	180
ENSG00000157895	C12orf43	180
ENSG00000137198	GMPR	179.5714286
ENSG00000164142	FAM160A1	178.5
ENSG00000182858	ALG12	178.5
ENSG00000165689	SDCCAG3	178.3333333
ENSG00000167969	ECI1	177.5
ENSG00000231365	RP11-418J17.1	177
ENSG00000184371	CSF1	176.4083333
ENSG00000219665	CTD-2006C1.2	176
ENSG00000237550	RPL9P9	175.8571429
ENSG00000128833	MYO5C	175
ENSG00000259974	LINC00261	174

ENSG00000137642	SORL1	174
ENSG00000179242	CDH4	173.5
ENSG00000117877	CD3EAP	173
ENSG00000100104	SRRD	172.6666667
ENSG00000131697	NPHP4	172.5
ENSG00000188002	RP11-43F13.1	172
ENSG00000165832	TRUB1	171.5
ENSG00000112584	FAM120B	171.2222222
ENSG00000214087	ARL16	170.6666667
ENSG00000272888	LINC01578	170.375
ENSG00000078114	NEBL	170
ENSG00000166278	C2	169.75
ENSG00000067606	PRKCZ	169.5
ENSG00000160299	PCNT	169.1111111
ENSG00000167840	ZNF232	169
ENSG00000101017	CD40	168.6
ENSG00000065054	SLC9A3R2	168.0625
ENSG00000171224	C10orf35	168
ENSG00000213366	GSTM2	168
ENSG00000107159	CA9	167
ENSG00000115339	GALNT3	166.6666667
ENSG00000131944	FAAP24	166
ENSG00000164211	STARD4	165.8461538
ENSG00000172493	AFF1	165.7916667
ENSG00000135451	TROAP	165.6666667
ENSG00000175764	TTLL11	164
ENSG00000110721	CHKA	163.3888889
ENSG00000126709	IFI6	163
ENSG00000189057	FAM111B	163
ENSG00000269001	CTD-2245F17.3	162
ENSG00000161249	DMKN	162
ENSG00000232434	C9orf172	162
ENSG00000096080	MRPS18A	161.6666667
ENSG00000254470	AP5B1	161.4285714
ENSG00000131389	SLC6A6	161.2368421
ENSG00000163082	SGPP2	161

ENSG00000154447	SH3RF1	160.5
ENSG00000170915	PAQR8	160
ENSG00000006432	MAP3K9	159.6
ENSG00000073350	LLGL2	157.6666667
ENSG00000197037	ZSCAN25	157.3333333
ENSG00000221947	XKR9	156.5
ENSG00000233461	RP11-295G20.2	155
ENSG00000276180	HIST1H4I	155
ENSG00000146648	EGFR	154.8288288
ENSG00000102181	CD99L2	154.5714286
ENSG00000179403	VWA1	154.5
ENSG00000167695	FAM57A	154.4545455
ENSG00000184792	OSBP2	154
ENSG00000167880	EVPL	153.6666667
ENSG00000223552	RP11-24F11.2	153.5
ENSG00000165983	PTER	153.3333333
ENSG00000157911	PEX10	153.25
ENSG00000095383	TBC1D2	153.1666667
ENSG00000171450	CDK5R2	153
ENSG00000213066	FGFR1OP	153
ENSG00000178229	ZNF543	153
ENSG00000147509	RGS20	153
ENSG00000164828	SUN1	152.2083333
ENSG00000132326	PER2	152
ENSG00000147996	CBWD5	151
ENSG00000169084	DHRX	151
ENSG00000114541	FRMD4B	151
ENSG00000204386	NEU1	150.877193
ENSG00000156587	UBE2L6	150.7777778
ENSG00000083812	ZNF324	150.5
ENSG00000125814	NAPB	150
ENSG00000083457	ITGAE	150
ENSG00000241258	CRCP	149.2222222
ENSG00000204642	HLA-F	149
ENSG00000131037	EPS8L1	149
ENSG00000100263	RHBDD3	148
ENSG00000158792	SPATA2L	148
ENSG00000272620	AFAP1-AS1	147

ENSG00000243701	DUBR	147
ENSG00000183914	DNAH2	146.7142857
ENSG00000168159	RNF187	146.4074074
ENSG00000171792	RHNO1	146.2857143
ENSG00000160179	ABCG1	146
ENSG00000281376	ABALON	146
ENSG00000197903	HIST1H2BK	145.8
ENSG00000164284	GRPEL2	144.75
ENSG00000133313	CNDP2	144.21875
ENSG00000184898	RBM43	144
ENSG00000186272	ZNF17	144
ENSG00000140464	PML	143.5
ENSG00000213853	EMP2	143.3333333
ENSG00000244509	APOBEC3C	143.1666667
ENSG00000264247	LINC00909	143
ENSG00000165891	E2F7	142.55
ENSG00000137393	RNF144B	142
ENSG00000103145	HCFC1R1	141.3571429
ENSG00000197321	SVIL	141
ENSG00000185838	GNB1L	141
ENSG00000128849	CGNL1	141
ENSG00000165264	NDUFB6	140.8571429
ENSG00000104892	KLC3	140
ENSG00000166578	IQCD	139
ENSG00000175906	ARL4D	139
ENSG00000204438	GPANK1	138.8333333
ENSG00000135749	PCNX2	138.5
ENSG00000141934	PLPP2	138
ENSG00000074621	SLC24A1	138
ENSG00000204540	PSORS1C1	138
ENSG00000090971	NAT14	137.75
ENSG00000175756	AURKAIP1	137.7
ENSG00000168734	PKIG	137.25
ENSG00000158169	FANCC	136.8
ENSG00000125482	TTF1	136.3333333
ENSG00000010278	CD9	136.2777778
ENSG00000165661	QSOX2	136.2142857
ENSG00000174607	UGT8	135.3333333

ENSG00000130803	ZNF317	135.1818182
ENSG00000169306	IL1RAPL1	135
ENSG00000188242	PP7080	135
ENSG00000244187	TMEM141	134.4
ENSG00000189306	RRP7A	133.7272727
ENSG00000187957	DNER	133.6666667
ENSG00000170017	ALCAM	133.6470588
ENSG00000153395	LPCAT1	133.25
ENSG00000182118	FAM89A	133
ENSG00000177427	MIEF2	133
ENSG00000087269	NOP14	133
ENSG00000139209	SLC38A4	132.25
ENSG00000149218	ENDOD1	132.1666667
ENSG00000101160	CTSZ	132.1
ENSG00000174059	CD34	132
ENSG00000243667	WDR92	132
ENSG00000005884	ITGA3	131.7424658
ENSG00000064199	SPA17	131.5
ENSG00000169241	SLC50A1	131
ENSG00000136866	ZFP37	130
ENSG00000101400	SNTA1	129.3
ENSG00000152464	RPP38	129.25
ENSG00000165102	HGSNAT	129.2307692
ENSG00000143842	SOX13	129.1
ENSG00000214013	GANC	128.3333333
ENSG00000110921	MVK	128.25
ENSG00000215548	FRG1JP	128
ENSG00000069974	RAB27A	128
ENSG00000058085	LAMC2	127.4761905
ENSG00000030582	GRN	127.3762376
ENSG00000099337	KCNK6	127
ENSG00000161653	NAGS	126.5
ENSG00000185414	MRPL30	126.5
ENSG00000107281	NPDC1	126.5
ENSG00000134202	GSTM3	126.4090909
ENSG00000154146	NRGN	126
ENSG00000211445	GPX3	125.09375

ENSG00000203780	FANK1	125
ENSG00000198515	CNGA1	125
ENSG00000175592	FOSL1	124.6973684
ENSG00000139725	RHOF	124.5
ENSG00000130638	ATXN10	124.3636364
ENSG00000169710	FASN	124
ENSG00000182578	CSF1R	124
ENSG00000167889	MGAT5B	124
ENSG00000111775	COX6A1	123.96
ENSG00000183077	AFMID	123.6666667
ENSG00000055332	EIF2AK2	123.6
ENSG00000148396	SEC16A	123.2285714
ENSG00000118193	KIF14	122.6
ENSG00000178719	GRINA	122.5555556
ENSG00000174547	MRPL11	122.4444444
ENSG00000116661	FBXO2	122.4285714
ENSG00000180537	RNF182	122
ENSG00000183691	NOG	121.7058824
ENSG00000103150	MLYCD	121.5
ENSG00000103528	SYT17	121
ENSG00000161692	DBF4B	120.8
ENSG00000125430	HS3ST3B1	120.75
ENSG00000147394	ZNF185	120.6666667
ENSG00000185033	SEMA4B	120.3478261
ENSG00000177951	BET1L	120.2777778
ENSG00000165215	CLDN3	120
ENSG00000254614	AP003068.23	120
ENSG00000171608	PIK3CD	119.625
ENSG00000158156	XKR8	119.625
ENSG00000197763	TXNRD3	119.5
ENSG00000184428	TOP1MT	119.375
ENSG00000144821	MYH15	119
ENSG00000237039	AC018738.2	119
ENSG00000161267	BDH1	118.7142857
ENSG00000196227	FAM217B	118.3333333
ENSG00000163950	SLBP	118.1538462
ENSG00000164687	FABP5	118
ENSG00000176953	NFATC2IP	118

ENSG00000158062	UBXN11	117.5
ENSG00000116213	WRAP73	117.3333333
ENSG00000164751	PEX2	117.25
ENSG00000228672	PROB1	117
ENSG00000260822	GS1-358P8.4	116.5
ENSG00000130921	C12orf65	116.125
ENSG00000106868	SUSD1	116
ENSG00000138821	SLC39A8	115.8888889
ENSG00000076944	STXBP2	115.75
ENSG00000114626	ABTB1	115.4285714
ENSG00000082458	DLG3	115.3333333
ENSG00000119705	SLIRP	115.1666667
ENSG00000171729	TMEM51	115.1428571
ENSG00000172889	EGFL7	115
ENSG00000163701	IL17RE	115
ENSG00000234745	HLA-B	114.7589577
ENSG00000169908	TM4SF1	114.75
ENSG00000176720	BOK	114.6
ENSG00000119632	IFI27L2	114.4
ENSG00000267673	FDX2	114
ENSG00000228594	FNDC10	114
ENSG00000146109	ABT1	113.8181818
ENSG00000123240	OPTN	113.7010309
ENSG00000273604	C17orf96	113.5
ENSG00000049089	COL9A2	113.5
ENSG00000125898	FAM110A	113.5
ENSG00000007520	TSR3	113.25
ENSG00000213694	S1PR3	113
ENSG00000004961	HCCS	113
ENSG00000144026	ZNF514	113
ENSG00000167685	ZNF444	113
ENSG00000188636	LDOC1L	113
ENSG00000112208	BAG2	112.6666667
ENSG00000077157	PPP1R12B	112.5
ENSG00000132591	ERAL1	112.4705882
ENSG00000156313	RPGR	112.3333333
ENSG00000050730	TNIP3	112
ENSG00000231721	LINC-PINT	112

ENSG00000215908	CROCCP2	112
ENSG00000150281	CTF1	112
ENSG00000172171	TEFM	111.5
ENSG00000128626	MRPS12	111.4444444
ENSG00000112541	PDE10A	111
ENSG00000145439	CBR4	110.75
ENSG00000152291	TGOLN2	110.3571429
ENSG00000148082	SHC3	110.3333333
ENSG00000183474	GTF2H2C	110
ENSG00000251602	RP11-521B24.3	110
ENSG00000177150	FAM210A	109.9375
ENSG00000167740	CYB5D2	109.75
ENSG00000166016	ABTB2	109.4
ENSG00000130164	LDLR	109.2515723
ENSG00000070882	OSBPL3	109.1896552
ENSG00000170684	ZNF296	109
ENSG00000103168	TAF1C	109
ENSG00000167680	SEMA6B	109
ENSG00000011132	APBA3	108.7142857
ENSG00000184860	SDR42E1	108.5
ENSG00000196150	ZNF250	108.5
ENSG00000157538	DSCR3	108.48
ENSG00000180758	GPR157	108.3333333
ENSG00000173276	ZBTB21	108.2307692
ENSG00000217165	ANKRD18EP	108
ENSG00000149054	ZNF215	108
ENSG00000185860	CCDC190	108
ENSG00000133027	PEMT	107.75
ENSG00000075407	ZNF37A	107.6153846
ENSG00000025772	TOMM34	107.52
ENSG00000116096	SPR	107.5
ENSG00000112902	SEMA5A	107.2
ENSG00000145819	ARHGAP26	107.1666667
ENSG00000178026	LRRC75B	107
ENSG00000197016	ZNF470	106.875
ENSG00000165752	STK32C	106.8571429
ENSG00000111412	C12orf49	106.75

ENSG00000119899	SLC17A5	106.1886792
ENSG00000092929	UNC13D	105.8787879
ENSG00000160209	PDXK	105.6333333
ENSG00000111674	ENO2	105.5576923
ENSG00000124787	RPP40	105.5
ENSG00000182253	SYNM	105.5
ENSG00000174327	SLC16A13	105.4615385
ENSG00000171552	BCL2L1	105.3555556
ENSG00000101255	TRIB3	105.3018868
ENSG00000172164	SNTB1	105.25
ENSG00000103005	USB1	104.862069
ENSG00000134324	LPIN1	104.7222222
ENSG00000163251	FZD5	104.6666667
ENSG00000130731	METTL26	104.3333333
ENSG00000164808	SPIDR	104.3
ENSG00000125648	SLC25A23	104.2
ENSG00000141198	TOM1L1	104.1
ENSG00000146555	SDK1	104
ENSG00000215193	PEX26	103.9
ENSG00000196730	DAPK1	103.6666667
ENSG00000204611	ZNF616	103.5
ENSG00000071655	MBD3	103.4230769
ENSG00000176978	DPP7	103.3157895
ENSG00000001630	CYP51A1	103.2857143
ENSG00000143674	RP5-862P8.2	103.2
ENSG00000212747	FAM127C	103.2
ENSG00000178971	CTC1	103.1666667
ENSG00000277462	ZNF670	103
ENSG00000160298	C21orf58	103
ENSG00000242802	AP5Z1	102.5454545
ENSG00000054148	PHPT1	102.4375
ENSG00000090530	P3H2	102.2727273
ENSG00000165806	CASP7	102.125
ENSG00000176046	NUPR1	102
ENSG00000215375	MYL5	102
ENSG00000196550	FAM72A	102
ENSG00000099624	ATP5D	101.8181818
ENSG00000153046	CDYL	101.5454545

ENSG00000120306	CYSTM1	101.2
ENSG00000168994	PXDC1	101.1428571
ENSG00000257093	KIAA1147	100.8333333
ENSG00000148296	SURF6	100.7777778
ENSG00000143774	GUK1	100.7241379
ENSG00000112378	PERP	100.6363636
ENSG00000086232	EIF2AK1	100.4444444
ENSG00000100302	RASD2	100.25
ENSG00000072042	RDH11	100.1578947
ENSG00000116883	RP11-268J15.5	100
ENSG00000125731	SH2D3A	100

REFERENCES

1. Cowin, P. & Wysolmerski, J. Molecular mechanisms guiding embryonic mammary gland development. *Cold Spring Harb Perspect Biol* **2**, (2010).
2. Macias, H. & Hinck, L. Mammary gland development. *Wiley Interdiscip. Rev. Dev. Biol.* **1**, 533–557 (2012).
3. Howard, B. A. & Gusterson, B. A. Human breast development. *J. Mammary Gland Biol. Neoplasia* **5**, 119–37 (2000).
4. Robinson, G. W. Cooperation of signalling pathways in embryonic mammary gland development. *Nat Rev Genet* **8**, (2007).
5. Ahmed, M. I., Elias, S., Mould, A. W., Bikoff, E. K. & Robertson, E. J. The transcriptional repressor Blimp1 is expressed in rare luminal progenitors and is essential for mammary gland development. *Development* **143**, (2016).
6. Hiremath, M. & Wysolmerski, J. Role of PTHrP in Mammary Gland Development and Breast Cancer. *Clin. Rev. Bone Miner. Metab.* **12**, 178–189 (2014).
7. Inman, J. L., Robertson, C., Mott, J. D. & Bissell, M. J. Mammary gland development: cell fate specification, stem cells and the microenvironment. *Development* **142**, (2015).
8. Richert, M. M., Schwertfeger, K. L., Ryder, J. W. & Anderson, S. M. An atlas of mouse mammary gland development. *J Mammary Gland Biol Neoplasia* (2000).
9. Ewald, A. J. *et al.* Mammary collective cell migration involves transient loss of epithelial features and individual cell migration within the epithelium. *J. Cell Sci.* **125**, (2012).
10. Watson, C. J. & Khaled, W. T. Mammary development in the embryo and adult: a journey of morphogenesis and commitment. *Development* **135**, (2008).
11. Beaudry, K. L., Parsons, C. L. M., Ellis, S. E. & Akers, R. M. Localization and quantitation of macrophages, mast cells, and eosinophils in the developing bovine mammary gland1. *J. Dairy Sci.* **99**, 796–804 (2016).

12. Meyer, M. J., Capuco, A. V., Ross, D. A., Lintault, L. M. & Van Amburgh, M. E. Developmental and Nutritional Regulation of the Prepubertal Bovine Mammary Gland: II. Epithelial Cell Proliferation, Parenchymal Accretion Rate, and Allometric Growth. *J. Dairy Sci.* (2006). doi:10.3168/jds.S0022-0302(06)72476-6
13. Capuco, A. V. & Ellis, S. E. Comparative Aspects of Mammary Gland Development and Homeostasis. *Annu. Rev. Anim. Biosci.* (2013). doi:10.1146/annurev-animal-031412-103632
14. Feinberg, T. Y., Rowe, R. G., Saunders, T. L. & Weiss, S. J. Functional roles of MMP14 and MMP15 in early postnatal mammary gland development. *Development* **143**, 3956–3968 (2016).
15. Zardavas, D., Baselga, J. & Piccart, M. Emerging targeted agents in metastatic breast cancer. *Nat. Rev. Clin. Oncol.* **10**, 191+ (2013).
16. Ekblom, A., Adami, H. O., Trichopoulos, D., Hsieh, C. C. & Lan, S. J. Evidence of prenatal influences on breast cancer risk. *Lancet* (1992). doi:10.1016/0140-6736(92)93019-J
17. Howard, B. & Ashworth, A. Signalling pathways implicated in early mammary gland morphogenesis and breast cancer. *PLoS Genet.* **2**, 1121+ (2006).
18. Hens, J. R. & Wysolmerski, J. J. Key stages of mammary gland development: molecular mechanisms involved in the formation of the embryonic mammary gland. *Breast Cancer Res* **7**, (2005).
19. Hennighausen, L. & Robinson, G. W. Signaling pathways in mammary gland development. *Dev Cell* **1**, (2001).
20. Robinson, G. W., Karpf, A. B. & Kratochwil, K. Regulation of mammary gland development by tissue interaction. *J Mammary Gland Biol Neoplasia* **4**, (1999).
21. Hennighausen, L. & Robinson, G. W. Information networks in the mammary gland. *Nat. Rev. Mol. Cell Biol.* **6**, 715+ (2005).
22. Sakakura, T. Mammary embryogenesis. in *The Mammary Gland: Development, Regulation and Function* (eds. Neville, M. C. & Daniel, C. W.) (Plenum, 1987). doi:10.1007/978-1-4899-5043-7_2

23. Chu, E. Y. *et al.* Canonical WNT signaling promotes mammary placode development and is essential for initiation of mammary gland morphogenesis. *Development* **131**, (2004).
24. Veltmaat, J. M., Van Veelen, W., Thiery, J. P. & Bellusci, S. Identification of the mammary line in mouse by Wnt10b expression. *Dev Dyn* **229**, (2004).
25. Veltmaat, J. M., Mailleux, A. A., Thiery, J. P. & Bellusci, S. Mouse embryonic mammosgenesis as a model for the molecular regulation of pattern formation. *Differentiation* **71**, (2003).
26. Chu, E. Y. *et al.* Canonical WNT signaling promotes mammary placode development and is essential for initiation of mammary gland morphogenesis. *Development* **131**, (2004).
27. van Genderen, C. *et al.* Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1-deficient mice. *Genes Dev* **8**, (1994).
28. Dunbar, M. E. *et al.* Parathyroid hormone-related protein signaling is necessary for sexual dimorphism during embryonic mammary development. *Development* **126**, (1999).
29. Foley, J. *et al.* Parathyroid hormone-related protein maintains mammary epithelial fate and triggers nipple skin differentiation during embryonic breast development. *Development* **128**, (2001).
30. Mailleux, A. A. *et al.* Role of FGF10/FGFR2b signaling during mammary gland development in the mouse embryo. *Development* **129**, (2002).
31. Pond, A. C. *et al.* Fibroblast Growth Factor Receptor Signaling Is Essential for Normal Mammary Gland Development and Stem Cell Function. *Stem Cells* **31**, 178–189 (2013).
32. Villegas, E. *et al.* Plk2 regulates mitotic spindle orientation and mammary gland development. *Development* **141**, (2014).
33. Richert, M. M., Schwertfeger, K. L., Ryder, J. W. & Anderson, S. M. An atlas of mouse mammary gland development. *J Mammary Gland Biol Neoplasia* **5**, (2000).

34. Gouon-Evans, V., Rothenberg, M. E. & Pollard, J. W. Postnatal mammary gland development requires macrophages and eosinophils. *Development* **127**, (2000).
35. Lilla, J. N. & Werb, Z. Mast cells contribute to the stromal microenvironment in mammary gland branching morphogenesis. *Dev. Biol.* (2010). doi:10.1016/j.ydbio.2009.10.021
36. Schwertfeger, K. L., Rosen, J. M. & Cohen, D. A. Mammary gland macrophages: Pleiotropic functions in mammary development. *J. Mammary Gland Biol. Neoplasia* (2006). doi:10.1007/s10911-006-9028-y
37. Gouon-Evans, V., Lin, E. Y. & Pollard, J. W. Requirement of macrophages and eosinophils and their cytokines/chemokines for mammary gland development. *Breast Cancer Res.* (2002). doi:10.1186/bcr441
38. Capuco, A. V., Ellis, S., Wood, D. L., Akers, R. M. & Garrett, W. Postnatal mammary ductal growth: Three-dimensional imaging of cell proliferation, effects of estrogen treatment, and expression of steroid receptors in prepubertal calves. *Tissue Cell* (2002). doi:10.1016/S0040-8166(02)00024-1
39. Hogg, N. A., Harrison, C. J. & Tickle, C. Lumen formation in the developing mouse mammary gland. *J. Embryol. Exp. Morphol.* **73**, 39–57 (1983).
40. Petersen, O. W. & Polyak, K. Stem cells in the human breast. (2011).
41. Gjorevski, N. & Nelson, C. M. Integrated morphodynamic signalling of the mammary gland. *Nat. Rev. Mol. Cell Biol.* **12**, 581+ (2011).
42. Visvader, J. E. Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. *Genes and Development* (2009). doi:10.1101/gad.1849509
43. Jolicoeur, F. Intrauterine Breast Development and the Mammary Myoepithelial Lineage. *J. Mammary Gland Biol. Neoplasia* **10**, 199–210 (2005).
44. Oakes, S. R., Hilton, H. N. & Ormandy, C. J. Key stages in mammary gland development - The alveolar switch: coordinating the proliferative cues and cell fate decisions that drive the formation of lobuloalveoli from ductal epithelium. *Breast Cancer Res.* (2006). doi:10.1186/bcr1411

45. Nelson, C. M., Vanduijn, M. M., Inman, J. L., Fletcher, D. A. & Bissell, M. J. Tissue geometry determines sites of mammary branching morphogenesis in organotypic cultures. *Science* (80-.). **314**, (2006).
46. Teixeira, L. K. *et al.* Cyclin E deregulation promotes loss of specific genomic regions. *Curr. Biol.* **25**, 1327–33 (2015).
47. Wang, Q. A. *et al.* Distinct regulatory mechanisms governing embryonic versus adult adipocyte maturation. *Nat. Cell Biol.* (2015). doi:10.1038/ncb3217
48. Howard, B. A. & Lu, P. Stromal regulation of embryonic and postnatal mammary epithelial development and differentiation. *Semin. Cell Dev. Biol.* **25**, 43–51 (2014).
49. Topper, Y. J. & Freeman, C. S. Multiple hormone interactions in the developmental biology of the mammary gland. *Physiol. Rev.* **60**, 1049–106 (1980).
50. Kim, H. Y. & Nelson, C. M. Extracellular matrix and cytoskeletal dynamics during branching morphogenesis. *Organogenesis* **8**, 56–64 (2012).
51. Cardiff, R. D. & Wellings, S. R. The comparative pathology of human and mouse mammary glands. *J Mammary Gland Biol Neoplasia* (1999).
52. Brisken, C. & Ataca, D. Endocrine hormones and local signals during the development of the mouse mammary gland. *Wiley Interdiscip. Rev. Dev. Biol.* **4**, 181–195 (2015).
53. Tong, Q. & Hotamisligil, G. S. Developmental biology: Cell fate in the mammary gland. *Nature* **445**, 724+ (2007).
54. Cadigan, K. M. Wnt – β -catenin signaling. *Curr. Biol.* (2008). doi:10.1016/j.cub.2008.08.017
55. Gurusamy, D., Ruiz-Torres, S. J., Johnson, A. L., Smith, D. A. & Waltz, S. E. Hepatocyte growth factor-like protein is a positive regulator of early mammary gland ductal morphogenesis. *Mech. Dev.* **133**, 11–22 (2014).
56. Brantley, D. M. *et al.* Nuclear factor-kappaB (NF-kappaB) regulates proliferation and branching in mouse mammary epithelium. *Mol. Biol. Cell* **12**, 1445–55 (2001).

57. Wickenden, J. A. & Watson, C. J. Key signalling nodes in mammary gland development and cancer. Signalling downstream of PI3 kinase in mammary epithelium: a play in 3 Akts. *Breast Cancer Res* **12**, (2010).
58. Cantley, L. The phosphoinositide 3-kinase pathway. *Sci.* (2002). doi:10.1126/science.296.5573.1655
59. LaRocca, J., Pietruska, J., Hixon, M., Clarke, R. & Briegel, K. Akt1 Is Essential for Postnatal Mammary Gland Development, Function, and the Expression of Btn1a1. *PLoS One* **6**, e24432 (2011).
60. Watson, C. J. & Neoh, K. The Stat family of transcription factors have diverse roles in mammary gland development. *Semin. Cell Dev. Biol.* (2008). doi:10.1016/j.semcdb.2008.07.021
61. Wagner, K.-U. & Schmidt, J. W. The two faces of Janus kinases and their respective STATs in mammary gland development and cancer. *J. Carcinog.* (2011). doi:10.4103/1477-3163.90677
62. Shillingford, J. M. *et al.* Jak2 is an essential tyrosine kinase involved in pregnancy-mediated development of mammary secretory epithelium. *Mol. Endocrinol. (Baltimore, Md)* (2002). doi:10.1210/me.16.3.563
63. Wagner, K.-U. *et al.* Impaired alveologenesis and maintenance of secretory mammary epithelial cells in Jak2 conditional knockout mice. *Mol. Cell. Biol.* (2004). doi:10.1128/MCB.24.12.5510-5520.2004
64. Sakamoto, K. *et al.* Janus Kinase 1 Is Essential for Inflammatory Cytokine Signaling and Mammary Gland Remodeling. *Mol. Cell. Biol.* **36**, 1673–90 (2016).
65. Haricharan, S. & Li, Y. STAT signaling in mammary gland differentiation, cell survival and tumorigenesis. *Mol. Cell. Endocrinol.* **382**, 560–569 (2014).
66. Eblaghie, M. C. *et al.* Interactions between FGF and Wnt signals and Tbx3 gene expression in mammary gland initiation in mouse embryos. *J Anat* **205**, (2004).
67. Davenport, T. G., Jerome-Majewska, L. & Papaioannou, V. E. Mammary gland, limb and yolk sac defects in mice lacking Tbx3, the gene mutated in human ulnar mammary syndrome. *Development* **130**, (2003).

68. Chen, J. *et al.* Microarray analysis of Tbx2-directed gene expression: a possible role in osteogenesis. *Mol Cell Endocrinol* **177**, (2001).
69. Douglas, N. C. & Papaioannou, V. E. The T-box Transcription Factors TBX2 and TBX3 in Mammary Gland Development and Breast Cancer. *J. Mammary Gland Biol. Neoplasia* **18**, 143–147 (2013).
70. Koledova, Z. *et al.* SPRY1 regulates mammary epithelial morphogenesis by modulating EGFR-dependent stromal paracrine signaling and ECM remodeling. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5731-40 (2016).
71. Wood, T. L., Richert, M. M., Stull, M. A. & Allar, M. A. The Insulin-Like Growth Factors (IGFs) and IGF Binding Proteins in Postnatal Development of Murine Mammary Glands. *J. Mammary Gland Biol. Neoplasia* **5**, 31–42 (2000).
72. Wicha, M. S., Liotta, L. A., Vonderhaar, B. K. & Kidwell, W. R. Effects of inhibition of basement membrane collagen deposition on rat mammary gland development. *Dev. Biol.* (1980). doi:10.1016/0012-1606(80)90402-9
73. Barcellos-Hoff, M. H., Aggeler, J., Ram, T. G. & Bissell, M. J. Functional differentiation and alveolar morphogenesis of primary mammary cultures on reconstituted basement membrane. *Development* **105**, 223–35 (1989).
74. Lu, P., Weaver, V. M. & Werb, Z. The extracellular matrix: a dynamic niche in cancer progression. *J. Cell Biol.* **196**, 395–406 (2012).
75. Rozario, T. & DeSimone, D. W. The extracellular matrix in development and morphogenesis: a dynamic view. *Dev. Biol.* **341**, 126–40 (2010).
76. Hu, G., Li, L. & Xu, W. Extracellular matrix in mammary gland development and breast cancer progression. *Front. Lab. Med.* **1**, 36–39 (2017).
77. Daley, W. P. & Yamada, K. M. ECM-modulated cellular dynamics as a driving force for tissue morphogenesis. *Curr. Opin. Genet. Dev.* **23**, 408–14 (2013).
78. Wiseman, B. S. & Werb, Z. Stromal effects on mammary gland development and breast cancer. (Review : Development). *Science (80-.).* **296**, 1046+ (2002).
79. Nelson, D. A. & Larsen, M. Heterotypic control of basement membrane dynamics during branching morphogenesis. *Dev. Biol.* **401**, 103–109 (2015).

80. Booth, B. W. *et al.* The mammary microenvironment alters the differentiation repertoire of neural stem cells. *Proc. Natl. Acad. Sci. U. S. A.* (2008). doi:10.1073/pnas.0803214105
81. Boulanger, C. A., Mack, D. L., Booth, B. W. & Smith, G. H. Interaction with the mammary microenvironment redirects spermatogenic cell fate in vivo. *Proc. Natl. Acad. Sci. U. S. A.* (2007). doi:10.1073/pnas.0611637104
82. Rørth, P. Fellow travellers: emergent properties of collective cell migration. *EMBO Rep.* (2012). doi:10.1038/embor.2012.149
83. Ewald, A. J., Brenot, A., Duong, M., Chan, B. S. & Werb, Z. Collective epithelial migration and cell rearrangements drive mammary branching morphogenesis. *Dev. Cell* **14**, 570–81 (2008).
84. Martinson, H. A., Jindal, S., Durand-Rougely, C., Borges, V. F. & Schedin, P. Wound healing-like immune program facilitates postpartum mammary gland involution and tumor progression. *Int. J. Cancer* **136**, 1803–1813 (2015).
85. Beaudry, K. L. Innate Immune Cells may be Involved in Prepubertal Bovine Mammary Development. (2015).
86. Plaks, V. *et al.* Adaptive Immune Regulation of Mammary Postnatal Organogenesis. *Dev. Cell* **34**, 493–504 (2015).
87. Reed, J. R. & Schwertfeger, K. L. Immune cell location and function during post-natal mammary gland development. *Journal of Mammary Gland Biology and Neoplasia* (2010). doi:10.1007/s10911-010-9188-7
88. Coussens, L. M. & Pollard, J. W. Leukocytes in mammary development and cancer. *Cold Spring Harb. Perspect. Biol.* (2011). doi:10.1101/cshperspect.a003285
89. Khaled, W. T. *et al.* The IL-4/IL-13/Stat6 signalling pathway promotes luminal mammary epithelial cell development. *Development* (2007). doi:10.1242/dev.003194
90. Horton, M. A. *Macrophages and Related Cells, 5th Edition Blood Cell Biochemistry, Volume 5* (ed. M. A. Horton). New York: Plenum Press. (1996).

91. Brady, N. J., Chuntova, P. & Schwertfeger, K. L. Macrophages: Regulators of the Inflammatory Microenvironment during Mammary Gland Development and Breast Cancer. *Mediators Inflamm.* **2016**, 1–13 (2016).
92. King, B. L. *et al.* Immunocytochemical analysis of breast cells obtained by ductal lavage. *Cancer* (2002). doi:10.1002/cncr.10719
93. Sun, X. & Ingman, W. V. Cytokine Networks That Mediate Epithelial Cell-Macrophage Crosstalk in the Mammary Gland: Implications for Development and Cancer. *J. Mammary Gland Biol. Neoplasia* **19**, 191–201 (2014).
94. Ryan, G. R. *et al.* Rescue of the colony-stimulating factor 1 (CSF-1)-nullizygous mouse (Csf1op/Csf1op) phenotype with a CSF-1 transgene and identification of sites of local CSF-1 synthesis. *Blood* (2001). doi:10.1182/blood.V98.1.74
95. Ingman, W. V., Wyckoff, J., Gouon-Evans, V., Condeelis, J. & Pollard, J. W. Macrophages promote collagen fibrillogenesis around terminal end buds of the developing mammary gland. *Dev. Dyn.* (2006). doi:10.1002/dvdy.20972
96. Van Nguyen, A. & Pollard, J. W. Colony Stimulating Factor-1 Is Required to Recruit Macrophages into the Mammary Gland to Facilitate Mammary Ductal Outgrowth. *Dev. Biol.* (2002). doi:10.1006/dbio.2002.0669
97. Lyons, R. M., Gentry, L. E., Purchio, A. F. & Moses, H. L. Mechanism of activation of latent recombinant transforming growth factor beta 1 by plasmin. *J Cell Biol.* (1990).
98. Kacinski, B. M. *et al.* FMS (CSF-1 receptor) and CSF-1 transcripts and protein are expressed by human breast carcinomas in vivo and in vitro. *Oncogene* (1991).
99. Sapi, E., Flick, M. B., Rodov, S., Carter, D. & Kacinski, B. M. Expression of CSF-I and CSF-I receptor by normal lactating mammary epithelial cells. *J. Soc. Gynecol. Investig.* (1998). doi:10.1016/S1071-5576(97)00108-1
100. Sapi, E. The role of CSF-1 in normal physiology of mammary gland and breast cancer: an update. *Exp. Biol. Med. (Maywood)*. (2004). doi:10.1258/ebm.2009.009e01

101. Mantovani, A., Sozzani, S., Locati, M., Allavena, P. & Sica, A. Macrophage polarization: Tumor-associated macrophages as a paradigm for polarized M2 mononuclear phagocytes. *Trends in Immunology* (2002). doi:10.1016/S1471-4906(02)02302-5
102. Sica, A. & Mantovani, A. Macrophage plasticity and polarization: In vivo veritas. *Journal of Clinical Investigation* (2012). doi:10.1172/JCI59643
103. Correia, A. L., Mori, H., Chen, E. I., Schmitt, F. C. & Bissell, M. J. The hemopexin domain of MMP3 is responsible for mammary epithelial invasion and morphogenesis through extracellular interaction with HSP90 β . *Genes Dev.* **27**, 805–17 (2013).
104. Fata, J. E., Werb, Z. & Bissell, M. J. Regulation of mammary gland branching morphogenesis by the extracellular matrix and its remodeling enzymes. *Breast Cancer Res.* **6**, 1–11 (2003).
105. Hotary, K., Allen, E., Punturieri, A., Yana, I. & Weiss, S. J. Regulation of cell invasion and morphogenesis in a three-dimensional type I collagen matrix by membrane-type matrix metalloproteinases 1, 2, and 3. *J. Cell Biol.* (2000). doi:10.1083/jcb.149.6.1309
106. Silberstein, G. B. & Daniel, C. W. Reversible inhibition of mammary gland growth by transforming growth factor-beta. *Science* **237**, 291–3 (1987).
107. Witty, J. P., Wright, J. H. & Matrisian, L. M. Matrix metalloproteinases are expressed during ductal and alveolar mammary morphogenesis, and misregulation of stromelysin-1 in transgenic mice induces unscheduled alveolar development. *Mol. Biol. Cell* (1995). doi:10.1091/MBC.6.10.1287
108. Kessenbrock, K. *et al.* A role for matrix metalloproteinases in regulating mammary stem cell function via the Wnt signaling pathway. *Cell Stem Cell* **13**, 300–13 (2013).
109. Tan, J. *et al.* Stromal matrix metalloproteinase-11 is involved in the mammary gland postnatal development. *Oncogene* **33**, 4050–9 (2014).
110. Bingle, L., Brown, N. J. & Lewis, C. E. The role of tumour-associated macrophages in tumour progression: Implications for new anticancer therapies. *Journal of Pathology* (2002). doi:10.1002/path.1027

111. Leek, R. D. *et al.* Association of macrophage infiltration with angiogenesis and prognosis in invasive breast carcinoma. *Cancer Res.* (1996).
112. Mahmoud, S. M. *et al.* Tumour-infiltrating macrophages and clinical outcome in breast cancer. *J Clin Pathol* (2012). doi:jclinpath-2011-200355 [pii]r10.1136/jclinpath-2011-200355
113. Qian, B. & Pollard, J. Macrophage diversity enhances tumor progression and metastasis. *Cell* (2010). doi:10.1016/j.cell.2010.03.014
114. Wilson, G. J. *et al.* Atypical chemokine receptor ACKR2 controls branching morphogenesis in the developing mammary gland. *Development* **144**, (2017).
115. Mikaelian, I. *et al.* Expression of terminal differentiation proteins defines stages of mouse mammary gland development. *Vet. Pathol.* (2006). doi:10.1354/vp.43-1-36
116. Yamamoto, M., Nakata, H., Kumchantuek, T., Sakulsak, N. & Iseki, S. Immunohistochemical localization of keratin 5 in the submandibular gland in adult and postnatal developing mice. *Histochem. Cell Biol.* **145**, 327–339 (2016).
117. Sun, P., Yuan, Y., Li, A., Li, B. & Dai, X. Cytokeratin expression during mouse embryonic and early postnatal mammary gland development. *Histochem. Cell Biol.* **133**, 213–221 (2010).
118. Ramovs, V., te Molder, L. & Sonnenberg, A. The opposing roles of laminin-binding integrins in cancer. *Matrix Biol.* **57**, 213–243 (2017).
119. Hynes, R. O. Integrins: versatility, modulation, and signaling in cell adhesion. *Cell* **69**, (1992).
120. Ginsberg, M. H., Du, X. & Plow, E. F. Inside-out integrin signalling. *Curr Opin Cell Biol* **4**, (1992).
121. Zaidel-Bar, R., Itzkovitz, S., Ma'ayan, A., Iyengar, R. & Geiger, B. Functional atlas of the integrin adhesome. *Nat. Cell Biol.* (2007). doi:10.1038/ncb0807-858
122. Legate, K. R., Wickström, S. A. & Fässler, R. Genetic and cell biological analysis of integrin outside-in signaling. *Genes and Development* (2009). doi:10.1101/gad.1758709

123. Akhtar, N. *et al.* Molecular dissection of integrin signalling proteins in the control of mammary epithelial development and differentiation. *Development* **136**, (2009).
124. Moreno-Layseca, P. & Streuli, C. H. Signalling pathways linking integrins with cell cycle progression. *Matrix Biol.* **34**, 144–153 (2014).
125. Walker, J. L. & Assoian, R. K. Integrin-dependent signal transduction regulating cyclin D1 expression and G1 phase cell cycle progression. *Cancer Metastasis Rev.* (2005). doi:DOI 10.1007/s10555-005-5130-7
126. Falcioni, R. *et al.* Alpha 6 beta 4 and alpha 6 beta 1 integrins associate with ErbB-2 in human carcinoma cell lines. *Exp Cell Res* (1997).
127. Nisticò, P., Di Modugno, F., Spada, S. & Bissell, M. J. β 1 and β 4 integrins: from breast development to clinical practice. *Breast Cancer Res.* **16**, 459 (2014).
128. Taddei, I. *et al.* Integrins in mammary gland development and differentiation of mammary epithelium. *J Mammary Gland Biol Neoplasia* (2003). doi:10.1023/B:JOMG.0000017426.74915.b9\r481709 [pii]
129. Lahlou, H. & Muller, W. J. β 1-integrins signaling and mammary tumor progression in transgenic mouse models: implications for human breast cancer. *Breast Cancer Res* **13**, (2011).
130. Rooney, N., Wang, P., Brennan, K., Gilmore, A. P. & Streuli, C. H. The Integrin-Mediated ILK-Parvin- α 5 β 1 Signaling Axis Controls Differentiation in Mammary Epithelial Cells. *J. Cell. Physiol.* **231**, 2408–2417 (2016).
131. Li, J., Sun, H., Feltri, M. L. & Mercurio, A. M. Integrin β 4 regulation of PTHrP underlies its contribution to mammary gland development. *Dev. Biol.* **407**, 313–320 (2015).
132. Mercurio, A. M. Laminin receptors: achieving specificity through cooperation. *Trends in Cell Biology* (1995). doi:10.1016/S0962-8924(00)89100-X
133. Borradori, L. & Sonnenberg, A. Structure and function of hemidesmosomes: More than simple adhesion complexes. *Journal of Investigative Dermatology* (1999). doi:10.1046/j.1523-1747.1999.00546.x

134. Zhao, F. *et al.* Up-regulation of integrin $\alpha 6\beta 4$ expression by mitogens involved in dairy cow mammary development. *Vitr. Cell. Dev. Biol. - Anim.* **51**, 287–299 (2015).
135. Klinowska, T. C. *et al.* Epithelial development and differentiation in the mammary gland is not dependent on alpha 3 or alpha 6 integrin subunits. *Dev Biol* **233**, (2001).
136. Stahl, S., Weitzman, S. & Jones, J. C. The role of laminin-5 and its receptors in mammary epithelial cell branching morphogenesis. *J. Cell Sci.* 55–63 (1997).
137. NisticA[sup.2], P., Di Modugno, F., Spada, S. & Bissell, M. J. beta]1 and beta]4 integrins: from breast development to clinical practice. *Breast Cancer Res.* **16**, (2014).
138. Alford, D. & Taylor-Papadimitriou, J. Cell adhesion molecules in the normal and cancerous mammary gland. *J Mammary Gland Biol Neoplasia* (1996).
139. Weaver, V. M. *et al.* Reversion of the Malignant Phenotype of Human Breast Cells in Three-Dimensional Culture and In Vivo by Integrin Blocking Antibodies. *J Cell Biol* **137**, 231 (1997).
140. Chen, C. S., Alonso, J. L., Ostuni, E., Whitesides, G. M. & Ingber, D. E. Cell shape provides global control of focal adhesion assembly. *Biochem. Biophys. Res. Commun.* (2003). doi:10.1016/S0006-291X(03)01165-3
141. Durbeej, M. Laminins. *Cell Tissue Res* (2010). doi:10.1007/s00441-009-0838-2
142. Srichai, M. B. & Zent, R. Integrin structure and function. in *Cell-Extracellular Matrix Interactions in Cancer* (2010). doi:10.1007/978-1-4419-0814-8_2
143. Klinowska, T. C. M. *et al.* Laminin and $\beta 1$ Integrins Are Crucial for Normal Mammary Gland Development in the Mouse. *Dev. Biol.* **215**, 13–32 (1999).
144. Chomwisarutkun, K., Murani, E., Ponsuksili, S. & Wimmers, K. Gene expression analysis of mammary tissue during fetal bud formation and growth in two pig breeds “ indications of prenatal initiation of postnatal phenotypic differences. *BMC Dev. Biol.* **12**, 13 (2012).
145. Anbazhagan, R., Bartek, J., Monaghan, P. & Gusterson, B. A. Growth and development of the human infant breast. *Am. J. Anat.* **192**, 407–417 (1991).

146. Qu, Y. *et al.* Evaluation of MCF10A as a Reliable Model for Normal Human Mammary Epithelial Cells. *PLoS One* **10**, e0131285 (2015).
147. Sullivan, T. P., Eaglstein, W. H., Davis, S. C. & Mertz, P. THE PIG AS A MODEL FOR HUMAN WOUND HEALING. *Wound Repair Regen.* **9**, 66–76 (2001).
148. Miller, E. R. & Ullrey, D. E. The Pig as a Model for Human Nutrition. *Annu. Rev. Nutr.* **7**, 361–382 (1987).
149. Meurens, F., Summerfield, A., Nauwynck, H., Saif, L. & Gerdt, V. The pig: a model for human infectious diseases. *Trends Microbiol.* **20**, 50–57 (2012).
150. Lunney, J. K. Advances in Swine Biomedical Model Genomics. *Int. J. Biol. Sci.* *Int. J. Biol. Sci.* www.biolsci.org **3**, 179–184 (2007).
151. Fairbairn, L., Kapetanovic, R., Sester, D. P. & Hume, D. A. The mononuclear phagocyte system of the pig as a model for understanding human innate immunity and disease. *J. Leukoc. Biol.* **89**, 855–71 (2011).
152. Kouros-Mehr, H. & Werb, Z. Candidate regulators of mammary branching morphogenesis identified by genome-wide transcript analysis. *Dev. Dyn.* **235**, 3404–3412 (2006).
153. Ozsolak, F. & Milos, P. M. RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* **12**, 87–98 (2011).
154. Vijay, N., Poelstra, J. W., K?nstner, A. & Wolf, J. B. W. Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive *in silico* assessment of RNA-seq experiments. *Mol. Ecol.* **22**, 620–634 (2013).
155. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* (2008). doi:10.1038/nmeth.1226
156. Morozova, O., Hirst, M. & Marra, M. A. Applications of new sequencing technologies for transcriptome analysis. *Annu Rev Genomics Hum Genet* (2009). doi:10.1146/annurev-genom-082908-145957

157. Sultan, M. *et al.* A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**, 956–60 (2008).
158. Metzker, M. L. Sequencing technologies ? the next generation. *Nat. Rev. Genet.* **11**, 31–46 (2010).
159. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
160. McGettigan, P. A. Transcriptomics in the RNA-seq era. *Curr. Opin. Chem. Biol.* **17**, 4–11 (2013).
161. Chopra-Dewasthaly, R. *et al.* Comprehensive RNA-Seq Profiling to Evaluate the Sheep Mammary Gland Transcriptome in Response to Experimental *Mycoplasma agalactiae* Infection. *PLoS One* **12**, e0170015 (2017).
162. Cánovas, A. *et al.* Comparison of five different RNA sources to examine the lactating bovine mammary gland transcriptome using RNA-Sequencing. **4**, 5297 (2014).
163. Bergamo, A. *et al.* RNA-seq analysis of the whole transcriptome of MDA-MB-231 mammary carcinoma cells exposed to the antimetastatic drug NAMI-A. *Metallomics* **7**, 1439–1450 (2015).
164. Mooney, M. *et al.* Comparative RNA-Seq and microarray analysis of gene expression changes in B-cell lymphomas of *Canis familiaris*. *PLoS One* **8**, e61088 (2013).
165. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Meth* **5**, 621–628 (2008).
166. Zhao, S. *et al.* Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. *PLoS One* **9**, e78644 (2014).
167. Mantione, K. J. *et al.* Comparing bioinformatic gene expression profiling methods: microarray and RNA-Seq. *Med. Sci. Monit. Basic Res.* **20**, 138–42 (2014).
168. Seyednasrollah, F., Laiho, A. & Elo, L. L. Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief. Bioinform.* **16**, 59–70 (2015).

169. Soneson, C. & Delorenzi, M. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics* **14**, 91 (2013).
170. Giorgi, F. M., Del Fabbro, C. & Licausi, F. Comparative study of RNA-seq- and Microarray-derived coexpression networks in *Arabidopsis thaliana*. *Bioinformatics* **29**, 717–724 (2013).
171. Avril-Sassen, S. *et al.* Characterisation of microRNA expression in post-natal mouse mammary gland development. *BMC Genomics* **10**, 548 (2009).
172. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).
173. D'Urso, P. Exploratory multivariate analysis for empirical information affected by uncertainty and modeled in a fuzzy manner: a review. *Granul. Comput.* 1–23 (2017). doi:10.1007/s41066-017-0040-y
174. Paliy, O. & Shankar, V. Application of multivariate statistical techniques in microbial ecology. *Mol. Ecol.* (2016). doi:10.1111/mec.13536
175. Warton, D. I., Foster, S. D., De'ath, G., Stoklosa, J. & Dunstan, P. K. Model-based thinking for community ecology. *Plant Ecol.* **216**, 669–682 (2015).
176. Warton, D. I. *et al.* So Many Variables: Joint Modeling in Community Ecology. *Trends Ecol. Evol.* **30**, 766–779 (2015).
177. Buttigieg, P. L. *et al.* A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses. *FEMS Microbiol. Ecol.* **90**, 543–550 (2014).
178. Rigsbee, L. *et al.* Quantitative Profiling of Gut Microbiota of Children With Diarrhea-Predominant Irritable Bowel Syndrome. *Am. J. Gastroenterol.* **107**, 1740–1751 (2012).
179. Shankar, V. *et al.* Species and genus level resolution analysis of gut microbiota in *Clostridium difficile* patients following fecal microbiota transplantation. *Microbiome* **2**, 13 (2014).

180. Gopalakrishnan Radhika, A., Chawla, S., Nanda, P., Yadav, G. & Radhakrishnan, G. A Multivariate Analysis of Correlation between Severity and Duration of Symptoms, Patient Profile and Stage of Endometriosis. *Open J. Obstet. Gynecol.* **6**, 615–622 (2016).
181. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer sSiegel, R. L., Miller, K. D., & Jemal, A. (2016). Cancer statistics, 2016. *CA Cancer J Clin*, 66(1), 7–30. <https://doi.org/10.3322/caac.21332>tatistics, 2016. *CA Cancer J Clin* (2016). doi:10.3322/caac.21332
182. Badve, S. *et al.* Basal-like and triple-negative breast cancers: a critical review with an emphasis on the implications for pathologists and oncologists. *Mod. Pathol.* **24**, 157–167 (2011).
183. Sorlie, T. *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl. Acad. Sci.* (2003). doi:10.1073/pnas.0932692100
184. Hu, Z. *et al.* The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* (2006). doi:10.1186/1471-2164-7-96
185. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* (2000). doi:10.1038/35021093
186. Sørli, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci.* (2001).
187. Dent R1, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, Lickley LA, Rawlinson E, Sun P, N. S. Triple-negative breast cancer: clinical features and patterns of recurrence. *Clin Cancer Res.* (2007).
188. Hergueta-Redondo, M., Palacios, J., Cano, A. & Moreno-Bueno, G. ‘New’ molecular taxonomy in breast cancer. *Clin.Transl.Oncol.* (2008).
189. Carey, L. A. *et al.* Race, Breast Cancer Subtypes, and Survival in the Carolina Breast Cancer Study. *JAMA* (2006). doi:10.1001/jama.295.21.2492
190. Penault-Llorca, F. & Viale, G. Pathological and molecular diagnosis of triple-negative breast cancer: A clinical perspective. *Ann. Oncol.* (2012). doi:10.1093/annonc/mds190

191. Colleoni, M. & Montagna, E. Neoadjuvant therapy for ER-positive breast cancers. *Ann. Oncol.* (2012). doi:10.1093/annonc/mds305
192. Vici, P. *et al.* Triple positive breast cancer: A distinct subtype? *Cancer Treatment Reviews* (2015). doi:10.1016/j.ctrv.2014.12.005
193. Lucas, F. V. & Perez-Mesa, C. Inflammatory carcinoma of the breast. *Cancer* (1978). doi:10.1002/1097-0142(197804)41:4<1595::AID-CNCR2820410450>3.0.CO;2-Y
194. K.W., H., W.F., A., S.S., D., H.A., Y. & P.H., L. Trends in inflammatory breast carcinoma incidence and survival: The surveillance, epidemiology, and end results program at the National Cancer Institute. *Journal of the National Cancer Institute* (2005).
195. Robertson, F. M. *et al.* Inflammatory breast cancer: the disease, the biology, the treatment. *CA. Cancer J. Clin.* (2010). doi:10.3322/caac.20082; 10.3322/caac.20082
196. Anderson, W. F., Schairer, C., Chen, B. E., Hance, K. W. & Levine, P. H. Epidemiology of inflammatory breast cancer (IBC). *Breast Dis.* (2010). doi:10.1007/978-94-007-3907-9_2
197. Fernandez, S. V. *et al.* Inflammatory breast cancer (IBC): Clues for targeted therapies. *Breast Cancer Res. Treat.* (2013). doi:10.1007/s10549-013-2600-4
198. Bertucci, F. *et al.* Gene expression profiles of inflammatory breast cancer: Correlation with response to neoadjuvant chemotherapy and metastasis-free survival. *Ann. Oncol.* (2014). doi:10.1093/annonc/mdt496
199. Haffty, B. G. *et al.* Locoregional relapse and distant metastasis in conservatively managed triple negative early-stage breast cancer. *J. Clin. Oncol.* (2006). doi:10.1200/JCO.2006.06.5664
200. Carey, L. A. *et al.* The triple negative paradox: Primary tumor chemosensitivity of breast cancer subtypes. *Clin. Cancer Res.* (2007). doi:10.1158/1078-0432.CCR-06-1109
201. Kreike, B. *et al.* Gene expression profiling and histopathological characterization of triple negative/basal-like breast carcinomas. *Breast Cancer Res.* (2007).

202. Jacquemier, J. *et al.* Typical medullary breast carcinomas have a basal/myoepithelial phenotype. *J. Pathol.* (2005). doi:10.1002/path.1845
203. Diaz, L. K., Cryns, V. L., Symmans, W. F. & Sneige, N. Triple negative breast carcinoma and the basal phenotype: from expression profiling to clinical practice. *Adv Anat Pathol* (2007). doi:10.1097/PAP.0b013e3181594733
204. Rakha, E. A. & Ellis, I. O. Triple-negative/basal-like breast cancer: review. *Pathology* **41**, 40–47 (2009).
205. Rouzier, R. *et al.* Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clin Cancer Res* (2005). doi:10.1158/1078-0432.CCR-04-2421
206. Calza, S. *et al.* Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients. *Breast Cancer Res.* (2006). doi:10.1186/bcr1517
207. Jumppanen, M. *et al.* Basal-like phenotype is not associated with patient survival in estrogen-receptor-negative breast cancers. *Breast Cancer Res* (2007). doi:10.1186/bcr1649
208. Bertucci, F. *et al.* How basal are triple-negative breast cancers? *Int J Cancer* (2008). doi:10.1002/ijc.23518
209. Tischkowitz, M. *et al.* Use of immunohistochemical markers can refine prognosis in triple negative breast cancer. *BMC Cancer* (2007). doi:10.1186/1471-2407-7-134
210. Tan, D. S. *et al.* Triple negative breast cancer: molecular profiling and prognostic impact in adjuvant anthracycline-treated patients. *Breast Cancer Res Treat* (2008). doi:10.1007/s10549-007-9756-8
211. Rakha, E. A. *et al.* Prognostic markers in triple-negative breast cancer. *Cancer* (2007). doi:10.1002/cncr.22381
212. Lakhani, S. R. *et al.* Prediction of BRCA1 status in patients with breast cancer using estrogen receptor and basal phenotype. *Clin. Cancer Res.* (2005). doi:10.1158/1078-0432.CCR-04-2424

213. van de Rijn, M. *et al.* Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. *Am. J. Pathol.* (2002). doi:10.1016/S0002-9440(10)64476-8
214. Rakha, E. A. *et al.* Morphological and immunophenotypic analysis of breast carcinomas with basal and myoepithelial differentiation. *J. Pathol.* (2006). doi:10.1002/path.1916
215. J., K. Possible treatment strategies for triple-negative breast cancer on the basis of molecular characteristics. *Breast Cancer* (2009).
216. Brady-West, D. C. & McGrowder, D. a. Triple negative breast cancer: therapeutic and prognostic implications. *Asian Pacific J. Cancer Prev.* (2011).
217. Podo, F. *et al.* Triple-negative breast cancer: Present challenges and new perspectives. *Themat. Issue Mol. Biol. Breast Cancer* (2010). doi:http://dx.doi.org/10.1016/j.molonc.2010.04.006
218. Piccart-Gebhart, M. *et al.* Trastuzumab after Adjuvant Chemotherapy in HER2-Positive Breast Cancer. *N Engl J Med* **353**, 1659–1672 (2005).
219. Valabrega, G., Montemurro, F. & Aglietta, M. Trastuzumab: mechanism of action, resistance and future perspectives in HER2-overexpressing breast cancer. *Ann. Oncol.* **18**, 977–984 (2007).
220. Bianchini, G., Balko, J. M., Mayer, I. A., Sanders, M. E. & Gianni, L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nat Rev Clin Oncol* **advance on**, (2016).
221. Ueda, Y. *et al.* Overexpression of HER2 (erbB2) in human breast epithelial cells unmasks transforming growth factor beta-induced cell motility. *J Biol Chem* **279**, (2004).
222. Bonotto, M. *et al.* Measures of outcome in metastatic breast cancer: insights from a real-world scenario. *Oncologist* (2014). doi:10.1634/theoncologist.2014-0002
223. Dogan, B. E. & Turnbull, L. W. Imaging of triple-negative breast cancer. *Ann. Oncol.* **23**, vi23-vi29 (2012).
224. Weigelt, B. *et al.* Metastatic breast carcinomas display genomic and transcriptomic heterogeneity. *Mod. Pathol.* **28**, 340–351 (2015).

225. Malorni, L. *et al.* Clinical and biologic features of triple-negative breast cancers in a large cohort of patients with long-term follow-up. *Breast Cancer Res Treat* (2012). doi:10.1007/s10549-012-2315-y
226. Conlin, A. K. & Seidman, A. D. Beyond cytotoxic chemotherapy for the first-line treatment of HER2-negative, hormone-insensitive metastatic breast cancer: current status and future opportunities. *Clin Breast Cancer* (2008). doi:H1244Q816751WR14 [pii]r10.3816/CBC.2008.n.024
227. Carey, L. A. Directed Therapy of Subtypes of Triple-Negative Breast Cancer. *Oncologist* (2010). doi:10.1634/theoncologist.2010-S5-49
228. Venkitaraman, A. R. Linking the cellular functions of BRCA genes to cancer pathogenesis and treatment. *Annu. Rev. Pathol.* (2009). doi:10.1146/annurev.pathol.3.121806.151422
229. Atchley, D. P. *et al.* Clinical and pathologic characteristics of patients with BRCA-positive and BRCA-negative breast cancer. *J. Clin. Oncol.* (2008). doi:10.1200/JCO.2008.16.6231
230. Turner, N., Tutt, A. & Ashworth, A. Opinion: Hallmarks of 'BRCAness' in sporadic cancers. *Nat. Rev. Cancer* (2004). doi:10.1038/nrc1457
231. Reis-Filho, J. S. *et al.* Cyclin D1 protein overexpression and CCND1 amplification in breast carcinomas: an immunohistochemical and chromogenic in situ hybridisation analysis. *Mod. Pathol.* (2006). doi:10.1038/modpathol.3800621
232. Foulkes, W. D. *et al.* The prognostic implication of the basal-like (cyclin E high/p27 low/p53+/glomeruloid-microvascular-proliferation+) phenotype of BRCA1-related breast cancer. *Cancer Res.* (2004). doi:10.1158/0008-5472.CAN-03-2970
233. Vaziri, S. a, Tubbs, R. R., Darlington, G. & Casey, G. Absence of CCND1 gene amplification in breast tumours of BRCA1 mutation carriers. *Mol. Pathol.* (2001). doi:10.1136/mp.54.4.259
234. Palacios, J. *et al.* Phenotypic characterization of BRCA1 and BRCA2 tumors based in a tissue microarray study with 37 immunohistochemical markers. *Breast Cancer Res. Treat.* (2005). doi:10.1007/s10549-004-1536-0

235. Børresen, A. L. *et al.* Screening for germ line TP53 mutations in breast cancer patients. *Cancer Res.* (1992).
236. Scully, R. *et al.* Association of BRCA1 with Rad51 in mitotic and meiotic cells. *Cell* (1997). doi:10.1016/S0092-8674(00)81847-4
237. Taniguchi, T. *et al.* S-phase-specific interaction of the Fanconi anemia protein, FANCD2, with BRCA1 and RAD51. *Blood* (2002). doi:10.1182/blood-2002-01-0278
238. Li, F. P. *et al.* A Cancer Family Syndrome in Twenty-four Kindreds. *Cancer Res.* (1988).
239. Gao, N., Zhang, Z., Jiang, B.-H. & Shi, X. Role of PI3K/AKT/mTOR signaling in the cell cycle progression of human prostate cancer. *Biochem. Biophys. Res. Commun.* (2003). doi:10.1016/j.bbrc.2003.09.132
240. Ghosh, S. *et al.* PI3K/mTOR signaling regulates prostatic branching morphogenesis. *Dev. Biol.* (2011). doi:10.1016/j.ydbio.2011.09.027
241. Emerling, B. M. & Akcakanat, A. Targeting PI3K/mTOR signaling in cancer. in *Cancer Research* (2011). doi:10.1158/0008-5472.CAN-11-1699
242. Hanahan & Weinberg. Cell - Hallmarks of Cancer: The Next Generation. *Cell, Volume 144, Issue 5, 646-674, 4 March 2011* (2011).
243. Engelman, J. A. Targeting PI3K signalling in cancer: opportunities, challenges and limitations. *Nat. Rev. Cancer* (2009). doi:10.1038/nrc2664
244. Wong, K.-K., Engelman, J. A. & Cantley, L. C. Targeting the PI3K signaling pathway in cancer. *Curr. Opin. Genet. Dev.* **20**, 87–90 (2010).
245. DiMeo, T. A. *et al.* A Novel Lung Metastasis Signature Links Wnt Signaling with Cancer Cell Self-Renewal and Epithelial-Mesenchymal Transition in Basal-like Breast Cancer. *Cancer Res* **69**, (2009).
246. DC, K. *et al.* Comprehensive molecular portraits of human breast tumours TCGA Research Network Nature 2012 490 7418 61 70 10.1038/nature11412. *Nature* (2012). doi:10.1038/nature11412

247. Whitman, M., Downes, C. P., Keeler, M., Keller, T. & Cantley, L. Type I phosphatidylinositol kinase makes a novel inositol phospholipid, phosphatidylinositol-3-phosphate. *Nature* (1988). doi:10.1038/332644a0
248. Auger, K. R., Serunian, L. A., Soltoff, S. P., Libby, P. & Cantley, L. C. PDGF-dependent tyrosine phosphorylation stimulates production of novel polyphosphoinositides in intact cells. *Cell* (1989). doi:10.1016/0092-8674(89)90182-7
249. Toker, A. Phosphoinositide 3-kinases—a historical perspective. *Subcell. Biochem.* (2015). doi:10.1007/978-94-007-3012-0_4
250. Toker, a & Cantley, L. C. Signalling through the lipid products of phosphoinositide-3-OH kinase. *Nature* (1997). doi:10.1038/42648
251. Vanhaesebroeck, B. & Waterfield, M. D. Signaling by Distinct Classes of Phosphoinositide 3-Kinases. *Exp. Cell Res.* (1999). doi:10.1006/excr.1999.4701
252. Manning, B. D. *et al.* AKT/PKB signaling: navigating downstream. *Cell* (2007). doi:10.1016/j.cell.2007.06.009
253. Bellacosa, A., Kumar, C. C., Cristofano, A. D. & Testa, J. R. *Activation of AKT kinases in cancer: Implications for therapeutic targeting. Advances in Cancer Research* (2005). doi:10.1016/S0065-230X(05)94002-5
254. Cantley, L. C. & Neel, B. G. New insights into tumor suppression: PTEN suppresses tumor formation by restraining the phosphoinositide 3-kinase/AKT pathway. *Proc. Natl. Acad. Sci. U. S. A.* (1999). doi:10.1073/pnas.96.8.4240
255. Li, J. *et al.* PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science* (80-.). (1997). doi:10.1126/science.275.5308.1943
256. Steck, P. a *et al.* Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. *Nat. Genet.* (1997). doi:10.1038/ng0497-356
257. Diehl, J. A., Cheng, M., Roussel, M. F. & Sherr, C. J. Glycogen synthase kinase-3 β regulates cyclin D1 proteolysis and subcellular localization. *Genes Dev.* (1998). doi:10.1101/gad.12.22.3499

258. Sears, R. *et al.* Multiple Ras-dependent phosphorylation pathways regulate Myc protein stability. *Genes Dev.* (2000). doi:10.1101/gad.836800
259. Burgering, B. M. T. & Medema, R. H. Decisions on life and death: FOXO Forkhead transcription factors are in command when PKB/Akt is off duty. *J. Leukoc. Biol.* (2003). doi:10.1189/jlb.1202629
260. Roy, P. G. & Thompson, A. M. Cyclin D1 and breast cancer. *The Breast* (2006). doi:10.1016/j.breast.2006.02.005
261. Pelengaris, S., Khan, M. & Evan, G. c-MYC: more than just a matter of life and death. *Nat. Rev. Cancer* (2002). doi:10.1038/nrc904
262. Levine, A. J., Hu, W. & Feng, Z. The P53 pathway: what questions remain to be explored? *Cell Death Differ.* **13**, 1027–1036 (2006).
263. Marcel, V., Catez, F. & Diaz, J.-J. p53, a translational regulator: contribution to its tumour-suppressor activity. *Oncogene* **34**, 5513–5523 (2015).
264. Liu, J., Zhang, C., Hu, W. & Feng, Z. Tumor suppressor p53 and its mutants in cancer metabolism. *Cancer Lett.* **356**, 197–203 (2015).
265. Levine, A. J. & Oren, M. The first 30 years of p53: growing ever more complex. *Nat. Rev. Cancer* (2009). doi:10.1038/nrc2723
266. Olivier, M., Hussain, S. P., Caron de Fromentel, C., Hainaut, P. & Harris, C. C. TP53 mutation spectra and load: a tool for generating hypotheses on the etiology of cancer. *IARC Sci. Publ.* (2004).
267. Luo, J., Manning, B. D. & Cantley, L. C. Targeting the PI3K-Akt pathway in human cancer. *Cancer Cell* **4**, 257–262 (2003).
268. Lien, E. C., Dibble, C. C. & Toker, A. PI3K signaling in cancer: beyond AKT. *Curr. Opin. Cell Biol.* **45**, 62–71 (2017).
269. Wertheimer, E. *et al.* Rac signaling in breast cancer: A tale of GEFs and GAPs. *Cellular Signalling* (2012). doi:10.1016/j.cellsig.2011.08.011
270. Sanz-Moreno, V. *et al.* Rac Activation and Inactivation Control Plasticity of Tumor Cell Movement. *Cell* (2008). doi:10.1016/j.cell.2008.09.043

271. Taylor-Harding, B. *et al.* Cyclin E1 and RTK/RAS signaling drive CDK inhibitor resistance via activation of E2F and ETS. *Oncotarget* **6**, 696–714 (2015).
272. Cousin, S. *et al.* Clinical impact of extensive molecular profiling in advanced cancer patients. *J. Hematol. Oncol.* **10**, 45 (2017).
273. Curigliano, G. *et al.* Ribociclib plus letrozole in early breast cancer: A presurgical, window-of-opportunity study. *Breast* **28**, 191–8 (2016).
274. Dancau, A.-M. *et al.* PPFIA1 and CCND1 are frequently coamplified in breast cancer. *Genes. Chromosomes Cancer* **49**, 1–8 (2010).
275. Ross, J. S. *et al.* Genomic profiling of advanced-stage, metaplastic breast carcinoma by next-generation sequencing reveals frequent, targetable genomic abnormalities and potential new treatment options. *Arch. Pathol. Lab. Med.* **139**, 642–9 (2015).
276. Dimeo, T. A. *et al.* A Novel Lung Metastasis Signature Links Wnt Signaling with Cancer Cell Self-Renewal and Epithelial-Mesenchymal Transition in Basal-like Breast Cancer. *Cancer Res.* (2009). doi:Doi 10.1158/0008-5472.Can-08-4135
277. Fouad, T. M., Kogawa, T., Reuben, J. M. & Ueno, N. T. *The role of inflammation in inflammatory breast cancer. Advances in Experimental Medicine and Biology* (2014). doi:10.1007/978-3-0348-0837-8_3
278. Li, L. *et al.* Notch-1 signaling activates NF- κ B in human breast carcinoma MDA-MB-231 cells via PP2A-dependent AKT pathway. *Med. Oncol.* **33**, 1–11 (2016).
279. Katoh, M. Networking of WNT, FGF, Notch, BMP, and Hedgehog signaling pathways during carcinogenesis. *Stem Cell Rev* **3**, (2007).
280. Brennan, K. R. & Brown, A. M. C. Wnt Proteins in Mammary Development and Cancer. *J. Mammary Gland Biol. Neoplasia* **9**, 119–131 (2004).
281. Micalizzi, D. S., Farabaugh, S. M. & Ford, H. L. Epithelial-Mesenchymal Transition in Cancer: Parallels Between Normal Development and Tumor Progression. *J. Mammary Gland Biol. Neoplasia* **15**, 117–134 (2010).
282. Yook, J. I. *et al.* A Wnt-Axin2-GSK3 β cascade regulates Snail1 activity in breast cancer cells. *Nat Cell Biol* **8**, (2006).

283. Neth, P. *et al.* The Wnt signal transduction pathway in stem cells and cancer cells: influence on cellular invasion. *Stem Cell Rev* **3**, (2007).
284. Bachelder, R. E., Marchetti, A., Falcioni, R., Soddu, S. & Mercurio, A. M. Activation of p53 Function in Carcinoma Cells by the $\alpha 6\beta 4$ Integrin. *J. Biol. Chem.* **274**, 20733–20737 (1999).
285. Carico, E. *et al.* Integrin beta 4 expression in the neoplastic progression of cervical epithelium. *Gynecologic oncology* (1993). doi:10.1006/gyno.1993.1087
286. Seguin, L., Desgrosellier, J. S., Weis, S. M. & Cheresch, D. A. Integrins and cancer: regulators of cancer stemness, metastasis, and drug resistance. *Trends Cell Biol.* **25**, 234–240 (2015).
287. Desgrosellier, J. S. & Cheresch, D. A. Integrins in cancer: biological implications and therapeutic opportunities. *Nat Rev Cancer* **10**, 9–22 (2010).
288. Lu, S., Simin, K., Khan, A. & Mercurio, A. M. Analysis of integrin beta4 expression in human breast cancer: association with basal-like tumors and prognostic significance. *Clin Cancer Res* **14**, (2008).
289. Diaz, L. K. *et al.* Beta4 integrin subunit gene expression correlates with tumor size and nuclear grade in early breast cancer. *Mod Pathol* (2005). doi:10.1038/modpathol.3800411
290. Stewart, R. L. & O'Connor L., K. Clinical significance of the integrin $\alpha 6\beta 4$ in human malignancies. *Lab. Invest.* **95**, 976–986 (2015).
291. Guo, W. & Giancotti, F. G. Integrin signalling during tumour progression. *Nat. Rev. Mol. Cell Biol.* **5**, 816+ (2004).
292. Marinkovich, M. P. Tumour microenvironment: Laminin 332 in squamous-cell carcinoma. *Nat. Rev. Cancer* (2007). doi:10.1038/nrc2089
293. Chao, C., Lotz, M. M., Clarke, A. C. & Mercurio, A. M. A function for the integrin $\alpha 6\beta 4$ in the invasive properties of colorectal carcinoma cells. *Cancer Res.* (1996).
294. Kim, B. G. *et al.* Laminin-332-rich tumor microenvironment for tumor invasion in the interface zone of breast cancer. *Am. J. Pathol.* (2011). doi:10.1016/j.ajpath.2010.11.028

295. Savoia, P., Cremona, O., Trusolino, L., Pepino, E. & Marchisio, P. C. Integrins and basement membrane proteins in skin carcinomas. *Pathol.Res.Pract.* (1994).
296. Savoia, P., Trusolino, L., Pepino, E., Cremona, O. & Marchisio, P. C. Expression and topography of integrins and basement membrane proteins in epidermal carcinomas: basal but not squamous cell carcinomas display loss of alpha 6 beta 4 and BM-600/nicein. *J Invest Dermatol* (1993).
297. Downer, C. S., Watt, F. M. & Speight, P. M. Loss of alpha 6 and beta 4 integrin subunits coincides with loss of basement membrane components in oral squamous cell carcinomas. *J. Pathol.* (1993). doi:10.1002/path.1711710306
298. Skubitz, A. P., Bast, R. J., Wayner, E. A., Letourneau, P. C. & Wilke, M. S. Expression of alpha 6 and beta 4 integrins in serous ovarian carcinoma correlates with expression of the basement membrane protein laminin. *Am J Pathol* (1996).
299. Cress, A. E., Rabinovitz, I., Zhu, W. & Nagle, R. B. The alpha 6 beta 1 and alpha 6 beta 4 integrins in human prostate cancer progression. *Cancer Metastasis Rev* (1995).
300. Liebert, M., Washington, R., Wedemeyer, G., Carey, T. E. & Grossman, H. B. Loss of co-localization of alpha 6 beta 4 integrin and collagen VII in bladder cancer. *Am. J. Pathol.* (1994).
301. Hall, P. A., Coates, P., Lemoine, N. R. & Horton, M. A. Characterization of integrin chains in normal and neoplastic human pancreas. *J. Pathol.* (1991). doi:10.1002/path.1711650107
302. Falcioni, R. *et al.* Alpha 6 beta 4 and alpha 6 beta 1 integrins associate with ErbB-2 in human carcinoma cell lines. *Exp Cell Res* **236**, (1997).
303. Lipscomb, E. A. & Mercurio, A. M. Mobilization and activation of a signaling competent $\alpha 6 \beta 4$ integrin underlies its contribution to carcinoma progression. *Cancer metastasis Rev.* **24**, 413–423 (2005).
304. Ramovs, V., Te Molder, L. & Sonnenberg, A. The opposing roles of laminin-binding integrins in cancer. (2016). doi:10.1016/j.matbio.2016.08.007
305. Guo, W. *et al.* $\beta 4$ Integrin Amplifies ErbB2 Signaling to Promote Mammary Tumorigenesis. *Cell* **126**, 489–502 (2006).

306. Rabinovitz, I., Toker, A. & Mercurio, A. M. Protein kinase C-dependent mobilization of the $\alpha 6 \beta 4$ integrin from hemidesmosomes and its association with actin-rich cell protrusions drive the chemotactic migration of carcinoma cells. *J Cell Biol* **146**, (1999).
307. Koukoulis, G. K. *et al.* Immunohistochemical localization of integrins in the normal, hyperplastic, and neoplastic breast. Correlations with their functions as receptors and cell adhesion molecules. *Am. J. Pathol.* **139**, 787–799 (1991).
308. Sehgal, B. U. *et al.* Integrin $\beta 4$ regulates migratory behavior of keratinocytes by determining laminin-332 organization. *J. Biol. Chem.* (2006). doi:10.1074/jbc.M606317200
309. Rabinovitz, I., Gipson, I. K. & Mercurio, a M. Traction forces mediated by $\alpha 6 \beta 4$ integrin: implications for basement membrane organization and tumor invasion. *Mol. Biol. Cell* (2001).
310. Carpenter, B. L. *et al.* Integrin $\alpha 6 \beta 4$ promotes autocrine Epidermal Growth Factor Receptor (EGFR) signaling to stimulate migration and invasion toward hepatocyte growth factor (HGF). *J. Biol. Chem.* (2015). doi:10.1074/jbc.M115.686873
311. Daemi, N. *et al.* Anti- $\beta 4$ integrin antibodies enhance migratory and invasive abilities of human colon adenocarcinoma cells and their MMP-2 expression. *Int. J. Cancer* (2000). doi:10.1002/(SICI)1097-0215(20000315)85:6<850::AID-IJC19>3.0.CO;2-B
312. Shaw, L. M., Rabinovitz, I., Wang, H. H. F., Toker, A. & Mercurio, A. M. Activation of phosphoinositide 3-OH kinase by the $\alpha 6 \beta 4$ integrin promotes carcinoma invasion. *Cell* **91**, 949+ (1997).
313. Guo, W. *et al.* Beta 4 integrin amplifies ErbB2 signaling to promote mammary tumorigenesis. *Cell* **126**, (2006).
314. Nagata, M. *et al.* ITGA3 and ITGB4 expression biomarkers estimate the risks of locoregional and hematogenous dissemination of oral squamous cell carcinoma. *BMC Cancer* (2013). doi:10.1186/1471-2407-13-410
315. Masugi, Y. *et al.* Upregulation of integrin $\beta 4$ promotes epithelial–mesenchymal transition and is a novel prognostic marker in pancreatic ductal adenocarcinoma. *Lab. Investig.* (2015). doi:10.1038/labinvest.2014.166

316. Cadigan, K. M. & Nusse, R. Wnt signaling: A common theme in animal development. *Genes and Development* (1997). doi:10.1101/gad.11.24.3286
317. Howe, L. R. & Brown, A. M. C. Wnt Signaling and Breast Cancer. *Cancer Biol. Ther.* **3**, 36–41 (2004).
318. Bergstein, I. & Brown, A. M. C. WNT genes and breast cancer. (1999).
319. Huguet, E. L., McMahon, J. A., McMahon, A. P., Bicknell, R. & Harris, A. L. Differential expression of human Wntgenes 2, 3, 4, and 7B in human breast cell lines and normal and diseased breast tissue. *Cancer Res* **54**, (1994).
320. Dale, T. C. *et al.* Compartment switching of WNT-2 expression in human breast tumors. *Cell Regul.* **56**, (1996).
321. Lejeune, S., Huguet, E. L., Hamby, A., Poulson, R. & Harris, A. L. Wnt5a cloning, expression, and up-regulation in human primary breast cancers. *Clin Cancer Res* **1**, (1995).
322. Bui, T. D. *et al.* A novel human Wnt gene, WNT10B, maps to 12q13 and is expressed in human breast carcinomas. *Oncogene* **14**, (1997).
323. Congdon, K. L. *et al.* Activation of Wnt Signaling in Hematopoietic Regeneration. *Stem Cells* **26**, (2008).
324. Reguart, N. *et al.* The role of Wnt signaling in cancer and stem cells. *Futur. Oncol.* (2005). doi:10.2217/14796694.1.6.787
325. Zeng, Y. A. *et al.* Wnt proteins are self-renewal factors for mammary stem cells and promote their long-term expansion in culture. *Cell Stem Cell* **6**, 568–77 (2010).
326. Liu, Bob Y., *et al.* The transforming activity of Wnt effectors correlates with their ability to induce the accumulation of mammary progenitor cells. *Proc. Natl. Acad. Sci. U. S. A.* (2004). doi:10.1073/pnas.0400699101
327. Polakis, P. Wnt signaling and cancer. *Genes Dev* **14**, (2000).
328. Mao, J. *et al.* Low-density lipoprotein receptor-related protein-5 binds to Axin and regulates the canonical Wnt signaling pathway. *Mol Cell* **7**, (2001).

329. Mao, B. *et al.* LDL-receptor-related protein 6 is a receptor for Dickkopf proteins. *Nature* **411**, (2001).
330. Lin, S. Y. *et al.* Beta-catenin, a novel prognostic marker for breast cancer: its roles in cyclin D1 expression and cancer progression. *Proc Natl Acad Sci USA* **97**, (2000).
331. Ryo, a, Nakamura, M., Wulf, G., Liou, Y. C. & Lu, K. P. Pin1 regulates turnover and subcellular localization of beta-catenin by inhibiting its interaction with APC. *Nat. Cell Biol.* (2001). doi:10.1038/ncb0901-793
332. Michaelson, J. S. & Leder, P. beta-catenin is a downstream effector of Wnt-mediated tumorigenesis in the mammary gland. *Oncogene* (2001). doi:10.1038/sj.onc.1204586
333. Imbert, A., Eelkema, R., Jordan, S., Feiner, H. & Cowin, P. Δ N89 β -catenin induces precocious development, differentiation, and neoplasia in mammary gland. *J Cell Biol* **153**, (2001).
334. Brown, A. M. C. Wnt signaling in breast cancer: have we come full circle? *Breast Cancer Res.* **3**, 351 (2001).
335. Wissmann, C. *et al.* WIF1, a component of the Wnt pathway, is down-regulated in prostate, breast, lung, and bladder cancer. *J. Pathol.* (2003). doi:10.1002/path.1449
336. Ugolini, F. *et al.* WNT pathway and mammary carcinogenesis: loss of expression of candidate tumor suppressor gene SFRP1 in most invasive carcinomas except of the medullary type. *Oncogene* (2001). doi:10.1038/sj.onc.1204706
337. Petrocelli, T. & Slingerland, J. M. PTEN deficiency: a role in mammary carcinogenesis. *Breast Cancer Res.* **3**, 356 (2001).
338. Persad, S., A.Troussard, A., McPhee, T. R., Mulholland, D. J. & Dedhar, S. Tumor Suppressor Pten Inhibits Nuclear Accumulation of β -Catenin and T Cell/Lymphoid Enhancer Factor 1-Mediated Transcriptional Activation. *J. Cell Biol.* **153**, 1161–1174 (2001).
339. Hashimoto, M. *et al.* Fibroblast growth factor 1 regulates signaling via the glycogen synthase kinase-3 β pathway. Implications for neuroprotection. *J. Biol. Chem.* **277**, 32985–91 (2002).

340. Børresen-Dale, A.-L. *TP53* and breast cancer. *Hum. Mutat.* **21**, 292–300 (2003).
341. Lamberti, C. *et al.* Regulation of beta-catenin function by the IKK β kinases. *J Biol Chem* **276**, (2001).
342. Li, Y. *et al.* Evidence that transgenes encoding components of the Wnt signaling pathway preferentially induce mammary cancers from progenitor cells. *Proc. Natl. Acad. Sci. U. S. A.* (2003). doi:10.1073/pnas.2136825100
343. Jang, G.-B. *et al.* Blockade of Wnt/ β -catenin signaling suppresses breast cancer metastasis by inhibiting CSC-like phenotype. *Sci. Rep.* **5**, 12465 (2015).
344. King, T. D., Suto, M. J. & Li, Y. The wnt/ β -catenin signaling pathway: A potential therapeutic target in the treatment of triple negative breast cancer. *J. Cell. Biochem.* **113**, 13–18 (2012).
345. Chrenek, M. a, Wong, P. & Weaver, V. M. Tumour-stromal interactions. Integrins and cell adhesions as modulators of mammary cell survival and transformation. *Breast Cancer Res.* (2001). doi:10.1186/bcr300
346. Elgert, K. D., Alleva, D. G. & Mullins, D. W. Tumor-induced immune dysfunction: the macrophage connection. *J. Leukoc. Biol.* (1998).
347. Lin, E. Y., Gouon-Evans, V., Nguyen, A. V. & Pollard, J. W. The Macrophage Growth Factor CSF-1 in Mammary Gland Development and Tumor Progression. *J. Mammary Gland Biol. Neoplasia* **7**, 147–162 (2002).
348. Scholl, S. M. *et al.* Circulating levels of the macrophage colony stimulating factor CSF-1 in primary and metastatic breast cancer patients. A pilot study. *Breast Cancer Res. Treat.* (1996). doi:10.1007/BF01806155
349. Goswami, S. *et al.* Macrophages Promote the Invasion of Breast Carcinoma Cells via a Colony-Stimulating Factor-1 / Epidermal Growth Factor Paracrine Loop. *Cancer Res.* (2005). doi:10.1158/0008-5472.CAN-04-1853
350. Jakowlew, S. B. Transforming growth factor-beta in cancer and metastasis. *Cancer Metastasis Rev.* (2006). doi:10.1007/s10555-006-9006-2
351. Moustakas, A., Pardali, K., Gaal, A. & Heldin, C. H. Mechanisms of TGF-beta signaling in regulation of cell growth and differentiation. *Immunol. Lett.* (2002).

352. Pardali, K. & Moustakas, A. Actions of TGF-beta as tumor suppressor and pro-metastatic factor in human cancer. *Biochim. Biophys. Acta - Rev. Cancer* (2007). doi:10.1016/j.bbcan.2006.06.004
353. Soria, G. & Ben-Baruch, A. The inflammatory chemokines CCL2 and CCL5 in breast cancer. *Cancer Letters* (2008). doi:10.1016/j.canlet.2008.03.018
354. Arendt, L. M. *et al.* Obesity promotes breast cancer by CCL2-mediated macrophage recruitment and angiogenesis. *Cancer Res.* (2013). doi:10.1158/0008-5472.CAN-13-0926
355. Fujimoto, H. *et al.* Stromal MCP-1 in mammary tumors induces tumor-associated macrophage infiltration and contributes to tumor progression. *Int. J. Cancer* (2009). doi:10.1002/ijc.24378
356. Qian, B.-Z. *et al.* CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. *Nature* (2011). doi:10.1038/nature10138
357. Ueno, T. *et al.* Significance of macrophage chemoattractant protein-1 in macrophage recruitment, angiogenesis, and survival in human breast cancer. *Clin. Cancer Res.* (2000).
358. Tiede, B. & Kang, Y. From milk to malignancy: the role of mammary stem cells in development, pregnancy and breast cancer. *Cell Res.* **21**, 245–257 (2011).
359. Kühl, S. J. & Kühl, M. On the role of Wnt/ β -catenin signaling in stem cells. *Biochimica et Biophysica Acta - General Subjects* (2013). doi:10.1016/j.bbagen.2012.08.010
360. Polyak, K. & Hahn, W. C. Roots and stems: stem cells in cancer. *Nat. Med.* **12**, 296–300 (2006).
361. Smalley, M. & Ashworth, A. Stem cells and breast cancer: A field in transit. *Nat. Rev. Cancer* **3**, 832+ (2003).
362. Yu, Q., Verheyen, E. & Zeng, Y. Mammary Development and Breast Cancer: A Wnt Perspective. *Cancers (Basel)*. **8**, 65 (2016).

363. Sca-1 Delineates an Estrogen Responsive Stem Cell within the Mammary Gland : Nuclear Receptors and Co-Regulators in Health and Disease (posters). Available at: <http://press.endocrine.org/doi/abs/10.1210/endo-meetings.2016.NRSHA.3.FRI-206>. (Accessed: 30th May 2017)
364. Soady, K. J. *et al.* Mouse mammary stem cells express prognostic markers for triple-negative breast cancer. *Breast Cancer Res.* **17**, 31 (2015).
365. Liu, S., Dontu, G. & Wicha, M. S. Mammary stem cells, self-renewal pathways, and carcinogenesis. *Breast Cancer Res.* (2005). doi:10.1186/bcr1021
366. An, S. M., Ding, Q. P. & Li, L. Stem cell signaling as a target for novel drug discovery: recent progress in the WNT and Hedgehog pathways. *Acta Pharmacol. Sin.* **34**, 777–783 (2013).
367. Reedijk, M. *et al.* JAG1 expression is associated with a basal phenotype and recurrence in lymph node-negative breast cancer. *Breast Cancer Res. Treat.* (2008). doi:10.1007/s10549-007-9805-3
368. Khramtsov, A. I. *et al.* Wnt/beta-catenin pathway activation is enriched in basal-like breast cancers and predicts poor outcome. *Am J Pathol* (2010). doi:10.2353/ajpath.2010.091125
369. Matos, I., Dufloth, R., Alvarenga, M., Zeferino, L. C. & Schmitt, F. p63, cytokeratin 5, and P-cadherin: Three molecular markers to distinguish basal phenotype in breast carcinomas. *Virchows Arch.* (2005). doi:10.1007/s00428-005-0010-7
370. Koker, M. M. & Kleer, C. G. p63 expression in breast cancer: a highly sensitive and specific marker of metaplastic carcinoma. *Am J Surg Pathol* **28**, (2004).
371. Ripka, S. *et al.* WNT5A--target of CUTL1 and potent modulator of tumor cell migration and invasion in pancreatic cancer. *Carcinogenesis* **28**, 1178–87 (2007).
372. Dawood, S., Broglio, K., Buzdar, A. U., Hortobagyi, G. N. & Giordano, S. H. Prognosis of women with metastatic breast cancer by HER2 status and trastuzumab treatment: An institutional-based review. *J. Clin. Oncol.* (2010). doi:10.1200/JCO.2008.19.9844
373. Draheim, K. M. *et al.* ARRDC3 suppresses breast cancer progression by negatively regulating integrin $\beta 4$. *Oncogene* (2010). doi:10.1038/onc.2010.250

374. Howlett AR, Bailey N, Damsky C, Petersen OW, Bissell MJ. Cellular growth and survival are mediated by beta 1 integrins in normal human breast epithelium but not in breast carcinoma. *J. Cell Sci.* (1995).
375. Lipscomb, E. A. & Mercurio, A. M. Mobilization and activation of a signaling competent $\alpha 6\beta 4$ integrin underlies its contribution to carcinoma progression. *Cancer Metastasis Rev.* (2005). doi:10.1007/s10555-005-5133-4
376. Mukhopadhyay, R., Theriault, R. L. & Price, J. E. Increased levels of $\alpha 6$ integrins are associated with the metastatic phenotype of human breast cancer cells. *Clin. Exp. Metastasis* (1999). doi:10.1023/A:1006659230585
377. Rabinovitz, I. & Mercurio, A. M. The integrin $\alpha 6\beta 4$ functions in carcinoma cell migration on laminin-1 by mediating the formation and stabilization of actin-containing motility structures. *J. Cell Biol.* (1997).
378. Holliday, D. L. & Speirs, V. Choosing the right cell line for breast cancer research. *Breast Cancer Res.* (2011).
379. Kalluri, R. & Zeisberg, M. Fibroblasts in cancer. *Nat. Rev. Cancer* (2006). doi:10.1038/nrc1877
380. Balko, J. M. *et al.* Molecular profiling of the residual disease of triple-negative breast cancers after neoadjuvant chemotherapy identifies actionable therapeutic targets. *Cancer Discov.* (2014). doi:10.1158/2159-8290.CD-13-0286
381. Reis-Filho, J. S. & Tutt, A. N. J. Triple negative tumours: A critical review. *Histopathology* (2008). doi:10.1111/j.1365-2559.2007.02889.x
382. J., M., J., B., Mendelsohn, J. & Baselga, J. Epidermal growth factor receptor targeting in cancer. *Semin. Oncol.* (2006). doi:10.1053/j.seminoncol.2006.04.003
383. Ciardiello, F. & Tortora, G. A novel approach in the treatment of cancer: targeting the epidermal growth factor receptor. *Clin Cancer Res* (2001).
384. Burness, M. L., Grushko, T. A. & Olopade, O. I. Epidermal growth factor receptor in triple-negative and basal-like breast cancer: Promising clinical target or only a marker? *Cancer J.* (2010).
385. a., H. *et al.* Vascular endothelial growth factor and angiogenesis. *Pharmacol. Rev.* (2004). doi:10.1124/pr.56.4.3.549

386. Linderholm, B., Tavelin, B., Grankvist, K. & Henriksson, R. Vascular endothelial growth factor is of high prognostic value in node-negative breast carcinoma. *J Clin Oncol* (1998).
387. Rayson, D., Vantyghem, S. A. & Chambers, A. F. Angiogenesis as a target for breast cancer therapy. *J Mammary Gland Biol Neoplasia* (1999).
388. Kariya, Y., Kariya, Y. & Gu, J. Roles of laminin-332 and alpha6beta4 integrin in tumor progression. *Mini Rev. Med. Chem.* (2009).
389. Kwon, S.-Y., Chae, S. W., Wilczynski, S. P., Arain, A. & Carpenter, P. M. Laminin 332 Expression in Breast Carcinoma. *Appl. Immunohistochem. Mol. Morphol.* (2012). doi:10.1097/PAI.0b013e3182329e8f
390. Kang, S. G. *et al.* Effect of laminin 332 on motility and invasion in bladder cancer. *Kaohsiung J. Med. Sci.* (2013). doi:10.1016/j.kjms.2012.12.003
391. Lu, S., Simin, K., Khan, A. & Mercurio, A. M. Analysis of Integrin $\alpha 4$ Expression in Human Breast Cancer: Association with Basal-like Tumors and Prognostic Significance. *Clin Cancer Res* **14**, 1050 (2008).
392. Miner, J. H. & Yurchenco, P. D. LAMININ FUNCTIONS IN TISSUE MORPHOGENESIS. *Annu. Rev. Cell Dev. Biol.* **20**, 255–284 (2004).
393. Nguyen, N. M. & Senior, R. M. Laminin isoforms and lung development: All isoforms are not equal. *Dev. Biol.* **294**, 271–279 (2006).
394. Giannelli, G. Induction of Cell Migration by Matrix Metalloprotease-2 Cleavage of Laminin-5. *Science* (80-.). (1997). doi:10.1126/science.277.5323.225
395. Tsuruta, D. *et al.* Laminin-332-integrin interaction: a target for cancer therapy? *Curr. Med. Chem.* (2008). doi:10.2174/092986708785132834
396. Baker, S. E. *et al.* Laminin-5 and hemidesmosomes: role of the alpha 3 chain subunit in hemidesmosome stability and assembly. *J. Cell Sci.* **109** (Pt 10), 2509–20 (1996).
397. Okamoto, O. *et al.* Normal Human Keratinocytes Bind to the $\beta 3$ LG4/5 Domain of Unprocessed Laminin-5 through the Receptor Syndecan-1. *J. Biol. Chem.* **278**, 44168–44177 (2003).

398. Fontao, L., Dirrig, S., Owaribe, K., Kedinger, M. & Launay, J. F. Polarized Expression of HD1: Relationship with the Cytoskeleton in Cultured Human Colonic Carcinoma Cells. *Exp. Cell Res.* **231**, 319–327 (1997).
399. Uematsu, J., Nishizawa, Y., Sonnenberg, A. & Owaribe, K. Demonstration of type II hemidesmosomes in a mammary gland epithelial cell line, BMGE-H. *J. Biochem.* **115**, 469–76 (1994).
400. Kim, J. M., Park, W. H. & Min, B. M. The PPFLMLLLKGSTR motif in globular domain 3 of the human laminin-5 γ 3 chain is crucial for integrin α 3 β 1 binding and cell adhesion. *Exp. Cell Res.* (2005). doi:10.1016/j.yexcr.2004.11.009
401. Hirosaki, T., Mizushima, H., Tsubota, Y., Moriyama, K. & Miyazaki, K. Structural requirement of carboxyl-terminal globular domains of laminin γ 3 chain for promotion of rapid cell adhesion and migration by laminin-5. *J. Biol. Chem.* (2000). doi:10.1074/jbc.M001326200
402. Engbring, J. A. & Kleinman, H. K. The basement membrane matrix in malignancy. *J. Pathol.* **200**, 465–470 (2003).
403. Sathyanarayana, U. G. *et al.* Epigenetic inactivation of laminin-5-encoding genes in lung cancers. *Clin. Cancer Res.* **9**, 2665–72 (2003).
404. Sathyanarayana, U. G. *et al.* Aberrant promoter methylation and silencing of laminin-5-encoding genes in breast carcinoma. *Clin. Cancer Res.* **9**, 6389–94 (2003).
405. Hlubek, F., Jung, A., Kotzor, N., Kirchner, T. & Brabletz, T. Expression of the invasion factor laminin gamma2 in colorectal carcinomas is regulated by beta-catenin. *Cancer Res.* **61**, 8089–93 (2001).
406. Hlubek, F., Spaderna, S., Jung, A., Kirchner, T. & Brabletz, T. β -Catenin activates a coordinated expression of the proinvasive factors laminin-5 γ 2 chain and MT1-MMP in colorectal carcinomas. *Int. J. Cancer* **108**, 321–326 (2004).
407. Lyons, A. J. & Jones, J. Cell adhesion molecules, the extracellular matrix and oral squamous carcinoma. *Int. J. Oral Maxillofac. Surg.* **36**, 671–679 (2007).
408. Nakayama, M., Sato, Y., Okamoto, M. & Hirohashi, S. Increased Expression of Laminin-5 and Its Prognostic Significance in Hypopharyngeal Cancer. *Laryngoscope* **114**, 1259–1263 (2004).

409. Moriya, Y. *et al.* Increased expression of laminin-5 and its prognostic significance in lung adenocarcinomas of small size. An immunohistochemical analysis of 102 cases. *Cancer* **91**, 1129–41 (2001).
410. Lohi, J. *et al.* Basement membrane laminin-5 is deposited in colorectal adenomas and carcinomas and serves as a ligand for alpha3beta1 integrin. *APMIS* **108**, 161–72 (2000).
411. Fukushima, Y., Ohnishi, T., Arita, N., Hayakawa, T. & Sekiguchi, K. Integrin alpha3beta1-mediated interaction with laminin-5 stimulates adhesion, migration and invasion of malignant glioma cells. *Int. J. cancer* **76**, 63–72 (1998).
412. Tani, T. *et al.* Pancreatic carcinomas deposit laminin-5, preferably adhere to laminin-5, and migrate on the newly deposited basement membrane. *Am. J. Pathol.* **151**, 1289–302 (1997).
413. Rabinovitz, I. & Mercurio, A. M. The integrin alpha 6 beta 4 and the biology of carcinoma. *Biochem. Cell Biol.* **74**, 811–21 (1996).
414. Giannelli, G. *et al.* Human hepatocellular carcinoma (HCC) cells require both alpha3beta1 integrin and matrix metalloproteinases activity for migration and invasion. *Lab. Invest.* **81**, 613–27 (2001).
415. Wang, H. *et al.* Tumor cell $\alpha_3\beta_1$ integrin and vascular laminin-5 mediate pulmonary arrest and metastasis. *J. Cell Biol.* **164**, 935–941 (2004).
416. Homan, S. M., Mercurio, A. M. & LaFlamme, S. E. Endothelial cells assemble two distinct alpha6beta4-containing vimentin-associated structures: roles for ligand binding and the beta4 cytoplasmic tail. *J. Cell Sci.* **111** (Pt 18), 2717–28 (1998).
417. Giancotti, F. G. Targeting integrin beta4 for cancer and anti-angiogenic therapy. *Trends Pharmacol. Sci.* (2007). doi:10.1016/j.tips.2007.08.004
418. Akita, N. *et al.* Identification of oligopeptides binding to peritoneal tumors of gastric cancer. *Cancer Sci.* **97**, 1075–1081 (2006).
419. Seftor, R. E. B., Seftor, E. A., Kirschmann, D. A. & Hendrix, M. J. C. Targeting the tumor microenvironment with chemically modified tetracyclines: inhibition of laminin 5 gamma2 chain promigratory fragments and vasculogenic mimicry. *Mol. Cancer Ther.* **1**, 1173–9 (2002).

420. Sroka, T. C., Pennington, M. E. & Cress, A. E. Synthetic D-amino acid peptide inhibits tumor cell motility on laminin-5. *Carcinogenesis* **27**, 1748–1757 (2006).
421. Turner, B. P. *Recombinant G3 domain protein of the rat laminin-5 alpha3 chain binds to integrins on tumorogenic breast cancer cells to induce apoptosis*. (2008).
422. Schalkoff, K. CHANGES IN EXPRESSION OF AKT PATHWAY PROTEINS FOLLOWING TREATMENT WITH rG3 IN VITRO. *All Theses* (2011).
423. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* (2014). doi:10.1093/bioinformatics/btu170
424. Wu, T. D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* (2010). doi:10.1093/bioinformatics/btq057
425. Li, H. *et al.* The Sequence Alignment / Map format and SAMtools. *Bioinformatics* (2009). doi:10.1093/bioinformatics/btp352
426. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–30 (2014).
427. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108 (2013).
428. Sturn, A. *et al.* Genesis: cluster analysis of microarray data. *Bioinformatics* (2002). doi:10.1093/bioinformatics/18.1.207
429. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, I. T. Cytoscape: An Open Source Platform for Complex Network Analysis and Visualization. *Genome Research* (2003). doi:10.1101/gr.1239303
430. Croft, D., Mundo, A., Haw, R. & Milacic, M. The Reactome pathway knowledgebase. *Nucleic acids* (2014). doi:10.1093/nar/gkt1102
431. Fabregat, A. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res.* (2016). doi:10.1093/nar/gkv1351

432. Borick, H. Analysis of cell adhesion and proliferation of human epithelial tumorigenic cells treated with recombinant laminin-5 α -3 chain G3 domain protein. (Clemson University, 2004).
433. Keller, A., Nesvizhskii, A. I., Kolker, E. & Aebersold, R. Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. doi:10.1021/ac025747h
434. Nesvizhskii, A. I., Keller, A., Kolker, E. & Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646–58 (2003).
435. van Iersel, M. P. *et al.* Presenting and exploring biological pathways with PathVisio. *BMC Bioinformatics* **9**, 399 (2008).
436. Kutmon, M. *et al.* PathVisio 3: An Extendable Pathway Analysis Toolbox. *PLOS Comput. Biol.* **11**, e1004085 (2015).
437. Kelder, T. *et al.* WikiPathways: building research communities on biological pathways. *Nucleic Acids Res.* **40**, D1301–D1307 (2012).
438. Kutmon, M. *et al.* WikiPathways: capturing the full diversity of pathway knowledge. *Nucleic Acids Res.* **44**, D488–D494 (2016).
439. Zdobnov, E. M. *et al.* OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* **45**, D744–D749 (2017).
440. Oxender, W. D., Askew, E. W., Benson, J. D. & Emery, R. S. Biopsy of Liver, Adipose Tissue and Mammary Gland of Lactating Cows. *J. Dairy Sci.* **54**, 286–288 (1971).
441. VanKlompenberg, M. K., McMicking, H. F. & Hovey, R. C. Technical note: A vacuum-assisted approach for biopsying the mammary glands of various species. *J. Dairy Sci.* **95**, 243–246 (2012).
442. Pullan, S. E. & Streuli, C. H. The Mammary gland epithelial cell. in *Epithelial Cell Culture* (1996).

443. Epithelial Tissue and Mammary Gland | histology. Available at: <http://histology.medicine.umich.edu/resources/epithelial-tissue-mammary-gland>. (Accessed: 4th July 2017)
444. Fleige, S. & Pfaffl, M. W. RNA integrity and the effect on the real-time qRT-PCR performance. *Mol. Aspects Med.* **27**, 126–139 (2006).
445. Storey, J. D. *et al.* Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci.* **100**, 9440–9445 (2003).
446. Gallego Romero, I., Pai, A. A., Tung, J. & Gilad, Y. *Impact of RNA degradation on measurements of gene expression*. *bioRxiv* (2014). doi:10.1101/002261
447. Vermeulen, J. *et al.* Measurable impact of RNA quality on gene expression results from quantitative PCR. *Nucleic Acids Res.* (2011). doi:10.1093/nar/gkr065
448. Schroeder, A. *et al.* The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.* (2006). doi:10.1186/1471-2199-7-3
449. Holland, S. M. PRINCIPAL COMPONENTS ANALYSIS (PCA). (2016).
450. Yeung, K. Y. & Ruzzo, W. L. Principal component analysis for clustering gene expression data. *BIOINFORMATICS* **17**, 763–774 (2001).
451. Ringnér, M. What is principal component analysis? *Nat. Biotechnol.* **26**, 303–304 (2008).
452. Borcard, D. Canonical ordination: redundancy analysis (RDA) and Canonical correspondence analysis (CCA). *Multivar. Anal.* (2006).
453. Makarenkov, V. & Legendre, P. NONLINEAR REDUNDANCY ANALYSIS AND CANONICAL CORRESPONDENCE ANALYSIS BASED ON POLYNOMIAL REGRESSION. *Ecol. Lett.* (2002).
454. Zuur, A. F., Ieno, E. N. & Smith, G. M. 12 Principal component analysis and redundancy analysis. *Anal. Ecol. Data Stat. Biol. Heal.* (2007). doi:10.1007/978-0-387-45972-1_12
455. Wollenberg, A. Van Den. Redundancy analysis an alternative for canonical correlation analysis. *Psychometrika* (1977).

456. Shin, J. *et al.* Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades underlying Adult Neurogenesis. *Cell Stem Cell* **17**, 360–372 (2015).
457. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–96 (2016).
458. Liu, Y. *et al.* RNA-Seq identifies novel myocardial gene expression signatures of heart failure. *Genomics* **105**, 83–89 (2015).
459. Saliba, A.-E. E. A.-E. *et al.* Single-cell RNA-seq ties macrophage polarization to growth rate of intracellular Salmonella. *Nat. Microbiol.* (2016). doi:10.1038/nmicrobiol.2016.206
460. Nelson, A. C., Mould, A. W., Bikoff, E. K. & Robertson, E. J. Single-cell RNA-seq reveals cell type-specific transcriptional signatures at the maternal–foetal interface during pregnancy. *Nat. Commun.* (2016). doi:10.1038/ncomms11414
461. FISHER, R. A. THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS. *Ann. Eugen.* (1936). doi:10.1111/j.1469-1809.1936.tb02137.x
462. Pérez-Enciso, M. & Tenenhaus, M. Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach. *Hum. Genet.* (2003). doi:10.1007/s00439-003-0921-9
463. Redis, R. *et al.* Allele-Specific Reprogramming of Cancer Metabolism by the Long Non-coding RNA CCAT2. *Mol. Cell* **61**, 520–534 (2016).
464. Gorden, D. L. *et al.* Biomarkers of NAFLD progression: a lipidomics approach to an epidemic. *J. Lipid Res.* **56**, 722–36 (2015).
465. Murakami, Y. *et al.* Comparison of hepatocellular carcinoma miRNA expression profiling as evaluated by next generation sequencing and microarray. *PLoS One* (2014). doi:10.1371/journal.pone.0106314
466. Zhang, N. *et al.* The RNA-seq approach to discriminate gene expression profiles in response to melatonin on cucumber lateral root formation. *J. Pineal Res.* **56**, 39–50 (2014).
467. Lindholm, M. E. *et al.* The human skeletal muscle transcriptome: sex differences, alternative splicing, and tissue homogeneity assessed with RNA sequencing. *FASEB J.* **28**, 4571–4581 (2014).

468. Wang, S. *et al.* The potent tumor suppressor miR-497 inhibits cancer phenotypes in nasopharyngeal carcinoma by targeting ANLN and HSPA4L. *Oncotarget* **6**, 35893–907 (2015).
469. Held, T. *et al.* Hspa4l-deficient mice display increased incidence of male infertility and hydronephrosis development. *Mol. Cell. Biol.* **26**, 8099–108 (2006).
470. Held, T. *et al.* Heat-shock protein HSPA4 is required for progression of spermatogenesis. *Reproduction* **142**, 133–44 (2011).
471. Takahashi, H. *et al.* Identification of an overexpressed gene, HSPA4L, the product of which can provoke prevalent humoral immune responses in leukemia patients. *Exp. Hematol.* **35**, 1091–9 (2007).
472. GeneCards - Human Genes | Gene Database | Gene Search. Available at: <https://genecards.weizmann.ac.il/v3/>. (Accessed: 10th July 2017)
473. Yamakage, M. & Namiki, A. Calcium channels — basic aspects of their structure, function and gene encoding; anesthetic action on the channels — a review. *Can. J. Anesth. Can. d'anesthésie* **49**, 151–164 (2002).
474. Schramedei, K. *et al.* MicroRNA-21 targets tumor suppressor genes ANP32A and SMARCA4. *Oncogene* **30**, 2975–2985 (2011).
475. Williams, T. K. *et al.* pp32 (ANP32A) expression inhibits pancreatic cancer cell growth and induces gemcitabine resistance by disrupting HuR binding to mRNAs. *PLoS One* **5**, e15455 (2010).
476. Velmurugan, B. K. *et al.* Acidic leucine-rich nuclear phosphoprotein-32A (ANP32A) association with lymph node metastasis predicts poor survival in oral squamous cell carcinoma patients. *Oncotarget* **7**, 10879–90 (2016).
477. Reilly, P. T., Yu, Y., Hamiche, A. & Wang, L. Cracking the ANP32 whips: Important functions, unequal requirement, and hints at disease implications. *BioEssays* (2014). doi:10.1002/bies.201400058
478. Matilla, A. & Radrizzani, M. The Anp32 family of proteins containing leucine-rich repeats. *The Cerebellum* **4**, 7–18 (2005).
479. Kobe, B. The leucine-rich repeat as a protein recognition motif. *Curr. Opin. Struct. Biol.* **11**, 725–732 (2001).

480. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Y. I. Sequence and organization of the human mitochondrial genome. *Nature* (1981). doi:10.1038/290457A0
481. Yu, X. *et al.* The mtDNA nt7778 G/T polymorphism affects autoimmune diseases and reproductive performance in the mouse. *Hum. Mol. Genet.* (2009). doi:10.1093/hmg/ddp432
482. Lee, J. H. *et al.* Initiation factor eIF5B catalyzes second GTP-dependent step in eukaryotic translation initiation. *Proc. Natl. Acad. Sci. U. S. A.* (2002). doi:10.1073/pnas.262569399
483. Lee, S. *et al.* Upregulation of eIF5B controls cell-cycle arrest and specific developmental stages. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E4315-22 (2014).
484. Pestova, T. V *et al.* The joining of ribosomal subunits in eukaryotes requires eIF5B. *Nature* (2000). doi:10.1038/35002118
485. Lee, S. *et al.* Upregulation of eIF5B controls cell-cycle arrest and specific developmental stages. *Proc. Natl. Acad. Sci. U. S. A.* (2014). doi:10.1073/pnas.1320477111
486. Lee, J. H., Choi, S. K., Roll-Mecak, A., Burley, S. K. & Dever, T. E. Universal conservation in translation initiation revealed by human and archaeal homologs of bacterial translation initiation factor IF2. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 4342–7 (1999).
487. Shuttleworth, C. A. Type VIII collagen. *Int. J. Biochem. Cell Biol.* **29**, 1145–1148 (1997).
488. Hou, G., Mulholland, D., Gronska, M. A. & Bendeck, M. P. Type VIII collagen stimulates smooth muscle cell migration and matrix metalloproteinase synthesis after arterial injury. *Am. J. Pathol.* (2000). doi:10.1016/S0002-9440(10)64751-7
489. Iruela-Arispe, M. L. & Sage, E. H. Expression of type VIII collagen during morphogenesis of the chicken and mouse heart. *Dev. Biol.* (1991). doi:10.1016/0012-1606(91)90483-J

490. MacBeath, J. R., Kielty, C. M. & Shuttleworth, C. A. Type VIII collagen is a product of vascular smooth-muscle cells in development and disease. *Biochem. J.* (1996).
491. Illidge, C., Kielty, C. & Shuttleworth, A. The $\alpha 1(\text{VIII})$ and $\alpha 2(\text{VIII})$ chains of type VIII collagen can form stable homotrimeric molecules. *J. Biol. Chem.* (1998). doi:10.1074/jbc.273.34.22091
492. Kittelberger, R., Davis, P. F., Flynn, D. W. & Greenhill, N. S. Distribution of type VIII collagen in tissues: an immunohistochemical study. *Connect. Tissue Res.* (1990). doi:10.3109/03008209009152157
493. Illidge, C., Kielty, C. & Shuttleworth, A. Type VIII collagen: Heterotrimeric chain association. *Int. J. Biochem. Cell Biol.* (2001). doi:10.1016/S1357-2725(01)00013-9
494. Hansen, N. U. B. & Karsdal, M. A. Type VIII Collagen. in *Biochemistry of Collagens, Laminins and Elastin: Structure, Function and Biomarkers* (2016). doi:10.1016/B978-0-12-809847-9.00008-8
495. Rüger, B. *et al.* Human mast cells produce type VIII collagen in vivo. *Int. J. Exp. Pathol.* (1994).
496. Xu, R. *et al.* NC1 domain of human type VIII collagen ($\alpha 1$) inhibits bovine aortic endothelial cell proliferation and causes cell apoptosis. *Biochem. Biophys. Res. Commun.* (2001). doi:10.1006/bbrc.2001.5970
497. Basbaum, C. B. & Werb, Z. Focalized proteolysis: spatial and temporal regulation of extracellular matrix degradation at the cell surface. *Curr. Opin. Cell Biol.* (1996).
498. Zhao, Y. *et al.* siRNA-Targeted COL8A1 Inhibits Proliferation, Reduces Invasion and Enhances Sensitivity to D-Limonene Treatment in Hepatocarcinoma Cells. doi:10.1002/iub.151
499. Takeyama, K. *et al.* The BAL-binding protein BBAP and related Deltex family members exhibit ubiquitin-protein isopeptide ligase activity. *J. Biol. Chem.* **278**, 21930–7 (2003).

500. Bachmann, S. B. *et al.* DTX3L and ARTD9 inhibit IRF1 expression and mediate in cooperation with ARTD8 survival and proliferation of metastatic prostate cancer cells. *Mol. Cancer* **13**, 125 (2014).
501. Camicia, R. *et al.* BAL1/ARTD9 represses the anti-proliferative and pro-apoptotic IFN γ -STAT1-IRF1-p53 axis in diffuse large B-cell lymphoma. *J Cell Sci* **126**, (2013).
502. Juszczynski, P. *et al.* BAL1 and BBAP are regulated by a gamma interferon-responsive bidirectional promoter and are overexpressed in diffuse large B-cell lymphomas with a prominent inflammatory infiltrate. *Mol. Cell. Biol.* **26**, 5348–59 (2006).
503. Barbarulo, A. *et al.* Poly(ADP-ribose) polymerase family member 14 (PARP14) is a novel effector of the JNK2-dependent pro-survival signal in multiple myeloma. *Oncogene* **32**, (2012).
504. Obiero, J., Walker, J. R. & Dhe-Paganon, S. Fold of the conserved DTC domain in Deltex proteins. *Proteins* **80**, (2012).
505. Grunewald, T. G. *et al.* Prazeres da Costa O, Gorlach A, Cossarizza A, Butt E, Richter GH, Burdach S: STEAP1 is associated with the invasive and oxidative stress phenotype of Ewing tumors. *Mol Cancer Res* **10**, (2012).
506. Wilting, S. M. *et al.* Integrated genomic and transcriptional profiling identifies chromosomal loci with altered gene expression in cervical cancer. *Genes Chromosom. Cancer* **47**, (2008).
507. Sun, W. *et al.* Activation of the NOTCH pathway in head and neck cancer. *Cancer Res* **74**, (2014).
508. Yan, Q. *et al.* BBAP monoubiquitylates histone H4 at lysine 91 and selectively modulates the DNA damage response. *Mol. Cell* **36**, 110–20 (2009).
509. Dinh Thang, N. *et al.* Deltex-3-like (DTX3L) stimulates metastasis of melanoma through FAK/PI3K/AKT but not MEK/ERK pathway. *Oncotarget* **6**, 14290–14299 (2015).
510. Thu, Y. M. & Richmond, A. NF- κ B inducing kinase: A key regulator in the immune system and in cancer. *Cytokine Growth Factor Rev.* (2010). doi:10.1016/j.cytogfr.2010.06.002

511. Xia, Y., Shen, S. & Verma, I. M. NF- κ B, an active player in human cancers. *Cancer Immunol. Res.* (2014). doi:10.1158/2326-6066.CIR-14-0112
512. Staudt, L. M. Oncogenic activation of NF-kappaB. *Cold Spring Harbor perspectives in biology* (2010). doi:10.1101/cshperspect.a000109
513. Sun, S.-C. The noncanonical NF-kB pathway. *Immunol. Rev.* (2012). doi:10.1111/j.1600-065X.2011.01088.x.The
514. Sun, S. C. Controlling the Fate of NIK: A Central Stage in Noncanonical NF- B Signaling. *Sci. Signal.* (2010). doi:10.1126/scisignal.3123pe18
515. Razani, B., Reichardt, A. D. & Cheng, G. Non-canonical NF- κ B signaling activation and regulation: Principles and perspectives. *Immunol. Rev.* (2011). doi:10.1111/j.1600-065X.2011.01059.x
516. Razani, B. *et al.* Negative feedback in noncanonical NF-kappaB signaling modulates NIK stability through IKKalpha-mediated phosphorylation. *Sci. Signal.* **3**, ra41 (2010).
517. Oeckinghaus, Hayden & Ghosh. Crosstalk in NF- κ B signaling pathways. *Nat. Immunol.* (2011). doi:10.1038/ni.2065
518. Odqvist, L. *et al.* NIK controls classical and alternative NF- κ B activation and is necessary for the survival of human T-cell lymphoma cells. *Clin. Cancer Res.* (2013). doi:10.1158/1078-0432.CCR-12-3151
519. Ramakrishnan, P., Wang, W. & Wallach, D. Receptor-specific signaling for both the alternative and the canonical NF-kappaB activation pathways by NF-kappaB-inducing kinase. *Immunity* (2004). doi:S1074761304002456 [pii]n10.1016/j.immuni.2004.08.009
520. Nidai Ozes, O. *et al.* NF- κ B activation by tumour necrosis factor requires the Akt serine-threonine kinase. *Nature* **401**, 82–85 (1999).
521. Yamamoto, M. *et al.* Epigenetic alteration of the NF- κ B-inducing kinase (NIK) gene is involved in enhanced NIK expression in basal-like breast cancer. *Cancer Sci.* **101**, 2391–7 (2010).

522. Yamaguchi, N. *et al.* Constitutive activation of nuclear factor-kappaB is preferentially involved in the proliferation of basal-like subtype breast cancer cell lines. *Cancer Sci.* **100**, 1668–74 (2009).
523. Yamamoto, M. *et al.* NF-kappaB non-cell-autonomously regulates cancer stem cell populations in the basal-like breast cancer subtype. *Nat. Commun.* (2013). doi:10.1038/ncomms3299
524. Vazquez-Santillan, K. *et al.* NF-kappaB-inducing kinase regulates stem cell phenotype in breast cancer. *Sci. Rep.* **6**, 37340 (2016).
525. Furukawa, Y. *et al.* Isolation of a Novel Human Gene, ARHGAP9, Encoding a Rho-GTPase Activating Protein. *Biochem. Biophys. Res. Commun.* **284**, 643–649 (2001).
526. Hancock, C. *et al.* Mitogen Activated Protein (MAP) Kinases: Development of ATP and Non- ATP Dependent Inhibitors. *Med. Chem. (Los. Angeles)*. **2**, 213–222 (2006).
527. Seino, H. *et al.* Basic helix-loop-helix transcription factor DEC1 regulates the cisplatin-induced apoptotic pathway of human esophageal cancer cells. *Biomed. Res.* **36**, 89–96 (2015).
528. Zheng, Y. *et al.* The increased expression of DEC1 gene is related to HIF-1 α protein in gastric cancer cell lines. *Mol. Biol. Rep.* **39**, 4229–36 (2012).
529. Wu, Y. *et al.* The BHLH transcription factor DEC1 plays an important role in the epithelial-mesenchymal transition of pancreatic cancer. *Int. J. Oncol.* **41**, 1337–46 (2012).
530. Bi, H. *et al.* DEC1 regulates breast cancer cell proliferation by stabilizing cyclin E protein and delays the progression of cell cycle S phase. *Cell Death Dis.* **6**, e1891 (2015).
531. Liu, Y. *et al.* DEC1 is positively associated with the malignant phenotype of invasive breast cancers and negatively correlated with the expression of claudin-1. *Int. J. Mol. Med.* **31**, 855–60 (2013).
532. Liu, Y. *et al.* Anti-apoptotic effect of the basic helix-loop-helix (bHLH) transcription factor DEC2 in human breast cancer cells. *Genes Cells* **15**, 315–25 (2010).

533. Jia, Y.-F. *et al.* Differentiated embryonic chondrocyte-expressed gene 1 is associated with hypoxia-inducible factor 1 α and Ki67 in human gastric cancer. *Diagn. Pathol.* **8**, 37 (2013).
534. Liu, Y. *et al.* The transcription factor DEC1 (BHLHE40/STRA13/SHARP-2) is negatively associated with TNM stage in non-small-cell lung cancer and inhibits the proliferation through cyclin D1 in A549 and BE1 cells. *Tumour Biol.* **34**, 1641–50 (2013).
535. Zheng, Y. *et al.* The hypoxia-regulated transcription factor DEC1 (Stra13, SHARP-2) and its expression in gastric cancer. *OMICS* **13**, 301–6 (2009).
536. Wu, Y. *et al.* Basic helix-loop-helix transcription factors DEC1 and DEC2 regulate the paclitaxel-induced apoptotic pathway of MCF-7 human breast cancer cells. *Int. J. Mol. Med.* **27**, 491–5 (2011).
537. Giatromanolaki, A. *et al.* DEC1 (STRA13) protein expression relates to hypoxia-inducible factor 1- α and carbonic anhydrase-9 overexpression in non-small cell lung cancer. *J. Pathol.* **200**, 222–8 (2003).
538. Peinado, H., Olmeda, D. & Cano, A. Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nat Rev Cancer* **7**, (2007).
539. Riedel, H. Grb10 exceeding the boundaries of a common signaling adapter. *Frontiers in bioscience : a journal and virtual library* (2004). doi:10.2741/1227
540. HOLT, L. J. & SIDDLE, K. Grb10 and Grb14: enigmatic regulators of insulin action – and more? *Biochem. J.* (2005). doi:10.1042/BJ20050216
541. Kabir, N. N. & Kazi, J. U. Grb10 is a dual regulator of receptor tyrosine kinase signaling. *Molecular Biology Reports* (2014). doi:10.1007/s11033-014-3046-4
542. Ramos, F. J., Langlais, P. R., Hu, D., Dong, L. Q. & Liu, F. Grb10 mediates insulin-stimulated degradation of the insulin receptor: a mechanism of negative regulation. *Am. J. Physiol. Endocrinol. Metab.* **290**, E1262-6 (2006).
543. Morrión, A. *et al.* Grb10: A new substrate of the insulin-like growth factor I receptor. *Cancer Res.* **56**, 3165–7 (1996).

544. Tezuka, N., Brown, A. M. C. & Yanagawa, S. GRB10 binds to LRP6, the Wnt co-receptor and inhibits canonical Wnt signaling pathway. *Biochem. Biophys. Res. Commun.* **356**, 648–54 (2007).
545. Mroue, R., Huang, B., Braunstein, S., Firestone, A. J. & Nakamura, J. L. Monoallelic Loss of the Imprinted Gene Grb10 Promotes Tumor Formation in Irradiated Nf1^{+/-} Mice. *PLOS Genet.* **11**, e1005235 (2015).
546. Wei, J. *et al.* Identification of gene sets and pathways associated with lactation performance in mice.
547. Plasschaert, R. N., Bartolomei, M. S. & Keverne, E. B. Tissue-specific regulation and function of Grb10 during growth and neuronal commitment. doi:10.1073/pnas.1411254111
548. Lee, J. T. & Bartolomei, M. S. X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* (2013). doi:10.1016/j.cell.2013.02.016
549. Garfield, A. S. *et al.* Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. *Nature* (2011). doi:10.1038/nature09651
550. Shiura, H. *et al.* Paternal deletion of Meg1/Grb10 DMR causes maternalization of the Meg1/ Grb10 cluster in mouse proximal Chromosome 11 leading to severe pre- and postnatal growth retardation. *Hum. Mol. Genet.* (2009). doi:10.1093/hmg/ddp049
551. Smith, F. M. *et al.* Mice with a disruption of the imprinted Grb10 gene exhibit altered body composition, glucose homeostasis, and insulin signaling during postnatal life. *Mol. Cell. Biol.* (2007). doi:10.1128/MCB.02087-06
552. Wang, L. *et al.* Peripheral disruption of the Grb10 gene enhances insulin signaling and sensitivity in vivo. *Mol. Cell. Biol.* (2007). doi:10.1128/MCB.00679-07
553. Piñeiro, M. *et al.* ITIH4 serum concentration increases during acute-phase processes in human patients and is up-regulated by interleukin-6 in hepatocarcinoma HepG2 cells. *Biochem. Biophys. Res. Commun.* **263**, 224–9 (1999).
554. Banerjee, A., Gugasyan, R., McMahon, M. & Gerondakis, S. Diverse Toll-like receptors utilize Tpl2 to activate extracellular signal-regulated kinase (ERK) in hemopoietic cells. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 3274–9 (2006).

555. Liu, Y. *et al.* Mitogen-Activated Protein Kinase 8 (MAP3K8) Mediates the Signaling Pathway of Estradiol Stimulating Progesterone Production Through G Protein-Coupled Receptor 30 (GPR30) in Mouse Corpus Luteum. *Mol. Endocrinol.* **29**, 703–15 (2015).
556. Sugimoto, K. *et al.* A serine/threonine kinase, Cot/Tpl2, modulates bacterial DNA-induced IL-12 production and Th cell differentiation. *J. Clin. Invest.* (2004). doi:10.1172/JCI200420014
557. Kim, K. *et al.* Interleukin-22 promotes epithelial cell transformation and breast tumorigenesis via MAP3K8 activation. *Carcinogenesis* **35**, 1352–1361 (2014).
558. Kamiyama, M., Naguro, I. & Ichijo, H. *In vivo* gene manipulation reveals the impact of stress-responsive MAPK pathways on tumor progression. *Cancer Sci.* **106**, 785–796 (2015).
559. Narlis, M., Grote, D., Gaitan, Y., Boualia, S. K. & Bouchard, M. Pax2 and pax8 regulate branching morphogenesis and nephron differentiation in the developing kidney. *J. Am. Soc. Nephrol.* **18**, 1121–9 (2007).
560. Bouchard, M. Transcriptional control of kidney development. *Differentiation* **72**, 295–306 (2004).
561. Grote, D., Souabni, A., Busslinger, M. & Bouchard, M. Pax2/8-regulated Gata3 expression is necessary for morphogenesis and guidance of the nephric duct in the developing kidney. *Development* **133**, 53–61 (2006).
562. Bouchard, M., Souabni, A., Mandler, M., Neubüser, A. & Busslinger, M. Nephric lineage specification by Pax2 and Pax8. *Genes Dev.* **16**, 2958–2970 (2002).
563. Mansouri, A., Chowdhury, K. & Gruss, P. Follicular cells of the thyroid gland require Pax8 gene function. *Nat. Genet.* **19**, 87–90 (1998).
564. Poleev, A. *et al.* PAX8, a human paired box gene: isolation and expression in developing thyroid, kidney and Wilms' tumors. *Development* **116**, 611–23 (1992).
565. Plachov, D. *et al.* Pax8, a murine paired box gene expressed in the developing excretory system and thyroid gland. *Development* **110**, 643–51 (1990).
566. Mansouri, A., Hallonet, M. & Gruss, P. Pax genes and their roles in cell differentiation and development. *Curr. Opin. Cell Biol.* **8**, 851–7 (1996).

567. Dahl, E., Koseki, H. & Balling, R. Pax genes and organogenesis. *BioEssays* **19**, 755–765 (1997).
568. Ozcan, A. *et al.* PAX 8 expression in non-neoplastic tissues, primary tumors, and metastatic tumors: a comprehensive immunohistochemical study. *Mod. Pathol.* **24**, 751–64 (2011).
569. Shirai, M., Takihara, Y. & Morisaki, T. Pcgf5 Contributes to PRC1 (Polycomb Repressive Complex 1) in Developing Cardiac Cells. in *Etiology and Morphogenesis of Congenital Heart Disease* 305–312 (Springer Japan, 2016). doi:10.1007/978-4-431-54628-3_43
570. Oliviero, G. *et al.* The variant Polycomb Repressor Complex 1 component PCGF1 interacts with a pluripotency sub-network that includes DPPA4, a regulator of embryogenesis. *Sci. Rep.* **5**, 18388 (2015).
571. Gearhart, M. D., Corcoran, C. M., Wamstad, J. A. & Bardwell, V. J. Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets. *Mol. Cell. Biol.* **26**, 6880–9 (2006).
572. Junco, S. E. *et al.* Structure of the Polycomb Group Protein PCGF1 in Complex with BCOR Reveals Basis for Binding Selectivity of PCGF Homologs. *Structure* **21**, 665–671 (2013).
573. Oliviero, G. *et al.* The variant Polycomb Repressor Complex 1 component PCGF1 interacts with a pluripotency sub-network that includes DPPA4, a regulator of embryogenesis. *Sci. Rep.* **5**, 18388 (2016).
574. Li, H., Fan, R., Sun, M., Jiang, T. & Gong, Y. Nspc1 regulates the key pluripotent Oct4–Nanog–Sox2 axis in P19 embryonal carcinoma cells via directly activating Oct4. *Biochem. Biophys. Res. Commun.* **440**, 527–532 (2013).
575. Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Cavalli, G. Genome Regulation by Polycomb and Trithorax Proteins. *Cell* **128**, 735–745 (2007).
576. Schuettengruber, B. & Cavalli, G. Recruitment of Polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development* **136**, 3531–3542 (2009).

577. Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873–878 (2004).
578. Cao, R., Tsukada, Y. & Zhang, Y. Role of Bmi-1 and Ring1A in H2A Ubiquitylation and Hox Gene Silencing. *Mol. Cell* **20**, 845–854 (2005).
579. Grau, D. J., Antao, J. M. & Kingston, R. E. Functional Dissection of Polycomb Repressive Complex 1 Reveals the Importance of a Charged Domain. *Cold Spring Harb. Symp. Quant. Biol.* **75**, 61–70 (2010).
580. Si, S. *et al.* Loss of Pcgf5 Affects Global H2A Monoubiquitination but Not the Function of Hematopoietic Stem and Progenitor Cells. *PLoS One* **11**, e0154561 (2016).
581. Dupret, B. *et al.* The Polycomb Group Protein Pcgf1 Is Dispensable in Zebrafish but Involved in Early Growth and Aging. *PLoS One* **11**, e0158700 (2016).
582. Salameh, A. *et al.* PRUNE2 is a human prostate cancer suppressor regulated by the intronic long noncoding RNA PCA3. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 8403–8 (2015).
583. Iwama, E. *et al.* Cancer-related PRUNE2 protein is associated with nucleotides and is highly expressed in mature nerve tissues. *J. Mol. Neurosci.* **44**, 103–14 (2011).
584. Yu, W. *et al.* Whole-exome sequencing studies of parathyroid carcinomas reveal novel PRUNE2 mutations, distinctive mutational spectra related to APOBEC-catalyzed DNA mutagenesis and mutational enrichment in kinases associated with cell migration and invasion. *J. Clin. Endocrinol. Metab.* **100**, E360–4 (2015).
585. de Sauvage, F. J. *et al.* Stimulation of megakaryocytopoiesis and thrombopoiesis by the c-Mpl ligand. *Nature* (1994). doi:10.1038/369533a0
586. Kaushansky, K. *et al.* Thrombopoietin, the Mpl ligand, is essential for full megakaryocyte development. *Proc. Natl. Acad. Sci. U. S. A.* (1995). doi:10.1073/pnas.92.8.3234
587. Pallard, C. *et al.* Thrombopoietin activates a STAT5-like factor in hematopoietic cells. *EMBO J.* **14**, 2847–56 (1995).

588. Miyakawa, Y., Rojnuckarin, P., Habib, T. & Kaushansky, K. Thrombopoietin induces phosphoinositol 3-kinase activation through SHP2, Gab, and insulin receptor substrate proteins in BAF3 cells and primary murine megakaryocytes. *J. Biol. Chem.* **276**, 2494–502 (2001).
589. Campus, F. *et al.* Thrombopoietin complements G(i)- but not G(q)-dependent pathways for integrin {alpha}(IIb){beta}(3) activation and platelet aggregation. *J. Biol. Chem.* **280**, 24386–95 (2005).
590. Matsumura, I. *et al.* Thrombopoietin-induced differentiation of a human megakaryoblastic leukemia cell line, CMK, involves transcriptional activation of p21(WAF1/Cip1) by STAT5. *Mol. Cell. Biol.* **17**, 2933–43 (1997).
591. de Graaf, C. A. & Metcalf, D. Thrombopoietin and hematopoietic stem cells. *Cell Cycle* **10**, 1582–9 (2011).
592. Fox, N., Priestley, G., Papayannopoulou, T. & Kaushansky, K. Thrombopoietin expands hematopoietic stem cells after transplantation. *J. Clin. Invest.* (2002). doi:10.1172/JCI15430
593. ter Braak, C. J. F. & Šmilauer, P. Topics in constrained and unconstrained ordination. *Plant Ecology* (2015). doi:10.1007/s11258-014-0356-5
594. Van den Brink, P. J. & Braak, C. J. F. Ter. Principal response curves: Analysis of time-dependent multivariate responses of biological community to stress. *Environ. Toxicol. Chem.* (1999). doi:10.1002/etc.5620180207
595. van den Brink, P. J., den Besten, P. J., bij de Vaate, A. & ter Braak, C. J. F. Principal response curves technique for the analysis of multivariate biomonitoring time series. *Environ. Monit. Assess.* (2009). doi:10.1007/s10661-008-0314-6
596. van den Brink, P. & ter Braak, C. Multivariate analysis of stress in experimental ecosystems by Principal Response Curves and similarity analysis. *Aquat. Ecol.* (1998). doi:10.1023/a:1009944004756
597. Frampton, G. K., Van den Brink, P. J. & Gould, P. J. L. Effects of spring precipitation on a temperate arable collembolan community analysed using Principal Response Curves. *Appl. Soil Ecol.* (2000). doi:10.1016/S0929-1393(00)00051-2

598. Moser, T., Römcke, J., Schallnass, H. J. & Van Gestel, C. A. M. The use of the multivariate Principal Response Curve (PRC) for community level analysis: A case study on the effects of carbendazim on enchytraeids in Terrestrial Model Ecosystems (TME). *Ecotoxicology* (2007). doi:10.1007/s10646-007-0169-6
599. Heegaard, E. & Vandvik, V. Climate change affects the outcome of competitive interactions - an application of principal response curves. *Oecologia* (2004). doi:10.1007/s00442-004-1523-5
600. Van Den Brink, P. J. & Ter Braak, C. J. F. Principal Response Curves (Prc). *Environ. Monit. Assess.* (2009). doi:10.1007/s10661-008-0314-6
601. Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene function analysis with the PANTHER classification system. *Nat. Protoc.* (2013). doi:10.1038/nprot.2013.092
602. Mi, H., Guo, N., Kejariwal, A. & Thomas, P. D. PANTHER version 6: Protein sequence and function evolution data with expanded representation of biological pathways. *Nucleic Acids Res.* (2007). doi:10.1093/nar/gkl869
603. Thomas, P. D. *et al.* PANTHER: A browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Research* (2003). doi:10.1093/nar/gkg115
604. Mi, H. *et al.* The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.* (2005). doi:10.1093/nar/gki078
605. Thomas, P. D. *et al.* PANTHER: A library of protein families and subfamilies indexed by function. *Genome Res.* (2003). doi:10.1101/gr.772403
606. Kremer, A. N. *et al.* Human Leukocyte Antigen–DO Regulates Surface Presentation of Human Leukocyte Antigen Class II–Restricted Antigens on B Cell Malignancies. *Biol. Blood Marrow Transplant.* **20**, 742–747 (2014).
607. Thibodeau, J., Bourgeois-Daigneault, M.-C. & Lapointe, R. Targeting the MHC Class II antigen presentation pathway in cancer immunotherapy. *Oncoimmunology* **1**, 908–916 (2012).

608. Tabibzadeh, S. S., Sivarajah, A., Carpenter, D., Ohlsson-Wilhelm, B. M. & Satyaswaroop, P. G. Modulation of HLA-DR expression in epithelial cells by interleukin 1 and estradiol-17 beta. *J Clin Endocrinol Metab* (1990). doi:10.1210/jcem-71-3-740
609. Durrant, L. G. *et al.* Quantitation of MHC antigen expression on colorectal tumours and its association with tumour progression. *Br. J. Cancer* (1987). doi:10.1038/bjc.1987.218
610. Souwer, Y. *et al.* Detection of aberrant transcription of major histocompatibility complex class II antigen presentation genes in chronic lymphocytic leukaemia identifies HLA-DOA mRNA as a prognostic factor for survival. *Br. J. Haematol.* (2009). doi:10.1111/j.1365-2141.2009.07625.x
611. Van Lith, M., Van Ham, M. & Neefjes, J. Novel polymorphisms in HLA-DOA and HLA-DOB in B-cell malignancies. *Immunogenetics* (2002). doi:10.1007/s00251-002-0500-6
612. Yi, Z., Lin, W. W., Stunz, L. L. & Bishop, G. A. Survey Roles for TNF-receptor associated factor 3 (TRAF3) in lymphocyte functions. (2014). doi:10.1016/j.cytogfr.2013.12.002
613. Xie, P., Hostager, B. S. & Bishop, G. A. Requirement for TRAF3 in signaling by LMP1 but not CD40 in B lymphocytes. *J. Exp. Med.* (2004). doi:10.1084/jem.20031255
614. Chan, H. & Reed, J. C. TRAF-dependent association of protein kinase Tpl2/COT1 (MAP3K8) with CD40. *Biochem. Biophys. Res. Commun.* **328**, 198–205 (2005).
615. Xie, P., Stunz, L. L., Larison, K. D., Yang, B. & Bishop, G. A. TRAF3 is a critical regulator of B cell homeostasis in secondary lymphoid organs. *Immunity* (2007). doi:10.1016/j.immuni.2007.07.012
616. Gardam, S., Sierro, F., Basten, A., Mackay, F. & Brink, R. TRAF2 and TRAF3 Signal Adapters Act Cooperatively to Control the Maturation and Survival Signals Delivered to B Cells by the BAFF Receptor. *Immunity* (2008). doi:10.1016/j.immuni.2008.01.009
617. Mackay, F., Schneider, P., Rennert, P. & Browning, J. BAFF AND APRIL: a tutorial on B cell survival. *Annu. Rev. Immunol.* (2003). doi:10.1146/annurev.immunol.21.120601.141152

618. Xie, P., Kraus, Z. J., Stunz, L. L., Liu, Y. & Bishop, G. a. TNF receptor-associated factor 3 is required for T cell-mediated immunity and TCR/CD28 signaling. *J. Immunol.* (2011). doi:10.4049/jimmunol.1000290
619. Hildebrand, J. M. *et al.* Roles of tumor necrosis factor receptor associated factor 3 (TRAF3) and TRAF5 in immune cell functions. *Immunol. Rev.* (2011).
620. Franzoso, G. *et al.* Mice deficient in nuclear factor (NF)-kappa B/p52 present with defects in humoral responses, germinal center reactions, and splenic microarchitecture. *J. Exp. Med.* (1998). doi:10.1084/jem.187.2.147
621. Keats, J. J. *et al.* Promiscuous Mutations Activate the Noncanonical NF-??B Pathway in Multiple Myeloma. *Cancer Cell* (2007). doi:10.1016/j.ccr.2007.07.003
622. Annunziata, C. M. *et al.* Frequent Engagement of the Classical and Alternative NF-κB Pathways by Diverse Genetic Abnormalities in Multiple Myeloma. *Cancer Cell* (2007). doi:10.1016/j.ccr.2007.07.004
623. Otto, C. *et al.* Genetic lesions of the TRAF3 and MAP3K14 genes in classical Hodgkin lymphoma. *Br. J. Haematol.* (2012). doi:10.1111/j.1365-2141.2012.09113.x
624. Moore, C. R. *et al.* Specific deletion of TRAF3 in B lymphocytes leads to B-lymphoma development in mice. *Leukemia* (2012). doi:10.1038/leu.2011.309
625. Hostager, B. S. & Bishop, G. A. Cutting edge: contrasting roles of TNF receptor-associated factor 2 (TRAF2) and TRAF3 in CD40-activated B lymphocyte differentiation. *J. Immunol.* (1999).
626. Khare, S. D. & Hsu, H. The role of TALL-1 and APRIL in immune regulation. *Trends in Immunology* (2001). doi:10.1016/S1471-4906(00)01843-3
627. Cancro, M. P. The BLyS/BAFF family of ligands and receptors: key targets in the therapy and understanding of autoimmunity. *Ann. Rheum. Dis.* (2006). doi:10.1136/ard.2006.058412
628. Cancro, M. P. The BLyS family of ligands and receptors: An archetype for niche-specific homeostatic regulation. *Immunological Reviews* (2004). doi:10.1111/j.0105-2896.2004.00212.x

629. Do, R. K. G. & Chen-Kiang, S. Mechanism of BlyS action in B cell immunity. *Cytokine and Growth Factor Reviews* (2002). doi:10.1016/S1359-6101(01)00025-9
630. Rothe, M., Pan, M. G., Henzel, W. J., Ayres, T. M. & Goeddel, D. V. The TNFR2-TRAF signaling complex contains two novel proteins related to baculoviral-inhibitor of apoptosis proteins. *Cell* (1995). doi:10.1016/0092-8674(95)90149-3
631. Häcker, H. *et al.* Specificity in Toll-like receptor signalling through distinct effector functions of TRAF3 and TRAF6. *Nature* (2006). doi:10.1038/nature04369
632. Oganessian, G., Saha, S., Guo, B. & He, J. Critical role of TRAF3 in the Toll-like receptor-dependent and-independent antiviral response. *Nature* (2005). doi:10.1038/nature04374
633. Perkins, D. J. *et al.* Reprogramming of Murine Macrophages through TLR2 Confers Viral Resistance via TRAF3-Mediated, Enhanced Interferon Production. *PLoS Pathog.* (2013). doi:10.1371/journal.ppat.1003479
634. Xie, P. *et al.* Enhanced Toll-like receptor (TLR) responses of TNFR-associated factor 3 (TRAF3)-deficient B lymphocytes. *J Leukoc Biol* (2011). doi:10.1189/jlb.0111044
635. Yi, Z., Stunz, L. L. & Bishop, G. A. TNF receptor associated factor 3 plays a key role in development and function of invariant natural killer T cells. *J. Exp. Med.* **210**, (2013).
636. Niranjana, B. *et al.* HGF/SF: a potent cytokine for mammary growth, morphogenesis and development. *Development* **121**, 2897–908 (1995).
637. Kamalati, T., Niranjana, B., Yant, J. & Buluwela, L. HGF/SF in Mammary Epithelial Growth and Morphogenesis: In Vitro and In Vivo Models. *J. Mammary Gland Biol. Neoplasia* **4**, 69–77 (1999).
638. Gherardi, E. & Stoker, M. Hepatocyte growth factor--scatter factor: mitogen, motogen, and met. *Cancer Cells* **3**, 227–32 (1991).
639. Plath, A. *et al.* Expression of transforming growth factors alpha and beta-1 messenger RNA in the bovine mammary gland during different stages of development and lactation. *J. Endocrinol.* **155**, 501–511 (1997).

640. Borowiak, M. *et al.* Met provides essential signals for liver regeneration. *Proc. Natl. Acad. Sci. U. S. A.* (2004). doi:10.1073/pnas.0403412101
641. Huh, C. G. *et al.* Hepatocyte growth factor/c-met signaling pathway is required for efficient liver regeneration and repair. *Proc Natl Acad Sci U S A* (2004). doi:10.1073/pnas.0306068101r0306068101 [pii]
642. Neuss, S., Becher, E., Woltje, M., Tietze, L. & Jahnen-Dechent, W. Functional expression of HGF and HGF receptor/c-met in adult human mesenchymal stem cells suggests a role in cell mobilization, tissue repair, and wound healing. *Stem Cells* (2004). doi:10.1634/stemcells.22-3-405
643. Forte, G. *et al.* Hepatocyte growth factor effects on mesenchymal stem cells: proliferation, migration, and differentiation. *Stem Cells* (2006). doi:2004-0176 [pii]r10.1634/stemcells.2004-0176
644. Sonnenberg, E., Meyer, D., Weidner, K. M. & Birchmeier, C. Scatter factor/hepatocyte growth factor and its receptor, the c-met tyrosine kinase, can mediate a signal exchange between mesenchyme and epithelia during mouse development. *J. Cell Biol.* (1993). doi:10.1083/jcb.123.1.223
645. Schmidt, C. *et al.* Scatter factor/hepatocyte growth factor is essential for liver development. *Nature* (1995). doi:10.1038/373699a0
646. Uehara, Y. *et al.* Placental defect and embryonic lethality in mice lacking hepatocyte growth factor/scatter factor. *Nature* (1995). doi:10.1038/373702a0
647. Streit, A. *et al.* A role for HGF/SF in neural induction and its expression in Hensen's node during gastrulation. *Development* (1995).
648. Stern, C. D. *et al.* Epithelial scatter factor and development of the chick embryonic axis. *Development* (1990). doi:10.1016/0168-9525(91)90056-V
649. Myokai, F. *et al.* Expression of the hepatocyte growth factor gene during chick limb development. *Dev. Dyn.* (1995). doi:10.1002/aja.1002020108
650. Ronen, D. *et al.* Met-HGF/SF mediates growth arrest and differentiation in T47D breast cancer cells. *Cell Growth Differ.* **10**, 131–40 (1999).

651. Yang, Y. *et al.* Sequential requirement of hepatocyte growth factor and neuregulin in the morphogenesis and differentiation of the mammary gland. *J. Cell Biol.* **131**, 215–226 (1995).
652. Bertotti, A., Comoglio, P. M. & Trusolino, L. β 4 integrin activates a Shp2-Src signaling pathway that sustains HGF-induced anchorage-independent growth. *J. Cell Biol.* (2006). doi:10.1083/jcb.200605114
653. Bertotti, A., Comoglio, P. M. & Trusolino, L. β 4 integrin is a transforming molecule that unleashes Met tyrosine kinase tumorigenesis. *Cancer Res.* (2005). doi:10.1158/0008-5472.CAN-05-2827
654. Trusolino, L., Bertotti, A. & Comoglio, P. M. A signaling adapter function for α 6 β 4 integrin in the control of HGF-dependent invasive growth. *Cell* (2001). doi:10.1016/S0092-8674(01)00567-0
655. Ishizawa, R. & Parsons, S. J. C-Src and cooperating partners in human cancer. *Cancer Cell* (2004). doi:10.1016/j.ccr.2004.09.001
656. Wojcik, E. J. *et al.* A novel activating function of c-Src and Stat3 on HGF transcription in mammary carcinoma cells. *Oncogene* (2006). doi:10.1038/sj.onc.1209306
657. Liu, S. HGF-MET as a breast cancer biomarker. *Aging (Albany, NY)*. **7**, 150–1 (2015).
658. Charafe-Jauffret, E. *et al.* Gene expression profiling of breast cell lines identifies potential new basal markers. *Oncogene* **25**, (2006).
659. Goncalves, A. *et al.* Protein profiling of human breast tumor cells identifies novel biomarkers associated with molecular subtypes. *Mol Cell Proteomics* (2008). doi:M700487-MCP200 [pii]r10.1074/mcp.M700487-MCP200
660. Smolen, G. A. *et al.* Frequent Met oncogene amplification in a Brca1/Trp53 mouse model of mammary tumorigenesis. *Cancer Res.* (2006). doi:10.1158/0008-5472.CAN-05-4181
661. Sarkar, A. & Hochedlinger, K. The Sox Family of Transcription Factors: Versatile Regulators of Stem and Progenitor Cell Fate. *Cell Stem Cell* **12**, 15–30 (2013).

662. Roose, J. *et al.* High expression of the HMG box factor Sox-13 in arterial walls during embryonic development. *Nucleic Acids Res.* **26**, 469–476 (1998).
663. Roose, J. *et al.* *The Sox-13 Gene: Structure, Promoter Characterization, and Chromosomal Localization.* *Genomics* **57**, (1999).
664. Melichar, H. J. *et al.* Regulation of γ Versus δ T Lymphocyte Differentiation by the Transcription Factor SOX13. *Science (80-.)*. **315**, 230–233 (2007).
665. Kane, L. A. *et al.* PINK1 phosphorylates ubiquitin to activate parkin E3 ubiquitin ligase activity. *J. Cell Biol.* (2014). doi:10.1083/jcb.201402104
666. van der Merwe, C., Jalali Sefid Dashti, Z., Christoffels, A., Loos, B. & Bardien, S. Evidence for a common biological pathway linking three Parkinson's disease-causing genes: Parkin, PINK1 and DJ-1. *European Journal of Neuroscience* (2015). doi:10.1111/ejn.12872
667. Cookson, M. R. Parkinsonism due to mutations in PINK1, Parkin, and DJ-1 and oxidative stress and mitochondrial pathways. *Cold Spring Harb. Perspect. Med.* (2012). doi:10.1101/cshperspect.a009415
668. Ishihara-Paul, L. *et al.* PINK1 mutations and parkinsonism. *Neurology* (2008). doi:10.1212/01.wnl.0000323812.40708.1f
669. Unoki, M. & Nakamura, Y. Growth-suppressive effects of BPOZ and EGR2, two genes involved in the PTEN signaling pathway. *Oncogene* (2001). doi:10.1038/sj.onc.1204608
670. Yin, Y. & Shen, W. H. PTEN: a new guardian of the genome. *Oncogene* (2008). doi:10.1038/onc.2008.241
671. Nakajima, A., Kataoka, K., Hong, M., Sakaguchi, M. & Huh, N. H. BRPK, a novel protein kinase showing increased expression in mouse cancer cell lines with higher metastatic potential. *Cancer Lett.* (2003). doi:10.1016/S0304-3835(03)00443-9
672. Valente, E. M. *et al.* Hereditary early-onset Parkinson's disease caused by mutations in PINK1. *Science* (2004). doi:10.1126/science.1096284
673. Arena, G. *et al.* PINK1 protects against cell death induced by mitochondrial depolarization, by phosphorylating Bcl-xL and impairing its pro-apoptotic cleavage. *Cell Death Differ.* (2013). doi:10.1038/cdd.2013.19

674. Trempe, J. F. & Fon, E. A. Structure and function of Parkin, PINK1, and DJ-1, the three musketeers of neuroprotection. *Front. Neurol.* (2013). doi:10.3389/fneur.2013.00038
675. O’Flanagan, C. H. & O’Neill, C. PINK1 signalling in cancer biology. *Biochimica et Biophysica Acta - Reviews on Cancer* (2014). doi:10.1016/j.bbcan.2014.10.006
676. Fujiwara, M. *et al.* Parkin as a tumor suppressor gene for hepatocellular carcinoma. *Oncogene* (2008). doi:10.1038/onc.2008.199
677. Berthier, A. *et al.* PINK1 displays tissue-specific subcellular location and regulates apoptosis and cell growth in breast cancer cells. *Hum. Pathol.* (2010). doi:10.1016/j.humpath.2010.05.016
678. Jiménez-Sainz, J., Berthier, A. & Pulido, R. PINK1: An anti-apoptotic and anti-proliferative protein in human breast cancer cells. *FEBS J.* (2012). doi:10.1111/j.1742-4658.2010.08705.x
679. Matsuda, S., Nakanishi, A., Minami, A., Wada, Y. & Kitagishi, Y. Functions and characteristics of PINK1 and Parkin in cancer. *Front. Biosci. (Landmark Ed.* (2015). doi:10.2741/4321
680. Akundi, R. S., Zhi, L. & Büeler, H. PINK1 enhances insulin-like growth factor-1-dependent Akt signaling and protection against apoptosis. *Neurobiol. Dis.* (2012). doi:10.1016/j.nbd.2011.08.034
681. Sánchez-Mora, R. M., Arboleda, H. & Arboleda, G. PINK1 overexpression protects against C2-ceramide-induced CAD cell death through the PI3K/AKT pathway. *J. Mol. Neurosci.* (2012). doi:10.1007/s12031-011-9687-z
682. Murata, H. *et al.* A new cytosolic pathway from a Parkinson disease-associated kinase, BRPK/PINK1: Activation of AKT via MTORC2. *J. Biol. Chem.* (2011). doi:10.1074/jbc.M110.179390
683. Mei, Y. *et al.* FOXO3a-dependent regulation of Pink1 (Park6) mediates survival signaling in response to cytokine deprivation. *Proc. Natl. Acad. Sci. U. S. A.* (2009). doi:10.1073/pnas.0901104106
684. Storz, P., Doppler, H., Copland, J. A., Simpson, K. J. & Toker, A. FOXO3a Promotes Tumor Cell Invasion through the Induction of Matrix Metalloproteinases. *Mol. Cell. Biol.* (2009). doi:10.1128/MCB.00077-09

685. Martin, S. A., Hewish, M., Sims, D., Lord, C. J. & Ashworth, A. Parallel high-throughput RNA interference screens identify PINK1 as a potential therapeutic target for the treatment of DNA mismatch repair-deficient cancers. *Cancer Res.* (2011). doi:10.1158/0008-5472.CAN-10-2836
686. O’Flanagan, C. H., Morais, V. a, Wurst, W., De Strooper, B. & O’Neill, C. The Parkinson’s gene PINK1 regulates cell cycle progression and promotes cancer-associated phenotypes. *Oncogene* (2014). doi:10.1038/onc.2014.81
687. Berthier, A. *et al.* PINK1 displays tissue-specific subcellular location and regulates apoptosis and cell growth in breast cancer cells. *Hum. Pathol.* **42**, 75–87 (2011).
688. Jauhainen, A. *et al.* Distinct Cytoplasmic and Nuclear Functions of the Stress Induced Protein DDIT3/CHOP/GADD153. *PLoS One* **7**, e33208 (2012).
689. Yu, M. *et al.* Ddit3 suppresses the differentiation of mouse chondroprogenitor cells. *Int. J. Biochem. Cell Biol.* (2016). doi:10.1016/j.biocel.2016.11.009
690. Tsai, S.-F. *et al.* Isochahalactone-induced DDIT3 causes ER stress-PERK independent apoptosis in glioblastoma multiforme cells. *Oncotarget* **8**, 4051–4061 (2017).
691. Wu, Y., Sun, H., Song, F., Fu, D. & Wang, J. DDIT3 overexpression increases odontoblastic potential of human dental pulp cells. *Cell Prolif.* (2014). doi:10.1111/cpr.12104
692. Pina, C. *et al.* Single-Cell Network Analysis Identifies DDIT3 as a Nodal Lineage Regulator in Hematopoiesis. *Cell Rep.* (2015). doi:10.1016/j.celrep.2015.05.016
693. Liu, T. *et al.* Hypoxia induces p53-dependent transactivation and Fas/CD95-dependent apoptosis. *Cell Death Differ* (2007). doi:10.1038/sj.cdd.4402022
694. Tang, J. R. *et al.* Mechanism of oxidative stress-induced GADD153 gene expression in vascular smooth muscle cells. *Biochem.Biophys.Res.Comm.* (2002).
695. Ma, Y., Brewer, J. W., Alan Diehl, J. & Hendershot, L. M. Two distinct stress signaling pathways converge upon the CHOP promoter during the mammalian unfolded protein response. *J. Mol. Biol.* (2002). doi:10.1016/S0022-2836(02)00234-6

696. Jousse, C. *et al.* Amino acid limitation regulates CHOP expression through a specific pathway independent of the unfolded protein response. *FEBS Lett.* (1999). doi:10.1016/S0014-5793(99)00373-7
697. Jackman, A., Alamo, A. & Fornace, F. Genotoxic Stress Confers Preferential and Coordinate Messenger RNA Stability on the Five gadd Genes. *Cancer Res.* (1994).
698. Gately, D. P. *et al.* Induction of the growth arrest and DNA damage-inducible gene GADD153 by cisplatin in vitro and in vivo. *Br. J. Cancer* (1994). doi:10.1038/bjc.1994.455
699. Luethy, J. D. & Holbrook, N. J. Activation of the gadd153 promoter by genotoxic agents: A rapid and specific response to DNA damage. *Cancer Res.* (1992).
700. Oyadomari, S. & Mori, M. Roles of CHOP/GADD153 in endoplasmic reticulum stress. *Cell Death Differ.* (2004). doi:10.1038/sj.cdd.4401373
701. Yamaguchi, H. & Wang, H.-G. CHOP is involved in endoplasmic reticulum stress-induced apoptosis by enhancing DR5 expression in human carcinoma cells. *J. Biol. Chem.* (2004). doi:10.1074/jbc.M406933200
702. Chen, Y. *et al.* Down-regulation of PERK-ATF4-CHOP pathway by astragaloside IV is associated with the inhibition of endoplasmic reticulum stress-induced podocyte apoptosis in diabetic rats. *Cell. Physiol. Biochem.* (2014). doi:10.1159/000362974
703. Tang, J. *et al.* ER Stress via CHOP Pathway is Involved in FK506-Induced Apoptosis in Rat Fibroblasts. *Cell. Physiol. Biochem.* (2016). doi:10.1159/000447894
704. Ohoka, N., Yoshii, S., Hattori, T., Onozaki, K. & Hayashi, H. TRB3, a novel ER stress-inducible gene, is induced via ATF4-CHOP pathway and is involved in cell death. *EMBO J.* (2005). doi:10.1038/sj.emboj.7600596
705. Vij, N., Amoako, M. O., Mazur, S. & Zeitlin, P. L. CHOP transcription factor mediates IL-8 signaling in cystic fibrosis bronchial epithelial cells. *Am. J. Respir. Cell Mol. Biol.* (2008). doi:10.1165/rcmb.2007-0197OC
706. Oliveira, S. J. *et al.* ER stress-inducible factor CHOP affects the expression of hepcidin by modulating C/EBPalpha activity. *PLoS One* (2009). doi:10.1371/journal.pone.0006618

707. Park, S. H. *et al.* Endoplasmic reticulum stress-activated C/EBP homologous protein enhances nuclear factor-kappaB signals via repression of peroxisome proliferator-activated receptor gamma. *J. Biol. Chem.* (2010). doi:10.1074/jbc.M110.136259
708. Goodall, J. C. *et al.* Endoplasmic reticulum stress-induced transcription factor, CHOP, is crucial for dendritic cell IL-23 expression. *Proc. Natl. Acad. Sci. U. S. A.* (2010). doi:10.1073/pnas.1011736107
709. Engström, K. *et al.* The myxoid/round cell liposarcoma fusion oncogene FUS-DDIT3 and the normal DDIT3 induce a liposarcoma phenotype in transfected human fibrosarcoma cells. *Am. J. Pathol.* (2006). doi:10.2353/ajpath.2006.050872
710. Thorp, E. *et al.* Reduced Apoptosis and Plaque Necrosis in Advanced Atherosclerotic Lesions of Apoe^{-/-} and Ldlr^{-/-} Mice Lacking CHOP. *Cell Metab.* (2009). doi:10.1016/j.cmet.2009.03.003
711. Batchvarova, N., Wang, X. Z. & Ron, D. Inhibition of adipogenesis by the stress-induced protein CHOP (Gadd153). *EMBO J.* (1995).
712. Ron, D. & Habener, J. F. CHOP, a novel developmentally regulated nuclear protein that dimerizes with transcription factors C/EBP and LAP and functions as a dominant-negative inhibitor of gene transcription. *Genes Dev.* (1992). doi:10.1101/gad.6.3.439
713. Shirakawa, K. *et al.* CCAAT/enhancer-binding protein homologous protein (CHOP) regulates osteoblast differentiation. *Mol. Cell. Biol.* (2006). doi:10.1128/MCB.02429-05
714. Cui, K., Coutts, M., Stahl, J. & Sytkowski, A. J. Novel interaction between the transcription factor CHOP (GADD153) and the ribosomal protein FTE/S3a modulates erythropoiesis. *J. Biol. Chem.* (2000). doi:10.1074/jbc.275.11.7591
715. Vittoria Barone, M., Crozat, A., Tabae, A., Philipson, L. & Ron, D. CHOP (GADD153) and its oncogenic variant, TLS-CHOP, have opposing effects on the induction of G1/S arrest. *Genes Dev.* (1994). doi:10.1101/gad.8.4.453
716. Su, N. & Kilberg, M. S. C/EBP homologous protein (CHOP) interacts with activating transcription factor 4 (ATF4) and negatively regulates the stress-dependent induction of the asparagine synthetase gene. *J. Biol. Chem.* (2008). doi:10.1074/jbc.M806874200

717. Fleming, J. V., Fontanier, N., Harries, D. N. & Rees, W. D. The growth arrest genes *gas5*, *gas6*, and *CHOP-10* (*gadd153*) are expressed in the mouse preimplantation embryo. *Mol. Reprod. Dev.* **48**, 310–316 (1997).
718. Horndasch, M. *et al.* The C/EBP homologous protein CHOP (GADD153) is an inhibitor of Wnt/TCF signals. *Oncogene* **25**, 3397–3407 (2006).
719. Thevenot, P. T. *et al.* The stress-response sensor chop regulates the function and accumulation of myeloid-derived suppressor cells in tumors. *Immunity* **41**, 389–401 (2014).
720. Kobayashi, M., Harada, K., Negishi, M. & Katoh, H. Dock4 forms a complex with SH3YL1 and regulates cancer cell migration. *Cell. Signal.* **26**, 1082–1088 (2014).
721. Yajnik, V. *et al.* DOCK4, a GTPase Activator, Is Disrupted during Tumorigenesis. *Cell* **112**, 673–684 (2003).
722. Wu, Y. C. & Horvitz, H. R. C. elegans phagocytosis and cell-migration protein CED-5 is similar to human DOCK180. *Nature* (1998). doi:10.1038/33163
723. Nolan, K. M. *et al.* Myoblast city, the Drosophila homolog of DOCK180/CED-5, is required in a Rac signaling pathway utilized for multiple developmental processes. *Genes Dev.* (1998). doi:10.1101/gad.12.21.3337
724. Hasegawa, H. *et al.* DOCK180, a major CRK-binding protein, alters cell morphology upon translocation to the cell membrane. *Mol. Cell. Biol.* (1996).
725. Rathjen, F. G. & Schachner, M. Immunocytological and biochemical characterization of a new neuronal cell surface component (L1 antigen) which is involved in cell adhesion. *EMBO J.* (1984).
726. Arlt, A. *et al.* Autocrine production of interleukin 1beta confers constitutive nuclear factor kappaB activity and chemoresistance in pancreatic carcinoma cell lines. *Cancer Res.* (2002).
727. Brummendorf, T., Kenwrick, S. & Rathjen, F. G. Neural cell recognition molecule L1: from cell biology to human hereditary brain malformations. *Curr. Opin. Neurobiol.* (1998). doi:10.1016/S0959-4388(98)80012-3
728. Kiefel, H. *et al.* L1CAM. *Cell Adh. Migr.* **6**, 374–384 (2012).

729. Kiefel, H., Pfeifer, M., Bondong, S., Hazin, J. & Altevogt, P. Linking L1CAM-mediated signaling to NF- κ B activation. *Trends Mol. Med.* **17**, 178–187 (2011).
730. Gast, D. *et al.* The RGD integrin binding site in human L1-CAM is important for nuclear signaling. *Exp. Cell Res.* (2008). doi:10.1016/j.yexcr.2008.04.004
731. Schmid, R. S. & Maness, P. F. L1 and NCAM adhesion molecules as signaling coreceptors in neuronal migration and process outgrowth. *Current Opinion in Neurobiology* (2008). doi:10.1016/j.conb.2008.07.015
732. Mechtersheimer, S. *et al.* Ectodomain shedding of L1 adhesion molecule promotes cell migration by autocrine binding to integrins. *J. Cell Biol.* (2001). doi:10.1083/jcb.200101099
733. Muërkoter, S. S. *et al.* Drug-induced expression of the cellular adhesion molecule L1CAM confers anti-apoptotic protection and chemoresistance in pancreatic ductal adenocarcinoma cells. *Oncogene* (2007). doi:10.1038/sj.onc.1210076
734. Nishimune, H. *et al.* Neural adhesion molecules L1 and CHL1 are survival factors for motoneurons. *J. Neurosci. Res.* (2005). doi:10.1002/jnr.20517
735. Chen, S., Mantei, N., Dong, L. & Schachner, M. Prevention of neuronal cell death by neural adhesion molecules L1 and CHL1. *J. Neurobiol.* (1999). doi:10.1002/(SICI)1097-4695(19990215)38:3<428::AID-NEU10>3.0.CO;2-6
736. Hall, H. & Hubbell, J. A. Matrix-bound sixth Ig-like domain of cell adhesion molecule L1 acts as an angiogenic factor by ligating $\alpha\beta 3$ -integrin and activating VEGF-R2. *Microvasc. Res.* (2004). doi:10.1016/j.mvr.2004.07.001
737. Duczmal, A. *et al.* The L1 adhesion molecule supports alpha v beta 3-mediated migration of human tumor cells and activated T lymphocytes. *Biochem Biophys Res Commun* (1997). doi:10.1006/bbrc.1997.6265
738. Ruppert, M., Aigner, S., Hubbe, M., Yagita, H. & Altevogt, P. The L1 adhesion molecule is a cellular ligand for VLA-5. *J. Cell Biol.* (1995). doi:10.1083/jcb.131.6.1881
739. Voura, E. B., Ramjeeasingh, R. a, Montgomery, a M. & Siu, C. H. Involvement of integrin alpha(v)beta(3) and cell adhesion molecule L1 in transendothelial migration of melanoma cells. *Mol. Biol. Cell* (2001).

740. Montgomery, A. M. P. *et al.* Human neural cell adhesion molecule L1 and rat homologue NILE are ligands for integrin $\alpha_5\beta_1$. *J. Cell Biol.* (1996). doi:10.1083/jcb.132.3.475
741. Schmid, R. S., Midkiff, B. R., Kedar, V. P. & Maness, P. F. Adhesion molecule L1 stimulates neuronal migration through Vav2-Pak1 signaling. *Neuroreport* (2004). doi:00001756-200412220-00021 [pii]
742. Schaefer, A. W. *et al.* Activation of the MAPK signal cascade by the neural cell adhesion molecule L1 requires L1 internalization. *J. Biol. Chem.* (1999). doi:10.1074/jbc.274.53.37965
743. Watanabe, H. *et al.* Phospholipase D2 functions as a downstream signaling molecule of MAP kinase pathway in L1-stimulated neurite outgrowth of cerebellar granule neurons. *J. Neurochem.* (2004). doi:10.1111/j.1471-4159.2004.02308.x
744. Schmid, R. S., Pruitt, W. M. & Maness, P. F. A MAP kinase-signaling pathway mediates neurite outgrowth on L1 and requires Src-dependent endocytosis. *J. Neurosci.* (2000). doi:20/11/4177 [pii]
745. Gavert, N., Ben-Shmuel, A., Raveh, S. & Ben-Ze'ev, A. L1-CAM in cancerous tissues. *Expert Opin. Biol. Ther.* (2008). doi:10.1517/14712598.8.11.1749
746. Kiefel, H. *et al.* L1CAM-integrin interaction induces constitutive NF-kappaB activation in pancreatic adenocarcinoma cells by enhancing IL-1beta expression. *Oncogene* (2010). doi:10.1038/onc.2010.230
747. Gavert, N., Ben-Shmuel, A., Lemmon, V., Brabletz, T. & Ben-Ze'ev, A. Nuclear factor-kappaB signaling and ezrin are essential for L1-mediated metastasis of colon cancer cells. *J. Cell Sci.* (2010). doi:10.1242/jcs.069542
748. Shtutman, M., Levina, E., Ohouo, P., Baig, M. & Roninson, I. B. Cell adhesion molecule L1 disrupts E-cadherin-containing adherens junctions and increases scattering and motility of MCF7 breast carcinoma cells. *Cancer Res.* (2006). doi:10.1158/0008-5472.CAN-06-2106
749. Gavert, N. *et al.* L1, a novel target of β -catenin signaling, transforms cells and is expressed at the invasive front of colon cancers. *J. Cell Biol.* (2005). doi:10.1083/jcb.200408051

750. Teuliere, J. *et al.* Targeted activation of beta-catenin signaling in basal mammary epithelial cells affects mammary development and leads to hyperplasia. *Development* **132**, (2005).
751. Jensen, K. B. & Watt, F. M. Single-cell expression profiling of human epidermal stem and transit-amplifying cells: Lrig1 is a regulator of stem cell quiescence. *Proc. Natl. Acad. Sci.* (2006). doi:10.1073/pnas.0601886103
752. Ikenouchi, J. & Umeda, M. FRMD4A regulates epithelial polarity by connecting Arf6 activation with the PAR complex. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 748–53 (2010).
753. Tepass, U. FERM proteins in animal morphogenesis. *Curr. Opin. Genet. Dev.* **19**, 357–367 (2009).
754. Chishti, A. H. *et al.* The FERM domain: a unique module involved in the linkage of cytoplasmic proteins to the membrane. *Trends Biochem. Sci.* **23**, 281–282 (1998).
755. Terawaki, S. I., Kitano, K., Aoyama, M. & Hakoshima, T. Crystallographic characterization of the radixin FERM domain bound to the cytoplasmic tail of membrane-type 1 matrix metalloproteinase (MT1-MMP). *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* (2008). doi:10.1107/S1744309108026869
756. Fiévet, B., Louvard, D. & Arpin, M. ERM proteins in epithelial cell organization and functions. *Biochim. Biophys. Acta - Mol. Cell Res.* **1773**, 653–660 (2007).
757. Clucas, J. & Valderrama, F. ERM proteins in cancer progression. *J. Cell Sci.* **127**, (2014).
758. Tanentzapf, G. & Brown, N. H. An interaction between integrin and the talin FERM domain mediates integrin activation but not linkage to the cytoskeleton. *Nat. Cell Biol.* (2006). doi:10.1038/ncb1411
759. Frame, M. C., Patel, H., Serrels, B., Lietha, D. & Eck, M. J. The FERM domain: organizing the structure and function of FAK. *Nat. Rev. Mol. Cell Biol.* (2010). doi:10.1038/nrm2996
760. Papusheva, E. *et al.* Dynamic conformational changes in the {FERM} domain of {FAK} are involved in focal-adhesion behavior during cell spreading and motility. *J. Cell. Sci.* (2009). doi:10.1242/jcs.028738

761. Fan, Y. *et al.* Increased expression of FERM domain-containing 4A protein is closely associated with the development of rectal cancer. *Exp. Ther. Med.* **11**, 421–426 (2016).
762. Goldie, S. J. *et al.* FRMD4A upregulation in human squamous cell carcinoma promotes tumor growth and metastasis and is associated with poor prognosis. *Cancer Res.* **72**, 3424–36 (2012).
763. Durkin, M. E., Ullmannova, V., Guan, M. & Popescu, N. C. Deleted in liver cancer 3 (DLC-3), a novel Rho GTPase-activating protein, is downregulated in cancer and inhibits tumor cell growth. *Oncogene* (2007). doi:10.1038/sj.onc.1210244
764. Sjöblom, T., Jones, S., Wood, L. & Al., E. The Consensus Coding Sequences of Human Breast and Colorectal Cancers. *AAAS Sci.* (2006). doi:10.1126/science.1133427
765. Kawai, K., Kiyota, M., Seike, J., Deki, Y. & Yagisawa, H. START-GAP3/DLC3 is a GAP for RhoA and Cdc42 and is localized in focal adhesions regulating cell morphology. *Biochem. Biophys. Res. Commun.* **364**, 783–789 (2007).
766. Holeiter, G. *et al.* The RhoGAP protein Deleted in Liver Cancer 3 (DLC3) is essential for adherens junctions integrity. *Oncogenesis* (2012). doi:10.1038/oncsis.2012.13
767. Zhang, Y., Chen, K., Guo, L. & Wu, C. Characterization of PINCH-2, a new focal adhesion protein that regulates the PINCH-1-ILK interaction, cell spreading, and migration. *J. Biol. Chem.* **277**, 38328–38 (2002).
768. Zhang, Y., Guo, L., Chen, K. & Wu, C. A critical role of the PINCH-integrin-linked kinase interaction in the regulation of cell shape change and migration. *J Biol Chem* (2002). doi:10.1074/jbc.M108257200
769. Wu, C. Integrin-linked kinase and PINCH: partners in regulation of cell-extracellular matrix interaction and signal transduction. *J. Cell Sci.* (1999).
770. Tu, Y., Li, F., Goicoechea, S. & Wu, C. The LIM-only protein PINCH directly interacts with integrin-linked kinase and is recruited to integrin-rich sites in spreading cells. *Mol. Cell. Biol.* (1999). doi:10.1128/MCB.19.3.2425

771. Rooney, N. & Streuli, C. H. How integrins control mammary epithelial differentiation: A possible role for the ILK-PINCH-Parvin complex. *FEBS Lett.* **585**, 1663–1672 (2011).
772. Wu-Baer, F., Ludwig, T. & Baer, R. The UBXN1 protein associates with autoubiquitinated forms of the BRCA1 tumor suppressor and inhibits its enzymatic function. *Mol. Cell. Biol.* **30**, 2787–98 (2010).
773. Wang, Y.-B. *et al.* Ubiquitin-associated domain-containing ubiquitin regulatory X (UBX) protein UBXN1 is a negative regulator of nuclear factor κ B (NF- κ B) signaling. *J. Biol. Chem.* **290**, 10395–405 (2015).
774. Lockwood, W. W., Chandel, S. K., Stewart, G. L., Erdjument-Bromage, H. & Beverly, L. J. The Novel Ubiquitin Ligase Complex, SCFFbxw4, Interacts with the COP9 Signalosome in an F-Box Dependent Manner, Is Mutated, Lost and Under-Expressed in Human Cancers. *PLoS One* **8**, e63610 (2013).
775. Randle, S. J. & Laman, H. F-box protein interactions with the hallmark pathways in cancer. *Semin. Cancer Biol.* **36**, 3–17 (2016).
776. Carpenter, A. E. & Sabatini, D. M. Systematic genome-wide screens of gene function. *Nat Rev Genet* **5**, (2004).
777. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* **47**, (2015).
778. Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. Is my network module preserved and reproducible? *PLoS Comput Biol* **7**, (2011).
779. Zhao, W. *et al.* Weighted gene coexpression network analysis: state of the art. *J Biopharm Stat* (2010). doi:10.1080/10543400903572753
780. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* (2008). doi:10.1186/1471-2105-9-559
781. Botía, J. A. *et al.* An additional k-means clustering step improves the biological features of WGCNA gene co-expression networks. *BMC Syst. Biol.* (2017). doi:10.1186/s12918-017-0420-6

782. Piccolo, S. R. *et al.* Integrative analyses reveal signaling pathways underlying familial breast cancer susceptibility. *Mol. Syst. Biol.* (2016). doi:10.15252/msb.20156506
783. Liu, R., Guo, C.-X. & Zhou, H.-H. Network-based approach to identify prognostic biomarkers for estrogen receptor–positive breast cancer treatment with tamoxifen. *Cancer Biol. Ther.* **16**, 317–324 (2015).
784. Liu, R. *et al.* Correlating transcriptional networks with pathological complete response following neoadjuvant chemotherapy for breast cancer. *Breast Cancer Res. Treat.* (2015). doi:10.1007/s10549-015-3428-x
785. Diermeier, S. D. *et al.* Identification and characterization of long non-coding RNAs as targets of mammary tumor cell proliferation and migration. (2016). doi:10.1101/036418
786. Pendse, S. N. *et al.* Information-dependent enrichment analysis reveals time-dependent transcriptional regulation of the estrogen pathway of toxicity. *Arch. Toxicol.* **91**, 1749–1762 (2017).
787. Hu, W., Yang, Y., Li, X. & Zheng, S. Pan-organ transcriptome variation across 21 cancer types. *Oncotarget* (2016). doi:10.18632/oncotarget.14303
788. Bainer, R. *et al.* Gene expression in local stroma reflects breast tumor states and predicts patient outcome. *Sci. Rep.* (2016). doi:10.1038/srep39240
789. Pouladi, N., Cowper-Sallari, R. & Moore, J. H. Combining functional genomics strategies identifies modular heterogeneity of breast cancer intrinsic subtypes. *BioData Min.* (2014). doi:10.1186/1756-0381-7-27
790. Yeung, N., Cline, M. S., Kuchinsky, A., Smoot, M. E. & Bader, G. D. Exploring biological networks with cytoscape software. *Current Protocols in Bioinformatics* (2008). doi:10.1002/0471250953.bi0813s23
791. Bader, G. & Hogue, C. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* (2003). doi:10.1186/1471-2105-4-2
792. Su, G., Morris, J. H., Demchak, B. & Bader, G. D. Biological Network Exploration with Cytoscape 3. *Curr. Protoc. Bioinforma.* (2014). doi:10.1002/0471250953.bi0813s47

793. Gao, B. *et al.* A multidimensional integration analysis reveals potential bridging targets in the process of colorectal cancer liver metastasis. *PLoS One* **12**, e0178760 (2017).
794. Yang, B., Chen, Z., Huang, Y., Han, G. & Li, W. Identification of potential biomarkers and analysis of prognostic values in head and neck squamous cell carcinoma by bioinformatics analysis. *Onco. Targets. Ther.* **10**, 2315–2321 (2017).
795. Russo, L. C., Araujo, C. B., Iwai, L. K., Ferro, E. S. & Forti, F. L. A Cyclin D2-derived peptide acts on specific cell cycle phases by activating ERK1/2 to cause the death of breast cancer cells. *J. Proteomics* (2017). doi:10.1016/j.jprot.2016.06.028
796. Chen, I.-H. *et al.* Phosphoproteins in extracellular vesicles as candidate markers for breast cancer. *Proc. Natl. Acad. Sci.* (2017). doi:10.1073/pnas.1618088114
797. Hu, Z. *et al.* Genome co-amplification upregulates a mitotic gene network activity that predicts outcome and response to mitotic protein inhibitors in breast cancer. *Breast Cancer Res.* (2016). doi:10.1186/s13058-016-0728-y
798. Crowley, M. R., Head, K. L., Kwiatkowski, D. J., Asch, H. L. & Asch, B. B. The Mouse Mammary Gland Requires the Actin-Binding Protein Gelsolin for Proper Ductal Morphogenesis. *Dev. Biol.* (2000). doi:10.1006/dbio.2000.9844
799. Keely, P. J. Mechanisms by which the extracellular matrix and integrin signaling act to regulate the switch between tumor suppression and tumor promotion. *J. Mammary Gland Biol. Neoplasia* **16**, 205–19 (2011).
800. Weymouth, N., Shi, Z. & Rockey, D. C. Smooth muscle α actin is specifically required for the maintenance of lactation. *Dev. Biol.* **363**, 1–14 (2012).
801. Deugnier, M.-A., Moiseyeva, E. P., Thiery, J. P. & Glukhova, M. Myoepithelial cell differentiation in the developing mammary gland: Progressive acquisition of smooth muscle phenotype. *Dev. Dyn.* **204**, 107–117 (1995).
802. Griffiths, M. & Slater, E. The significance of striated muscle in the mammary glands of marsupials. *J. Anat.* **156**, 141–56 (1988).
803. Rodgers, L. S. & Fanning, A. S. Regulation of epithelial permeability by the actin cytoskeleton. *Cytoskeleton* (2011). doi:10.1002/cm.20547

804. Schmidt, A. & Hall, M. N. Signaling to the actin cytoskeleton. *Annu. Rev. Cell Dev. Biol.* (1998). doi:10.1146/annurev.cellbio.14.1.305
805. Green, K. & Jones, J. Desmosomes and hemidesmosomes: Structure and function of molecular components. *Faseb J.* (1996).
806. Tang, D. D. & Gerlach, B. D. The roles and regulation of the actin cytoskeleton, intermediate filaments and microtubules in smooth muscle cell migration. *Respir. Res.* **18**, 54 (2017).
807. Shankar, J. & Nabi, I. R. Actin cytoskeleton regulation of epithelial mesenchymal transition in metastatic cancer cells. *PLoS One* **10**, e0119954 (2015).
808. DeMali, K. A., Wennerberg, K. & Burridge, K. Integrin signaling to the actin cytoskeleton. *Curr. Opin. Cell Biol.* **15**, 572–582 (2003).
809. Rooney, C. *et al.* The Rac activator STEF (Tiam2) regulates cell migration by microtubule-mediated focal adhesion disassembly. *EMBO Rep.* **11**, 292–298 (2010).
810. Defilippi, P. *et al.* Actin cytoskeleton organization in response to integrin-mediated adhesion. *Microsc. Res. Tech.* **47**, 67–78 (1999).
811. Rodgers, L. S. & Fanning, A. S. Regulation of epithelial permeability by the actin cytoskeleton. *Cytoskeleton* **68**, 653–660 (2011).
812. Stingl, J., Eaves, C. J., Zandieh, I. & Emerman, J. T. Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue. *Breast Cancer Res. Treat.* **67**, 93–109 (2001).
813. Raouf, A. *et al.* Transcriptome Analysis of the Normal Human Mammary Cell Commitment and Differentiation Process. *Cell Stem Cell* **3**, 109–118 (2008).
814. Rodgers, L. S. & Fanning, A. S. Regulation of epithelial permeability by the actin cytoskeleton. *Cytoskelet. (Hoboken)*. (2011). doi:10.1002/cm.20547
815. Mayya, V. & Dustin, M. L. Actin cytoskeleton and the dynamics of immunological synapse. in *Actin-based Motility: Cellular, Molecular and Physical Aspects* (2010). doi:10.1007/978-90-481-9301-1_5

816. Maruyama, T., Nara, K., Yoshikawa, H. & Suzuki, N. Txk, a member of the non-receptor tyrosine kinase of the Tec family, forms a complex with poly(ADP-ribose) polymerase 1 and elongation factor 1alpha and regulates interferon-gamma gene transcription in Th1 cells. *Clin. Exp. Immunol.* **147**, 164–75 (2007).
817. Jatav, P. *et al.* Identification of internal control genes in milk-derived mammary epithelial cells during lactation cycle of Indian zebu cow. *Anim. Sci. J.* **87**, 344–353 (2016).
818. Scaggiante, B. *et al.* Dissecting the expression of EEF1A1/2 genes in human prostate cancer cells: the potential of EEF1A2 as a hallmark for prostate transformation and progression. *Br. J. Cancer* **106**, 166–173 (2012).
819. Greenberg, A. S. & Obin, M. S. Obesity and the role of adipose tissue in inflammation and metabolism. *Am. J. Clin. Nutr.* **83**, 461S–465S (2006).
820. Couldrey, C. *et al.* Adipose tissue: A vital in vivo role in mammary gland development but not differentiation. *Dev. Dyn.* **223**, 459–468 (2002).
821. Hovey, R. C. & Aimo, L. Diverse and active roles for adipocytes during mammary gland growth and function. *J. Mammary Gland Biol. Neoplasia* **15**, 279–90 (2010).
822. Zhu, W. & Nelson, C. M. Adipose and mammary epithelial tissue engineering. *Biomatter* **3**, (2013).
823. Landskroner-Eiger, S., Park, J., Israel, D., Pollard, J. W. & Scherer, P. E. Morphogenesis of the developing mammary gland: stage-dependent impact of adipocytes. *Dev. Biol.* **344**, 968–78 (2010).
824. Laurent, V. *et al.* Mammary adipocytes stimulate breast cancer invasion through metabolic remodeling of tumor cells. *Nat Commun* **7**, (2017).
825. Spiegelman, B. M., Hu, E., Kim, J. B. & Brun, R. PPAR gamma and the control of adipogenesis. *J. Control. Release* (1997). doi:S0300-9084(97)81500-3 [pii]
826. Andarawewa, K. L. *et al.* Stromelysin-3 is a potent negative regulator of adipogenesis participating to cancer cell-adipocyte interaction/crosstalk at the tumor invasive front. *Cancer Res.* (2005). doi:10.1158/0008-5472.CAN-05-1231

827. Wu, Z., Puigserver, P. & Spiegelman, B. M. Transcriptional activation of adipogenesis. *Current Opinion in Cell Biology* (1999). doi:10.1016/S0955-0674(99)00037-X
828. Walenta, E. *et al.* α/β -Hydrolase Domain Containing Protein 15 (ABHD15) – an Adipogenic Protein Protecting from Apoptosis. *PLoS One* **8**, e79134 (2013).
829. Prokesch, A. *et al.* Molecular Aspects of Adipoepithelial Transdifferentiation in Mouse Mammary Gland. *Stem Cells* **32**, 2756–2766 (2014).
830. Lin, Z. *et al.* Adenosine A1 receptor, a target and regulator of estrogen receptor α action, mediates the proliferative effects of estradiol in breast cancer. *Oncogene* **29**, 1114–1122 (2010).
831. Mirza, A. *et al.* RNA interference targeting of A1 receptor-overexpressing breast carcinoma cells leads to diminished rates of cell proliferation and induction of apoptosis. *Cancer Biol. Ther.* **4**, 1355–60 (2005).
832. Hansen, C. *et al.* Wnt-5a-induced phosphorylation of DARPP-32 inhibits breast cancer cell migration in a CREB-dependent manner. *J Biol Chem* **284**, (2009).
833. Christenson, J. L. & Kane, S. E. Darpp-32 and t-Darpp are differentially expressed in normal and malignant mouse mammary tissue. *Mol. Cancer* **13**, 192 (2014).
834. Vangamudi, B. *et al.* t-DARPP regulates phosphatidylinositol-3-kinase-dependent cell growth in breast cancer. *Mol Cancer* **9**, (2010).
835. Yger, M. & Girault, J. A. DARPP-32, jack of all trades {...} master of which? *Front Behav Neurosci* **5**, (2011).
836. Borowsky, M. L. & Hynes, R. O. Layilin, A Novel Talin-binding Transmembrane Protein Homologous with C-type Lectins, is Localized in Membrane Ruffles. *J. Cell Biol.* **143**, (1998).
837. Ahn, R. S. *et al.* Transcriptional landscape of epithelial and immune cell populations revealed through FACS-seq of healthy human skin. *Sci. Rep.* **7**, 1343 (2017).
838. Kikkawa, Y. *et al.* The lutheran/basal cell adhesion molecule promotes tumor cell migration by modulating integrin-mediated cell attachment to laminin-511 protein. *J. Biol. Chem.* (2013). doi:10.1074/jbcM113.486456

839. Akiyama, H. *et al.* The FBI1/Akirin2 Target Gene, BCAM, Acts as a Suppressive Oncogene. *PLoS One* **8**, e78716 (2013).
840. Kannan, K. *et al.* Recurrent BCAM-AKT2 fusion gene leads to a constitutively activated AKT2 fusion kinase in high-grade serous ovarian carcinoma. doi:10.1073/pnas.1501735112
841. Gosling, J. *et al.* Cutting edge: identification of a novel chemokine receptor that binds dendritic cell- and T cell-active chemokines including ELC, SLC, and TECK. *J. Immunol.* **164**, 2851–6 (2000).
842. Vinet, J. *et al.* Inhibition of CXCR3-mediated chemotaxis by the human chemokine receptor-like protein CCX-CKR. *Br. J. Pharmacol.* **168**, 1375–1387 (2013).
843. Harata-Lee, Y. *et al.* The atypical chemokine receptor CCX-CKR regulates metastasis of mammary carcinoma via an effect on EMT. *Immunol. Cell Biol.* **92**, 815–824 (2014).
844. Comerford, I., Milasta, S., Morrow, V., Milligan, G. & Nibbs, R. The chemokine receptor CCX-CKR mediates effective scavenging of CCL19 in vitro. *Eur. J. Immunol.* **36**, 1904–1916 (2006).
845. Cassel, T. N. & Nord, M. C/EBP transcription factors in the lung epithelium. *Am. J. Physiol. - Lung Cell. Mol. Physiol.* **285**, (2003).
846. Centrella, M., Christakos, S. & McCarthy, T. L. Skeletal hormones and the C/EBP and Runx transcription factors: interactions that integrate and redefine gene expression. *Gene* **342**, 13–24 (2004).
847. Darlington, G. J., Ross, S. E. & MacDougald, O. A. The Role of C/EBP Genes in Adipocyte Differentiation. *J. Biol. Chem.* **273**, 30057–30060 (1998).
848. Friedman, A. D. Transcriptional control of granulocyte and monocyte development. *Oncogene* **26**, 6816–6828 (2007).
849. Zhang, P. *et al.* Induction of granulocytic differentiation by 2 pathways. *Blood* **99**, 4406–12 (2002).
850. Friedman, A. D. *et al.* Regulation of granulocyte and monocyte differentiation by CCAAT/enhancer binding protein alpha. *Blood Cells. Mol. Dis.* **31**, 338–41

851. Heath, V. *et al.* C/EBP α deficiency results in hyperproliferation of hematopoietic progenitor cells and disrupts macrophage development in vitro and in vivo. *Blood* **104**, (2004).
852. Zhang, D. E. *et al.* CCAAT enhancer-binding protein (C/EBP) and AML1 (CBF α 2) synergistically activate the macrophage colony-stimulating factor receptor promoter. *Mol. Cell. Biol.* **16**, 1231–40 (1996).
853. Lee, B. *et al.* C/EBP α regulates macrophage activation and systemic metabolism. *Am. J. Physiol. - Endocrinol. Metab.* **306**, (2014).
854. Kimura, T. *et al.* Hypoglycemia-associated hyperammonemia caused by impaired expression of ornithine cycle enzyme genes in C/EBP α knockout mice. *J. Biol. Chem.* **273**, 27505–10 (1998).
855. Darlington, G. J., Ross, S. E. & MacDougald, O. A. The role of C/EBP genes in adipocyte differentiation. *J. Biol. Chem.* **273**, 30057–60 (1998).
856. Robinson, G. W., Johnson, P. F., Hennighausen, L. & Sterneck, E. The C/EBP β transcription factor regulates epithelial cell proliferation and differentiation in the mammary gland. *Genes Dev.* (1998). doi:10.1101/gad.12.12.1907
857. Seagroves, T. N. *et al.* C/EBP β , but not C/EBP α , is essential for ductal morphogenesis, lobuloalveolar proliferation, and functional differentiation in the mouse mammary gland. *Genes Dev.* (1998). doi:10.1101/gad.12.12.1917
858. Sabatakos, G., Davies, G. E., Grosse, M., Cryer, A. & Ramji, D. P. Expression of the genes encoding CCAAT-enhancer binding protein isoforms in the mouse mammary gland during lactation and involution. *Biochem. J.* (1998).
859. van der Wal, a C., Das, P. K., Tigges, A. J. & Becker, A. E. Macrophage differentiation in atherosclerosis. An in situ immunohistochemical analysis in humans. *Am. J. Pathol.* (1992).
860. Ross, R. The pathogenesis of atherosclerosis: a perspective for the 1990s. *Nature* (1993). doi:10.1038/362801a0
861. Chang, M. Y., Olin, K. L., Tsoi, C., Wight, T. N. & Chait, A. Human monocyte-derived macrophages secrete two forms of proteoglycan-macrophage colony-stimulating factor that differ in their ability to bind low density lipoproteins. *J. Biol. Chem.* **273**, 15985–92 (1998).

862. Jacob, S. S. & Sudhakaran, P. R. Monocyte–macrophage differentiation in three dimensional collagen lattice. *Biochim. Biophys. Acta - Mol. Cell Res.* **1540**, 50–58 (2001).
863. Williams, C. B., Yeh, E. S. & Soloff, A. C. Tumor-associated macrophages: unwitting accomplices in breast cancer malignancy. *npj Breast Cancer* (2016). doi:10.1038/npjbcancer.2015.25
864. Chang, M. Y. *et al.* Monocyte-to-macrophage differentiation: synthesis and secretion of a complex extracellular matrix. *J. Biol. Chem.* **287**, 14122–35 (2012).
865. Liu, H., Shi, B., Huang, C. C., Eksarko, P. & Pope, R. M. Transcriptional diversity during monocyte to macrophage differentiation. *Immunol. Lett.* (2008). doi:10.1016/j.imlet.2007.12.012
866. Geering, K. *et al.* FXYD proteins: new tissue- and isoform-specific regulators of Na,K-ATPase. *Ann. N. Y. Acad. Sci.* **986**, 388–94 (2003).
867. Liu, C.-C., Teh, R., Mozar, C. A., Baxter, R. C. & Rasmussen, H. H. Silencing overexpression of FXYD3 protein in breast cancer cells amplifies effects of doxorubicin and γ -radiation on Na⁺/K⁺-ATPase and cell survival. *Breast Cancer Res. Treat.* **155**, 203–213 (2016).
868. Yamamoto, H. *et al.* FXYD3 protein involved in tumor cell proliferation is overproduced in human breast cancer tissues. *Biol. Pharm. Bull.* **32**, 1148–54 (2009).
869. Li, Y. *et al.* Expression and clinical significance of FXYD3 in endometrial cancer. *Oncol. Lett.* **8**, 517–522 (2014).
870. Kiyamova, R. *et al.* Preliminary study of thyroid and colon cancers-associated antigens and their cognate autoantibodies as potential cancer biomarkers. *Biomarkers* **17**, 362–371 (2012).
871. Seltsmann, K. *et al.* Keratins mediate localization of hemidesmosomes and repress cell motility. *J. Invest. Dermatol.* **133**, 181–90 (2013).
872. Wang, C.-C., Bajikar, S. S., Jamal, L., Atkins, K. A. & Janes, K. A. A time- and matrix-dependent TGFBR3-JUND-KRT5 regulatory circuit in single breast epithelial cells and basal-like premalignancies. *Nat. Cell Biol.* **16**, 345–56 (2014).

873. Yoon, S.-O., Shin, S. & Mercurio, A. M. Hypoxia Stimulates Carcinoma Invasion by Stabilizing Microtubules and Promoting the Rab11 Trafficking of the $\alpha 6\beta 4$ Integrin. *Cancer Res.* **65**, 2761–2769 (2005).
874. Seachrist, J. L. & Ferguson, S. S. G. Regulation of G protein-coupled receptor endocytosis and trafficking by Rab GTPases. *Life Sci.* **74**, 225–35 (2003).
875. Hoekstra, D., Tyteca, D. & van IJzendoorn, S. C. D. The subapical compartment: a traffic center in membrane polarity development. *J. Cell Sci.* **117**, 2183–2192 (2004).
876. Lapierre, L. A. *et al.* Myosin vb is associated with plasma membrane recycling systems. *Mol. Biol. Cell* **12**, 1843–57 (2001).
877. Wang, X., Kumar, R., Navarre, J., Casanova, J. E. & Goldenring, J. R. Regulation of Vesicle Trafficking in Madin-Darby Canine Kidney Cells by Rab11a and Rab25. *J. Biol. Chem.* **275**, 29138–29146 (2000).
878. Calhoun, B. C. & Goldenring, J. R. Two Rab proteins, vesicle-associated membrane protein 2 (VAMP-2) and secretory carrier membrane proteins (SCAMPs), are present on immunoisolated parietal cell tubulovesicles. *Biochem. J.* **325** (Pt 2), 559–64 (1997).
879. Casanova, J. E. *et al.* Association of Rab25 and Rab11a with the apical recycling system of polarized Madin-Darby canine kidney cells. *Mol. Biol. Cell* **10**, 47–61 (1999).
880. Agarwal, R., Jurisica, I., Mills, G. B. & Cheng, K. W. The Emerging Role of the RAB25 Small GTPase in Cancer. *Traffic* **10**, 1561–1568 (2009).
881. Cheng, J.-M. *et al.* Tumor suppressor function of Rab25 in triple-negative breast cancer. *Int. J. Cancer* **126**, NA-NA (2010).
882. Yoshimine, Y. *et al.* Hepatic expression of the Sptlc3 subunit of serine palmitoyltransferase is associated with the development of hepatocellular carcinoma in a mouse model of nonalcoholic steatohepatitis. *Oncol. Rep.* **33**, 1657–66 (2015).
883. Stepan, L. P. *et al.* Expression of Trop2 Cell Surface Glycoprotein in Normal and Tumor Tissues. *J. Histochem. Cytochem.* **59**, 701–710 (2011).

884. Nakatsukasa, M. *et al.* Tumor-Associated Calcium Signal Transducer 2 Is Required for the Proper Subcellular Localization of Claudin 1 and 7. *Am. J. Pathol.* **177**, 1344–1355 (2010).
885. Stoyanova, T. *et al.* Regulated proteolysis of Trop2 drives epithelial hyperplasia and stem cell self-renewal via β -catenin signaling. *Genes Dev.* **26**, 2271–85 (2012).
886. The switch-like behavior of TACSTD2 in breast cancer and its... - Figure 7 of 8. Available at: https://www.researchgate.net/figure/51770424_fig7_The-switch-like-behavior-of-TACSTD2-in-breast-cancer-and-its-regulation-A-The-scatter. (Accessed: 29th July 2017)
887. Huang, H. *et al.* Aberrant Expression of Novel and Previously Described Cell Membrane Markers in Human Breast Cancer Cell Lines and Tumors. *Clin. Cancer Res.* **11**, 4357–4364 (2005).
888. Wang, J. *et al.* Loss of Trop2 Promotes Carcinogenesis and Features of Epithelial to Mesenchymal Transition in Squamous Cell Carcinoma. *Mol. Cancer Res.* **9**, 1686–1695 (2011).
889. Li, X. *et al.* TROP2 promotes proliferation, migration and metastasis of gallbladder cancer cells by regulating PI3K/AKT pathway and inducing EMT. *Oncotarget* **8**, 47052–47063 (2017).
890. Winter, M. J., Cirulli, V., Briare-de Bruijn, I. H. & Litvinov, S. V. Cadherins are regulated by Ep-CAM via phosphatidylinositol-3 kinase. *Mol. Cell. Biochem.* **302**, 19–26 (2007).
891. Trerotola, M., Li, J., Alberti, S. & Languino, L. R. Trop-2 inhibits prostate cancer cell adhesion to fibronectin through the β 1 integrin-RACK1 axis. *J. Cell. Physiol.* **227**, 3670–3677 (2012).
892. Masuda, Y., Takahashi, H. & Hatakeyama, S. TRIM29 regulates the p63-mediated pathway in cervical cancer cells. *Biochim. Biophys. Acta - Mol. Cell Res.* **1853**, 2296–2305 (2015).
893. Liu, J. *et al.* TRIM29 functions as a tumor suppressor in nontumorigenic breast cells and invasive ER+ breast cancer. *Am. J. Pathol.* **180**, 839–47 (2012).

894. Ai, L. *et al.* TRIM29 suppresses TWIST1 and invasive breast cancer behavior. (2017). doi:10.1158/0008-5472.CAN-13-3579
895. Hatakeyama, S. Early evidence for the role of TRIM29 in multiple cancer models. *Expert Opin. Ther. Targets* **20**, 767–770 (2016).
896. Wang, L. *et al.* ATDC induces an invasive switch in KRAS-induced pancreatic tumorigenesis. *Genes Dev.* **29**, 171–83 (2015).
897. Zeng, S.-X. *et al.* High expression of TRIM29^{Δ1/2}(ATDC) contributes to poor prognosis and tumor metastasis by inducing epithelial-mesenchymal transition in osteosarcoma. *Oncol. Rep.* (2017). doi:10.3892/or.2017.5842
898. Zhou, X.-M. *et al.* Upregulated TRIM29 promotes proliferation and metastasis of nasopharyngeal carcinoma via PTEN/AKT/mTOR signal pathway. *Oncotarget* **7**, 13634–13650 (2016).
899. Debnath, J., Muthuswamy, S. K. & Brugge, J. S. Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods* **30**, 256–68 (2003).
900. Amaral, J. D., Xavier, J. M., Steer, C. J. & Rodrigues, C. M. The Role of p53 in Apoptosis. *Discov. Med.* **9**, 145–152 (2010).
901. fibroblast | anatomy | Britannica.com. Available at: <https://www.britannica.com/science/fibroblast>. (Accessed: 11th September 2017)
902. Kunz-Schughart, L. A. & Knuechel, R. Tumor-associated fibroblasts (part I): Active stromal participants in tumor development and progression? *Histol. Histopathol.* **17**, 599–621 (2002).
903. Oh, H.-Y. *et al.* Characteristics of primary and immortalized fibroblast cells derived from the miniature and domestic pigs. *BMC Cell Biol.* **8**, 20 (2007).
904. Gudjonsson, T. *et al.* Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties. *Genes Dev* **16**, (2002).
905. Gao, Q. *et al.* Mutant p53-induced immortalization of primary human mammary epithelial cells. *Cancer Res* **56**, (1996).

906. Fuchs, Y. & Steller, H. Programmed cell death in animal development and disease. *Cell* **147**, 742–58 (2011).
907. Gupta, S. Molecular steps of death receptor and mitochondrial pathways of apoptosis. *Life Sci.* **69**, 2957–2964 (2001).
908. Faull, R. J., Ginsberg², M. H. & Ginsberg, M. H. Inside-Out Signaling Through Integrins¹. *J. Am. Soc. Nephrol* **7**, 1091–1097
909. Wu, Z., Cho, H., Hampton, G. M. & Theodorescu, D. Cdc6 and cyclin E2 are PTEN-regulated genes associated with human prostate cancer metastasis. *Neoplasia* **11**, 66–76 (2009).
910. Chen, S. *et al.* Mutation analysis of the putative tumor suppression gene PTEN/MMAC1 in sporadic breast cancer. *Breast Cancer Res. Treat.* **55**, 85–89 (1999).
911. Wen, S. *et al.* PTEN controls tumor-induced angiogenesis. *Proc. Natl. Acad. Sci.* **98**, 4622–4627 (2001).
912. Elmore, S. Apoptosis: a review of programmed cell death. *Toxicol. Pathol.* **35**, 495–516 (2007).
913. Fulda, S. Cell death and survival signaling in oncogenesis. *Klin. Padiatr.* **222**, 340–4 (2010).
914. Fulda, S. & Debatin, K.-M. Extrinsic versus intrinsic apoptosis pathways in anticancer chemotherapy. *Oncogene* **25**, 4798–4811 (2006).
915. Roos, W. P. & Kaina, B. DNA damage-induced cell death: from specific DNA lesions to the DNA damage response and apoptosis. *Cancer Lett.* **332**, 237–48 (2013).
916. Hongmei, Z. Extrinsic and Intrinsic Apoptosis Signal Pathway Review. in *Apoptosis and Medicine* (InTech, 2012). doi:10.5772/50129
917. Fulda, S. & Debatin, K.-M. Targeting Apoptosis Pathways in Cancer Therapy. *Curr. Cancer Drug Targets* **4**, 569–576 (2004).

918. Degterev, A., Boyce, M. & Yuan, J. A decade of caspases. *Oncogene* **22**, 8543–8567 (2003).
919. Brown, J. M. & Wilson, G. Apoptosis genes and resistance to cancer therapy: what does the experimental and clinical data tell us? *Cancer Biol. Ther.* **2**, 477–90
920. Walczak, H. & Krammer, P. H. The CD95 (APO-1/Fas) and the TRAIL (APO-2L) Apoptosis Systems. *Exp. Cell Res.* **256**, 58–66 (2000).
921. Walczak, H., Bouchon, A., Stahl, H. & Krammer, P. H. Tumor necrosis factor-related apoptosis-inducing ligand retains its apoptosis-inducing capacity on Bcl-2- or Bcl-xL-overexpressing chemotherapy-resistant tumor cells. *Cancer Res.* **60**, 3051–7 (2000).
922. Walczak, H. *et al.* Tumoricidal activity of tumor necrosis factor-related apoptosis-inducing ligand in vivo. *Nat. Med.* **5**, 157–63 (1999).
923. Fridman, J. S. & Lowe, S. W. Control of apoptosis by p53. *Oncogene* **22**, 9030–9040 (2003).
924. Lowe, S. W., Ruley, H. E., Jacks, T. & Housman, D. E. p53-dependent apoptosis modulates the cytotoxicity of anticancer agents. *Cell* **74**, 957–67 (1993).
925. Johnstone, R. W., Ruefli, A. A. & Lowe, S. W. Apoptosis: a link between cancer genetics and chemotherapy. *Cell* **108**, 153–64 (2002).
926. Wallace-Brodeur, R. R. & Lowe, S. W. Clinical implications of p53 mutations. *Cell. Mol. Life Sci. C.* **55**, 64–75 (1999).
927. Schuler, M., Bossy-Wetzel, E., Goldstein, J. C., Fitzgerald, P. & Green, D. R. p53 induces apoptosis by caspase activation through mitochondrial cytochrome c release. *J. Biol. Chem.* **275**, 7337–42 (2000).
928. Bouillet, P. & Strasser, A. BH3-only proteins — evolutionarily conserved proapoptotic Bcl-2 family members essential for initiating programmed cell death. *J. Cell Sci.* **115**, (2002).
929. Haupt, S., Berger, M., Goldberg, Z. & Haupt, Y. Apoptosis - the p53 network. *J. Cell Sci.* **116**, (2003).

930. Sax, J. K. & El-Deiry, W. S. p53 downstream targets and chemosensitivity. *Cell Death Differ.* **10**, 413–417 (2003).
931. Nakano, K. & Vousden, K. H. PUMA, a novel proapoptotic gene, is induced by p53. *Mol. Cell* **7**, 683–94 (2001).
932. Green, D. R. & Kroemer, G. Cytoplasmic functions of the tumour suppressor p53. *Nature* **458**, 1127–30 (2009).
933. Cory, S. & Adams, J. M. The bcl2 family: regulators of the cellular life-or-death switch. *Nat. Rev. Cancer* **2**, 647–656 (2002).
934. Adams, J. M. & Cory, S. The Bcl-2 apoptotic switch in cancer development and therapy. *Oncogene* **26**, 1324–37 (2007).
935. Kuwana, T. *et al.* Bid, Bax, and lipids cooperate to form supramolecular openings in the outer mitochondrial membrane. *Cell* **111**, 331–42 (2002).
936. Skulachev, V. P. Cytochrome c in the apoptotic and antioxidant cascades. *FEBS Lett.* **423**, 275–80 (1998).
937. Yu, J., Wang, Z., Kinzler, K. W., Vogelstein, B. & Zhang, L. PUMA mediates the apoptotic response to p53 in colorectal cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 1931–6 (2003).
938. Oda, E. *et al.* Noxa, a BH3-Only Member of the Bcl-2 Family and Candidate Mediator of p53-Induced Apoptosis. *Science (80-.).* **288**, (2000).
939. Finucane, D. M., Bossy-Wetzel, E., Waterhouse, N. J., Cotter, T. G. & Green, D. R. Bax-induced caspase activation and apoptosis via cytochrome c release from mitochondria is inhibitable by Bcl-xL. *J. Biol. Chem.* **274**, 2225–33 (1999).
940. Okamoto, A. *et al.* Mutations and altered expression of p16INK4 in human cancer. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 11045–9 (1994).
941. Nagata, S. & Golstein, P. The Fas death factor. *Science (80-.).* **267**, (1995).
942. Muzio, M. Signalling by proteolysis: death receptors induce apoptosis. *Int. J. Clin. Lab. Res.* **28**, 141–7 (1998).

943. Müller, M. *et al.* p53 Activates the CD95 (APO-1/Fas) Gene in Response to DNA Damage by Anticancer Drugs. *J. Exp. Med.* **188**, (1998).
944. Bennett, M. *et al.* Cell Surface Trafficking of Fas: A Rapid Mechanism of p53-Mediated Apoptosis. *Science* (80-.). **282**, (1998).
945. Krammer, P. H. CD95's deadly mission in the immune system. *Nature* **407**, 789–795 (2000).
946. Nagata, S. Fas Ligand-Induced Apoptosis. *Annu. Rev. Genet.* **33**, 29–55 (1999).
947. Ivanov, V. N. *et al.* Cooperation between STAT3 and c-jun suppresses Fas transcription. *Mol. Cell* **7**, 517–28 (2001).
948. Ivanov, V. N., Krasilnikov, M. & Ronai, Z. Regulation of Fas Expression by STAT3 and c-Jun Is Mediated by Phosphatidylinositol 3-Kinase-AKT Signaling. *J. Biol. Chem.* **277**, 4932–4944 (2002).
949. Ivanov, V. N. *et al.* FAP-1 association with Fas (Apo-1) inhibits Fas expression on the cell surface. *Mol. Cell. Biol.* **23**, 3623–35 (2003).
950. Joshi, S. & Ryan, K. M. Autophagy chews Fap to promote apoptosis. *Nat. Cell Biol.* **16**, 23–25 (2013).
951. Yuan, J. & Rozengurt, E. PKD, PKD2, and p38 MAPK mediate Hsp27 serine-82 phosphorylation induced by neurotensin in pancreatic cancer PANC-1 cells. *J. Cell. Biochem.* **103**, 648–662 (2008).
952. Yuan, J. *et al.* Protein kinase D1 mediates NF- B activation induced by cholecystokinin and cholinergic signaling in pancreatic acinar cells. *AJP Gastrointest. Liver Physiol.* **295**, G1190–G1201 (2008).
953. Yuan, J. *et al.* Protein Kinase D Regulates Cell Death Pathways in Experimental Pancreatitis. *Front. Physiol.* **3**, 60 (2012).
954. Shabelnik, M. Y. *et al.* Differential expression of PKD1 and PKD2 in gastric cancer and analysis of PKD1 and PKD2 function in the model system. *Exp. Oncol.* **33**, 206–11 (2011).

955. Kisfalvi, K., Guha, S. & Rozengurt, E. Neurotensin and EGF induce synergistic stimulation of DNA synthesis by increasing the duration of ERK signaling in ductal pancreatic cancer cells. *J. Cell. Physiol.* **202**, 880–890 (2005).
956. Harikumar, K. B. *et al.* A Novel Small-Molecule Inhibitor of Protein Kinase D Blocks Pancreatic Cancer Growth In vitro and In vivo. *Mol. Cancer Ther.* **9**, 1136–1146 (2010).
957. Guha, S., Tanasanvimon, S., Sinnott-Smith, J. & Rozengurt, E. Role of protein kinase D signaling in pancreatic cancer. *Biochem. Pharmacol.* **80**, 1946–1954 (2010).
958. Wu, G. S. *et al.* KILLER/DR5 is a DNA damage-inducible p53-regulated death receptor gene. *Nat. Genet.* **17**, 141–3 (1997).
959. Davies, L., Spiller, D., White, M. R. H., Grierson, I. & Paraoan, L. PERP expression stabilizes active p53 via modulation of p53-MDM2 interaction in uveal melanoma cells. *Cell Death Dis.* **2**, e136 (2011).
960. Attardi, L. D. *et al.* PERP, an apoptosis-associated target of p53, is a novel member of the PMP-22/gas3 family. *Genes Dev.* **14**, 704–18 (2000).
961. Attardi, L. D. *et al.* PERP, an apoptosis-associated target of p53, is a novel member of the PMP-22/gas3 family. *Genes Dev.* **14**, 704–18 (2000).
962. Vivanco, I. & Sawyers, C. L. The phosphatidylinositol 3-Kinase–AKT pathway in human cancer. *Nat. Rev. Cancer* **2**, 489–501 (2002).
963. Gottlieb, T. M., Leal, J. F. M., Seger, R., Taya, Y. & Oren, M. Cross-talk between Akt, p53 and Mdm2: possible implications for the regulation of apoptosis. *Oncogene* **21**, 1299–1303 (2002).
964. Mayo, L. D. & Donner, D. B. The PTEN, Mdm2, p53 tumor suppressor-oncoprotein network. *Trends Biochem. Sci.* **27**, 462–7 (2002).
965. Rivière, J.-B. *et al.* De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes. *Nat. Genet.* **44**, 934–940 (2012).
966. Wu, D. & Pan, W. GSK3: a multifaceted kinase in Wnt signaling. *Trends Biochem. Sci.* **35**, 161–8 (2010).

- 967. Yang, A., Kaghad, M., Caput, D. & McKeon, F. On the shoulders of giants: p63, p73 and the rise of p53. *Trends Genet.* **18**, 90–5 (2002).
- 968. Yang, A. *et al.* p73-deficient mice have neurological, pheromonal and inflammatory defects but lack spontaneous tumours. *Nature* **404**, 99–103 (2000).
- 969. McKeon, F. *et al.* p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. *Nature* **398**, 714–718 (1999).