

# **Systems Biology approaches for identifying molecular mechanisms of epithelial cell phenotype**

By

Ian C. McLean

B.S., Washington State University, 2010

M.S., Colorado State University, 2020

A DISSERTATION

Presented to the Department of Biomedical Engineering

and the Oregon Health & Science University

School of Medicine

in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

February 2025



## Table of Contents

<b>Table of Contents</b>	<b><i>i</i></b>
<b>Table of Figures</b>	<b>1</b>
<b>Acknowledgements</b>	<b>2</b>
<b>Dissertation Abstract</b>	<b>3</b>
<b>1. Chapter I: Introduction</b>	<b>4</b>
<b>1.1 The breast microenvironment</b>	<b>4</b>
1.1.1 The architecture of mammary tissue	4
1.1.2 Mammary ductal epithelial cells and their role in normal and pathological physiology	5
1.1.3 Communication to and from mammary epithelial cells is required for normal development and function	6
1.1.4 Changes to microenvironmental signaling contributes to pathological tissue states	7
1.1.5 Therapeutics targeting microenvironmental signaling can improve outcomes in disease	9
<b>1.2 Cellular communication</b>	<b>10</b>
1.2.1 Modes of cellular communication	10
1.2.2 Classes of microenvironmental ligands	12
1.2.3 Signaling pathways of induced by select microenvironmental ligands	13
<b>1.3 Epithelial cell phenotypes and modulation by microenvironmental ligands</b>	<b>15</b>
1.3.1 Proliferation	15
1.3.2 Migration	17
1.3.3 The epithelial-to-mesenchymal transition	19
1.3.4 Other types of phenotypic modulation	21
<b>1.4 Models and methods for exploring microenvironmental perturbations</b>	<b>22</b>
1.4.1 Experimental models	22
1.4.2 Classical biochemical techniques	24
1.4.3 Unbiased high-throughput assays	25
1.4.4 Systems biology approaches	27
1.4.5 Microenvironmental signals in combination	28
<b>2. Chapter II: A multi-omic analysis of MCF10A cells provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses</b>	<b>29</b>
<b>2.1 Abstract</b>	<b>31</b>
<b>2.2 Introduction</b>	<b>31</b>
<b>2.3 Methods</b>	<b>32</b>
<b>2.4 Results</b>	<b>48</b>
2.4.1 Approach to generate a LINCS ME perturbation dataset	48
2.4.2 Overview of the ligand-induced cellular and molecular responses that comprise the LINCS ME perturbation dataset	52

2.4.3 Identification and analysis of ligand-induced molecular signatures _____	55
2.4.4 Discovery of candidate functional relationships between molecular features _____	61
2.4.5 Investigation of the relationship between molecular modules and cellular phenotype _____	67
2.4.6 Examination of module activity to elucidate the molecular basis of ligand-induced phenotypic responses _____	70
<b>2.5 Discussion _____</b>	<b>72</b>
<b>2.6 Additional Information _____</b>	<b>73</b>
2.6.1 Data Availability _____	73
2.6.2 Code Availability _____	74
2.6.3 Contributions _____	74
2.6.4 Acknowledgements and funding sources _____	74
2.6.5 Competing Interests _____	74
2.6.6 Supplemental Data _____	74
<b>3. Chapter III: The cytokine Oncostatin M induces HIF1A dependent collective cell migration in mammary epithelial cells _____</b>	<b>75</b>
<b>3.1 Abstract _____</b>	<b>75</b>
<b>3.2 Introduction _____</b>	<b>75</b>
<b>3.3 Methods _____</b>	<b>77</b>
<b>3.4 Results _____</b>	<b>80</b>
3.4.1 OSM treatment initiates MCF10A epithelial collective cell migration _____	80
3.4.2 Comparative network analysis identifies OSM-specific molecular regulators _____	82
3.4.3 Functional Validation Reveals OSM-Specific Phenotypic Regulators _____	85
3.4.4 HIF1A as a Key Mediator of OSM-Induced CCM _____	89
<b>3.5 Discussion _____</b>	<b>93</b>
<b>3.6 Additional Information _____</b>	<b>94</b>
4.6.1 Data Availability _____	94
4.6.2 Contributions _____	94
2.6.4 Acknowledgements and funding sources _____	95
2.6.5 Competing Interests _____	95
2.6.6 Supplemental Data _____	95
<b>4. Chapter IV: Integrative Assessment of Ligand Combinations Reveals Synergistic Molecular Mechanisms Driving Complex Phenotypic Responses _____</b>	<b>95</b>
<b>4.1 Abstract _____</b>	<b>96</b>
<b>4.2 Introduction _____</b>	<b>96</b>
<b>4.3 Methods _____</b>	<b>97</b>
<b>4.4 Results _____</b>	<b>101</b>
4.4.1 EGF, OSM and TGFB treatment combinations induce emergent phenotypic responses _____	101
4.4.2 Identification of molecular programs induced by single and combination ligand treatments _____	103
4.4.3 Combination treatments result in specific synergistic transcriptional programs _____	107
4.4.4 Partial Least Squares Regression uncovers transcriptional signatures driving cellular phenotype _____	110

4.4.5 Synergistic transcriptional upregulation of CXCR2 chemotactic signaling molecules via CREB activation promotes increased cell motility	114
<b>4.5 Discussion</b>	<b>118</b>
<b>4.6 Additional Information</b>	<b>121</b>
4.6.1 Data Availability	121
4.6.2 Contributions	121
2.6.4 Acknowledgements and funding sources	121
2.6.5 Competing Interests	121
2.6.6 Supplemental Data	121
<b>5. Chapter V: Conclusions and future work</b>	<b>122</b>
<b>5.1 Summary of contributions</b>	<b>122</b>
5.1.1 Three dimensional organoid experiments	122
5.1.1 Collaborative works not discussed	124
<b>5.2 Biological conclusions</b>	<b>124</b>
<b>5.3 Future directions</b>	<b>127</b>
<b>Citations</b>	<b>128</b>

## Table of Figures

<b>Figure 1-1: The architecture and cell types that comprise the mammary gland throughout development</b> .....	5
<b>Figure 1-2: Mammary epithelial cell differentiation and relationships to breast cancer subtypes</b> .....	6
<b>Figure 1-3: Modulation of the breast microenvironment contributes to cancer progression</b> .....	9
<b>Figure 1-4: Microenvironmental therapeutic targets approved and currently in development for the treatment of breast cancer treatment</b> .....	10
<b>Figure 1-5: Illustration of the three primary mechanisms of cellular communication</b> .....	12
<b>Figure 1-6: Mechanism of ligand-mediated cellular communication</b> .....	13
<b>Figure 1-7: Canonical signaling pathways of OSM, EGF, and TGFB</b> .....	15
<b>Figure 1-8: Cellular proliferation is controlled by the cell</b> .....	17
<b>Figure 1-9: Modes and mechanisms driving cellular migration and metastasis</b> .....	19
<b>Figure 1-10: The continuum of the epithelial-mesenchymal transition</b> .....	21
<b>Figure 1-11: Mammary epithelial cells differentiate into three main lineages which can be distinguished through scRNAseq</b> .....	22
<b>Figure 1-12: MCF10A cells cultured in 2D and 3D conditions</b> .....	24
<b>Figure 1-13: Single cell Oncostatin M gene expression in murine mammary tumors</b> .....	27
<b>Figure 1-14: General workflow for elucidating molecular mechanisms through network-based analysis</b> .....	28
<b>Figure 2-1: Overview of experimental approach to assess the impact of microenvironmental factors</b> .....	50
<b>Figure 2-2: Ligand treatments induce diverse phenotypic responses</b> .....	53
<b>Figure 2-3: Six molecular assays reveal diverse dynamic responses to treatments</b> .....	55
<b>Figure 2-4: Assessment of ligand-induced molecular change</b> .....	57
<b>Figure 2-5: Integrated analysis identifies co-regulated molecular modules</b> .....	64
<b>Figure 2-6: Module 10 is associated with cell cycle progression</b> .....	69
<b>Figure 2-7: Analysis of molecular modules identifies functional relationships between molecular and phenotypic responses to OSM</b> .....	71
<b>Figure 3-1: Treatment of MCF10A cells with OSM but not IFNG initiates cell clustering and collective cell migration</b> .....	81
<b>Figure 3-2: Comparative network analysis of omics data reveals putative drivers of OSM-induced CCM</b> .....	84
<b>Figure 3-3: Quantification and visualization of siRNA screen to identify drivers of OSM-induced CCM</b> .....	88
<b>Figure 3-4: scRNAseq reveals HIF1A transcriptional program driving OSM-induced CCM</b> .....	91
<b>Figure 4-1: Combination treatments induce diverse and emergent phenotypic behavior</b> .....	102
<b>Figure 4-2: Transcriptional programs induced by single and combination ligand treatments</b> .....	105
<b>Figure 4-3: Synergistic transcriptional programs induced by ligand combination treatments</b> .....	109
<b>Figure 4-4: Partial Least Squares Regression (PLSR) links transcriptional signatures to phenotypic metrics</b> .....	113
<b>Figure 4-5: Expression of CXCR2 Agonists Correlates with Cell Motility</b> .....	115
<b>Figure 4-6: Synergistic Transcriptional Upregulation of CXCR2 Chemotactic Signaling Molecules via CREB Activation Promotes Increased Cell Motility</b> .....	117
<b>Figure 5-1: OSM-induced CCM in 3D model systems of normal and malignant mammary tissue</b> .....	123

## Acknowledgements

I am deeply grateful to my advisor, Dr. Laura Heiser, for granting me the freedom to chart my own research path and encouraging me to fully explore my scientific interests. This mentorship has been instrumental in shaping me into an independent researcher. I also extend my thanks to all the members of the Heiser lab for their feedback, expertise, and camaraderie, which fostered an accepting, professional, and intellectually stimulating environment where I could learn and grow.

A special thank you goes to Tiera Liby for patiently sharing your tissue culture expertise and for your kindness and understanding as I navigated my foibles in the lab. I am also sincerely grateful to my dissertation advisory committee—Dr. Andrew Adey, Dr. Carolyn Ibsen, and Dr. Mara Sherman—for their valuable suggestions and insights, as well as to Dr. Alexander Davies for serving on my exam committee. I would like to thank the Zuckerman lab and Dr. Jeremy Copperman for providing opportunities for interdisciplinary collaboration and offering a unique perspective on my research.

To my family and friends, thank you for your unwavering support. My parents instilled in me a love of learning and the determination to question and push boundaries. To my wonderful partner, Siana, your support and patience throughout my doctoral journey have been invaluable—I could not have done this without your love and encouragement. And lastly, to my dogs, thank you for being a constant source of joy and comfort.

## Dissertation Abstract

The microenvironment plays a critical role in determining cellular phenotype, yet the mechanisms through which combinations of extracellular signals drive specific phenotypic and molecular changes remain incompletely understood. Growth factors, cytokines, and extracellular matrix proteins are key microenvironmental components that mediate cellular communication and drive processes such as proliferation, migration, and differentiation. Dysregulation of these signals is implicated in a wide range of diseases, including cancer, fibrosis, and immune disorders. Developing a deeper understanding of how cells respond to individual and combinatorial signals is essential for uncovering the molecular underpinnings of these pathologies and for identifying potential therapeutic targets. This dissertation aims to elucidate how mammary epithelial cells integrate and respond to diverse microenvironmental cues, using MCF10A cells as a model system. By integrating transcriptomic, proteomic, and phenotypic data, I aim to reveal the dynamic interplay between signaling pathways and cellular behaviors, providing a framework for understanding epithelial cell plasticity.

In Chapter II, I present the LINCS Microenvironment (ME) perturbation dataset, a comprehensive resource cataloging the transcriptional, proteomic, epigenomic, and phenotypic responses of MCF10A cells to a panel of extracellular ligands, including EGF, HGF, OSM, IFNG, TGFB, and BMP2. This dataset systematically characterizes how each ligand influences molecular states and phenotypic outcomes, capturing both conserved and ligand-specific effects. Through illustrative analyses, I demonstrate how this dataset can uncover functional relationships between molecular features and specific cellular phenotypes, such as motility and proliferation. Additionally, the dataset provides a valuable resource for the broader scientific community to explore molecular perturbations, compare signaling pathways, and develop novel computational methods for integrative data analysis.

In Chapter III, I focus on understanding the molecular mechanisms underlying collective cell migration (CCM) and how the cytokine Oncostatin M (OSM) induces this phenomenon in MCF10A mammary epithelial cells. CCM is a critical biological process wherein groups of cells migrate together as a cohesive unit, retaining intercellular junctions while navigating through the extracellular environment. This behavior is pivotal in normal and disease processes such as tissue repair, embryonic development, and cancer metastasis. To dissect the molecular networks driving CCM, I applied a systems-biology approach, integrating transcriptomic, proteomic, and network analyses. Through these methods, I identified molecular subnetworks activated by OSM and evaluated their phenotypic contributions using high-throughput experimental approaches and identified HIF1A as a critical regulator of OSM-induced CCM. Further analysis using single-cell RNA sequencing (scRNA-seq) and experimental approaches revealed that the complement activation pathway plays a key role in OSM-induced CCM. This chapter advances our understanding of how epithelial cells respond to extracellular signals to drive collective migration and identifies key molecular networks that could serve as therapeutic targets in diseases characterized by aberrant migration, such as metastatic cancer.

In Chapter IV, I investigate how combinations of ligands shape cellular phenotypes and transcriptional programs in MCF10A epithelial cells. Using cytokines Oncostatin M (OSM), Transforming Growth Factor Beta 1 (TGFB), and Epidermal Growth Factor (EGF), I demonstrate that combinatorial treatments produce emergent phenotypes, including changes in motility, proliferation, and clustering, distinct from individual ligand effects. Transcriptomic and proteomic analyses identified synergistic upregulation of genes and pathways, including CREB signaling, driving phenotypic changes such as increased cell motility via CXCR2

activation. This work highlights the importance of combinatorial signaling in reprogramming epithelial behaviors and provides a framework for understanding how cells integrate microenvironmental signals to produce distinct molecular and phenotypic outcomes.

In summary, this dissertation provides a deep exploration of how mammary epithelial cells respond to microenvironmental signals. Through a progression of increasingly complex models and integrative experimental approaches, we identified key regulatory networks and molecular mechanisms that link extracellular cues to cellular behavior. From cataloging transcriptional, proteomic, and phenotypic responses to uncovering the role of HIF1A in OSM-induced CCM and elucidating the synergistic effects of ligand combinations, this work advances our understanding of cellular signal integration.

## 1. Chapter I: Introduction

### 1.1 The breast microenvironment

#### 1.1.1 The architecture of mammary tissue

The mammary gland is a highly specialized tissue with a complex organization that supports its primary function: milk production and delivery[1]. It consists of multiple cell types arranged into distinct compartments that work together dynamically. These compartments are broadly categorized into the epithelial and stromal regions, each contributing uniquely to the organ's structure and function.

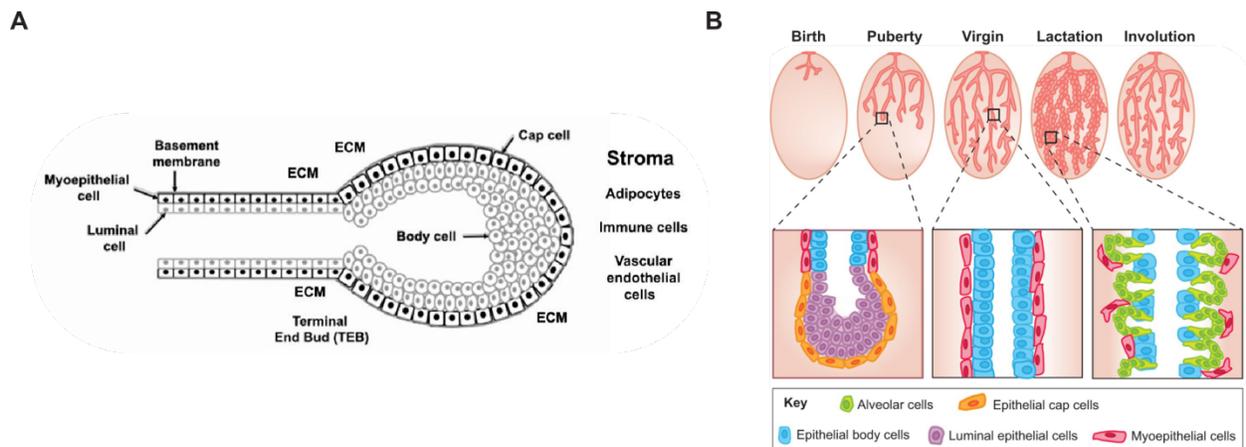
At the core of the mammary gland's functionality is the epithelial compartment, which is organized into ducts and lobules (**Figure 1-1A**) [2]. The ducts form a branching network that transports milk from its production sites to the nipple. This network is lined by a bilayer of epithelial cells. Luminal epithelial cells, positioned on the inner layer facing the ductal lumen, are directly involved in milk secretion during lactation [2]. Surrounding these luminal cells are myoepithelial cells, which provide structural support and facilitate milk ejection [3]. The terminal ends of the ductal system expand into lobules, which house clusters of alveoli. These alveolar structures are the primary sites of milk synthesis during lactation [1]. During periods of active ductal elongation, such as puberty, specialized structures known as terminal end buds form at the tips of growing ducts. These terminal end buds are dynamic, multilayered structures containing two key populations of cells: cap cells, which reside at the outer layer and give rise to myoepithelial cells, and body cells, which populate the interior and can differentiate into luminal cells [4]. The epithelial structures are anchored to a basement membrane, a thin layer of extracellular matrix proteins that provides mechanical support and acts as a barrier between the epithelium and the surrounding stroma.

The stromal compartment encompasses the connective tissue that supports and interacts with the epithelial structures. Fibroblasts within the stroma produce extracellular matrix (ECM) proteins which form a scaffold for the epithelium and regulate biomechanical and biochemical signals [5]. Adipocytes, or fat cells, are also abundant in the stroma, serving as an energy reservoir. Immune cells, such as macrophages, lymphocytes, and mast cells, are dispersed throughout the stroma, playing vital roles in tissue maintenance, remodeling, and defense [6], [7]. Additionally, vascular endothelial cells form the vascular networks essential for oxygen and nutrient delivery, as well as immune cell trafficking.

The development of the breast architecture unfolds across distinct life stages, each marked by hormonal and cellular changes (**Figure 1-1B**). During fetal development, the basic ductal structures begin to form under the influence of genetic programming and maternal hormones [2]. These initial epithelial

structures arise from a population of mammary progenitor cells, which have the capacity to differentiate into either luminal or myoepithelial and basal lineages [2]. At puberty, estrogen and growth hormone drive the elongation and branching of the ductal system, accompanied by increased stromal tissue deposition [8]. During pregnancy, the breast undergoes extensive remodeling, characterized by the proliferation and differentiation of alveolar cells to prepare for lactation [8]. This phase involves significant expansion of the epithelial compartment, supported by changes in the stroma to accommodate the increased functional demands. Following lactation, the breast enters a process of involution, during which alveolar cells undergo apoptosis, and the tissue returns to a baseline state through stromal remodeling and immune cell-mediated clearance [9].

The mammary gland is a complex and dynamic tissue, with diverse cell types interacting in intricate ways to fulfill its functional roles. Studying this ecosystem presents unique challenges, as it requires understanding the intricate signals that govern its organization, maintenance, and remodeling across different life stages. These signals, which coordinate interactions between epithelial, stromal, and immune components, are essential for the tissue's ability to adapt to physiological demands and maintain its functional integrity.



**Figure 1-1: The architecture and cell types that comprise the mammary gland throughout development**

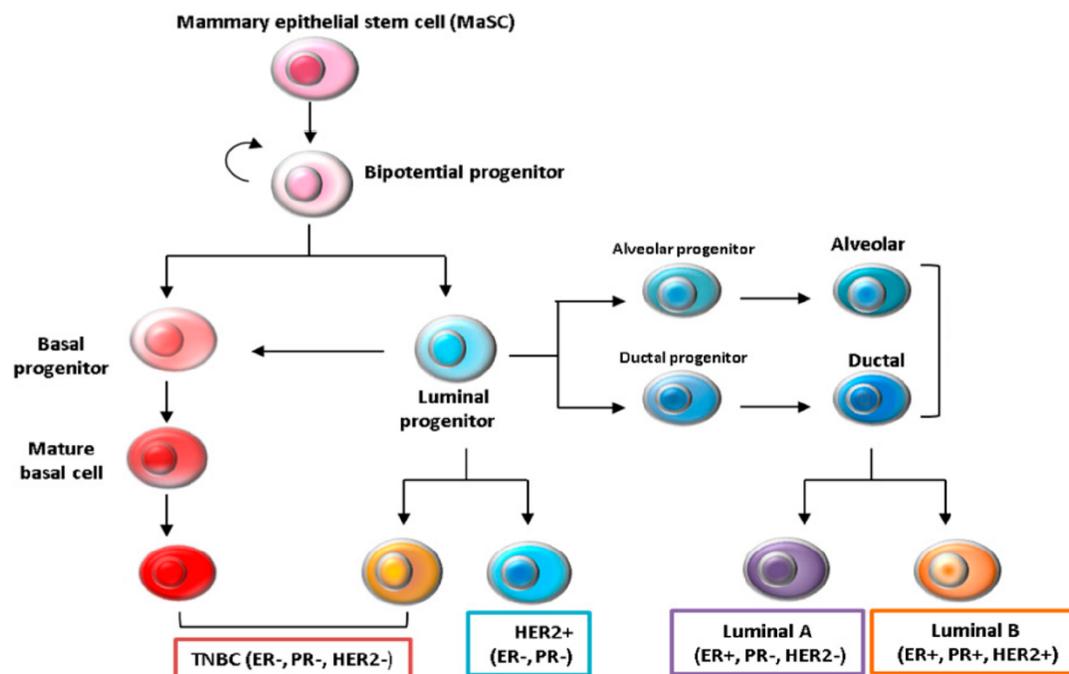
*Left: tissue architecture of a mammary terminal end bud. Right: mammary tissue morphology changes dramatically throughout development and pregnancy. Figure adapted from [2] and [10].*

### 1.1.2 Mammary ductal epithelial cells and their role in normal and pathological physiology

Mammary ductal epithelial cells (MDECs) are the primary structural and functional components of the mammary gland's ductal system. They are highly dynamic cells that undergo significant differentiation and remodeling during various physiological states and stages of development [10]. These processes contribute to their plasticity and ability to adapt to the functional demands of the mammary gland. However, this plasticity also renders them susceptible to pathological alterations, including cancer and other diseases [11]. MDECs originate from mammary epithelial stem cells which differentiate into progenitor cells. Progenitor cells then differentiate into two main lineages: luminal and basal/myoepithelial cells [10], [12]. Luminal epithelial cells line the ducts and alveoli, serving critical roles in milk production and secretion during lactation, while basal cells provide structural support and contractile function, aiding milk ejection (**Figure 1-2**).

MDECs are central to the development of breast cancer, one of the most common malignancies worldwide [12]. Genetic mutations, epigenetic changes, and microenvironmental factors can disrupt the tightly controlled processes of differentiation and proliferation, leading to malignant transformation. Luminal epithelial cells are the most frequent origin of breast cancer, giving rise to luminal A and B subtypes, which vary in their hormonal receptor status (estrogen receptor (ER), progesterone receptor (PR), and HER2 receptor) and responsiveness to therapy [13]. Basal-like or triple-negative breast cancers (TNBC), often more aggressive and challenging to treat, may arise from basal progenitor cells [14] (**Figure 1-2**).

Beyond cancer, MDECs are implicated in other pathological conditions, such as mastitis, an inflammatory disease often associated with lactation. Mastitis involves infection or injury-induced inflammation, leading to swelling, pain, and impaired milk secretion [15]. The epithelial barrier function of MDECs plays a crucial role in preventing pathogen invasion, and its disruption can exacerbate disease severity. Additionally, hormonal imbalances or genetic predispositions can lead to hyperplasia or benign breast diseases, further emphasizing the importance of maintaining epithelial cell homeostasis.



*Figure 1-2: Mammary epithelial cell differentiation and relationships to breast cancer subtypes*

Mature luminal and basal/myoepithelial lineages arise from a common progenitor. It is hypothesized that breast cancer subtypes, distinguished by hormone and HER2 receptor expression, arise from distinct mammary epithelial cell lineages. Figure adapted from [16].

### 1.1.3 Communication to and from mammary epithelial cells is required for normal development and function

Effective communication between mammary epithelial cells and their surrounding microenvironment is essential for the proper development and function of the mammary gland. This communication is mediated through a complex network of biochemical and mechanical signals that facilitate cell-to-cell and cell-to-matrix interactions. These interactions regulate processes such as proliferation,

differentiation, and tissue remodeling, ensuring the coordinated development and function of the gland [17].

The stromal compartment of the mammary gland provides critical support to mammary epithelial cells. Growth factors such as transforming growth factor- $\beta$  (TGF- $\beta$ ), epidermal growth factor (EGF), and insulin-like growth factor (IGF) are secreted by stromal cells and influence epithelial cell behavior [18]. These signals are particularly crucial during puberty, when ductal elongation and branching morphogenesis depend on epithelial-stromal crosstalk. Similarly, during pregnancy, stromal-derived factors coordinate with systemic hormonal signals to drive alveolar differentiation and milk production [18].

Hormonal signaling plays a pivotal role in orchestrating communication within the mammary gland. Estrogen, progesterone, and prolactin act on both epithelial and stromal cells to regulate ductal elongation, branching morphogenesis, and alveologenesis [18]. These hormones induce the expression of paracrine factors such as amphiregulin, which mediates epithelial-stromal crosstalk [19]. During pregnancy, prolactin is especially critical for inducing milk-producing alveolar cells, while estrogen and progesterone prepare the ductal network [18]. Hormonal withdrawal during involution triggers apoptosis and remodeling (**Figure 1-3**).

Mechanical cues, such as those arising from tissue stiffness and extracellular matrix tension, also contribute to mammary epithelial cell function. Changes in ECM stiffness, often mediated by stromal remodeling, influence epithelial cell differentiation and proliferation through mechanotransduction pathways involving integrins and focal adhesion complexes [5]. During puberty, mechanical forces guide ductal elongation and branching, while during lactation, ECM remodeling ensures glandular expansion [5].

The intricate communication between mammary epithelial cells and their surrounding microenvironment is fundamental to the normal development and function of the mammary gland. Specific environmental cues, including growth factors, hormonal signals, and mechanical forces are essential at distinct stages of development, such as puberty, pregnancy, and involution.

#### 1.1.4 Changes to microenvironmental signaling contributes to pathological tissue states

Alterations in the microenvironmental signaling pathways that regulate mammary epithelial cell behavior can significantly contribute to pathological tissue states. These changes may arise from genetic mutations, epigenetic modifications, chronic inflammation, or external factors such as exposure to environmental toxins [12]. Such disruptions not only impair normal glandular function but also create a permissive environment for disease development.

One major pathological consequence of altered signaling is the development of fibrosis, characterized by excessive deposition of extracellular matrix (ECM) components [15]. Fibrosis often results from chronic inflammation, during which activated fibroblasts secrete elevated levels of collagen and other ECM proteins [15]. Increased tissue stiffness can disrupt normal epithelial cell differentiation and promote a pro-tumorigenic phenotype [20]. For instance, in breast cancer, elevated ECM stiffness enhances epithelial cell proliferation and invasion through mechanotransduction pathways involving YAP/TAZ signaling [21].

Chronic inflammation and immune cell dysfunction are critical contributors to pathological states in the mammary gland [22]. Persistent activation of immune cells, such as macrophages and neutrophils, leads

to the release of pro-inflammatory cytokines and reactive oxygen species [22]. These factors can cause DNA damage and impair epithelial barrier function, thereby increasing susceptibility to infections and cancer. In the context of tumor progression, immune cells can adopt tumor-promoting roles by secreting growth factors that support angiogenesis and metastasis[23].

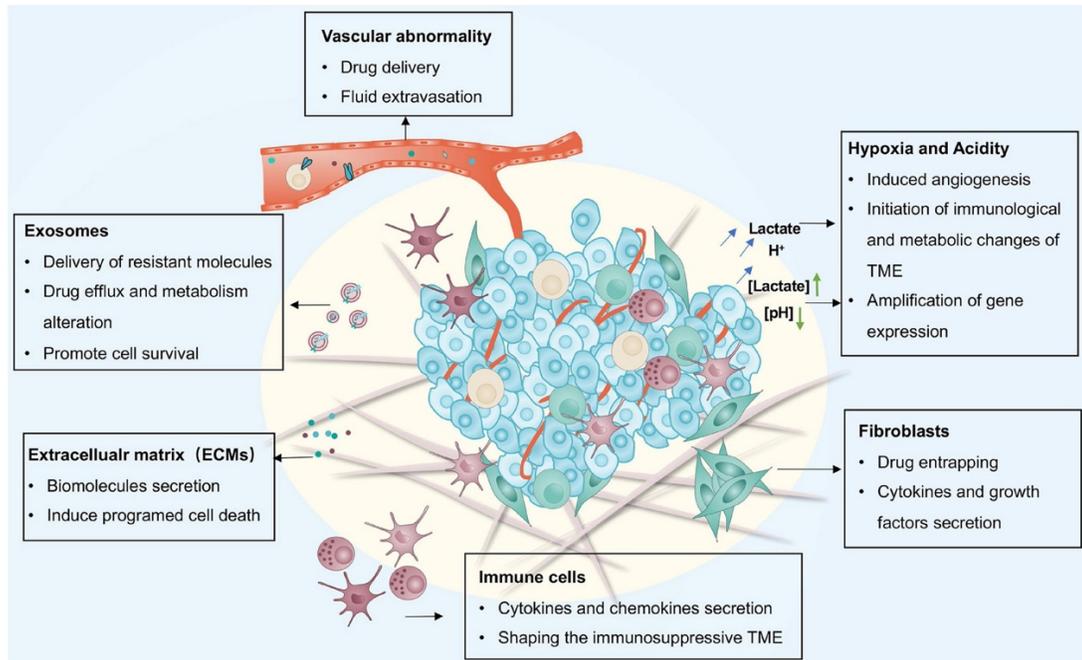
Disruptions in hormonal signaling also play a significant role in pathological tissue states. For example, excessive or prolonged estrogen exposure is associated with an increased risk of breast cancer [24]. Estrogen drives the proliferation of luminal epithelial cells and, when dysregulated, can lead to hyperplasia or malignant transformation. Similarly, hormonal imbalances can contribute to benign proliferative disorders, such as fibroadenomas or cystic changes, which may predispose the gland to further abnormalities [25].

The tumor microenvironment in breast cancer exemplifies how changes to microenvironmental signaling can promote pathological states (**Figure 1-3**). Cancer-associated fibroblasts (CAFs), reprogrammed from normal fibroblasts, secrete growth factors, chemokines, and ECM proteins that support tumor growth and invasion [26]. Additionally, metabolic alterations, such as hypoxia and acidosis within the tumor microenvironment, further drive malignant progression by promoting epithelial-mesenchymal transition (EMT) and immune evasion [27].

Exosomes, small extracellular vesicles secreted by both normal and cancerous epithelial cells, play a pivotal role in tumor progression and metastasis. These vesicles carry proteins, lipids, and nucleic acids that can reprogram recipient cells in the tumor microenvironment [28]. For instance, exosomes from cancer cells can promote angiogenesis by delivering pro-angiogenic factors such as VEGF to endothelial cells [29]. They also facilitate immune evasion by modulating the activity of immune cells, including T cells and macrophages [28]. Additionally, exosomes can prepare distant metastatic sites by altering the local microenvironment to support tumor cell colonization [30].

Vascular changes are another critical aspect of the microenvironment in breast cancer. The formation of abnormal, leaky blood vessels within tumors not only provides nutrients and oxygen to cancer cells but also facilitates the intravasation and dissemination of tumor cells into the bloodstream [31]. This process, known as the angiogenic switch, is driven by factors such as VEGF, secreted by both cancer cells and stromal components [32]. These vascular changes also contribute to creating hypoxic regions within the tumor, which further drive malignant progression and therapy resistance [27].

Immune cells play a dual role in the progression of breast cancer, serving both anti-tumorigenic and pro-tumorigenic functions depending on the context. Tumor-associated macrophage (TAMs), for example, are often reprogrammed by cancer cells to adopt an M2-like phenotype, which promotes tumor growth and suppresses immune responses [23], [33]. These TAMs secrete cytokines, chemokines, growth factors, and proteases that enhance angiogenesis, matrix remodeling, and metastasis. Similarly, regulatory T cells (Tregs) within the tumor microenvironment suppress cytotoxic T cell activity, enabling immune evasion [23]. Neutrophils, another immune cell type, can release neutrophil extracellular traps, which facilitate tumor cell migration and colonization at distant sites [34]. Understanding these immune interactions is critical for developing immunotherapeutic strategies aimed at reactivating anti-tumor immunity.



**Figure 1-3: Modulation of the breast microenvironment contributes to cancer progression**

The breast microenvironment is altered during breast cancer initiation and progression. Changes in stromal cell composition and phenotype, tissue architecture, and oxygen content can contribute to cancer progression and metastasis. Figure adapted from [35].

### 1.1.5 Therapeutics targeting microenvironmental signaling can improve outcomes in disease

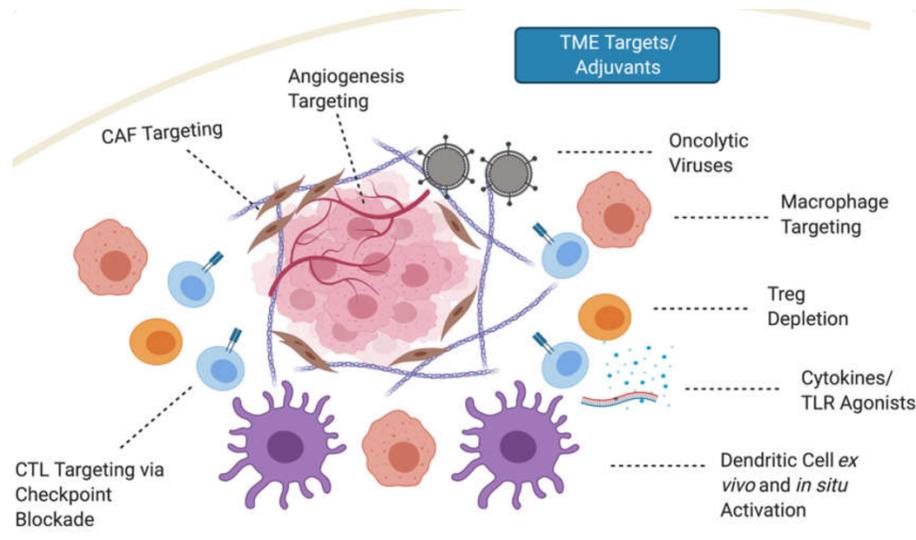
Therapeutic strategies targeting microenvironmental signaling have shown significant promise in improving outcomes in both cancer and non-cancer diseases of the mammary gland [36]. By modulating interactions within the tumor microenvironment or restoring normal signaling pathways, these therapies aim to halt disease progression and promote tissue recovery. In non-cancerous diseases, such as mastitis or fibrotic disorders, targeting the inflammatory and fibrotic components of the microenvironment has yielded promising results. Anti-inflammatory drugs, such as non-steroidal anti-inflammatory agents, can reduce inflammation and improve epithelial barrier function in mastitis, thereby restoring normal glandular function[37].

In cancer, immune checkpoint inhibitors targeting PD-1/PD-L1 or CTLA-4 pathways have revolutionized treatment by reactivating cytotoxic T cells against tumor cells [36]. These therapies have shown promise in triple-negative breast cancer (TNBC), where immune evasion is a hallmark of disease progression. Additionally, therapies targeting CAFs and Tregs are being developed to counteract the immunosuppressive microenvironment and restore effective immune surveillance [38].

Therapies targeting exosome production or uptake are emerging as a novel approach to disrupting cancer progression. Exosome inhibitors, such as GW4869 or neutralizing antibodies against exosome surface proteins, can prevent the transfer of oncogenic signals between cells [39]. Similarly, anti-angiogenic therapies targeting vascular endothelial growth factor (VEGF) or its receptors can normalize

tumor vasculature, reducing nutrient supply to tumors and limiting metastasis [40]. These strategies complement existing treatments by addressing critical microenvironmental drivers of cancer progression.

Despite the promising advances, challenges remain in developing microenvironment-targeted therapies. The heterogeneity of the tumor microenvironment, combined with the potential for off-target effects on normal tissues, necessitates a precision medicine approach. Moreover, understanding how microenvironmental signals influence epithelial cell behavior—such as proliferation, migration, and resistance to treatment—will enable the development of more effective therapeutic strategies.



**Figure 1-4: Microenvironmental therapeutic targets approved and currently in development for the treatment of breast cancer treatment**

Overview of key microenvironmental targets under investigation or in use for treating breast cancer. The illustration highlights critical components, including cancer-associated fibroblasts (CAFs), macrophages, and angiogenesis pathways. Therapeutic strategies such as immune checkpoint inhibitors are approved for use in breast cancer to normalize the breast microenvironment. Figure adapted from [41]

## 1.2 Cellular communication

Cellular communication is fundamental for the regulation of physiological processes, including growth, differentiation, and homeostasis. Ligands, such as growth factors, cytokines, and hormones, bind to specific receptors on target cells, initiating intracellular signaling cascades that modulate cellular behavior. This chapter provides a comprehensive overview of key signaling ligands—Epidermal Growth Factor (EGF), Transforming Growth Factor-beta (TGF- $\beta$ ), and Oncostatin M (OSM)—that play critical roles in mammary gland physiology. While extensive research has been conducted to understand how these ligands signal and influence normal physiology and disease, significant gaps remain in our understanding, particularly regarding the cellular context of their actions and their roles in cancer progression.

### 1.2.1 Modes of cellular communication

Cellular communication is a cornerstone of biological systems, enabling individual cells to coordinate their behavior with their neighbors and the surrounding environment. This communication is essential for maintaining tissue homeostasis, orchestrating developmental processes, and responding to environmental cues [42]. Cells generally communicate through three primary classes of mechanisms:

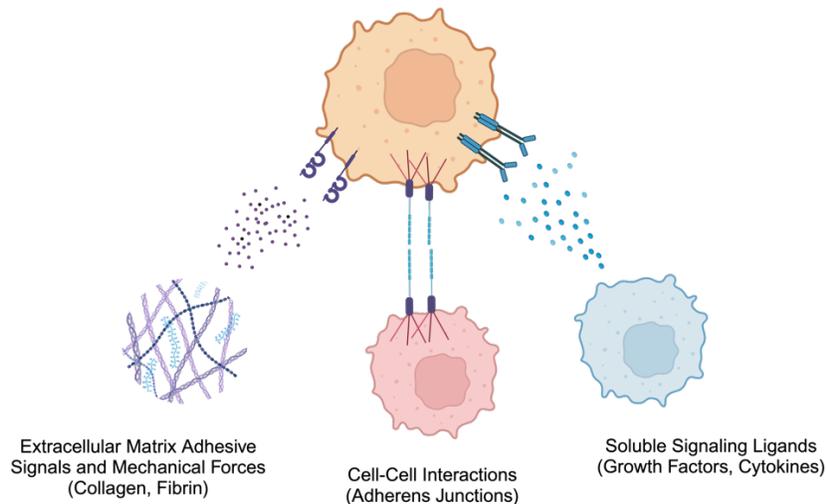
interactions with the extracellular matrix (ECM), direct physical interactions with other cells, and signaling mediated by soluble ligands [42] (**Figure 1-5**).

Interactions between cells and the ECM are fundamental for providing structural support and transmitting biochemical and mechanical signals. Cells attach to the ECM via transmembrane receptors, such as integrins, which connect the cytoskeleton to ECM components like collagen, fibronectin, and laminins [43]. These interactions regulate key cellular processes, including migration, differentiation, and proliferation. Integrin-mediated signaling can activate downstream pathways, such as focal adhesion kinase (FAK) and MAPK, which influence gene expression and cellular responses [44]. Mechanical properties of the ECM, such as stiffness, also play a role in directing cell behavior through mechanotransduction [5].

Direct physical interactions between cells are mediated by specialized structures such as cadherin-based adherens junctions, tight junctions, and desmosomes [42]. These connections enable mechanical coupling between cells and facilitate the transfer of signals that regulate tissue organization and function. Cadherins, a family of calcium-dependent adhesion molecules, play a critical role in maintaining tissue integrity by linking the actin cytoskeleton of adjacent cells [45]. Tight junctions, formed by claudins and occludins, regulate paracellular transport and maintain barriers in epithelial and endothelial tissues [46]. Additionally, gap junctions, composed of connexins, allow the direct exchange of ions and small molecules between cells, facilitating electrical and chemical coordination [47].

Soluble ligand signaling involves the secretion of signaling molecules that bind to receptors on the same cell (autocrine signaling) or nearby cells (paracrine signaling). This mode of communication enables cells to regulate both their own behavior and that of their local environment. Growth factors, cytokines, and hormones are common mediators of soluble signaling. For example, in autocrine signaling, a cell may secrete epidermal growth factor (EGF) to stimulate its own proliferation [48]. In paracrine signaling, fibroblasts in the stromal microenvironment may release transforming growth factor- $\beta$  (TGF- $\beta$ ) to influence adjacent epithelial cells [49]. Signal transduction pathways initiated by these ligands often involve receptor tyrosine kinases (RTKs), G-protein-coupled receptors (GPCRs), or other specialized receptors, leading to the activation of intracellular cascades that modulate gene expression and cellular function [50].

These three primary mechanisms of cellular communication—cell-ECM interactions, direct physical cell-cell interactions, and soluble ligand signaling—are integral to the regulation of cellular behavior and tissue function. By leveraging these diverse communication strategies, cells and tissues can adapt to complex and dynamic environments, ensuring the proper functioning of multicellular organisms.



**Figure 1-5: Illustration of the three primary mechanisms of cellular communication**

Cells can communicate with the microenvironment via cell-ECM interactions mediated by integrins, physical cell-cell interactions via cadherins and junctional complexes, and soluble ligand signaling, encompassing both autocrine and paracrine pathways. Figure created with Biorender.

### 1.2.2 Classes of microenvironmental ligands

Soluble ligands represent a diverse group of molecules capable of mediating cellular communication over short or long distances. These ligands can be secreted into the local extracellular environment or into systemic circulation, such as the bloodstream, to reach distant target cells. Additionally, some ligands function in juxtacrine signaling, remaining membrane-bound to directly engage receptors on adjacent cells [51]. Ligands influence cell behavior through a three-stage process: 1) binding to cell surface or intracellular receptors, 2) signal transduction and activation of complex signaling cascades, and 3) activation of cellular response that is often accompanied by changes in gene transcription and cell behavior [52] (**Figure 1-6**).

Cytokines are small proteins that play crucial roles in regulating immune responses, inflammation, and cell differentiation [53]. They are secreted by a wide range of cells, including immune cells, epithelial cells, and stromal cells. Cytokines typically signal through cytokine-specific receptors, such as the IL-6 receptor, which activates the JAK-STAT pathway [54]. Through these signaling cascades, cytokines influence processes such as cell proliferation, apoptosis, and immune cell recruitment.

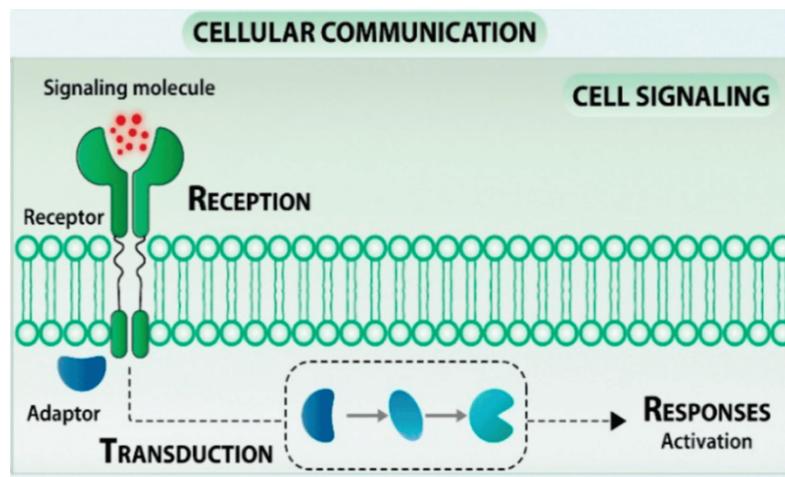
Chemokines are a subset of cytokines that primarily regulate cell migration by establishing chemical gradients [55]. These gradients direct the movement of cells, such as leukocytes, during immune surveillance and inflammatory responses. Chemokines are categorized into four main subfamilies (CXC, CC, CX3C, and C) based on the arrangement of conserved cysteine residues [55]. Examples include CXCL8 (IL-8), which recruits neutrophils, and CCL2 (MCP-1), which attracts monocytes [55]. Chemokine signaling is mediated by G-protein-coupled receptors (GPCRs), which activate intracellular pathways like FAK, PI3K-AKT and MAPK to orchestrate cellular responses [56].

Growth factors are essential molecules that regulate cell proliferation, survival, differentiation, and migration [57]. They include a diverse array of proteins such as epidermal growth factor (EGF), vascular

endothelial growth factor (VEGF), and transforming growth factor- $\beta$  (TGF- $\beta$ ). Growth factors bind to receptor tyrosine kinases (RTKs) like EGFR and VEGFR, initiating signal transduction pathways such as RAS-RAF-MEK-ERK and PI3K-AKT [50]. These pathways lead to changes in gene expression and cellular behavior. Growth factors are critical during development and tissue repair, and their dysregulation is a hallmark of cancer, contributing to tumor growth, angiogenesis, and metastasis[58].

Hormones are systemic signaling molecules that influence a wide range of physiological processes, including growth, metabolism, and reproduction [18]. Steroid hormones such as estrogen and progesterone are particularly relevant in the context of breast physiology, as they play crucial roles in mammary gland development and function. Estrogen, for instance, binds to nuclear receptors like the estrogen receptor (ER), modulating the transcription of genes involved in cell proliferation and differentiation [59]. Similarly, progesterone acts through its receptor (PR) to influence mammary epithelial cell growth and ductal branching [59]. Dysregulation of hormone signaling is a key factor in hormone-dependent cancers, such as certain subtypes of breast cancer, where therapies targeting ER or PR have shown significant clinical efficacy [60].

In addition to cytokines, chemokines, growth factors, and hormones, several other classes of soluble ligands contribute to cellular communication. Neurotransmitters are key mediators of neuronal signaling and influence processes beyond the nervous system, including immune regulation [61]. Leukotrienes, lipid-based signaling molecules derived from arachidonic acid, play roles in inflammation and allergic responses by binding to GPCRs like BLT1 and CysLT1 [62]. These diverse signaling molecules highlight the complexity and versatility of ligand-mediated communication in biological systems.



*Figure 1-6: Mechanism of ligand-mediated cellular communication*

Microenvironmental ligands signal through cell-surface receptors or nuclear receptors and activate signal transduction processes to alter cellular transcription and behavior. Figure adapted from[52].

### 1.2.3 Signaling pathways of induced by select microenvironmental ligands

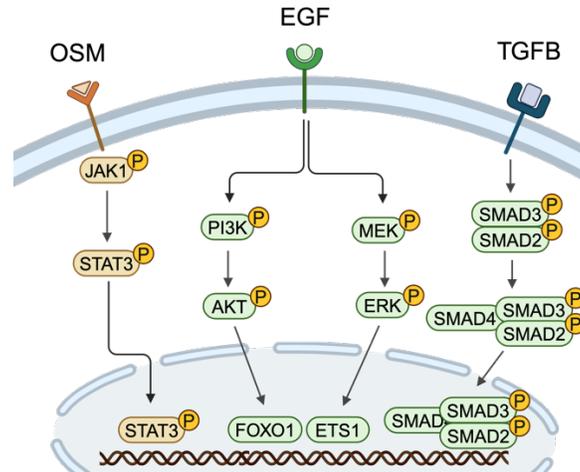
In the context of mammary gland physiology, ligand-mediated signaling plays a pivotal role in regulating cellular processes such as proliferation, differentiation, and survival [57]. Three key ligands—epidermal growth factor (EGF), transforming growth factor-beta (TGF- $\beta$ ), and Oncostatin M (OSM)—are chosen for this analysis due to their significant involvement in mammary epithelial cell function [63], [64], [65]. EGF, primarily secreted by cells such as fibroblasts and keratinocytes, promotes mammary epithelial cell

proliferation and survival [66]. TGF- $\beta$ , secreted by a variety of cell types including fibroblasts and immune cells, has a dual role in both inhibiting mammary epithelial cell growth and regulating differentiation [64]. OSM, produced by several cell types including macrophages, fibroblasts, and endothelial cells, is involved in the regulation of mammary epithelial cell differentiation and inflammatory responses [63].

EGF exerts its effects through binding to the epidermal growth factor receptor (EGFR), a receptor tyrosine kinase [67]. Upon ligand binding, EGFR undergoes dimerization and autophosphorylation on tyrosine residues, activating downstream signaling pathways such as the Ras/Raf/MEK/ERK pathway, which promotes cell proliferation and survival [68]. Additionally, the PI3K/Akt pathway is activated, contributing to cell survival and metabolic regulation [68]. The activation of these canonical pathways leads to changes in gene expression that support epithelial cell growth. Non-canonical signaling, such as the activation of Src family kinases, has also been shown to influence cellular responses to EGF, particularly in the context of cell migration and metastasis [69].

TGF- $\beta$  signals through its type I and type II serine/threonine kinase receptors, which upon ligand binding initiate the phosphorylation of Smad proteins, particularly Smad2 and Smad3 [70]. These phosphorylated Smads form a complex with Smad4 and translocate to the nucleus, where they regulate the transcription of target genes involved in cell cycle arrest, apoptosis, and extracellular matrix production [71]. In the mammary epithelium, TGF- $\beta$  functions as a potent inhibitor of cell proliferation and is critical for maintaining tissue architecture [64]. TGF- $\beta$  signaling also plays a crucial role in epithelial-to-mesenchymal transition (EMT), which is vital during both development and tumorigenesis [72]. While canonical Smad-dependent pathways dominate, non-canonical signaling pathways, including those involving MAPKs and Rho GTPases, also contribute to the diverse cellular outcomes driven by TGF- $\beta$  [73].

Oncostatin M signals through the Oncostatin M Receptor (OSMR) and has a weaker affinity to the Leukemia Inhibitory Factor Receptor (LIFR) which both form a heterodimer with the gp130 receptor, which is shared with other cytokines like interleukin-6 [74]. Binding of OSM to its receptor activates the JAK/STAT3 signaling pathway, leading to the phosphorylation and dimerization of STAT3, which translocates to the nucleus and regulates gene expression related to inflammation, differentiation, and cell survival [74]. In addition to the JAK/STAT3 axis, OSM has been shown to activate the MAPK and PI3K/Akt pathways, contributing to the regulation of cell survival and inflammation [63]. Non-canonical pathways involving Src family kinases and NF- $\kappa$ B can also be activated, influencing processes such as immune response and epithelial remodeling [75].



**Figure 1-7: Canonical signaling pathways of OSM, EGF, and TGFB**

EGF activates the AKT and MAPK pathways, TGFB induces SMAD signaling, and OSM canonically signals through JAK/STAT3 pathway. All three ligands can also activate non-canonical pathways, and potential crosstalk can occur between the signaling pathways. Figure created with Biorender.

### 1.3 Epithelial cell phenotypes and modulation by microenvironmental ligands

The mammary gland, like other epithelial tissues, relies on a diverse array of cellular phenotypes to maintain its structure and function [2]. Epithelial cells exhibit remarkable plasticity, dynamically adjusting their behavior in response to cues from their microenvironment [11]. Key phenotypes, including proliferation, migration, and epithelial-to-mesenchymal transition (EMT), contribute to critical processes such as development, tissue remodeling, and repair. These same mechanisms, however, are often co-opted during disease states such as cancer, contributing to tumor growth, invasion, and metastasis. Understanding the cellular phenotypes of epithelial cells and the molecular and environmental factors that regulate them is therefore essential for elucidating both normal mammary gland biology and the pathogenesis of diseases such as breast cancer.

This chapter explores the multifaceted nature of epithelial cell phenotypes, beginning with proliferation, the process by which cells divide to support tissue expansion and renewal. We then delve into migration and EMT, mechanisms critical for tissue remodeling and, in pathological contexts, cancer dissemination. Additionally, we examine the concept of partial EMT and its implications for tumor progression, as well as other phenotypes such as differentiation, which play pivotal roles in tissue function and architecture.

#### 1.3.1 Proliferation

Cellular proliferation is a fundamental process in both normal and pathological contexts of the mammary gland. During normal development, proliferation drives the expansion and remodeling of the mammary epithelium, facilitating processes such as ductal elongation, branching morphogenesis, and alveologenesis [1]. These tightly regulated events ensure proper development and function of the gland, particularly during puberty, pregnancy, and lactation. Conversely, in cancer, unchecked proliferation underlies tumor growth and progression, contributing to the formation of malignant masses and metastases [58].

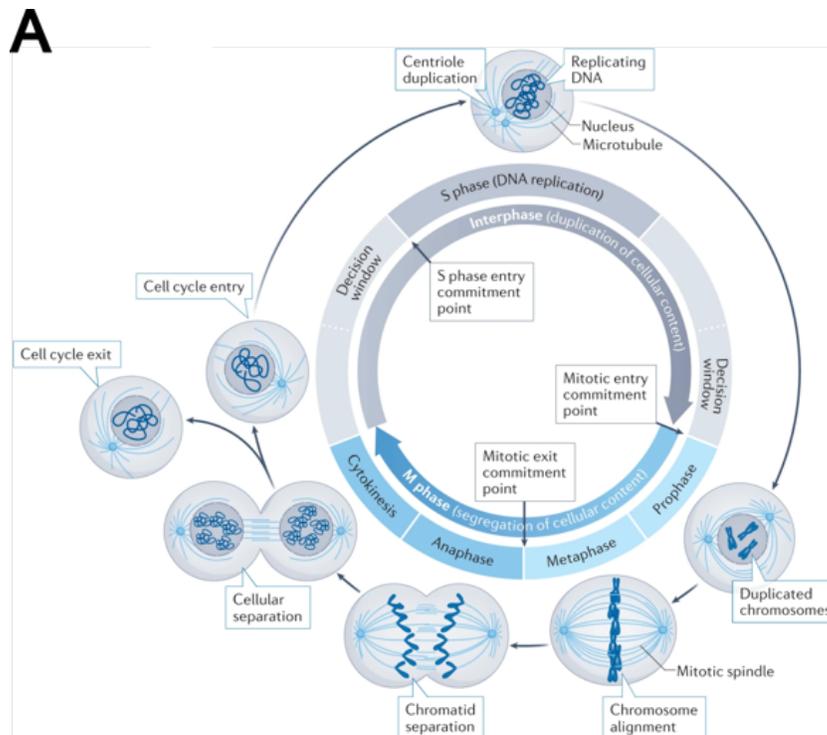
Proliferation occurs through the cell cycle, a tightly orchestrated sequence of events that ensure the accurate duplication and division of cellular contents [76]. The cell cycle is divided into four main phases: G1 (growth and preparation for DNA synthesis), S (DNA replication), G2 (preparation for mitosis), and M (mitosis and cytokinesis) [76]. Mitosis itself is a complex process involving the segregation of chromosomes into two daughter cells through prophase, metaphase, anaphase, and telophase, followed by physical cell division during cytokinesis (**Figure 1-8A**) [77]. Critical to the fidelity of the cell cycle are several checkpoints that monitor and control progression [78]. The G1/S checkpoint ensures that the cell has sufficient resources and proper signals to enter DNA synthesis. The G2/M checkpoint verifies that DNA replication is complete and free of damage before mitosis begins. During mitosis, the spindle assembly checkpoint ensures proper chromosome alignment and attachment to spindle fibers to prevent errors in chromosome segregation. Errors in checkpoint regulation often led to genomic instability, a hallmark of cancer [58].

The cell cycle is governed by a network of intrinsic regulators, including cyclins, cyclin-dependent kinases (CDKs), and CDK inhibitors (CKIs) [78]. Cyclins bind to and activate CDKs, driving progression through specific cell cycle phases. For example, cyclin D-CDK4/6 activity promotes the G1 to S phase transition, while cyclin B-CDK1 regulates entry into mitosis [79]. These activities are modulated by CKIs such as p21 and p27, which inhibit CDK activity under stress or DNA damage conditions, providing checkpoints to prevent uncontrolled proliferation [80]. Dysregulation of these intrinsic controls, often through mutations or alterations in gene expression, is a common feature in breast cancer.

The proliferation of mammary epithelial cells is profoundly influenced by signals from the surrounding microenvironment. Components such as the extracellular matrix (ECM), stromal fibroblasts, and immune cells secrete ligands that modulate proliferation [10]. For instance, insulin-like growth factor (IGF), secreted by stromal cells, promotes epithelial proliferation by binding to its receptor IGF-1R, activating downstream signaling pathways like PI3K-AKT and MAPK [81]. Mechanical cues, such as ECM stiffness, also regulate proliferation by modulating pathways like FAK and YAP/TAZ signaling [5].

Epidermal growth factor (EGF) exemplifies the role of microenvironmental ligands in regulating proliferation. EGF binds to the epidermal growth factor receptor (EGFR), a receptor tyrosine kinase, triggering dimerization and autophosphorylation of the receptor [82]. This activates downstream signaling cascades, including the RAS-RAF-MEK-ERK and PI3K-AKT pathways, which promote cell cycle progression and proliferation [66], [83]. Key transcription factors such as MYC and FOS are activated through these pathways, driving the expression of genes involved in DNA synthesis and mitotic entry [84]. In the mammary gland, EGF signaling is critical for ductal elongation and branching morphogenesis during development. Aberrant EGFR activation, whether through overexpression or mutation, is frequently observed in breast cancer, driving tumorigenesis and resistance to therapy [36]. This makes EGFR a critical target for therapeutic intervention.

A variety of techniques are employed to measure cell proliferation. Common methods include labeling proliferating cells with thymidine analogs such as BrdU or EdU, which incorporate into newly synthesized DNA during the S phase [85]. Immunostaining for proliferation markers, such as Ki-67 or PCNA, provides additional insights into cell cycle activity [86]. Advanced techniques like live-cell imaging allow for real-time monitoring of cell division, and computational pipelines are increasingly used to analyze imaging datasets [87]. These pipelines enable quantitative assessments of cell division dynamics, tracking individual cells over time to provide high-resolution insights into proliferation.



**Figure 1-8: Cellular proliferation is controlled by the cell**

Proliferation occurs through the cell cycle which is divided into phases controlled by molecular checkpoints. Figure panel A adapted from [77].

### 1.3.2 Migration

Cellular migration is required for normal development and physiology in the mammary gland. During development, migration orchestrates ductal elongation and branching, enabling the formation of the tissue architecture required for lactation [1]. In cancer, migration is a key step in metastasis, allowing tumor cells to traverse tissue barriers, invade surrounding stroma, and disseminate to distant organs [88].

Epithelial cells adopt distinct migration modes depending on cell type, cellular context, and microenvironmental cues (**Figure 1-9A**) [89]. Amoeboid migration involves rounded cells that navigate through ECM pores using actomyosin contractility and low adhesive interactions [90]. In contrast, mesenchymal migration relies on elongated cell morphology, strong adhesion to the ECM, and the formation of actin and integrin rich protrusions called lamellipodia and filopodia [91]. This mode requires ECM degradation via matrix metalloproteinases (MMPs) to facilitate movement through tissue barriers [92]. Collective cell migration (CCM), where groups of cells move cohesively while maintaining cell-cell junctions, is particularly significant in the mammary gland [93]. During branching morphogenesis, CCM is required for coordinated epithelial remodeling. In cancer, CCM may serve as a primary mechanism by which clusters of cells invade tissues and enter the vasculature. Evidence suggests that CCM is more efficient than single-cell migration for metastatic colonization (**Figure 1-9B**) [94].

Migration and invasion, while related, represent distinct processes. Migration refers to the movement of cells within or through a tissue, whereas invasion involves breaching and remodeling tissue barriers,

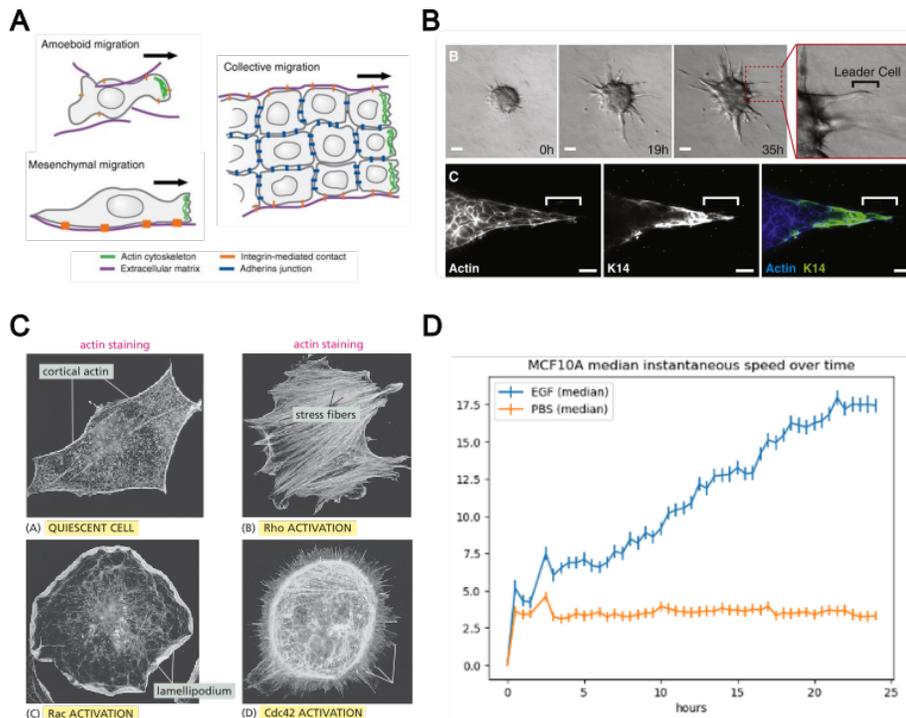
often accompanied by ECM degradation [95]. Both processes contribute to metastasis, but their mechanistic differences highlight the complexity of cancer cell dissemination. The transition from localized migration to systemic metastasis is a hallmark of cancer progression [58]. CCM, in particular, has been implicated in enhancing metastatic efficiency. Clusters of cancer cells exhibit improved survival during intravasation and extravasation out of blood vessels, potentially due to the retention of intercellular signaling and adhesion molecules that protect against anoikis [96]. Additionally, these clusters may carry supportive stromal or immune cells that facilitate colonization at distant sites [97], [98], [99]. Studies have shown that collective migration, rather than single-cell dissemination, often predominates during early metastatic events [100].

Cell migration is orchestrated by intricate molecular networks that regulate cytoskeletal dynamics, cell polarity, and adhesion. Rho family GTPases, including Rho, Rac, and Cdc42, are central to these processes (**Figure 1-9C**) [89]. Rho promotes actomyosin contractility, driving the mechanical forces needed for amoeboid migration [101]. Rac facilitates the formation of lamellipodia, critical for mesenchymal migration, while Cdc42 coordinates filopodia formation and establishes front-rear polarity [102]. These GTPases act through downstream effectors, such as ROCK, formins, and the Arp2/3 complex, to modulate actin and microtubule organization [103]. Additionally, GPCRs, including those activated by chemokines, can directly influence the RHO/ROCK axis, further modulating actomyosin contractility and migration dynamics [104]. Beyond the Rho family, integrins play a pivotal role in mediating adhesion to the ECM, translating external signals into intracellular pathways that regulate migration [105]. Dysregulation of these molecular pathways can enhance migratory capabilities, contributing to tumor invasiveness and metastasis.

The mammary microenvironment provides both biochemical and physical cues that shape epithelial cell migration [5]. ECM components such as collagen and fibronectin offer structural guidance, while mechanical properties like stiffness modulate integrin-mediated adhesion and downstream signaling pathways [5]. Stromal fibroblasts secrete growth factors, including hepatocyte growth factor (HGF) and EGF enhance epithelial cell motility (**Figure 1-9D**) [55], [66], [83]. Immune cells such as macrophages contribute by releasing matrix-degrading enzymes and pro-migratory chemokines, creating permissive tracks for migration [106].

Chemokines are critical regulators of epithelial cell migration, exerting their effects through G protein-coupled receptors (GPCRs) [55], [107]. CXCL12 (SDF-1) and its receptor CXCR4 represent a well-studied axis in mammary epithelial and cancer cell migration [107]. CXCL12 gradients guide cells during development and tumor metastasis by activating pathways such as PI3K-AKT and MAPK, which promote cytoskeletal reorganization and directional movement. Additionally, CXCR2 activation has been implicated in driving epithelial and cancer cell migration. CXCR2, responding to ligands like CXCL1 and CXCL8 (IL-8), enhances actin remodeling and promotes migratory phenotypes through similar

intracellular signaling cascades [108]. CXCR2 and other GPCRs also activate the RHO/ROCK axis, further regulating cytoskeletal contractility and cell motility.



**Figure 1-9: Modes and mechanisms driving cellular migration and metastasis**

**A)** Cells can migrate through three modes which involve distinct actin, integrin and cell-junction dynamics. **B)** Organoids derived from transgenic murine mammary tumors invade via CCM when cultured in collagen. The leading cells of the invading strands require expression of KRT14. **C)** Activation of Rho, Rac, and Cdc42 in fibroblast cells demonstrate the roles of each protein in control actin localization and contractility. **D)** EGF treatment induces cell migration in MCF10A mammary epithelial cells. Figure panel A adapted from [89]. Figure panel B adapted from [94]. Figure panel C adapted from [58]. Figure panel D adapted from [83].

### 1.3.3 The epithelial-to-mesenchymal transition

Epithelial-to-mesenchymal transition (EMT) is a fundamental biological process wherein epithelial cells lose their characteristic polarity and intercellular junctions to acquire a more mesenchymal phenotype, characterized by increased motility and invasiveness [58]. In normal physiology, EMT is pivotal during embryonic development, contributing to processes such as gastrulation, neural crest migration, and organogenesis [109], [110], [111]. In these contexts, the reversible nature of EMT allows cells to adopt transient migratory states before differentiating into specific lineages. In the mammary gland, EMT contributes to ductal remodeling and branching morphogenesis during development, as well as tissue repair and regeneration following injury. Despite its essential roles in normal physiology, aberrant activation of EMT is implicated in pathological states, particularly in cancer progression and metastasis [58].

The induction of EMT involves a network of transcription factors and signaling pathways that orchestrate the phenotypic shift from epithelial to mesenchymal states. Central to this process are transcription factors such as Snail, Slug, Twist, and members of the ZEB family (ZEB1 and ZEB2) [112]. These factors repress the expression of epithelial markers, including E-cadherin, a critical component of adherens

junctions, while promoting the expression of mesenchymal markers such as vimentin and N-cadherin. The downregulation of E-cadherin disrupts cell-cell adhesion, facilitating cellular detachment and migration [113].

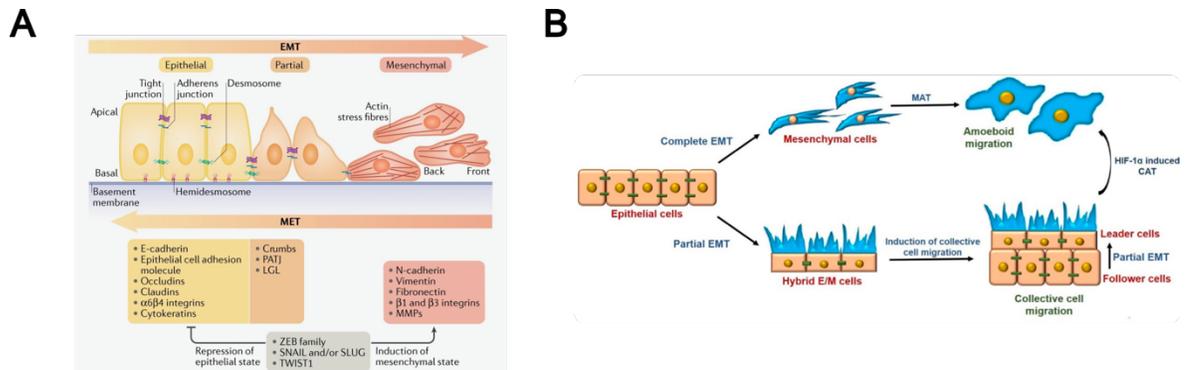
Key signaling pathways regulating EMT include the Wnt/ $\beta$ -catenin, Notch, and Hedgehog pathways [114]. Activation of the Wnt/ $\beta$ -catenin pathway stabilizes  $\beta$ -catenin, which translocates to the nucleus to drive the transcription of EMT-promoting genes. Similarly, Notch signaling enhances the expression of Snail and Slug, further promoting the mesenchymal phenotype [115]. Crosstalk among these pathways amplifies the EMT program [114].

The microenvironment plays a critical role in modulating EMT through both biochemical and biomechanical cues. Among the most potent and well-studied inducers of EMT is transforming growth factor- $\beta$  (TGF- $\beta$ ) [72]. TGF- $\beta$  signaling, through its canonical SMAD-dependent pathway, activates transcription factors like Snail and ZEB, driving the downregulation of epithelial markers and upregulation of mesenchymal genes [70]. Non-canonical TGF- $\beta$  signaling pathways, including those mediated by p38 MAPK and Rho/ROCK, further enhance cytoskeletal remodeling and motility, reinforcing the mesenchymal phenotype [73]. TGF- $\beta$ -induced EMT is not limited to cancer cells but also occurs in stromal and immune cells, shaping the tumor microenvironment to favor invasion and metastasis [64]. Additionally, components of the extracellular matrix (ECM), such as collagen and fibronectin, contribute to EMT by engaging integrin receptors and activating downstream signaling cascades [5]. Mechanical properties of the ECM, including stiffness, further modulate TGF- $\beta$  signaling, emphasizing the interplay between biochemical and physical cues in driving EMT [116].

It is believed that EMT is not a binary process, but rather a continuum of cell states [117]. Partial EMT refers to an intermediate state in which cells exhibit both epithelial and mesenchymal characteristics [118]. Unlike complete EMT, where cells fully transition to a mesenchymal state, partial EMT allows for enhanced plasticity, enabling cells to dynamically adapt to environmental cues. Hypoxia, a common feature of the tumor microenvironment, is a significant driver of partial EMT [118]. Hypoxia-inducible factors (HIFs) stabilize under low oxygen conditions and activate genes associated with mesenchymal traits while preserving certain epithelial features [119]. This hybrid phenotype is particularly relevant in cancer, as cells undergoing partial EMT retain some cell-cell adhesion properties, allowing collective migration [120]. Such plasticity enhances metastatic potential by enabling cells to adapt to varying microenvironmental conditions during dissemination and colonization.

EMT is a critical driver of cancer metastasis, enabling epithelial tumor cells to invade surrounding tissues and enter the bloodstream. During metastasis, EMT facilitates the detachment of cancer cells from the primary tumor, degradation of ECM barriers, and intravasation into the vasculature [109]. Once in circulation, mesenchymal-like cells exhibit increased resistance to anoikis, promoting survival during transit to distant sites [121]. Notably, EMT is often reversible, allowing cells to undergo mesenchymal-to-epithelial transition (MET) at metastatic sites to establish secondary tumors [122]. Collective migration, often driven by cells in a partial EMT state, further enhances metastatic efficiency by preserving intercellular communication and cooperative invasion [118].

In breast cancer, aberrant activation of EMT is associated with poor prognosis, increased tumor aggressiveness, and resistance to therapy [123]. Targeting EMT and its associated signaling pathways represents a potential strategy for mitigating metastasis and improving clinical outcomes.



**Figure 1-10: The continuum of the epithelial-mesenchymal transition**

**A)** Cells undergo EMT through downregulation or repression of epithelial state markers (particularly cell junction proteins) and upregulation of mesenchymal markers (notably Vimentin). Complete EMT results in epithelial cells dissociating from the cohesive tissue and adopting a more migratory cell state. The process is reversible, and a mesenchymal-epithelial transition can occur. **B)** EMT is a continuum of cell states. Partial EMT can occur where cells take on mesenchymal characteristics while retaining an epithelial identity. Adoption of a partial EMT cell state can lead to CCM. Hypoxia has been observed to induce a partial EMT state, mediated through the activation of the HIF1A transcription factor. Figure panel A adapted from [56]. Figure panel B adapted from [118].

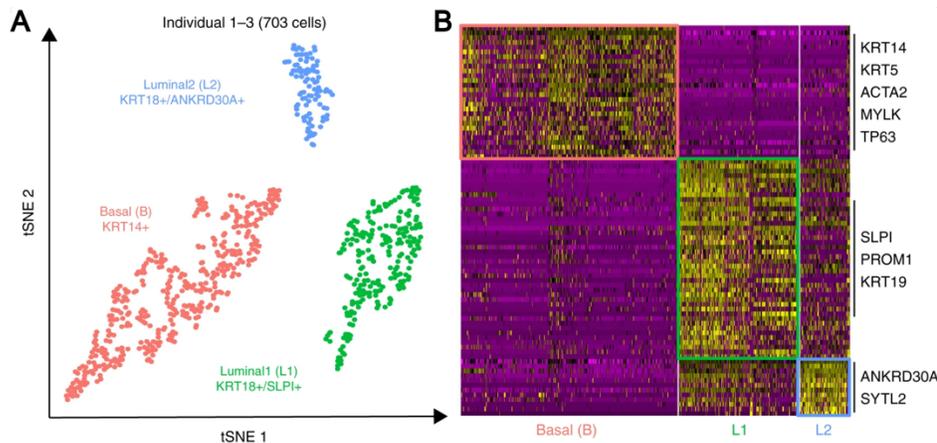
### 1.3.4 Other types of phenotypic modulation

Epithelial cells exhibit a wide array of phenotypes and behaviors beyond proliferation, motility, and epithelial-to-mesenchymal transition (EMT), each contributing uniquely to tissue function. Among these, differentiation is particularly significant, shaping cellular identity and behavior in ways that are essential for maintaining the structure and function of the mammary gland [4].

Differentiation refers to the process by which cells acquire specialized characteristics, enabling them to perform specific functions within a tissue (**Figure 1-11**). In the mammary gland, differentiation allows the gland to fulfill its role in lactation and tissue maintenance [4]. During pregnancy and lactation, mammary epithelial cells undergo extensive differentiation to form secretory alveoli capable of producing and secreting milk [124]. This process involves both structural and functional changes, such as the expansion of cytoplasmic volume to accommodate secretory organelles and the clustering of cells into functional units [125]. Differentiation also establishes polarity within epithelial cells, enabling the directional secretion of milk components into the ductal system [126].

The microenvironment plays a critical role in driving differentiation. Signals from the extracellular matrix (ECM), such as laminin and collagen, interact with integrins on the epithelial cell surface to activate intracellular pathways that promote differentiation [5]. Hormonal cues, particularly from prolactin, drive the expression of genes necessary for milk production and secretion [126]. Stromal fibroblasts and adipocytes contribute by releasing growth factors like insulin-like growth factor (IGF) and vascular endothelial growth factor (VEGF), which further support differentiation and tissue remodeling [2]. Soluble factors such as transforming growth factor-beta (TGF- $\beta$ ) and Wnt ligands also play pivotal roles in addition to ECM and soluble factors, cell-cell interactions are crucial [127]. E-cadherin-mediated junctions facilitate the transmission of differentiation-promoting signals, ensuring that epithelial cells maintain their cohesive and polarized architecture [45].

Beyond differentiation, epithelial cells exhibit a range of phenotypes that influence their behavior and tissue function. These include senescence, a state of irreversible growth arrest associated with aging and tissue repair, and apoptosis, which maintains tissue homeostasis by removing damaged or unnecessary cells [128]. Secretory activity, as seen in mammary epithelial cells during lactation, represents another specialized phenotype [126]. Furthermore, epithelial cells can transiently adopt stem-like properties, enabling tissue regeneration and repair [129]. These phenotypes, shaped by intrinsic programs and microenvironmental signals, enable epithelial cells to meet the diverse functional demands of the tissues they comprise.



**Figure 1-11: Mammary epithelial cells differentiate into three main lineages which can be distinguished through scRNAseq**

A) Dimensional reduction of single-cell RNAseq data collected from human mammary tissue cells reveals three main lineages: Basal, Luminal1, and Luminal2. B) Gene expression markers that define the three lineages of mammary epithelial cells. Figure adapted from [130].

## 1.4 Models and methods for exploring microenvironmental perturbations

Biological phenotypes arise from a complex and multiscale network of interwoven processes. Recent developments in sequencing technology, multiplexed imaging methods, and computational power have enabled the simultaneous measurement of dozens to thousands of features in a single assay with single cell resolution. These ‘high dimensional’ methods for characterizing biological systems necessitate the development of advanced analytics that scale adequately with the number of measured features or cells without loss in signal.

### 1.4.1 Experimental models

The study of microenvironmental perturbations and their impact on epithelial cells relies on a variety of experimental models. Each system offers unique advantages and limitations, which make them suitable for addressing specific biological questions. Commonly used models include immortalized cell lines, primary cells, organoids, and animal models, among others [131].

Immortalized cell lines, such as MCF10A, are widely used to study mammary epithelial cell biology [132]. MCF10A cells are non-tumorigenic mammary epithelial cells derived from human breast tissue. They are particularly valuable because they can be cultured in both two-dimensional (2D) monolayers and three-dimensional (3D) systems, allowing researchers to investigate how microenvironmental factors influence cell phenotype in different contexts [133]. In 2D cultures, MCF10A cells grow as a flat, single layer of

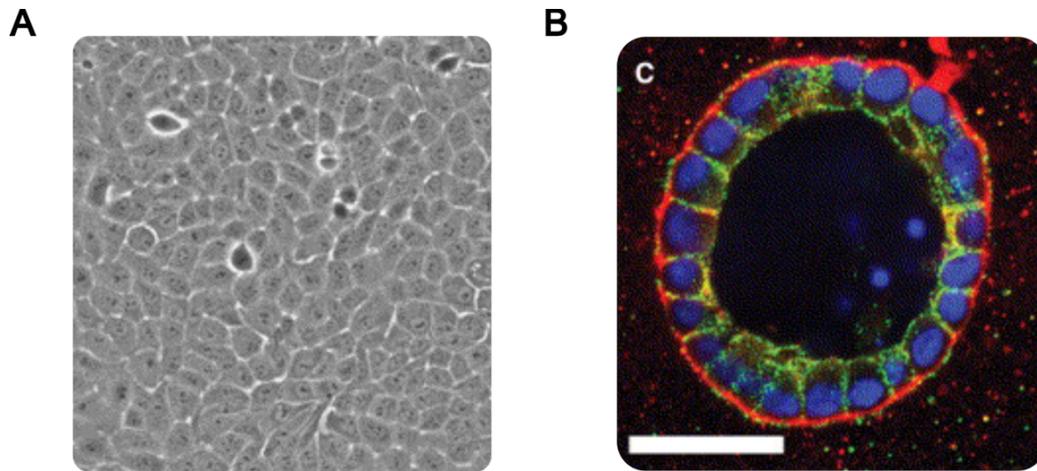
cells, in a characteristic cobblestone pattern, providing a simplified environment to study cell signaling and proliferation (**Figure 1-12A**) [134]. In 3D culture, these cells form acinar structures that better recapitulate in vivo tissue architecture, enabling studies of cell polarization, differentiation, and interaction with the extracellular matrix (ECM) (**Figure 1-12B**) [135], [136], [137]. However, cell lines like MCF10A lack the complexity of primary tissue and may not fully mimic in vivo conditions [134].

Organoids are 3D culture systems derived from primary cells or stem cells that self-organize into structures resembling the tissue of origin [138]. For mammary epithelial cells, organoids can model the branching morphogenesis and functional differentiation observed in vivo [139]. Organoids retain key cellular heterogeneity and microenvironmental interactions, making them powerful tools for studying ligand signaling, epithelial-to-mesenchymal transition (EMT), and other phenotypic changes [140]. Despite their advantages, organoids are technically demanding to culture and often lack the immune and stromal components present in the native tissue.

Mouse models are indispensable for studying microenvironmental perturbations in the context of a living organism [141]. Genetically engineered mouse models (GEMMs) allow for tissue-specific manipulation of genes involved in epithelial-microenvironment interactions, while xenograft models enable the study of human cells in vivo [142]. These systems provide insights into the systemic effects of microenvironmental changes, such as immune cell recruitment and vascular remodeling [143]. However, species-specific differences and the high cost of animal studies are significant limitations.

Interpreting epithelial cell function and phenotype relies on a diverse collection of tools and technologies. Microscopy is a cornerstone for analyzing cell phenotypes across experimental models [144]. The choice of imaging technique depends on the specific model and the biological questions being addressed. Confocal microscopy is commonly used in 2D and 3D cultures to visualize cell morphology, junctions, and protein localization with high spatial resolution [145]. Live-cell imaging enables the study of dynamic processes such as cell migration and division in real time. For example, time-lapse imaging of MCF10A cells in 3D culture can reveal how microenvironmental cues drive acinar formation or invasive behavior [146], [147]. Super-resolution microscopy provides enhanced detail for studying nanoscale structures, such as focal adhesions and cytoskeletal organization [148]. Multiphoton microscopy is particularly useful for imaging deeper into 3D organoids or tissue samples, capturing cellular interactions within the ECM [149]. Additionally, cyclic immunofluorescence (cyclIF) allows for multiplexed imaging of numerous biomarkers, offering a powerful tool to study cellular heterogeneity and signaling dynamics within tissues [150].

Quantifying imaging data to assess cellular phenotypes presents significant challenges, particularly in complex systems such as 3D organoids or tissue sections [87]. Computational tools have become essential for extracting meaningful insights from these datasets. Advanced algorithms for cell segmentation and tracking enable precise quantification of cell shape, movement, and division over time [87], [151]. Machine learning and deep learning approaches are increasingly being employed to identify patterns and classify phenotypes in high-dimensional datasets [152]. These tools are evolving rapidly, enabling researchers to address questions about cellular behavior and microenvironmental interactions with greater precision and scalability.



**Figure 1-12: MCF10A cells cultured in 2D and 3D conditions**

**A)** MCF10A cells cultured in 2D form a cobblestone pattern as they reach confluency. **B)** MCF10A cells can also be cultured in matrigel to investigate 3D structural organization of the mammary gland. Twenty days after cell seeding, MCF10A acinar structures form surrounded by a basement membrane (red) and hollow lumen. Cells are immunostained for laminin V (red), DAPI (blue) and activated caspase 3 (green). Bar = 25 $\mu$ m. Figure adapted from [133]

#### 1.4.2 Classical biochemical techniques

Biochemical approaches grounded in prior knowledge are fundamental to unraveling the molecular mechanisms underlying cell phenotype modulation. These methodologies often begin with hypothesis-driven investigations that utilize classical techniques to isolate and characterize signaling pathways and molecular interactions. For example, protein interaction studies using co-immunoprecipitation (Co-IP) and Western blotting enable the identification of signaling complexes, while kinase assays provide insights into pathway activation and downstream targets [153].

One of the most impactful applications of these approaches in mammary biology involves the study of transforming growth factor-beta (TGF $\beta$ ) [127]. TGF $\beta$  is a pivotal microenvironmental ligand that influences epithelial cell behavior, particularly in the context of EMT and cancer progression. Classic studies using mammary epithelial cells demonstrated that TGF $\beta$  signaling is mediated by its receptor complexes, which phosphorylate SMAD proteins to regulate gene transcription [70]. Biochemical techniques, such as SMAD-binding assays and transcriptomic profiling, have revealed how TGF $\beta$ -induced changes in gene expression drive the loss of cell-cell adhesion and the acquisition of migratory phenotypes [154]. These biochemical approaches have also led to translational advances, such as the development of small-molecule inhibitors targeting the TGF $\beta$ /SMAD pathway [155]. By blocking key steps in this signaling cascade, these inhibitors hold promise for limiting tumor progression and metastasis.

Through these classic biochemical experiments, researchers can begin to piece together the puzzle of how cells respond to their microenvironment and alter their behavior. However, this task is monumental. Cells can express over 20,000 proteins, each potentially modified by complex interactions, activities, and post-translational changes such as phosphorylation [156]. Each piece of evidence from these experiments offers valuable hints about how a protein or pathway fits into the broader signaling

network. Nevertheless, integrating these pieces into a comprehensive model of cellular behavior remains a significant and ongoing challenge for the field.

### 1.4.3 Unbiased high-throughput assays

Understanding how cells respond to microenvironmental perturbations requires methods capable of addressing the immense complexity of cellular signaling. Each cell expresses tens of thousands of proteins, many of which interact dynamically with one another and are influenced by extracellular cues [156]. Classical approaches, while powerful, often target individual proteins or pathways, leaving vast regions of the signaling network unexplored. To overcome these limitations, researchers have developed high-throughput, unbiased methodologies that can simultaneously assess numerous molecular interactions and cellular responses [157]. These approaches have become indispensable for uncovering how cells sense and adapt to their environments.

Among the most transformative advances in biological research has been the development of sequencing technologies [158]. High-throughput sequencing enables researchers to study gene expression, mutations, and epigenetic modifications across the entire genome. Bulk RNA sequencing (RNA-seq) analyzes average gene expression in a population of cells, providing insights into cellular responses to specific stimuli. However, bulk approaches mask the heterogeneity within cell populations. Single-cell RNA sequencing (scRNA-seq) has emerged as a powerful tool to address this challenge [159]. By analyzing gene expression in individual cells, scRNA-seq reveals how cells within a tissue or tumor differ in their signaling states, phenotypes, and interactions.

Single-cell sequencing has relevance to understanding microenvironmental perturbations in cancer. For example, scRNA-seq can uncover how immune cells interact with epithelial cells within the tumor microenvironment, revealing communication networks that drive tumor progression or immune evasion [160]. The generation of large-scale scRNA-seq atlases, such as The Cancer Genome Atlas (TCGA), has made it trivial to ask previously difficult questions, such as which cells express specific ligands in breast tumors, and which cells express the cognate receptors capable of responding to those ligands (**Figure 1-13A-B**) [161]. When combined with spatial transcriptomics, which retains the physical context of cells within tissues, researchers gain insights into how the microenvironment shapes cellular behavior in situ. These approaches allow for unprecedented resolution in studying complex cell-cell interactions and tissue organization.

The typical workflow for single-cell RNA sequencing involves several key steps: isolating individual cells, capturing and sequencing their RNA, and analyzing the resulting data [159]. This process begins with dissociating tissues into single cells, often using enzymatic digestion. Isolated cells are then encapsulated in microfluidic devices or sorted into individual wells, where their RNA is captured and reverse-transcribed into complementary DNA (cDNA) [159]. The cDNA is then amplified, sequenced, and analyzed using bioinformatics pipelines.

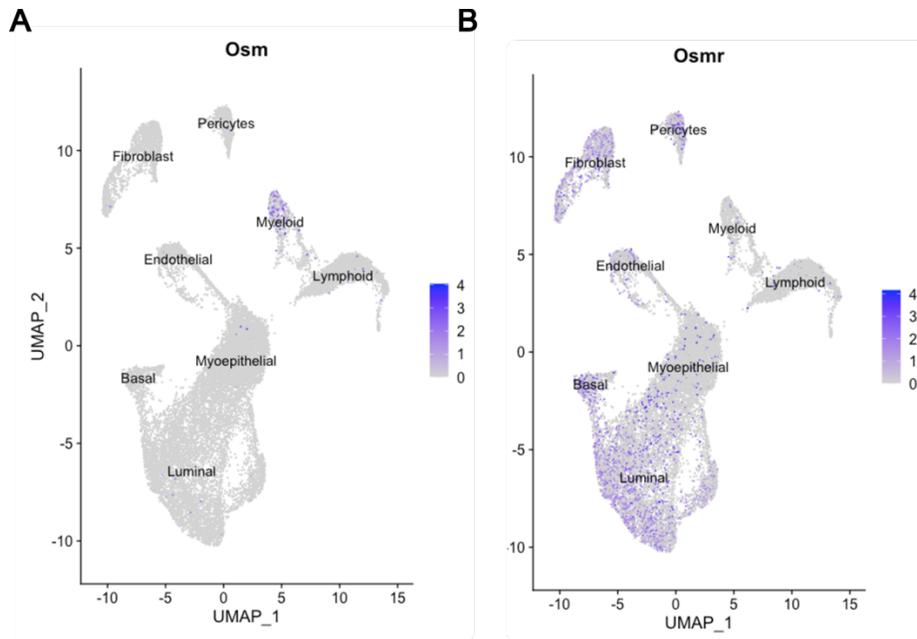
Once the sequencing data is generated, a series of analysis steps are undertaken to interpret the gene expression profiles of individual cells [162]. First, the raw data is filtered to remove low-quality cells, such as those with poor RNA integrity or low sequencing depth. Next, the data undergoes normalization to account for differences in sequencing depth between cells, followed by dimensionality reduction techniques like principal component analysis (PCA) or t-SNE (t-distributed stochastic neighbor embedding) [163]. These methods allow for the reduction of high-dimensional gene expression data into

a smaller number of principal components, making it easier to visualize and interpret. After dimensionality reduction, clustering algorithms such as k-means or graph-based methods (e.g., Louvain or Leiden clustering) are applied to identify subpopulations of cells with similar gene expression profiles [164]. The resulting clusters can be mapped onto known cell types or novel cell states, depending on the research question. Differential gene expression analysis is often performed to identify genes that are upregulated or downregulated in specific clusters, providing insights into the underlying biology of different cell types and states [165].

Despite its power, single-cell analysis presents numerous challenges. Cell dissociation can alter the transcriptome, introducing artifacts. Additionally, the stochastic nature of RNA capture leads to dropout effects, where some transcripts are undetected, complicating data interpretation [166]. The analysis of single-cell datasets also requires advanced computational tools to cluster cells, infer signaling pathways, and identify rare subpopulations [166].

Beyond RNA sequencing, a variety of high-throughput methodologies are used to probe cellular responses and molecular mechanisms. Chromatin accessibility assays, such as ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing), provide insights into the regulatory landscape of cells by identifying regions of open chromatin [167]. In ATAC-seq, transposases insert sequencing adapters into accessible regions of the genome, enabling the identification of active promoters and enhancers. Single-cell adaptations, like sci-ATAC-seq, allow researchers to study chromatin accessibility in individual cells, offering a deeper understanding of how epigenetic states influence cellular behavior [168]. Mass spectrometry-based proteomics represents another high-throughput approach, enabling the identification and quantification of thousands of proteins within a sample [169].

While these high-throughput assays are incredibly powerful, they generate vast datasets that must be carefully interpreted to uncover physiologically relevant signals. The ability to measure nearly all genes, proteins, or regulatory elements in a cell or tissue is a significant achievement, but identifying which components are critical for specific phenotypic outcomes remains a major challenge. Prioritizing signals for further investigation often rely on integrating data from multiple assays, leveraging computational models, and validating findings through focused biochemical or genetic experiments.



**Figure 1-13: Single cell Oncostatin M gene expression in murine mammary tumors**

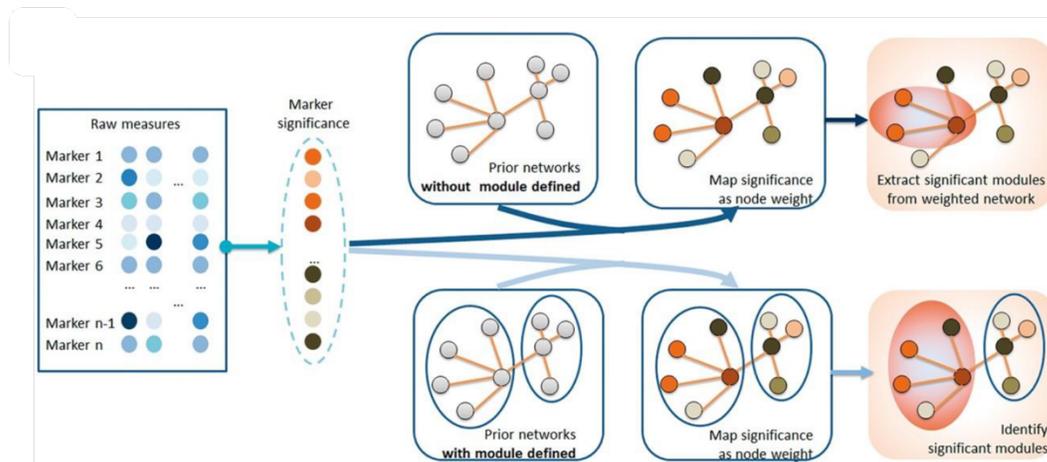
**A).** OSM is exclusively expressed by myeloid cells in mammary tumors harvested from MMTV-PyMT mice. Color bar represents scaled expression. **B).** The cognate receptor for OSM (OSMR) is expressed by the basal and luminal lineages of epithelial cancer cells, fibroblasts, pericytes, and endothelial cells, suggesting these cell types can respond to OSM secreted in the tumor microenvironment. Data represented in the figure adapted from [170].

#### 1.4.4 Systems biology approaches

Systems biology approaches aim to understand the molecular mechanisms underlying changes in cell phenotype by integrating data from multiple layers of biological information [171]. Unlike traditional reductionist methods that focus on single molecules or pathways, systems biology seeks to capture the complexity of cellular systems. This approach acknowledges that biological systems are highly interconnected, where small perturbations in one part of a network can lead to large-scale changes in cellular behavior. Systems biology, therefore, focuses on modeling the intricate networks of interactions between genes, proteins, and other biomolecules, ultimately providing a more holistic view of cellular responses to microenvironmental cues [171].

General strategies in systems biology often involve attempts to unite prior knowledge with high-throughput data to build comprehensive models of cellular behavior [172]. One common approach is the use of network analysis, where prior knowledge of molecular interactions is integrated with data from high-throughput assays such as RNA-seq, proteomics, or metabolomics (Figure 1-14). [173]. This allows researchers to generate network models that map how genes, proteins, and other molecules interact in response to perturbations. For instance, integrating gene expression data with known protein-protein interaction (PPI) networks can help identify key signaling hubs or pathways that drive changes in cell phenotype [174]. Another approach involves single-agent modeling, where the response of a cell to a single perturbation factor (like a drug, gene knockout, or environmental change) is modeled to predict how it will affect the behavior of individual cells or tissues [83]. Transcription factor (TF) enrichment analysis is also commonly employed, allowing researchers to identify which transcription factors are activated or repressed in response to specific stimuli [175].

One example of a systems biology approach is causal pathway analysis, which is designed to uncover the causal relationships between molecular components in complex biological systems [176]. Causal pathway analysis integrates data from high-throughput assays with prior knowledge of molecular interactions and network structures derived from literature sources to generate hypotheses about causal relationships within the system. This approach uses computational models to infer the directionality and strength of causal links between genes, proteins, and other molecular factors. By identifying key driver nodes and predicting how they influence cellular phenotypes, causal pathway analysis can provide deeper insights into the molecular underpinnings of cellular phenotypes.



**Figure 1-14: General workflow for elucidating molecular mechanisms through network-based analysis**

Raw data from high-throughput experimental assays is filtered to identify markers that define a perturbed condition or diseased state. Filtered data is then computationally analyzed using an unbiased approach (Top) or using known biological cell signaling networks (Bottom) to identify modules of molecular activity that are altered. Figure adapted from [177].

#### 1.4.5 Microenvironmental signals in combination

Cells *in vivo* are not exposed to a single signal at a time. Instead, they exist within a complex microenvironment filled with diverse biochemical and biophysical cues, including growth factors, extracellular matrix components, cytokines, and mechanical forces. These inputs are not isolated; cells must process and integrate multiple signals simultaneously to generate coherent responses. Understanding how cells combine and integrate multiple perturbations is critical to unraveling the mechanisms underlying cellular behavior. Is there crosstalk between signaling pathways activated by different inputs, or do these pathways operate independently? Do the effects of multiple signals exhibit synergy, producing outcomes greater than the sum of their individual effects? Conversely, are there cases of antagonism, where the combined effect is less than expected? Addressing these questions requires both experimental and computational approaches capable of disentangling the complexities of multi-signal integration.

One area where synergy has been studied extensively is in drug responses [178]. Phenotypic synergy, particularly in the context of drug combinations, refers to cases where the combined effect of two or more drugs exceeds what would be predicted based on their individual effects [179]. Several quantitative frameworks have been developed to define and measure drug synergy. The highest single agent (HSA) model compares the combined effect of two drugs to the effect of the most potent single drug in the

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

combination [179]. Synergy is observed if the combined response exceeds the response of the best-performing drug alone. The Bliss independence model, on the other hand, assumes that two drugs act independently and calculates the expected combined effect based on this assumption [180]. Excess over Bliss quantifies the deviation of the observed combined effect from the expected effect under Bliss independence, with positive deviations indicating synergy. Another common framework is Loewe additivity, which assumes that two drugs with similar mechanisms of action should have additive effects, and deviations from this baseline are used to infer synergy or antagonism [181]. Each of these models provides a unique lens through which to study drug combinations, helping researchers identify synergistic interactions with therapeutic potential.

While these frameworks are well-established for understanding synergy in simple phenotypic outcomes such as cell survival or death, they do not generalize easily to more complex cellular phenotypes. For example, how do we define synergy in processes like epithelial-to-mesenchymal transition, changes in cell morphology, or shifts in cellular signaling states? Similarly, there is no standardized framework for understanding **molecular synergy**—synergistic effects observed at the level of gene expression, protein activity, or signaling pathway activation. Yet, molecular synergy is highly relevant because it has been shown to correlate with phenotypic synergy in multiple studies [182]. Developing quantitative and conceptual frameworks to study molecular synergy is an important frontier in systems biology. Such frameworks could provide deeper mechanistic insights into how multiple inputs are integrated within cells and ultimately drive complex phenotypic changes.

## 2. Chapter II: A multi-omic analysis of MCF10A cells provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

The following chapter has been adapted from Gross et al (2022), [10.1038/s42003-022-03975-9](https://doi.org/10.1038/s42003-022-03975-9).

Ian McLean contributed to this manuscript through:

1. Analysis and phenotypic quantification of CyclF and live-cell imaging datasets
2. Assessment of assay variance
3. Comparisons between assays
4. Examination of module activity to elucidate the molecular basis of ligand-induced phenotypic responses
5. OSM validation experiments

**Authors:** Sean M. Gross<sup>1</sup>, Mark A. Dane<sup>1</sup>, Rebecca L. Smith<sup>1</sup>, Kaylyn L. Devlin<sup>1</sup>, **Ian McLean**<sup>1</sup>, Daniel S. Derrick<sup>1</sup>, Caitlin E. Mills<sup>2</sup>, Kartik Subramanian<sup>2</sup>, Alexandra B. London<sup>3</sup>, Denis Torre<sup>3</sup>, John Erol Evangelista<sup>3</sup>, Daniel J.B. Clarke<sup>3</sup>, Zhuorui Xie<sup>3</sup>, Cemal Erdem<sup>4</sup>, Nicholas Lyons<sup>5</sup>, Ted Natoli<sup>5</sup>, Sarah Pessa<sup>5</sup>, Xiaodong Lu<sup>5</sup>, James Mullahoo<sup>5</sup>, Jonathan Li<sup>6</sup>, Miriam Adam<sup>6</sup>, Brook Wassie<sup>6</sup>, Moqing Liu<sup>1</sup>, David F. Kilburn<sup>1</sup>, Tiera A. Liby<sup>1</sup>, Elmar Bucher<sup>1</sup>, Crystal Sanchez-Aguila<sup>1</sup>, Kenneth Daily<sup>7</sup>, Larsson Omberg<sup>7</sup>, Yunguan Wang<sup>2</sup>, Connor Jacobson<sup>2</sup>, Clarence Yapp<sup>2</sup>, Mirra Chung<sup>2</sup>, Dusica Vidovic<sup>8,9,10</sup>, Yiling Lu<sup>11</sup>, Stephan Schurer<sup>8,9,10</sup>, Albert Lee<sup>12</sup>, Ajay Pillai<sup>13</sup>, Aravind Subramanian<sup>5</sup>, Malvina Papanastasiou<sup>5</sup>, Ernest Fraenkel<sup>5,6</sup>, Heidi S. Feiler<sup>1</sup>, Gordon B. Mills<sup>14,15</sup>, Jake D. Jaffe<sup>5</sup>, Avi Ma'ayan<sup>3</sup>, Marc R. Birtwistle<sup>4</sup>, Peter K. Sorger<sup>2</sup>, James E. Korkola<sup>1,15</sup>, Joe W. Gray<sup>1,15</sup>, Laura M. Heiser<sup>1,15,\*</sup>

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

<sup>1</sup>Department of Biomedical Engineering, OHSU, Portland, OR, USA

<sup>2</sup>Laboratory of Systems Pharmacology, Department of Systems Biology, Harvard Program in Therapeutic Science, Harvard Medical School, Boston, MA, USA

<sup>3</sup>Department of Pharmacological Sciences, Mount Sinai Center for Bioinformatics, Icahn School of Medicine at Mount Sinai, New York, NY, USA

<sup>4</sup>Department of Chemical and Biomolecular Engineering, Clemson University, Clemson, SC, USA

<sup>5</sup>Broad Institute of MIT and Harvard, Cambridge, MA, USA

<sup>6</sup>Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

<sup>7</sup>Sage Bionetworks, Seattle, WA, USA

<sup>8</sup>Sylvester Comprehensive Cancer Center, University of Miami, FL 33136, USA

<sup>9</sup>Department of Molecular and Cellular Pharmacology, Miller School of Medicine, University of Miami, Miami, FL 33136, USA

<sup>10</sup>Institute for Data Science & Computing, University of Miami, FL 33136, USA

<sup>11</sup>Department of Genomic Medicine, Division of Cancer Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

<sup>12</sup>Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, USA

<sup>13</sup>Human Genome Research Institute, National Institutes of Health, Bethesda, USA

<sup>14</sup>Division of Oncological Sciences, OHSU, Portland, OR, USA

<sup>15</sup>Knight Cancer Institute, OHSU, Portland, OR, USA

\*Correspondence: [heiserl@ohsu.edu](mailto:heiserl@ohsu.edu) ORCID: 0000-0003-3330-0950

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

## 2.1 Abstract

The phenotype of a cell and its underlying molecular state is strongly influenced by extracellular signals, including growth factors, hormones, and extracellular matrix proteins. While these signals are normally tightly controlled, their dysregulation leads to phenotypic and molecular states associated with diverse diseases. To develop a detailed understanding of the linkage between molecular and phenotypic changes, we generated a comprehensive dataset that catalogs the transcriptional, proteomic, epigenomic and phenotypic responses of MCF10A mammary epithelial cells after exposure to the ligands EGF, HGF, OSM, IFNG, TGFB and BMP2. Systematic assessment of the molecular and cellular phenotypes induced by these ligands comprise the LINCS Microenvironment (ME) perturbation dataset, which has been curated and made publicly available for community-wide analysis and development of novel computational methods ([synapse.org/LINCS\\_MCF10A](https://synapse.org/LINCS_MCF10A)). In illustrative analyses, we demonstrate how this dataset can be used to discover functionally related molecular features linked to specific cellular phenotypes. Beyond these analyses, this dataset will serve as a resource for the broader scientific community to mine for biological insights, to compare signals carried across distinct molecular modalities, and to develop new computational methods for integrative data analysis.

## 2.2 Introduction

The function of cells and their organization into tissues is controlled by interactions between cell-intrinsic molecular networks and cell-extrinsic signals, while dysregulation of these signals is associated with various diseases [183]. Extracellular ligands activate cell surface receptors to modulate chromatin, RNA, and protein networks that induce changes in multiple cellular phenotypes including viability, growth rate [184], motility [185], polarization, and differentiation state [186]. Disease-specific studies—especially those focused on cancer—have concentrated on understanding phenotypes related to disease progression, resistance mechanisms, therapeutic vulnerabilities, and molecular predictors of response [187], [188], [189], [190], [191], [192], [193]. Several canonical signaling pathways have been linked to distinct normal and disease-associated cellular phenotypes, including MAPK [194], JAK/STAT, WNT [195], and TGFB [196]. However, a detailed mapping of the linkage between multi-modal molecular and phenotypic responses underlying cell state regulation, developmental processes and diverse diseases is lacking.

Two general approaches have been used to explore the role of extracellular signals in modulating cellular and molecular phenotypes. One approach involves systematic large-scale perturbation of panels of immortalized cell lines, which has yielded libraries of response signatures [187], [189], [190], [191], [192], [194], [197], [198], [199]. The other approach involves more focused assessment of phenotypic and molecular changes in more complex model systems, including engineered organoids [198], [199], flies [200], [201], worms [201], fish [202] and mice [203]. Together these studies indicate that comprehensive multi-omic assessment of perturbation responses are critical for gaining insights into molecular-phenotype relationships. From this work, module analysis of multi-omic molecular data has proven a powerful approach to identify co-regulated molecular features associated with normal [204], [205], [206] and disease-associated [207] phenotypes. Such data-driven approaches require comprehensive, systematically generated datasets, and in recognition of this, multiple data generation consortia have emerged over the past 20 years, including ENCODE [208], TCGA [209], GTEx [210], and HuBMAP [211].

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

The Library of Integrated Network-based Cellular Signatures (LINCS) consortium study presented here is a large-scale, cell line-based perturbation experiment designed to examine the molecular and phenotypic responses of normal cells to perturbations. Its uniqueness lies in the coordinated measurements of many different cellular and molecular responses to biologically relevant ligands that, when studied together, can be used for systems-level analysis of microenvironmental responses. Here we focused on the well-characterized human mammary epithelial MCF10A cell line [132], which is a non-transformed cell line that exhibits many of the key hallmarks of epithelial biology, including migration [212], [213] and organoid formation [214], [215]. It is also easily manipulated in a variety of assays including live-cell imaging [216], knock-down [217], and chemical perturbation [217], and therefore is commonly used for cell biology studies. The combination of molecular and cellular properties, as well as its wide adoption in the biomedical research community, provided the rationale for using MCF10A in these studies. Importantly, the focus on a single cell line provided a controlled cell-intrinsic genetic context, which afforded molecular and temporal density in experimental measurements and assessment of multiple perturbations across a variety of assays. We studied responses to six ligands that activate different canonical signaling pathways of biological and clinical relevance, enabling comparison of distinct molecular and phenotypic effects. These data are publicly available for community study at [synapse.org/LINCS\\_MCF10A](https://synapse.org/LINCS_MCF10A). The following sections describe and evaluate the information content of the LINCS ME perturbation dataset and present illustrative analyses showing how the dataset can be used to (a) elucidate molecular and cellular phenotypes that are influenced by the binding of specific ligands, (b) identify ligand-induced signatures that can be mined for biological insights, (c) discover candidate causal or functional relationships between molecular features with module analysis, and (d) identify molecular programs that control specific cellular phenotypes.

## 2.3 Methods

### General Considerations

The technical reproducibility of a data resource such as the one we described here is paramount. To support the development of a robust resource, we carefully planned all experiments to minimize technical artifacts and batch effects. Some aspects of the design of this data generation exercise were piloted in an earlier LINCS-wide study of reproducibility that we published jointly with co-authors of this manuscript (Niepel, et al. *Cell Systems* 2019 [218]). Specifically, in this study, we considered the following, which are described in more detail in the subsequent sections: (1) Cell line evolution (drift): Whenever possible, cell culture was performed at OHSU to minimize technical variation. Given the nature of the CyCIF assay, it was necessary for HMS to perform cell culture at their site. To control for cell line evolution, several cell aliquots were frozen down at OHSU prior to the start of the experiment. These aliquots were shared with HMS for CyCIF data generation. For each sample collection (described below), we used a fresh aliquot of cells and ensured that cells were minimally passaged during sample generation. (2) Reagent batch-to-batch variation: To minimize variation due to reagents, common stocks of media and ligands were used for all sample generation at OHSU and HMS. (3) Cell culture protocols: OHSU and HMS used common cell culture protocols to minimize technical and biological differences. (4) Experimental collections: The large number of cells required for each assay necessitated that we split the gathering of samples into different collections to ensure feasibility of cell culture, treatment, and harvest. Each collection had at least three biological replicates that took approximately a month to generate. Details about which assays were included in each collection are shown in **Figure 2-1E**. To test for consistency across collections, we performed functional analyses on each of the biological replicates

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

and found that they were broadly similar. Results of comparison from collections 1 and collections 2 are shown in **Figure 2-2** and indicate concordant responses.

### **Methodological Rationale**

A comprehensive study of how cells modulate their cellular and phenotypic responses to extracellular signals is critically important for understanding a variety of biological processes including cell state control, development, and diseases such as cancer. This includes identification of the molecular networks that are directly modulated, the duration and extent of modulation, how one perturbation compares to another, and identification of feedback mechanisms. Additionally, identification of the molecular networks that underlie phenotypic responses such as cell migration or proliferation remains challenging; for example, a TGFB network is not synonymous with a proliferation network despite TGFB treatment modulating proliferation.

These questions on ligand and phenotype networks have been difficult to address because they require identifying a sufficient range of perturbations that modify multiple phenotypes in a single cell type and then using experimental and quantitative approaches that can isolate the underlying networks from secondary responses (feedback) and multiple complex phenotypic responses (migration and proliferation are both stimulated by EGF). Furthermore, these experiments are difficult to conduct across cell lines because a ligand perturbation in one cell type may not be equivalent to a ligand perturbation in a second cell type due to differences in the receptors that are expressed, the abundance of those receptors, downstream signaling components, transcription factors, and the underlying state of the cell. In addition, without a large reference dataset it remains unclear the number and type of perturbations to analyze, the optimal time points to collect, the type of assays to measure, and what bioinformatic tools are necessary to integrate all this information identify these networks. An additional challenge is that it is inherently difficult to generate comprehensive multi-omic data as it requires expertise in the collection and analysis of each individual data type as well as development of methods to integrate data types together.

Motivated by this, we leveraged the LINCS consortium, comprised of multiple laboratories with diverse expertise, to create a comprehensive dataset on a single cell type that would be of broad use to the research community to mine for biological insights, develop novel computational analyses, and to serve as a guide of considerations for building multi-omic perturbation data sets. To maximize the richness of the resultant data resource, we decided to test multiple perturbations in a single cell line, which provides several advantages over testing fewer perturbations in multiple cell lines. First, this increases experimental tractability as perturbagens and assay growth conditions need only be optimized for a single sample, and second, the starting state of cells is the same in all samples, which enables a range of responses to be compared and leveraged against each other to isolate individual networks associated with different phenotypic responses. One disadvantage of using a single cell line is that it is not possible to directly address what portion of a perturbation response is cell type specific compared to the portion that is conserved across multiple cell types. Balancing these considerations with available resources, we chose to use a single cell type for this study. Our approach enabled isolation of primary from secondary response; for example, we were able to identify molecular changes specific to EGF and separate these from changes associated with secondary effects such as proliferation. This type of approach has been deployed for deep analysis of other model systems, including *Drosophila* [189] and *C. elegans* [201], [202].

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

### Cell Culture Methods

To decrease unwanted biological variation and ensure comparable results across data types, MCF10A cells were frozen in a single batch at the MD Anderson Cancer Center and used by both OHSU and HMS from the frozen batch with limited passaging. Cell identity was confirmed by short tandem repeat (STR) profiling and cells tested negative for mycoplasma.

Two media formulations were used in these experiments. For routine growth and passaging cells were cultured in growth media (GM) composed of DMEM/F12 (Invitrogen #11330-032), 5% horse serum (Sigma #H1138), 20 ng/ml EGF (R&D Systems #236-EG), 0.5 µg/ml hydrocortisone (Sigma #H-4001), 100 ng/ml cholera toxin (Sigma #C8052), 10 µg/ml insulin (Sigma #I9278), and 1% Pen/Strep (Invitrogen #15070-063). For perturbation experiments, we used growth factor free media—which we termed experimental media (EM)—that was composed of DMEM/F12, 5% horse serum, 0.5 µg/ml hydrocortisone (Sigma #H-4001), 100 ng/ml cholera toxin (Sigma #C8052), and 1% Pen/Strep (Invitrogen #15070-063). For each experiment, MCF10A cells were grown to 50-80% confluence in GM and detached using 0.05% trypsin-EDTA (Thermo Fisher Scientific 25300-054). Following detachment, 75,000 cells were seeded into collagen-1 (Cultrex #3442-050-01) coated 8-well plates (Thermo Fisher Scientific 267062) in GM. Six hours after seeding, cells were gently washed with PBS and EM was added. Following 18 hours of incubation in EM, cells were treated with ligand in fresh EM media as follows: 10 ng/ml EGF (R&D Systems #236-EG), 40 ng/ml HGF (R&D Systems #294-HG), 10 ng/ml OSM (R&D Systems #8475-OM), 20 ng/ml BMP2 (R&D Systems #355-BM) + 10 ng/ml EGF, 20 ng/ml IFN $\gamma$  (R&D Systems #258-IF) + 10 ng/ml EGF, 10 ng/ml TGF $\beta$  (R&D Systems #240-B) + 10 ng/ml EGF. The addition of ligand started the experimental clock. Samples were then collected at 1, 4, 8, 24 or 48H following ligand addition as shown in **Figure 2-1**.

Eight-well plates were coated with 20 µg/cm<sup>2</sup> collagen-1 in a mixture that mimicked the buffering and structural characteristics of MEMA spots: 200 µg/ml collagen-1 (Cultrex #3442-050-01), 10% v/v glycerol (Sigma G5516), 5 mM EDTA pH 8 (Invitrogen 15575), and 100 mM Tris-HCl pH 7.2 (Sigma T2069) in PBS. Plates were rocked at RT for 1 hour. Remaining coating mixture was gently aspirated, and plates were washed twice with sterile PBS. Wells were allowed to dry completely by leaving the plate uncovered in a laminar flow hood before being stored in a benchtop desiccator for a minimum of three days and maximum of six months before use.

After identification of the 6 ligand treatments, samples were generated over three collection periods. The first collection was completed at OHSU in the Fall of 2017 when RPPA, RNAseq, ATACseq, L1000, and IF samples were collected. The second collection was completed at OHSU in the Winter of 2018 and included GCP, L1000, and IF samples. The third collection was collected at HMS in the Summer of 2018 and included CyCIF and L1000 samples.

### Microenvironment Microarray (MEMA)

We used previously established high-throughput MEMA screens to identify microenvironmental factors that strongly influence growth [192], [219]. The key aspects of the MEMA assay are comprised of a set of printed insoluble proteins and a panel of soluble ligands. In brief, a panel of 48 insoluble proteins were printed into 8-well cell culture plates with an Aushon printer, forming 350 µm diameter spots on which cells can grow. Each matrix protein was mixed with collagen I to improve printing and cell attachment and printed in ~15 replicate pseudo-random locations. 22,000 cells per well were added to replicate

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

arrays and grown in experimental media for 18H. Following this, the media was exchanged and appropriate concentrations of a panel of 63 soluble ligands were added to each well. To account for the influence of EGF on MCF10A proliferation, we tested one set of arrays with 10 ng/ml EGF and the other without added EGF. Arrays were returned to the incubator for 71 hours, after which 1uM EdU was added to the medium for 1 hour. Cells were then fixed in 2% PFA at RT and stored at 4°C in PBS. After fixation, cells were permeabilized with 0.3% Triton X-100 for 25 minutes at RT. Array-bound cell staining was performed with KRT14 (Abcam, 1:200), CellMask, and DAPI (ThermoFisher, 1:10,000).

Arrays were imaged on a customized automated high content fluorescence microscope platform (Nikon HCA) and resultant image data was output to an OMERO image database [220]. Cells were segmented and intensity levels were calculated using CellProfiler [221]. The resulting MEMA data was preprocessed and normalized using open-source R software available from ([https://github.com/MEP-LINCS/MEP\\_Processing](https://github.com/MEP-LINCS/MEP_Processing)). The spot cell count was based on the DAPI stained nuclei. EdU intensity was auto gated to label cells as EdU+ and the proportion of EdU+ cells in each spot was reported to measure proliferation. Each intensity and morphology signal was independently RUV normalized in a series of matrices with arrays as the rows and spots as the columns [222]. The RUV controls were the residuals created by subtracting the replicate median from each spot value. After RUV normalization, bivariate LOESS normalization was applied to the normalized residuals using the array row and array column as the independent variables. After normalization, the ~15 replicates of each condition were median summarized to the MEP level.

### **MCF10A Dose Optimization**

We used a three-step process to identify ligands and optimize doses for this large-scale perturbation experiment. Importantly, rather than use the same dose concentration for each ligand, we ran pilot studies to identify functionally relevant concentrations. First, we used a high-throughput MEMA screen to identify ligands that modulated proliferation. Second, we prioritized hits from the MEMA screen by selecting a panel of ligands that target diverse receptor classes (cytokine, growth factor, TGFB family) and which targeted highly expressed receptors. Third, for each of the 6 candidate ligands, we performed dose-response studies to identify the relationship between ligand dose and change in cell numbers after perturbation. MCF10A cells were plated on collagen coated 24-well plates in full growth media for 7 hours at which point the media was exchanged for experimental media. Following 18 hours in experimental media, fresh experimental media was added with 7 doses of OSM, EGF, and HGF individually, or with seven doses of BMP2, IFNG, and TGFB in combination with 10ng/ml EGF. After 72 hours in ligand containing media, cells were fixed, stained with DAPI, and imaged on the ScanR microscope. Cell counts from the images were quantified using Cell Profiler and normalized based on the number of cells present in the 10ng/ml EGF condition. These dose-response experiments were performed in biological triplicate. From the resultant curves, we chose supramaximal doses for each ligand treatment, reasoning that this would ensure robust changes in cell number and minimize effects due to ligand depletion over the course of the 48H assay.

### **OSM validation experiments**

To assess responses to JAK/STAT inhibition MCF10A cells were plated in 24-well collagen coated plates. Following the media changes, cells were treated with 10 ng/ml OSM, 10  $\mu$ M ruxolitinib (Selleck Chemicals #S1378) and NuLight Rapid Red Dye (Essen Bioscience #4717) to label nuclei and count cells

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

across time. Cells were then placed in the IncuCyte S3 and imaged every 30 minutes for 48 hours using phase contrast and red fluorescent filter sets. Cell number was quantified in Cell Profiler by counting the number of fluorescent nuclei in each frame and normalizing counts to time 0H.

To assess cell responses to protease inhibitors cells were plated in 24-well collagen coated plates, underwent the standard media changes and then at time 0H treated with 10 ng/ml OSM and either a protease inhibitor cocktail at a 1:400 dilution (Sigma-Aldrich #P1860), 40  $\mu$ M bestatin (Sigma-Aldrich # B8385), 800 nM aprotinin (Sigma-Aldrich # A1153), 10  $\mu$ M E-64 (Sigma-Aldrich # 324890), 1.45  $\mu$ M pepstatin (Sigma-Aldrich # P5318 ). Cells were then placed in the IncuCyte S3 and imaged every 30 minutes for 48 hours.

Phase contrast images were registered using a custom ImageJ script and then imported into the Baxter Algorithms cell tracking software [223]. Clusters of cells with an area greater than 1000 pixels (~5 cells) were tracked using default parameters. Cell cluster tracks were then analyzed to quantify migration. Speed, displacement, mean squared displacement, and the cumulative distance traveled was calculated for cell clusters.

### **Live-cell imaging**

Well plates were placed in the IncuCyte FLR and phase contrast images were acquired every 30 minutes for 48 hours. Individual cells were manually tracked using the Fiji [224] plugin MtrackJ. Custom R scripts were used to quantify the migratory behavior of individual cell lineages. In brief, starting at the last time slot of each lineage, one cell was randomly selected and traced back through mitotic events until T0. Migration distance for each lineage was then calculated as the sum of the distances in pixels along the path between each image. To compare migratory behavior across different ligand treatments, we performed an ANOVA followed by Tukey's Honestly Significant Difference test in R. Ligand treatments with p-value < 0.05 were deemed significantly different.

### **Immunofluorescence**

Prior to fixation, cells were pulsed with 10  $\mu$ M EdU (Thermo Fisher Scientific C10357) for 1 hour under standard culture conditions. Cells were then fixed for 15 minutes with 2% paraformaldehyde (Electron Microscopy Sciences #15710) and permeabilized for 15 minutes with 0.01% Triton X-100 in PBS. Cells were then stained with CellMask (Thermo Fisher Scientific #H32713) for 30 minutes at RT, followed by fluorescent labeling of incorporated EdU for 1 hour at RT (Thermo Fisher Scientific C10357). Finally, cells were stained with a keratin 5 polyclonal antibody (BioLegend #905501) at 1:800 overnight at 4°C, followed by an anti-rabbit 488 secondary antibody (Thermo Fisher Scientific A21206) at 1:300 and Dapi (PromoKine PD-CA707-40043) at 0.5  $\mu$ g/ $\mu$ L for 1 hour at RT.

Fixed cells were imaged on an Olympus ScanR microscope. DAPI channel images were imported into Ilastik for pixel classification [221], [225]. A set of 20 images per plate were randomly selected and used for training. Pixels were classified as either nuclei or background using all default intensity, edge, and texture features, and with smoothing filters ranging from 0.3 – 10 pixels. Probability maps were then exported from Ilastik into CellProfiler version 3.1.8 for object segmentation [221], [226]. Nuclei were identified using the global Otsu method with a threshold smoothing scale of 1.35. Clumped nuclei were separated based on intensity with a smoothing filter of 12 pixels. Cytoplasm compartments were assigned to nuclei by a 10-pixel donut expansion from each nucleus. Cytoplasm and nuclear Intensity,

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

size, and morphology data was then exported into RStudio (RStudio Team, 2015). The values are analyzed as populations that have been median summarized from the cell-level data to the image or field level. The field level data are then median summarized to the well level. The EGF time course normalized values are the raw values divided by the corresponding EGF value at the same time point within the same replicate set. The preprocessing and QA script is at <https://github.com/MEP-LINCS/MDD/tree/master>. All samples passed qualitative QC inspection that the integrated DAPI intensity has the expected bimodal distribution.

### Phenotype analysis

All phenotypic quantifications were derived from immunofluorescent cell-level data. *Cell cycle phase* was determined by analysis DAPI intensity: each cell was classified into either G1 or G2M cell cycle phase by clustering cells into two groups based on total nuclear DAPI intensity. The Forgy k-means algorithm was used for clustering (R stats package), with the number of centers set to two. DAPI thresholds for classification were manually inspected, and multinucleated and poorly segmented cells were removed from further cell cycle analysis. *KRT5 intensity* was calculated as the mean intensity value of KRT5 in the cytoplasmic cell compartment.

Three spatial metrics were computed to quantify treatment induced changes in cell clustering and dispersal. The *number of neighbors* for each cell was calculated by quantifying the number of cell centroids within 100 pixels of a cell's centroid. Cells with coordinates less than 100 pixels from the image border were excluded. *Nearest neighbor distances* were determined by measuring the pixel Euclidean distances of each cell centroid to the centroids of the four nearest cells in the imaging field. To account for variations in image cell count, the mean nearest neighbor distances for each image were normalized by the expected mean distance to the nearest neighboring cell if the cells were distributed randomly [226]. The *number of cells per cluster* was computed in a two-step process: first performing mean shift clustering on the cell centroid coordinates for each image, using the R package LPCM (v 0.47), and then computing the mean number of cells per cluster.

To compare phenotypic responses across treatments, we analyzed quantifications of the immunofluorescent images 48 hours after treatment. The Kruskal-Wallis test was used to test for overall treatment dependent differences. Pairwise comparisons between treatments were then conducted using Pairwise Wilcoxon Rank Sum Tests followed by an FDR multiple comparisons correction. A stringent significance threshold of q-value < 0.05 was used to aid in identification of the most differentially induced phenotypic features.

### Reverse Phase Protein Array sample preparation

Cells were washed twice with ice-cold PBS followed by collection by manual scraping in 50-100  $\mu$ L of lysis buffer (1% Triton X-100, 50mM HEPES pH 7.4, 150mM NaCl, 1.5mM MgCl<sub>2</sub>, 1mM EGTA, 100mM Na pyrophosphate, 1mM Na<sub>3</sub>VO<sub>4</sub>, 10% glycerol, 1x cOMplete EDTA-free protease inhibitor cocktail (Roche #11873580001), 1x PhosSTOP phosphatase inhibitor cocktail (Roche #4906837001)). Lysates were incubated on ice for 20 minutes with gentle agitation every 5 minutes and then centrifuged at 14,000 rpm for 10 minutes at 4°C. Supernatant was collected into a fresh tube, quantitated by BCA assay, and the appropriate volume was combined with 4X SDS sample buffer (40% glycerol, 8% SDS, 0.25M Tris-HCl, 10%  $\beta$ -Me, pH 6.8), boiled for 5 minutes, and stored at -80°C. Three sets of replicates were collected over three weeks and submitted to MD Anderson Cancer Center for RPPA testing.

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

### **Reverse Phase Protein Array pre-processing and QC**

Samples underwent standard pre-processing using methods developed at the MD Anderson Cancer Center RPPA core [227]. In brief, the processing steps include the following: 1) Convert raw data from log<sub>2</sub> value to linear value. 2) Determine median for each antibody across the sample set. 3) Calculate the median-centered ratio by dividing each raw linear value by the median for each antibody. 4) Assess sample quality by computing a correction factor (CF.1) for protein loading adjustment for each sample as the median of the median-centered ratio values from Step 3 for all antibodies. Samples with correction factors above 2.5 or below 0.25 are considered outliers and discarded. 5) Compute the normalized linear value by dividing the median-centered ratio from Step 3 by CF.1. All samples passed MDACC's quality checks and are included in the dataset. The normalized RPPA log<sub>2</sub> values are joined with their experimental metadata and stored on Synapse as level 3 data. Replicates are median summarized and stored as Level 4 data.

### **RNAseq sample preparation and sequencing**

Following treatment protocols described, at the appropriate time point wells were aspirated and cells were harvested by scraping in 600 µl of RLT Plus buffer (Qiagen) plus 1% β-ME. Samples were flash frozen in liquid nitrogen and stored at -80°C prior to RNA extraction. Total RNA was extracted from frozen using a Qiagen RNeasy Mini kit. Columns were DNase treated following the recommended protocol of the manufacturer.

RNA concentration and purity was determined by UV absorption using a Nanodrop 1000 spectrophotometer. All samples had 260/280 absorption ratios of at least 2.0, indicating successful isolation of RNA from other nucleic acids. RNA integrity was assessed using an Agilent 2100 Bioanalyzer with an RNA 6000 Nano Chip. RNA integrity numbers (RIN) were calculated from Bioanalyzer electropherograms using the "Eukaryotic Total RNA Nano" program of the Bioanalyzer 2100 Expert software (B.02.08.SI648). RIN values were in the 8.5-10 range, indicating high-quality RNA, with one exception (TGFB\_48\_C1\_B; RIN = 6.9). UV absorption measurements and RIN values are available on Synapse (<https://doi.org/10.7303/syn12550434>).

cDNA libraries were prepared from polyA-selected RNA using an Illumina TruSeq Stranded mRNA library preparation kit. 100-bp single-end reads were sequenced on an Illumina HiSeq 2500 Sequencer, with a target of 60M reads per sample.

### **RNAseq pre-processing and QC**

Sequence preprocessing and alignment was performed using a Docker-based pipeline<sup>105</sup>. 100-bp single-end reads were trimmed of Illumina adapter sequences using TrimGalore (v. 0.4.3), a wrapper for CutAdapt (v. 1.10) and FastQC (v. 0.11.5). A minimum of 1-bp overlap with the adapter sequence (AGATCGGAAGAGC) was required for trimming. After trimming, reads with a length < 35 bp were discarded. Trimmed reads were aligned to the GENCODE V24 (GRCh38.p5) assembly of the human genome using the Kallisto pseudo-alignment software (v. 0.43.0). Kallisto, using the following parameters: --bias -b 30 --pseudobam.

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

Gene-level quantifications were produced from transcript-level abundance estimates using the R (v. 3.5.0) package tximport (v. 1.8.0). Mapping between gene/transcript identifiers was done using the biomaRt package (biomaRt v. 2.36.1) with the ENSEMBL\_MART\_ENSEMBL biomart and the hsapiens\_gene\_ensembl dataset. Gene-level quantifications were imported to DESeq2 (v. 1.24.0) [228]. The fpkm function of DESeq2 was used to normalize data for library size and gene length differences, and fpkm values were log<sub>2</sub> transformed with an added pseudocount of 1.

### **Transcription Factor enrichment scores**

Single-sample enrichment scores were calculated for 297 transcription factor target gene sets obtained from the CHEA3 ReMap\_ChIP-seq [229] using the R package GSVA (v. 1.32.0) [230]. A minimum expression filter was used for expressed genes; genes were retained only if expressed at a minimum of 0.5 log<sub>2</sub>(fpkm + 1) in a minimum of 3 samples. Enrichment scores were calculated from filtered RNAseq data, in units of log<sub>2</sub>(fpkm + 1), using the argument “method = ‘ssGSEA’”.

### **Identification of differentially expressed genes**

For each ligand treatment, we performed a differential expression analysis on the RNAseq gene-level summaries with the R package DESeq2 (1.24.0), with shrunken log<sub>2</sub> fold change estimates calculated using the apeglm method. We used the Benjamini-Hochberg method to correct p-values for multiple comparisons and set a threshold of q-value < 0.01 and shrunken log<sub>2</sub> fold change > 1.5 or < -1.5 to indicate significance.

### **Pathway enrichment of ligand-induced signatures**

We used Gene Set Enrichment Analysis (GSEA) to identify the pathways enriched by each ligand treatment. Specifically, we used Gene Set Enrichment Analysis 4.1.0 downloaded from <https://www.gsea-msigdb.org/gsea/index.jsp> to assess enrichment of the MSigDB Hallmark Pathways in the Level 3 data. For each 24H ligand treatment sample, we computed log<sub>2</sub> fold-change against CTRL\_0 from the Level 3 RNAseq data.

### **ATACseq sample preparation and sequencing**

ATACseq samples were collected following the Omni-ATAC protocol [231]. Briefly, MCF10A cells were washed once with PBS and detached from the plate using trypsin. Cells were then counted using a Countess (Invitrogen), and 50,000 cells per condition were distributed to 1.5 ml centrifuge tubes and spun at 500 RCF for 5 min. The supernatant was removed, and the cell pellet was resuspended in 500 µl of PBS and spun again at 500 RCF for 5 min. The supernatant was removed again, and the cell pellet was resuspended in 50 µl of cold ATAC resuspension buffer (RSB) containing 0.1% NP40, 0.1% Tween-20, and 0.01% digitonin by pipetting up and down three times. After 3 min on ice, 1 ml of cold RSB containing 0.1% Tween-20 was added, and the tube was inverted three times to mix. The nuclei were then pelleted by centrifugation at 500 RCF for 10 min at 4°C. The supernatant was then carefully aspirated, and the nuclei were resuspended in 50 µl of transposition buffer (25 µl 2x TD buffer (Illumina), 2.5 µl transposase (Illumina), 16.5 µl PBS, 0.5 µl 1% digitonin, 0.5 µl 10% Tween-20, and 5 µl H<sub>2</sub>O). Samples were then placed in a pre-warmed (37°C) thermomixer and mixed for 30 min at 100 RPM. Transposed fragments were then purified using a Qiagen MinElute column and frozen at -80°C for further processing.

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

The remaining steps of the Omni-ATAC protocol were performed by the OHSU Massively Parallel Sequencing Shared Resource. Transposed fragments were pre-amplified with 5 rounds of PCR. Afterward, 5  $\mu$ l of the pre-amplified mixture was used for a qPCR reaction to determine the concentration of tagmented DNA. After calculating the concentration of tagmented DNA, pre-amplified samples were diluted with elution buffer to a final concentration of 5  $\mu$ M. Six samples had an undiluted DNA concentration below 5  $\mu$ M and were not diluted. 5  $\mu$ M pre-amplified samples were amplified for 3 additional PCR cycles.

Tagmented DNA was pre-amplified with 5 rounds of PCR (72°C for 5 min, 98°C for 30 seconds, then 5 cycles of [98°C for 10 sec, 63°C for 30 sec, 72°C for 1 min]). PCR reactions contained 20  $\mu$ l eluate, 25  $\mu$ l NEBNext 2x MasterMix, 2.5  $\mu$ l 25  $\mu$ M i5 primer and 2.5  $\mu$ l 25  $\mu$ M i7 primer.

The DNA concentration of the pre-amplified samples was assessed by qPCR. 5  $\mu$ l of pre-amplified mix was added to 3.76  $\mu$ l sterile water, 0.5  $\mu$ l 25  $\mu$ M i5 primer, 0.5  $\mu$ l 25  $\mu$ M i7 primer, 5  $\mu$ l 2x NEBNext master mix, and 0.24  $\mu$ l 25x SYBR Gold (in DMSO). Samples were amplified for 20 cycles of [98°C for 10 sec, 63°C for 30 sec, 72°C for 1 min]. DNA concentration was calculated, and pre-amplified samples were diluted to a final concentration of 5  $\mu$ M. Six samples had an undiluted DNA concentration below 5  $\mu$ M and were not diluted. 5  $\mu$ M pre-amplified samples were amplified for 3 additional PCR cycles. 100bp PE reads were sequenced on an Illumina HiSeq 2500 Sequencer by the OHSU Massively Parallel Sequencing Shared Resource with a target of 20M reads per sample.

### **ATACseq pre-processing and QC**

ATACseq files were processed and aligned using the ATACseq (1 -> 3) workflow on the AnswerALS Galaxy server (answer.csbi.mit.edu). Reads were trimmed of adapter sequences and low-quality bases using Trimmomatic (Galaxy version 0.36.5). Reads were trimmed of low-quality bases (Phred score < 15) at the read start or end, and Nextera adapter sequences (CTGTCTCTTATA) were trimmed from read ends (minimum of a 2-bp overlap required for trimming). Reads were aligned to the human genome (hg38) using Bowtie2 (Galaxy version 2.3.4.1) in paired-end mode with otherwise default settings. BAM files were filtered to remove secondary alignments, unmapped reads, and mitochondrial DNA alignments using ngsutils bam filter (Galaxy version 0.5.9). PCR duplicates were detected and removed using Picard MarkDuplicates (Galaxy version 2.7.1.2). The de-duplicated, filtered BAM file was used for peak calling and quantification. Peaks were called using MACS2 (Galaxy Version 2.1.1.20160309.5) using the following parameters: -format BAMPE -nomodel -extsize 200 -shift -100 -qvalue 0.01.

ATACseq sample quality was assessed by calculating the fraction of reads in peaks (FRiP). Before calculating FRiP, a consensus peakset was generated for all samples by taking the union of all peaks called in all samples and merging any overlapping peaks, using the R (v. 3.6.1) package DiffBind (v. 2.12.0) [232]. For each sample, FRiP was then calculated by counting the proportion of reads in the de-duplicated, filtered BAM file that align within the consensus peakset. A minimum FRiP threshold of 0.15 was applied to remove samples with low levels of chromatin enrichment. Thirteen ATACseq samples did not pass the QC due to low FRiP scores; the fragment length distributions of these samples also lack the periodic peaks caused by nucleosome patterning. These low-quality samples likely are the result of fragment over-transposition due to a high Tn5-transpose-to-cell ratio [233], [234].

### **Construction of chromatin accessibility matrix**

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

DiffBind (v. 2.12.0) was used to generate a peak accessibility matrix for the QC-passing samples. First, a consensus peakset was re-generated after removal of low-FRiP samples. The `dba.count` function was then used to count the number of reads in the de-duplicated, filtered BAM files that overlap with each peak in the consensus peakset. The `dba.count` argument `"score = DBA_SCORE_TMM_READS_EFFECTIVE"` was used to output TMM counts normalized to each sample's effective library size, which is equal to the de-duplicated, filtered library size multiplied by the FRiP. A peak accessibility matrix in units of unnormalized counts was also generated using the `dba.count` function with the argument `"score = DBA_SCORE_READS"`.

### Motif Enrichment

Transcription factor motif enrichment scores were generated from the TMM-normalized chromatin accessibility data using the R package `chromVAR` (v. 1.6.0) [235]. ATACseq peaks were annotated with GC content using the `addGCBias` function of `chromVAR` and the `BSgenome.Hsapiens.UCSC.hg38` genome annotation package. Transcription factor motif position frequency matrices were obtained from the "JASPAR CORE 2018 Homo sapiens" set of motifs [236]. ATACseq peaks were matched to JASPAR motifs using the R package `motifmatchr` (v. 1.6.0). The expected fraction of reads per ATACseq peak was calculated using the `chromVAR` function `computeExpectations`, with the argument `"norm = TRUE"`. Each sample's deviation from the expected fraction of peaks in each annotated category was calculated using the function `computeDeviations`, and deviations were converted to Z-scores using the function `deviationScores`. Enrichment scores of individual transcription factors were mean summarized to the "family" level as annotated in JASPAR 2018.

### Global Chromatin Profiling

The GCP assay was performed as previously described in Creech *et al* [237] and Litichievskiy *et al* [189]. Cells were washed with ice-cold PBS, then collected by manual scraping in 200  $\mu$ l of cold PBS. Cells were then pelleted by centrifugation at 1500 RCF at 4°C for 5 min, resuspended in 1mL of cold PBS, and spun again as specified. The resultant cell pellets were then flash frozen in liquid nitrogen and stored at -80°C until further processing. Pellets were thawed and lysed with nucleus buffer, followed by histone extraction by sulfuric acid and precipitation using trichloroacetic acid. Sample input was normalized to 10  $\mu$ g of histone in H<sub>2</sub>O before being propionylated, desalted (Oasis HLB 5mg Plate) and digested by Promega trypsin overnight. A second round of propionylation, followed by desalting using C18 Sep-Pak cartridges (Waters) was employed after digestion. Propionylations and digestion were done in an automated fashion on an LT-Bravos system (Agilent). Isotopically labeled synthetic peptides from histones H3 and H4 were added as a reference to each sample prior to MS analysis. Peptides were separated on a C18 column (EASY-nLC 1000, Thermo Scientific) and analyzed by MS in a PRM mode (Q Exactive<sup>TM</sup>-plus, Thermo Scientific) [237]. Detailed protocols of sample preparation steps can be found in <https://panoramaweb.org/labkey/wiki/LINCS/Overview%20Information/page.view?name=sops>.

GCP data was merged with the experimental metadata and stored as level 3 data on Synapse. Replicates were median summarized and stored as level 4 data.

### L1000 sample preparation

L1000 samples were collected as part of three collections. The first L1000 sample collection was generated in parallel to the ATACseq samples. MCF10A cells were washed once with PBS and detached

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

from the plate using trypsin. Cells were then counted using a Countess (Invitrogen) and 50,000 cells per condition were distributed to 1.5 ml centrifuge tubes and spun at 500 RCF for 5 minutes. The supernatant was removed, and the cell pellet was resuspended in TCL buffer (Qiagen) containing 1%  $\beta$ -Me. For the second and third collections, cells were washed with PBS followed by the addition of TCL buffer (Qiagen) containing 1%  $\beta$ -Me. The cell and buffer mixture was allowed to sit for 30 minutes and then frozen at  $-80^{\circ}\text{C}$  for further processing. Samples from the first and second sample collections were frozen in 1.5ml tubes. Samples from the third data collection were frozen in their original 96-well plates. In total there were eighteen plates from the third HMS collection, which contained 21 samples per plate, and there were 190 samples from the first two OHSU collections. All samples were shipped to the BROAD for simultaneous processing on the L1000 platform. The source plates containing original samples were re-arrayed into six 96-well master plates. These master plates contained 21 samples from each of three original source plates, and 32 samples plated directly from tubes. In each of the six master plates, well A1 was left empty to accommodate for internal technical control spike-ins. The six 96-well master plates were then re-arrayed into the final 384 well v-bottom PCR Plates (Eppendorf #951020702).

### **L1000 Ligation Mediated Amplification**

For L1000 Ligation Mediated Amplification [238] crude cell lysates were transferred from source plates to 384 well v-bottom PCR Plates (Eppendorf #951020702) assay plates. Oligo dT coated magnetic particles (GE Healthcare #38152103010150) were added to capture mRNA. Plates were then incubated at room temperature on shaker tables for 10 minutes. The beads were then spun down onto flat magnets and unbound lysate was evacuated by centrifuging upside down on magnet to 800RPM for 30 seconds. 15 $\mu\text{l}$  of reverse transcription master mix containing SuperScript IV reverse transcriptase was added to the plates and the plates were incubated at  $55^{\circ}\text{C}$  for 10 minutes. Plates were again spun down, beads were pelleted on a flat magnet, and the remaining master mix was spun out. Probes were annealed to the first-strand cDNA by addition of 15 $\mu\text{l}$  of Probe Bind master mix, containing 100 fmole of each probe and Taq ligase buffer. Samples were denatured at  $95^{\circ}\text{C}$  for 5 minutes, then transferred to a ramping water bath that decreased temperature from  $70^{\circ}\text{C}$  to  $40^{\circ}\text{C}$  over six hours. The following day, beads were again spun down on a flat magnet and master mix was evacuated. To ligate probe pairs, 15  $\mu\text{L}$  of Ligation Master Mix was added, containing Taq DNA ligase and ligase buffer. Plates were sealed and incubated at  $45^{\circ}\text{C}$  for 60 minutes. Plates were spun down on magnets and ligation master mix was evacuated as with previous steps. 15 $\mu\text{l}$  PCR master mix containing 0.5 mmole of each primer (T3 and 50-biotinylated T7 universal primers), dNTPs, and PlatinumTaq polymerase in reaction buffer was added to each well, and plates were subjected to 29 cycle PCR. This process yielded biotinylated gene and bead (barcode) specific amplicons.

Each barcode corresponds to a complementary sequence on a Luminex bead, allowing the PCR product to be hybridized to a mixture containing per well  $\sim 100$  each of 500 Luminex analyte colors. The plate was then denatured at  $95^{\circ}\text{C}$  for 5 minutes and incubated at  $45^{\circ}\text{C}$  for 18 hours. Beads were pelleted and stained with streptavidin R-phycoerythrin conjugate for ten minutes. Finally, plates were read on Luminex FlexMap 3D Flow cytometers that detected analyte color (transcript identity) and fluorescence intensity (transcript abundance) for all analytes detected in all wells.

### **L1000 pre-processing**

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

To account for differences across the various cell collections, we adapted our standard data processing pipeline in several ways. L1000 data typically use a population-based normalization scheme, known as plate control, as described in Subramanian *et al* [238]. Here, the EGF treated wells served as the vehicle when conducting vehicle normalization. The standard data processing pipeline was followed, except for the changes at Level 1 and Level 4, described below. L1000 utilizes 10 sets of invariant genes, similar to ‘housekeeping’ genes, to assess quality and in later normalization steps. These gene sets, each containing 8 genes, represent control values that span the spectrum of gene expression, and are ordered according to their overall level of expression, the first level corresponding to the lowest expressing genes, and the 10th corresponding to the highest expressors.

Plates were computationally split at Level 1 (LXB) into subpopulations of wells, each containing only samples from a given time-point and collection combination. The fluorescence intensity values associated with each bead color were subjected to the peak deconvolution step, which separates the two genes associated with each bead color (Level 2). Data were then normalized via L1000 invariant set scaling (LISS), which scales the expression levels of the 978 measured landmarks in each well to the 80 control genes in the invariant gene set (Level 3). Next, we calculated differential expression using EGF as the vehicle control. Robust z-scoring was used to calculate differential expression values for each gene, where gene x is compared only to the vector of normalized gene expression of gene x across all EGF samples in that collection/time-point population (Level 4). Finally, individual biological and technical replicates were collapsed into a consensus signature by computing a pairwise Spearman correlation matrix between each replicate signature. The weights for each replicate were calculated by the sum of their correlations to the remaining replicates, summing to 1. The consensus signatures were generated by the linear combination of the replicate signatures using each signature’s weight as the coefficient (Level 5).

### **L1000 QC**

We used several approaches to assess data quality. First, to assess the quality in each detection plate, we visually inspected and measured the slope of the invariant gene calibration curve for each sample; outliers were omitted. Second, to assess plate effects, we plotted median fluorescence intensity and interquartile range of invariant set 10 across the entire plate. This allowed identification of failed (low signal) wells, tissue culture related plate effects, or wells with abnormally wide ranges in expression across each gene set. Third, to assess the efficacy of the deconvolution algorithm, we determined the number of well/analyte combinations where two peaks were clearly discernible.

In addition, we computed a transcriptional activity score (TAS) as a composite measure of L1000 transcriptional response. Here signature strength (SS) was computed as the number of genes with a z-score greater than or equal to 2 for each sample, and replicate correlation (CC) was computed as the 7th quantile of the spearman correlation between all pairwise combinations of replicates. TAS is calculated as the geometric mean of SS and CC for a signature and scaled by the square root of the number of landmark genes, yielding a final score between 1 and 0. QC metrics are available on Synapse (<https://doi.org/10.7303/syn19416843.1>). 2 L1000 samples (1 from C1 and 1 from C3) failed these QC metrics and were removed. Finally, within each sample collection (C1, C2, and C3), we clustered samples based on the Euclidian distances between expression of the 978 measured landmark genes in the Level 3 data, using the R function `hclust`. Each collection had a small number of outlier samples that showed markedly aberrant expression of the 978 landmark genes and clustered apart from all other samples, in a

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

pattern that was not explained by sample treatment; these 17 samples (3 from C1, 1 from C2, and 13 from C3) were removed. Additionally, 25 samples from Collection 2 lacked an appropriate EGF-treated control on the same 384-well plate and therefore were omitted from the final dataset. In total, 44 L1000 samples (4 from C1, 26 from C2, 14 from C3) were removed from the dataset.

### **Cyclic Immunofluorescence (CyCIF) sample preparation and imaging**

MCF10A cells were seeded 4000 cells/well in 200  $\mu$ l of GM in collagen coated (as described above) 96 well plates (NUNC, 165305) in technical (multiple wells on the same plate) and biological (experiments separated by a minimum of one cell passage) triplicates. Eight hours after seeding, the cells were washed once with PBS using an EL405x plate washer (BioTek), and 200  $\mu$ l of EM was added per well. Following an additional 16 hours (24 hours after initial plating), one plate was fixed (time = 0 hours) and EM was aspirated from all wells in the remaining plates using the plate washer and replaced with 200  $\mu$ l of the appropriate ligand or control treatment.

The treated plates were fixed following incubations of 1, 4, 8, 24, and 48 hours. Cells were fixed in 4% formaldehyde for one hour at room temperature and washed with PBS. Plates were sealed and stored at 4°C until all replicates were collected. Next, cells were permeabilized with ice cold methanol for ten minutes, blocked in Odyssey buffer (LI-COR) for one hour, pre-stained with secondary antibodies, bleached, and imaged to register background intensities prior to beginning CyCIF [150], [239]. For each cycle, cells were stained with three conjugated antibodies, unless otherwise specified, and Hoechst 33342 overnight at 4°C, washed with PBS, and imaged with an IN Cell Analyzer 6000 (nine fields of view per well, 20x/0.45NA air objective, 2x2 binning) (GE Healthcare Life Sciences). Following image acquisition, fluorophores were chemically inactivated as described [150], [239], and cells then entered the next staining cycle. Refer to **Supplementary Data 21** for antibody metadata.

### **CyCIF pre-processing and image analysis**

A flat field correction profile, generated from all fields on one plate using the BaSiC ImageJ plugin [240], was normalized to a mean value of one and each image was then divided by it. Image registration was performed with a custom ImageJ script. Segmentation of the nuclei (based on Hoechst staining), and cytoplasm (based on  $\beta$ -catenin staining) was performed with a custom MATLAB (MathWorks) script. Each cell was then divided into four subcellular masks: nucleus, peri-nuclear ring, cytoplasm, and cell membrane for feature extraction, a fifth region including all the cytoplasm (peri-nuclear ring, cytoplasm, and cell membrane together) was also defined. Segmentation was performed on the images acquired in cycle 4 only; the masks were then overlaid on all other cycles for feature extraction. Intensity, texture, and morphology features were extracted for each mask, as appropriate (see **Supplementary Data 22** for feature definitions).

### **CyCIF QC**

Quality control was performed in two steps. In the first step, cells that were washed away over the course of the experiment and those near the edges of the imaging fields that were incompletely captured cycle to cycle due to microscope stage drift were identified and excluded from subsequent analyses. These cells were identified by their high variation in nuclear Hoechst signal between successive cycles ([https://github.com/yunguan-wang/cyCIF\\_analysis\\_suite/blob/MCF10A/notebooks/Section2.1\\_Intensity%20based%20QC.ipynb](https://github.com/yunguan-wang/cyCIF_analysis_suite/blob/MCF10A/notebooks/Section2.1_Intensity%20based%20QC.ipynb)). If

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

more than 90% of the cells in a field of view failed this QC step, the entire field was removed. The median fraction of lost cells was ~15 % for fields 1-8 whereas an average of 60% of cells were lost from field 9, with a significant number of instances where the fraction of lost cells exceeded 90%. Field 9 was therefore excluded entirely from subsequent analyses. Additionally, for unknown reasons, most of the wells occupying row E on plate 18 exhibited cell loss in excess of 90% leading to the exclusion of all data from those wells in downstream analyses. In the second quality control step, cells with failed cytoplasm segmentation as identified by multinucleation were removed. Multi-nucleated cells were identified by re-segmenting each mask using the Python implementation of Opencv (<https://github.com/skvark/opencv-python>) and counting the nuclei; cells with two or more nuclei were excluded from downstream analyses ([https://github.com/yunguan-wang/cycif\\_analysis\\_suite/blob/MCF10A/notebooks/Section2.2\\_image\\_based\\_qc.ipynb](https://github.com/yunguan-wang/cycif_analysis_suite/blob/MCF10A/notebooks/Section2.2_image_based_qc.ipynb)). Although masks with two nuclei can represent failed segmentation or truly binucleated cells, visual inspection led us to conclude that these cases were primarily segmentation errors and were therefore excluded from downstream analyses.

### **Measuring association between variance and covariates**

We applied the Measuring Association between Variance and Covariates method to systematically assess the fractional variance explained by each experimental covariate of ligand, time, and replicate [241]. Briefly, each dataset was normalized by winsorization at 99% to remove extreme outliers and then median centering within replicate. Next, we performed principal component analysis to reduce the dimensionality of each data set while preserving the variability. A subspace of principal components (PCs) significantly associated with each covariate (ligand, time, replicate) was determined by lasso regression for continuous covariates and silhouette coefficient for categorical covariates. We then quantified the total variance explained by each covariate by summing the weighted variances of all principal components (PCs). Low variance PCs with an eigenvalue of less than 0.7 were unlikely to significantly correlate to any covariates and these discarded PCs were not included in the analysis.

### **L1000 drug signature comparison**

To compare our results to existing L1000 transcriptional drug signatures we used the L1000 FWD tool [238] available at <https://maayanlab.cloud/L1000FWD/>. We used as input the top 200 most significantly up-regulated and top 200 most significantly down-regulated genes at 24 H relative to CTRL\_0. We considered drug signatures with Fisher exact test q-values < 0.2 to be significantly correlated or anti-correlated with our ligand signatures. Finally, we summarized the number of drugs with similar mechanisms of action to identify common patterns.

### **Comparison of RNAseq and RPPA assays**

To examine the relationship between gene expression and protein abundance, we compared Z-scores calculated from our Level 3 RNAseq and RPPA data for the 222 genes/proteins measured by both assays. We also characterized the relationship between these assays by examining the concordance of genes and protein identified as differentially expressed (compared to time 0) after ligand treatment. Genes meeting an absolute fold-change threshold of 1.5 and an FDR-adjusted q-value of 0.01 were considered differentially expressed (as described in RNAseq methods). RPPA antibodies meeting an absolute log fold-change threshold of 0.5 and an FDR-adjusted q-value of 0.01 were considered differentially expressed. For this analysis, we used a more stringent alpha of  $q = 0.01$  (rather than  $q = 0.2$  used

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

elsewhere) to focus on the strongest and most robust signals in each assay. Measurements with differential expression in both assays were considered concordant. We visualized the concordance between these assays with paired heatmaps displaying upregulated and downregulated measurements. We summarized these results with a Euler diagram showing set relationships between upregulated and downregulated measurements across all ligand treatments.

### **Comparison of RNAseq and L1000 assays**

To assess the concordance between gene expression profiles generated by both the RNAseq and L1000 assays, we first filtered Collection 1 Level 3 data from both datasets to contain only samples and transcripts directly measured by both assays, then z-transformed the filtered datasets. We calculated the Pearson's correlation between the RNAseq and L1000 z-scores for all pairwise combinations of samples, then compared the distributions of treatment-matched and treatment-mismatched samples. Samples with the same ligand treatment and time point were considered treatment matched. We used a Mann-Whitney U test was used to test for differences in mean correlation between the treatment-matched and -mismatched groups.

### **Comparison of ATACseq and RNAseq assays**

To compare gene expression to chromatin accessibility at the respective transcriptional start site (TSS), we quantified chromatin accessibility using bedtools multiBamCov (v. 2.26.0). Chromatin accessibility was quantified in windows  $\pm 500$  bp from TSS coordinates provided by the R package TxDb.Hsapiens.UCSC.hg38.knownGene [PMID 20110278]. The most-accessible TSS was selected for genes with multiple TSS. Integer counts were transformed using the variance-stabilized transformation from the R package DESeq2 (v. 1.24.0). Genes within the MHC region of chromosome 6 (chr6: 28510120-33480577) were excluded from this analysis; ATACseq data from this region had poor alignment due to alternative contigs for this region in the hg38 genome assembly. Median VST-transformed TSS accessibility was compared to median Level 3 RNAseq data for the EGF\_48 condition.

We also compared the 10 most-variant ATACseq TF motifs (by standard deviation) to single-sample gene set enrichment scores computed for the same TFs from Level 3 RNAseq data, using the R package GSVA (v 1.32.0) and the TF-gene target mappings from the ReMap ChIP-seq library (as described above).

### **Multi-omic module detection**

To identify coordinately regulated multi-omic modules, we performed normalization, data scaling, feature selection and cluster analysis on molecular features induced by ligand treatments. For the GCP, RPPA and CyCIF datasets we used limma to normalize to CTRL\_0, summarize across the replicates and calculate adjusted p-values using Benjamini-Hochberg correction; we used DESeq2 to analyze the RNAseq data in a similar manner. We used chromVAR to aggregate chromatin accessibility peaks that share common motifs and then the individual motif enrichment scores of transcription factor families. We applied the rrscale transformation to each assay data set to minimize differences in the assay-specific data distributions [242]. In brief, each assay's TO CTRL-normalized data was rrscaled independently with Box Cox negative and asinh transformations using an infinite z score cutoff. This transformation yields data matrices for each assay that have symmetrical Gaussian-shaped distributions, making them suitable for parametric statistics. We selected a subset of highly variant and biologically interpretable features from the 24H and 48H samples from each assay. In GCP and RPPA assays, features in the lowest variance

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

quartile were removed. For the CyCIF, RNAseq, and GCP assays, features were retained if, for any condition, the absolute log fold change was greater than 1.5 and the p-value was less than 0.05. For the RPPA assay, we used a log fold-change threshold of 0.75 to account for differences in the RPPA data distribution. All ATACseq motif family scores were retained.

We performed k-means clustering using partitioning around medoids and a gap statistic analysis using the firstSEmax method to identify the optimal number of clusters (R package cluster, version 2.1.2). In brief, the gap statistic method runs PAM clustering on the integrated data matrix once for each k value, where  $k=2:25$ . Then for each k, we performed PAM clustering on 100 randomized permutations of the data that have structure similar to the actual data. At each k, the gap is calculated as the difference in the log of the within-groups sum of squares of the actual versus randomized data. To cluster the features, we use partitioning around medoids (PAM) clustering for the optimal number of clusters defined in the previous step ( $k=18$ ), with seeds randomly selected from the dataset. We repeated this 100 times to form an ensemble of partitions, then calculated consensus clusters from the ensemble using a hard euclidean (HE) method and 5 internal runs. We repeated this entire procedure 25 times and then calculated a final consensus clustering with the HE method from these 25 consensus clusters. We further refined these clusters by identifying and collapsing highly correlated clusters. In brief, we calculated the mean expression of features in each cluster for each condition and then computed Pearson correlations between all pairs of clusters. Next, we then used the R hclust function and the dendextend cutree function on the distance matrix of the correlations to identify highly correlated clusters. This resulted in combining 4 pairs of clusters to yield a final set of 14 modules for further analysis.

### **Consensus Principal Component Analysis**

To explore how our method compares against other published multiomics approaches [243], we performed a Consensus Principal Component Analysis (CPCA) using the R package MoCluster [244] and then compared the clusters to the refined molecular modules described above. In brief, the same features used in the consensus PAM clustering were input as separate blocks to the CPCA algorithm. For each Joint Latent Variable (JLV), the principal components of each assay (block) are calculated as the block latent variables (BLVs), normalized to 1, softly thresholded using a sparsity parameter (0.9) that controls the number of non-zero values and used to iteratively converge on a joint latent variable, which maximizes the correlation between the BLVs. Based on knee analysis of the CPCA pseudoeigenvalues, we kept the first 8 JLVs.

### **Module TF enrichment analysis**

We identified transcription factors enriched in the integrated modules by submitting all RNAseq features from each integrated module to the ChEA3 web-based transcription factor enrichment tool ChEA3 [229], which identifies transcription factors enriched for a list of genes using Fisher's exact test. We limited our analyses to transcription factor targets in the ReMap ChIP-Seq library and considered transcription factors significantly enriched if the FDR-corrected q-value was less than 0.2.

### **Module pathway enrichment analysis**

To identify pathways enriched in each module, we used the Reactome pathway enrichment analysis tool (<https://reactome.org/>) to analyze the genes in each module. In brief, this analysis performs a binomial

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

test of each gene set of 2516 curated pathways in the Reactome database. We identified significantly enriched pathways as those with FDR q-values (Benjamini-Hochberg method)  $< 0.2$ , gene ratios  $> 0.1$ , and pathways that included a minimum of 5 and maximum of 500 genes. To aid visual interpretation, only the top three pathways for each module sorted by FDR and descending gene ratio are shown in **Figure 2-5B**.

### Module expression scores

To calculate the expression of modules across different samples in our MCF10A dataset, we computed the mean expression of features in each module. To assess expression of the modules in external datasets (*e.g.* GTEx), we focused on the RNAseq features in each module and computed their mean expression. For our analysis of Module 10 gene expression in a panel of breast cancer cell lines, we processed and aligned raw sequence data using the Docker-based RNA-seq pipeline [245] described in **RNAseq pre-processing and QC**, then normalized the data using the variance-stabilizing transformation in the R package DESeq2 [228]. We used a Mann-Whitney U test to test for differences in mean Module 10 gene expression between groups.

### Set analysis

Set analysis was used to identify features significantly induced by a single ligand (ligand-specific) or multiple ligands (shared). The input to the set analysis was the integrated and scaled matrix of log fold change values derived from the multi-omic module analysis. Each feature in the multi-omic matrix was labelled either 'Unique' or 'Shared'. Features were defined as 'Unique' if they were significantly perturbed by only a single ligand, with log fold change greater than or equal to  $|1.5|$  and Benjamini-Hochberg adjusted p-value less than .05, relative to time 0. Features that were significantly regulated by two or more ligands were labelled 'Shared.'

### Statistics and Reproducibility

When testing for statistical significance, we adjusted for multiple testing using the Benjamini-Hochberg method. Assays were performed on samples in biological triplicate, as described in Cell Culture Methods. We used a threshold of  $q = 0.01$  for individual analyses of assay datasets (RNAseq and RPPA) and  $q = 0.05$  for phenotypic behavior comparisons to identify only the largest and most robust responses in the data, and a less stringent alpha of  $q = 0.2$  for all other analyses. The significance of list-based enrichment analyses (CHEA3, L1000 FWD) was evaluated using Fisher's exact tests. We used the nonparametric Mann-Whitney U test to test for between-group differences in RNA-L1000 correlation coefficients and Module 10 gene expression.

## 2.4 Results

### 2.4.1 Approach to generate a LINCS ME perturbation dataset

Eight laboratories supported by the NIH LINCS program contributed to the creation and analysis of an MCF10A perturbation dataset to enable community study of the molecular mechanisms engaged by microenvironmental signals to modulate specific cellular phenotypes (**Figure 2-1**). **Figure 2-1B** shows the experimental and computational steps involved in the creation of the database. The process began with screening and selection of ligands that strongly modulated phenotype. Both phenotypic and molecular responses to ligands were then measured over time and integrated computationally to identify the phenotypes and molecular modules engaged by each ligand. **Figure 2-1C** shows the experimental design

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

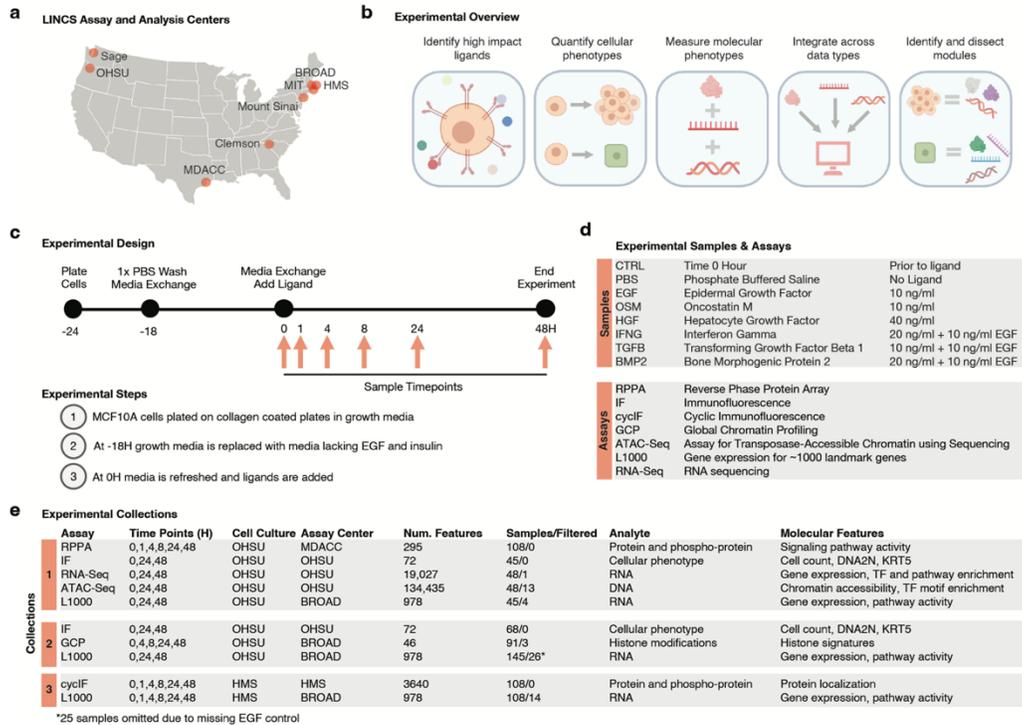
in which multiple endpoints were measured at several time points after the introduction of ligands. The ligands and experimental assays are summarized in **Figure 2-1D**.

The elucidation of phenotype-associated molecular networks requires study of multiple ligands that modulate cell behaviors through varied signaling pathways. To identify a panel of high-impact ligands, we performed two high-throughput microenvironment microarray (MEMA) screens of 3024 combinations of 63 soluble ligands and 48 insoluble extracellular matrix proteins [219]; one screen with and another without EGF, a typical component of MCF10A growth medium [132]. We focused on collagen-1 as the insoluble extracellular matrix component and identified EGF, HGF, and OSM as ligands that increased growth in the absence of EGF, while BMP2, IFNG, TGFB decreased growth in the presence of EGF (**Supplementary Figure 2-1A-B**). These ligands target highly expressed receptors that are members of different canonical receptor classes (**Supplementary Figure 2-1C**). Dose-response experiments identified the ligand doses necessary to yield maximal changes in cell numbers (**Supplementary Figure 2-1D-E**). Inclusion of EGF in combination with BMP2, IFNG, and TGFB ensured sufficient cell numbers for molecular profiling.

The participating LINCS consortium laboratories performed systematic and large-scale analyses of epigenomic, transcriptomic, proteomic and phenotypic responses to each ligand at several time points during a 48H period after treatment (**Figure 2-1B,D,E**). Experiments were carefully planned to minimize technical artifacts that are sometimes associated with large-scale experiments, such as cell line drift, variation in reagents, and protocol differences; a detailed description of considerations can be found in Methods. Cells for all analyses were grown and treated at OHSU and the treated cells or lysates were distributed to participating laboratories for analyses, except for those analyzed using cyclic immunofluorescence (CyCIF) [150], [239]. Cells for CyCIF were grown and treated at HMS using cells, culture media and ligands supplied by one laboratory at OHSU to minimize experimental variation [218] (**Figure 2-1E**). For each assay, MCF10A cells were plated on collagen-1-coated cell culture dishes in their standard growth medium, which contains the growth factors EGF and insulin [132]. After attachment, the growth medium was replaced with medium lacking EGF and insulin, and cells were then treated with the ligand panel at optimized concentrations (**Figure 2-1D**).

Samples were collected before and after treatment over the 48H time period beginning with a time 0H sample (referred to as control: CTRL, **Figure 2-1D**). Cellular responses were measured using live-cell imaging, four-color fluorescence imaging and CyCIF [150], [239]. Molecular responses were assessed for changes in protein expression with reverse phase protein arrays (RPPA)[227]; chromatin profiling using an Assay for Transposase-Accessible Chromatin using sequencing (ATACseq) and global chromatin profiling (GCP) [237]; RNA expression using RNAseq and the L1000 [238] transcriptomics panel designed to assess the levels of 1000 RNA transcripts. Samples for the different assays were collected in three experimental collections of at least three biological replicates each (**Figure 2-1E**). Logistical and cost constraints resulted in some assays being applied to only a subset of time points. Rigorous quality assessment (see methods) of all data led to the elimination of ~5% of samples (44/814). The resultant data and metadata are available at: [synapse.org/LINCS\\_MCF10A](https://synapse.org/LINCS_MCF10A).

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses



**Figure 2-1: Overview of experimental approach to assess the impact of microenvironmental factors.**

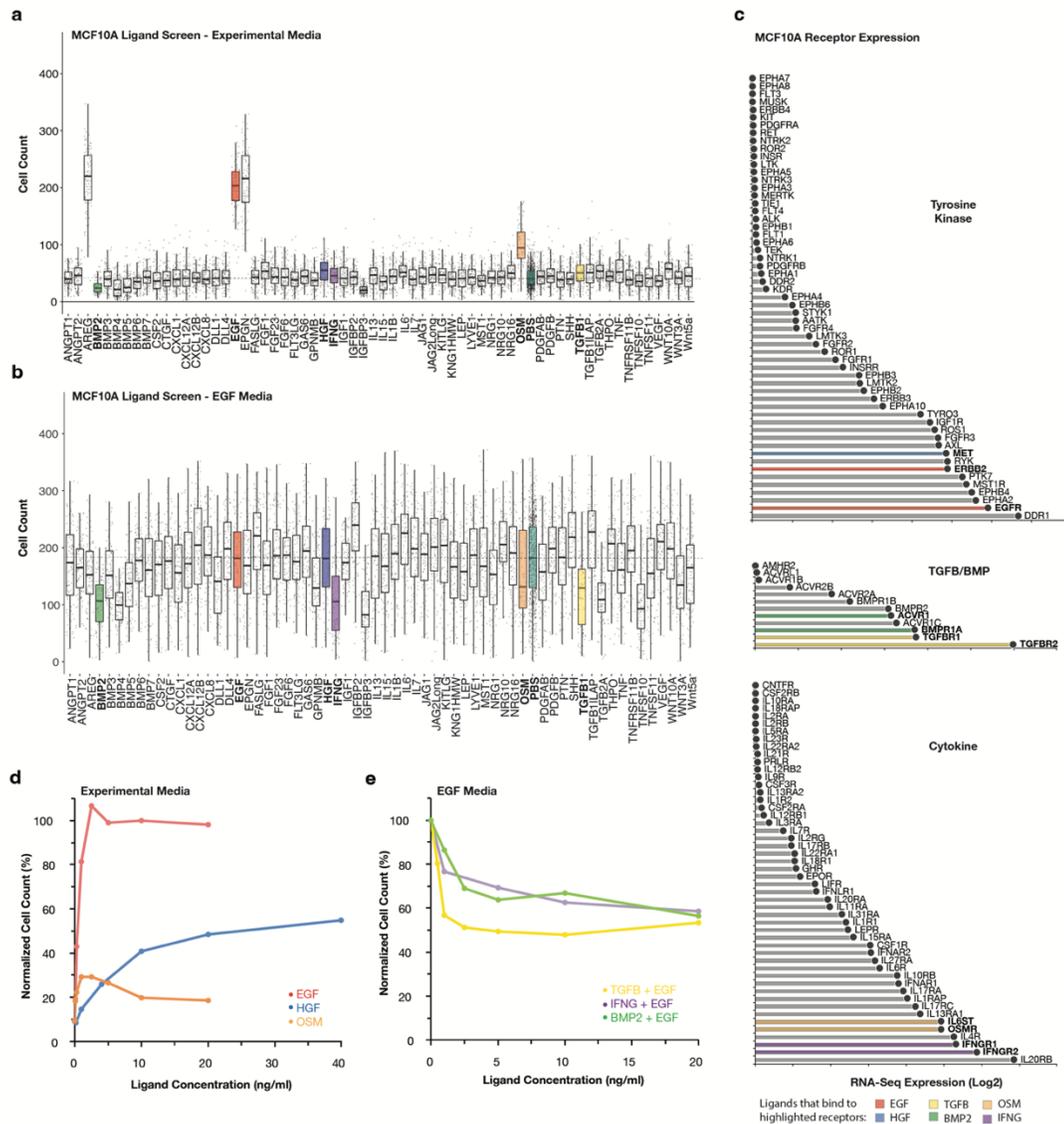
**A)** Map of LINCX data generation and analysis centers.

**B)** Schematic illustrating the experimental and analytical approaches to link molecular and cellular phenotypes.

**C)** Schematic of the experimental design, cell culture protocol, and sample harvest time points.

**D)** The experimental treatments, dosages, and assays deployed to generate the LINCX ME perturbation datasets.

**E)** Summary of the assays, time points, and features for the three experimental collections.



**Supplemental Figure 2-1: Experimental and bioinformatic approaches to identify high impact ligands.**

**A)** Microenvironmental assay (MEMA) to identify ligands that modulate MCF10A cell numbers. Cells were treated with ligands in experimental media lacking EGF and cell numbers were counted after 72H. In the boxplots, lower and upper hinges correspond to the first and third quartiles. The median is shown as the center line. Whiskers extend to no further than  $\pm 1.5 \times$  IQR from the hinge, where IQR is the inter-quartile range, or distance between the first and third quartiles.

**B)** MEMA assay results for MCF10A cells treated with ligands in experimental media containing EGF. Boxplots drawn as described in a.

**C)** MCF10A transcript expression from three receptor classes: Tyrosine kinase, cytokine, and TGFB/BMP. Transcript values are drawn from RNAseq measures for untreated cells in exponential growth. The primary receptors for the six ligands are highlighted: HGF: MET (Blue), EGF:EGFR/ERBB2 (Red), BMP2:BMPR1B/BMPR1A (Green), TGFB: TGFR1/TGFR2 (Yellow), OSM: IL6ST/OSMR (Orange), and IFNG:IFNGR1/IFNGR2 (Purple).

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

**D)** Cell count dose-responses after treatment with EGF, HGF, and OSM. Cell counts at 72H were normalized to the 10 ng/ml EGF condition.

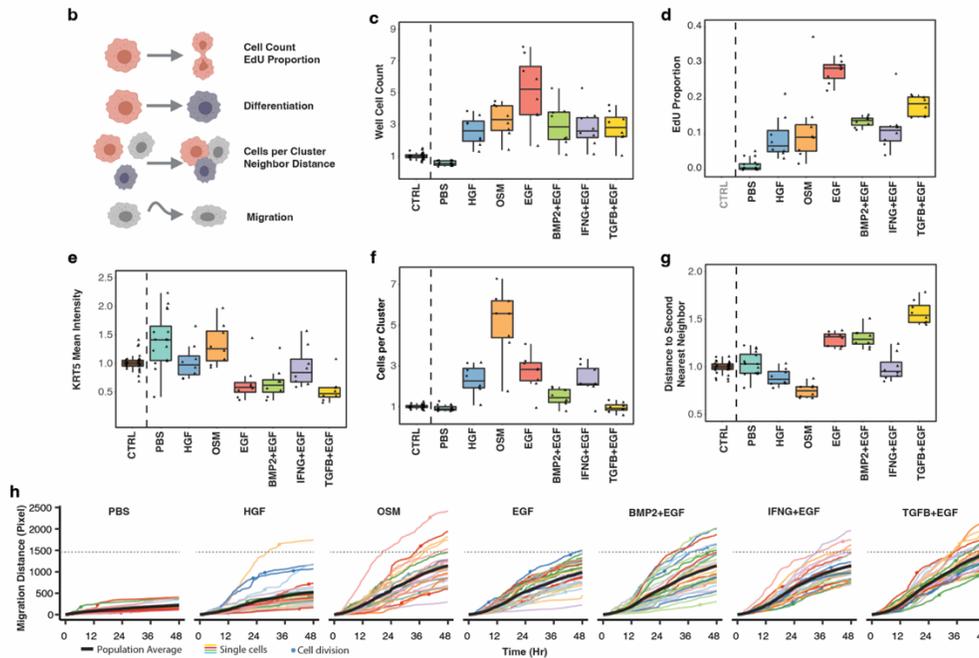
**E)** Cell count dose responses for TGFB1, IFNG, and BMP2. Each of the ligands were supplemented with 10 ng/ml EGF. Cells counts at 72H were normalized to the EGF condition with no secondary ligand.

#### 2.4.2 Overview of the ligand-induced cellular and molecular responses that comprise the LINCS ME perturbation dataset

*Cellular responses.* We quantified four-color immunofluorescence images from cells 24 and 48 hours after ligand treatment to assess cell clustering, cell density, shape, DNA content, and expression of proteins related to differentiation state, which revealed a constellation of changes following each treatment that were quantified with image analysis (**Figure 2-2A-B**). CyCIF collected at all time points revealed additional changes in cell state and pathway activity. Consistent with our MEMA screen, HGF, OSM and EGF increased cell numbers and EdU incorporation (a measure of proliferation). BMP2 and TGFB significantly suppressed growth relative to the EGF condition; IFNG also reduced growth (**Figure 2-2C-D**). HGF, OSM, and IFNG+EGF upregulated KRT5 expression, a marker of basal differentiation state in mammary epithelial cells [246] (**Figure 2-2E**). OSM caused cells to form tight clusters (**Figure 2-2F**). Lastly, TGFB+EGF induced evenly distributed cells with increased size, quantified as an increase in the distance to neighboring cells (**Figure 2-2G**). Together, these ligands constitute a powerful set of perturbations to probe molecular and phenotypic networks.

Analysis of live-cell images showed the emergence of each phenotype following ligand treatment. OSM induced cells to undergo collective migration, a unique phenotype among the tested ligands. We assessed cell migration by tracking individual cells across the 48 hour time period and quantified migration as the total distance traversed by each cell lineage (**Figure 2H**). In all ligand conditions, cell migration increased compared to the PBS condition, but to varying degrees: HGF-treated cells migrated the least while TGFB+EGF induced the greatest migration (Tukey's HSD,  $p$ -value  $< 9 \times 10^{-7}$ ). Together, the live cell imaging and migration analyses show the dynamic emergence of distinct phenotypic responses by each of the ligand treatments.

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses



**Figure 2-2: Ligand treatments induce diverse phenotypic responses.**

**A)** Representative immunofluorescent images of ligand-induced cellular phenotypes at 48H. MCF10A cells were stained with Cell Mask to visualize cytoplasm.

**B)** Cartoon showing the image-based cellular phenotypes assessed from the immunofluorescence and live cell imaging assays.

**C-G)** Boxplots summarizing cellular phenotypes at time 0H (CTRL) and 48H after ligand addition from 8 biological replicates. Individual datapoints represent well-level means normalized to 0H. Circles are from collection 1 and triangles are from collection 2. The interquartile range is indicated by the box, with whiskers extending to no further than 1.5 times the interquartile range. Note that EdU positive proportion was not measured at 0H.

**H)** Accumulated cell migration (colored lines) from 0-48H for 25 cell lineages (individual cells and one of their progeny if they divided). Circles indicate mitotic events. The solid black lines indicate the population average; the dotted gray line shows the average TGFB + EGF induced migration at 48H, which was the treatment that induced the greatest increase in cell migration.

**Molecular responses.** The responses to ligands involved numerous features in each of the molecular datasets. Here we demonstrate some of our key observations through analysis of the RPPA proteomic dataset as an exemplar use-case. We assessed the modulation of canonical signaling proteins downstream from each ligand (**Figure 2-3A**). These included: IRF1, a transcriptional target of STAT1 downstream of IFNG; pSTAT3, a signaling pathway component for OSM; and phosphorylation of MET, the receptor for HGF. PAI-1 provided an assessment of SMAD transcriptional activity, which is downstream of TGFB and BMP2. Additionally, phospho-HER2 provided a readout for conditions that contained EGF in the media. Each of these features were modulated as expected based on prior literature, validating the robustness of the dataset.

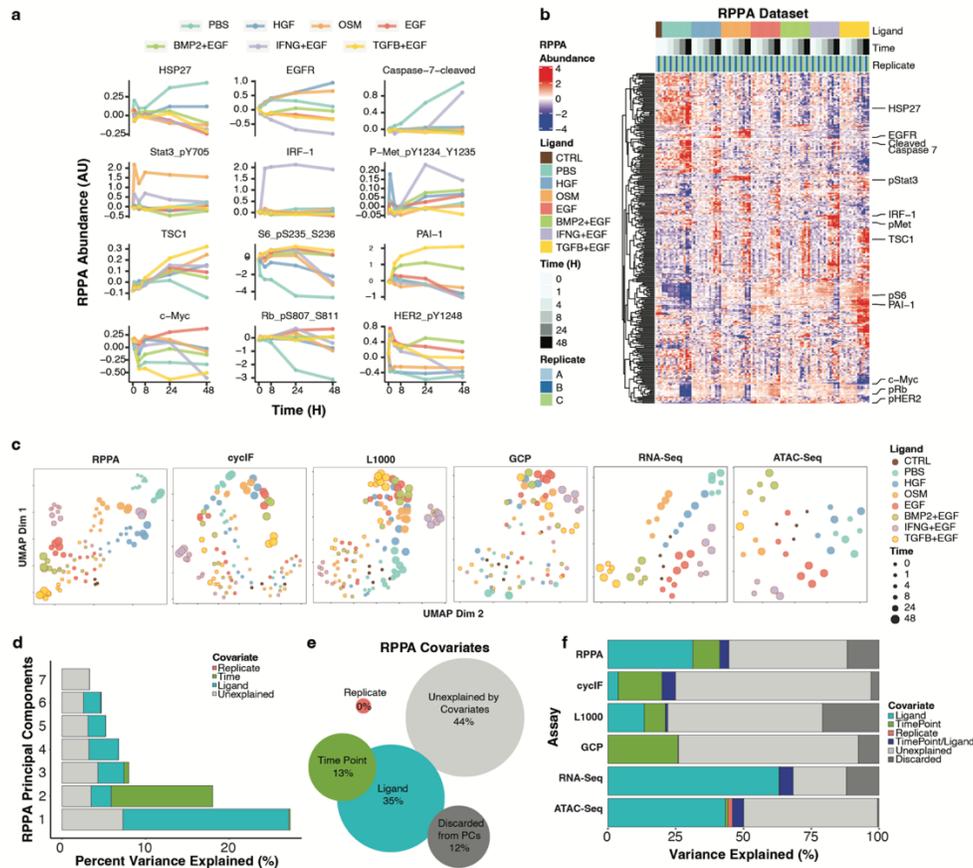
provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

Unsupervised hierarchical clustering of the RPPA data set revealed dynamic changes in the protein landscape over time, with some responses shared by multiple ligands and others that were uniquely induced (**Figure 2-3B**). All treatments that included EGF induced proteins related to growth factor signaling (*e.g.* pS6). The PBS condition, which lacks added growth factors, showed protein changes associated with reduced proliferation (*e.g.* decreased pRB) and induction of apoptosis (*e.g.* cleaved caspase 7), indicating that absence of growth factor signals strongly modulates phenotypic and molecular state.

To gain a high-level view of the six molecular assays, we performed Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction for all ligand-induced responses (**Figure 2-3C**). Most assays showed ligand-specific effects, as observed by samples from the same ligand treatment tending to group near one another. In addition, most datasets showed evolution over time from the starting state to another distinct state, captured by early time points clustering near the center of the UMAP and later time points for each ligand appearing in different UMAP regions. Principal Component Analysis revealed similar findings, though the variance was manifest in multiple components.

*Assessment of assay variance.* We applied the Measuring Association between Variance and Covariates method to systematically assess the fractional variance explained by ligand, time, and replicate [241]. In brief, we first performed principal component analysis to reduce the dimensionality of each data set while preserving the variability. Next, we quantified the total variance explained by each covariate (ligand, time, replicate) by summing the weighted variances of all statistically significant principal components (PCs). For example, in the RPPA dataset, the signal in the first PC was dominated by ligand while the second PC was dominated by time point (**Figure 2-3D**). We reasoned that PCs with an eigenvalue of less than 0.7 were unlikely to significantly correlate to any covariates and discarded these from the analysis. Summing across all significant PCs from the RPPA dataset revealed that 35% of the variance could be attributed to ligand and 13% to time point (**Figure 2-3E**). Variance explained by multiple co-variables is represented by overlap in the Venn diagram. Overall, 44% of the variance in the RPPA dataset could not be explained by one of these factors, suggesting signal in the data attributable to other factors, such as changes shared by multiple ligands. Similarly, all other assays carried signal attributable to ligand treatment, although to varying degrees: RNAseq (63.1%) and ATACseq (43.3%) contained the greatest ligand-associated signal while GCP (0.1%) contained the least (**Figure 2-3F**). Datasets with both early and late time points (RPPA, GCP, CyCIF) carried signal attributable to time. There was limited variation attributable to replicates across all assays, indicating modest biological and technical variation.

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses



**Figure 2-3: Six molecular assays reveal diverse dynamic responses to treatments.**

- A)** Line graphs show dynamic responses for 12 proteins measured in the RPPA assay under the different ligand treatments.
- B)** Heatmap of protein abundances as measured by RPPA. Rows represent abundance of 295 (phospho)proteins and are median-centered and hierarchically clustered. Columns represent individual replicate samples, ordered by treatment and time. Callouts show the 12 proteins from panel A.
- C)** UMAPs for each of the six molecular assays. Each dot represents data from an individual sample and is the 2-dimensional embedding of all features measured in the assay. Color indicates ligand treatment and size indicates time point.
- D)** Plot of the first two principal components (PCs) of RPPA assay. Variance in PC1 and PC2 is largely driven by ligand treatment and experimental time point, respectively.
- E)** Analysis of RPPA covariates reveals the proportion of variance explained by sample replicate, experimental time point, and ligand treatment for each of the top seven principal components of the RPPA dataset.
- F)** Stacked bar graph shows a comparison of the information content contained within each molecular assay.

### 2.4.3 Identification and analysis of ligand-induced molecular signatures

Here we present a systematic assessment of molecular signatures induced by each ligand and provide examples of how these signatures can be analyzed and mined. Specifically, we focus

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

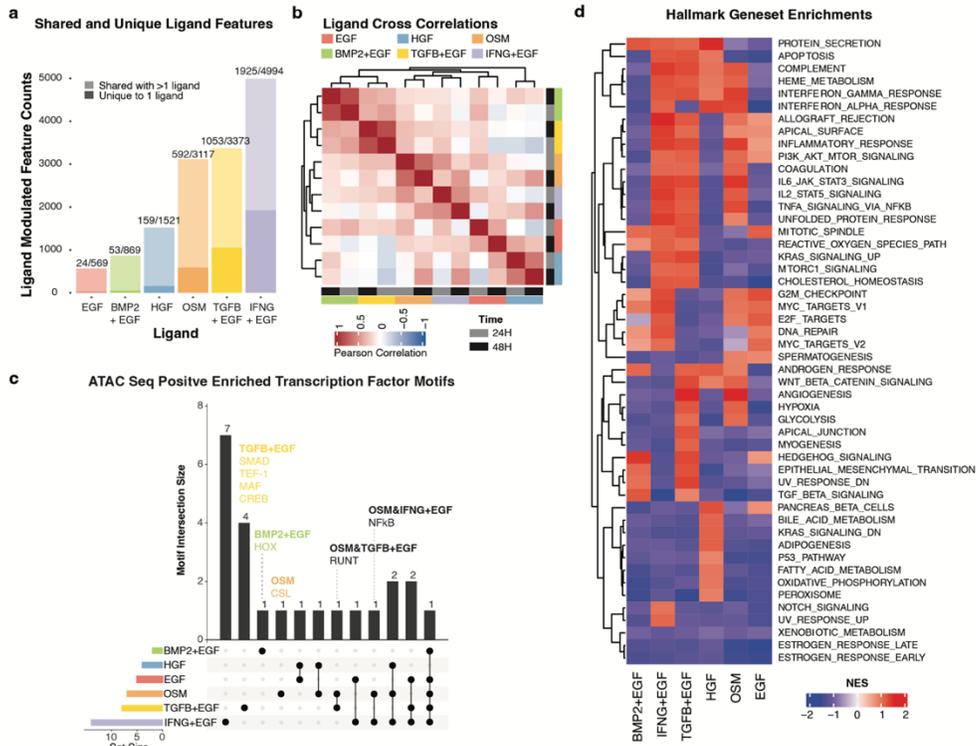
on IFNG+EGF to examine the temporal evolution of responses across modalities and to identify immune-related molecular features.

*Identification of ligand-induced signatures.* To create molecular signatures of ligand responses, we identified features from each of the 6 data types that were differentially expressed at 24H and 48H time points relative to the CTRL sample ( $q$ -value  $< 0.01$ ,  $|\logFC| \geq 1.5$ ) (**Figure 2-4A**). Features were classified as unique if they were modulated by a single treatment or shared if they were induced by more than one treatment. All treatments induced both unique and shared molecular responses. IFNG+EGF, TGFB+EGF and OSM induced the greatest shift in molecular state, as measured by the total number of features induced across the RNAseq, ATACseq, GCP, CyCIF and RPPA assays. In contrast, EGF, HGF and BMP2+EGF showed more modest effects, consistent with maintenance of MCF10A cells in a pre-treated state. Cross-correlation analysis of the molecular responses revealed that 24H and 48H responses were strongly correlated for each ligand and that responses to ligands from related families (BMP2/TGFB, OSM/IFNG, EGF/HGF) were more similar to one another than to other family classes (**Figure 2-4B**).

Motivated by our observation that the ATACseq and RNAseq datasets carried the strongest ligand signals, we more deeply interrogated these responses. We analyzed ATACseq transcription factor binding motif enrichment, a measure of transcription factor activity, and found that IFNG+EGF and TGFB+EGF induced the greatest number of enriched motifs. For example, TGFB+EGF induced SMAD, TEF-1, MAF and CREB motifs, while TGFB+EGF and OSM both induced changes in RUNT (**Figure 2-4C**). Gene set enrichment (GSEA) analysis [247] of the RNAseq dataset revealed a unique complement of gene programs associated with response to each ligand treatment (**Figure 2-4D**).

Ligand signatures that are strongly anti-correlated with drug-induced transcriptional signatures suggest environmental conditions that a therapeutic inhibitor could reverse and therefore may serve as a sensitizing signal, for example by inhibiting a ligand-activated pathway. Alternatively, if a ligand activates a pathway not affected by drug, this could serve as a possible bypass pathway to mediate resistance, which is captured as non-correlated responses.

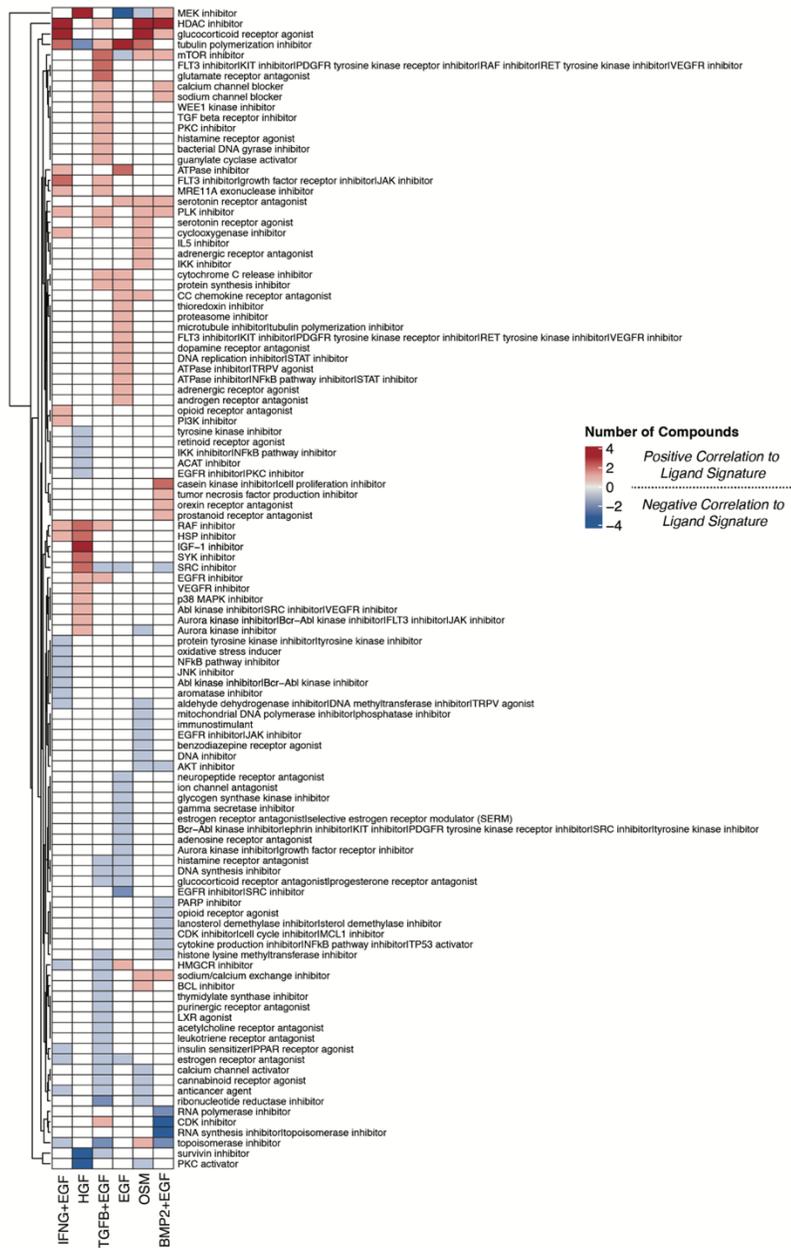
To test this, we compared our ligand signatures against the LINCS L1000 database [248], [249] of drug and other chemical response signatures (Fisher exact test,  $q$ -value $<0.2$ ). While some therapeutic inhibitor signatures were correlated with multiple ligands, the responses to most ligands were associated with a unique complement of inhibitor signatures (**Supplementary Figure 2-2**). For example, TGFB+EGF, BMP2+EGF, and EGF were negatively correlated with SRC inhibition, indicating that these ligands induce similar pathway activation along the SRC signaling axis. EGFR/JAK inhibitors were negatively correlated with OSM, suggesting that cells grown in OSM-rich environments may be particularly sensitive to JAK inhibition. All together, these findings indicate that extracellular ligands activate some of the same molecular programs as therapeutic inhibitors and that the impact of environmental signals on cellular and molecular state is an important consideration for identification of effective therapeutic regimens.



**Figure 2-4: Assessment of ligand-induced molecular change.**

- A)** Barplot showing the number of features significantly modulated by each ligand treatment at 24H or 48H. Shading indicates whether induced features are unique to a particular treatment (dark) or induced by multiple treatments (light). Numbers above bars indicate the number of features uniquely induced over the total number of features induced.
- B)** Heatmap showing pairwise correlations between molecular features induced by each ligand. Ligand responses from similar families are more highly correlated than those from unrelated families.
- C)** UpSet plot showing overlaps of induced transcription factor motifs among ligand treatments calculated from ATACseq data at 24H or 48H. Column heights represent the number of transcription factor motifs induced by the ligand(s) indicated with filled dots.
- D)** Hallmark Geneset enrichment scores computed from RNAseq data at 24H.

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses



**Supplemental Figure 2-2: Comparison of ligand and small molecule inhibitor signatures.**

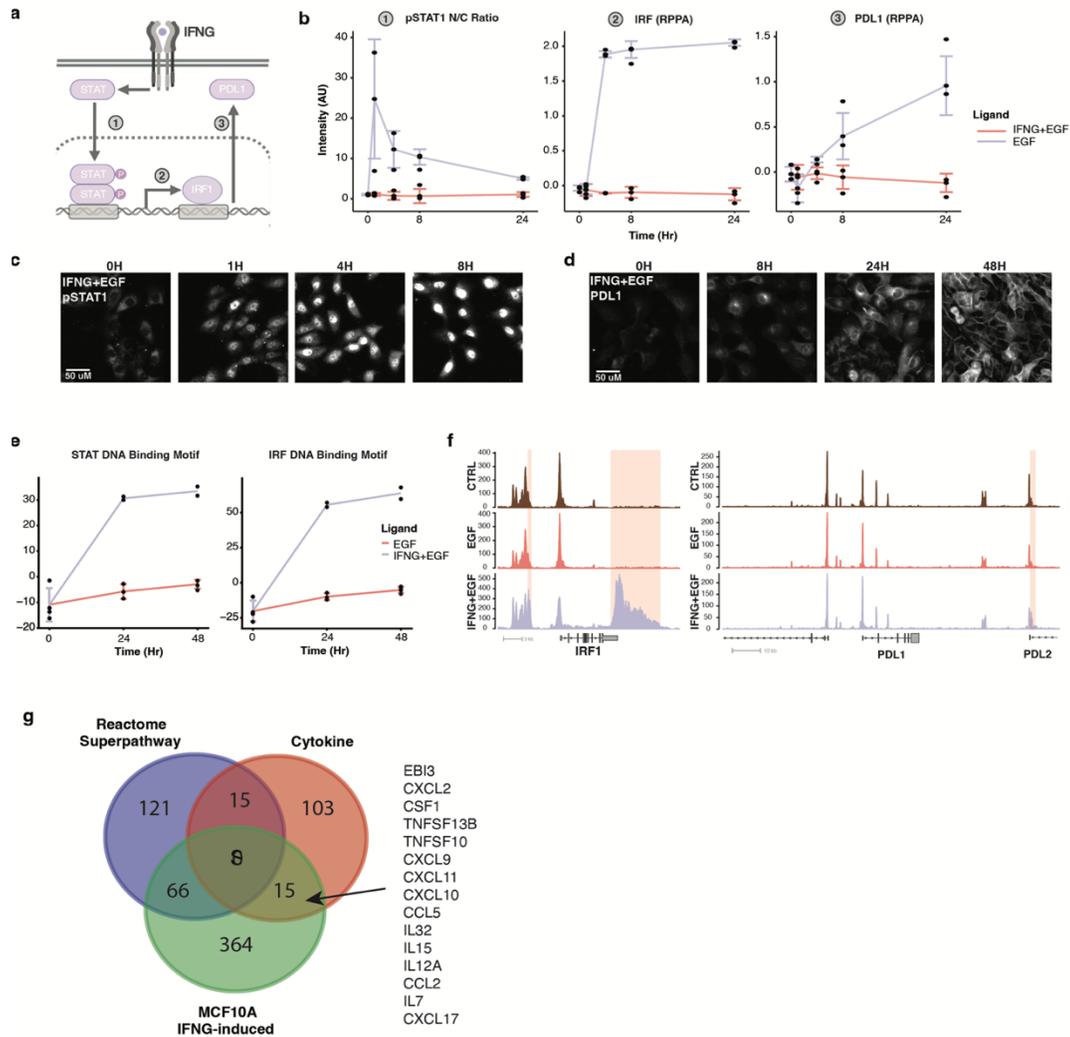
We leveraged the LINCS L1000 database of drug response signatures to identify targeted inhibitors that are shared by each ligand signature. Heatmap represents the number of compounds that have correlated (red) or anti-correlated (blue) signatures with each ligand (Fisher exact test,  $q$ -value < 0.2). The ligand panel activated many of the same signatures as small molecule inhibitors, indicating that shared molecular responses can be elicited by these distinct perturbagen classes.

*Identification of molecular features induced by IFNG.* We analyzed responses to IFNG+EGF to illustrate how the LINCS ME perturbation dataset can be used to study the molecular mechanisms associated with ligand responses across time. IFNG is a soluble cytokine secreted by cells of both the innate and adaptive immune systems and has become increasingly

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

scrutinized, owing to interest in understanding the role of the immune system in diverse pathophysiological [249] as well as cancer immunotherapies. IFNG+EGF treatment induced dynamic changes in canonical IFNG signaling molecules measured across assays and time, including: rapid nuclear translocation of STAT1, the resultant induction of IRF1 followed by upregulation of PDL1 at the membrane as well as associated epigenetic changes (**Supplementary Figure 2-2, 2-3A-F**). These findings indicate that the LINCS ME perturbation dataset enables the encoding of a stimulus to be traced across time and molecular modalities.

We observed that 66/202 Pathcards Reactome IFNG superpathway features [250] were among the most strongly modulated by IFNG+EGF treatment, indicating the induction of multiple known signaling responses (**Supplementary Figure 2-3G**). To gain deeper insight into the ability of IFNG to influence both adaptive and innate immune responses through altering cytokine production by malignant cells, we compared the MCF10A IFNG+EGF signature, the IFNG superpathway, and a curated cytokine gene list [251]. This comparison identified 15 cytokines not already included in the IFNG superpathway, suggesting additional cytokines produced by malignant cells in response to IFNG that may interact with various immune cell subsets, including: CSF1 [252], [253], IL15 [254], IL12A [255], CCL2 [256], and CXCL2 [257]. This demonstrates how the LINCS ME dataset can be mined to gain biological insights into immune-related signaling and to prioritize molecular features for future study.



**Supplemental Figure 2-3: IFNG responses are dynamically encoded across multiple molecular modalities.**

**A)** Cartoon of canonical STAT pathway activation after treatment with IFNG ligand.

**B)** Line graphs show induction of pSTAT1, IRF1 and PDL1 protein expression following IFNG treatment, as measured by CyCIF and RPPA assays. Error bars represent the standard deviation from the median value. Values from each replicate are plotted as individual data points.

**C-D)** Cyclic immunofluorescence images show changes in STAT1 and PDL1 protein abundance and localization induced by IFNG+EGF treatment.

**E)** Line graphs show enrichment of STAT-family and IRF-family motifs inferred from ATACseq chromatin accessibility data for IFNG+EGF and EGF conditions. Error bars represent standard deviation. Low quality EGF+IFNG samples were removed, yielding only two replicates for this treatment.

**F)** Chromatin accessibility near the IRF1 (chr5:132390440-191020960) and PDL1 (chr2: 190908460-191020960) gene loci. The local gene region for IRF1 showed a new peak in the promoter region and a large accessibility change in the 3' region. IFNG did not induce new ATACseq peaks in PDL1, however IFNG induced a new peak in the adjacent PDL2 gene (PCD1L1G2). DNA regions with changes in accessibility are marked with a red background. g

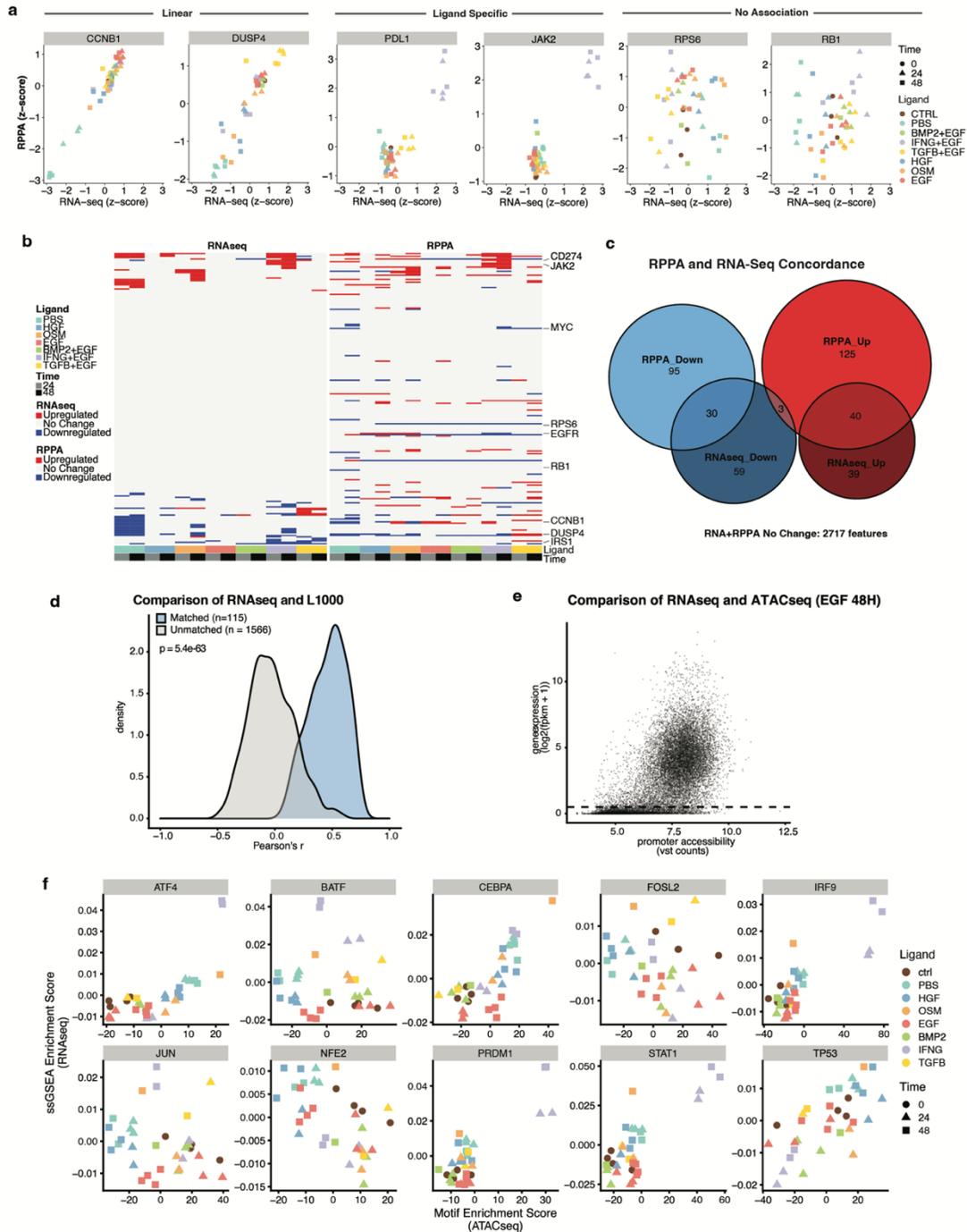
provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

Venn diagram showing the overlap between the Reactome IFNG pathway, curated cytokine gene lists and, and genes induced by IFNG+EGF treatment. The 15 cytokines induced by IFNG+EGF are listed on the right.

#### 2.4.4 Discovery of candidate functional relationships between molecular features

We reasoned that the patterns of robust multi-omic molecular changes induced across the panel of ligands could be analyzed together to discover coordinately regulated molecular programs. Importantly, our use of multiple ligands that perturb cells along various phenotypic and molecular axes enabled distinct molecular programs to be disentangled. Below we summarize our assessment of the relationships between different modalities, our approach to identify coordinately regulated biological modules, and also illustrate the utility of the modules to provide insights into the molecular programs active across diverse tissues.

*Identification of coordinately regulated modules.* We assessed coordinated responses in the RPPA, RNAseq, L1000, and ATACseq datasets by comparing molecular cognates across datasets that could be mapped through gene names (*e.g.* Cyclin B1 in RPPA and *CCNB1* in RNAseq). This revealed broad concordance, indicating conserved responses across molecular modalities (**Supplementary Figure 2-4**). For example, the relationships between RPPA and RNAseq showed several patterns: linear correlation (*CCNB1*, *DUSP4*); ligand-specific effects (*PDL1*, *JAK2*); or no association, which typically reflected only modest ligand-induced changes in abundance (*RPS6*, *RB1*). We assessed response concordance, which we defined as similar induction (up- or down-regulation) as compared to the CTRL samples, which revealed 40/207 features were concordantly up-regulated and 30/187 features were concordantly down-regulated in the RNAseq and RPPA datasets. Importantly, we also observed that 2717/3035 features were concordantly unchanged. Next, we measured Pearson correlation of RNAseq and L1000 gene expression measurements for matched and unmatched samples and found that matched samples were on average significantly better correlated than gene expression profiles from unmatched samples (Mann-Whitney U test;  $p < 2.2 \times 10^{-16}$ , **Supplementary Figure 2-4D**). In a third cross assay comparison, we found that chromatin accessibility was bimodal and that promoter accessibility was associated with transcriptional expression, consistent with prior studies [258] (**Supplementary Figure 2-4E**). Finally, we compared the 10 most-variant ATACseq transcription factor motifs to single-sample gene set enrichment scores for the same transcription factors from CyCIF data and found that they were generally concordant (**Supplementary Figure 2-4F**).



**Supplemental Figure 2-4: Comparison across assays reveals concordance in response to ligand treatment**

**A)** Scatter plots of paired RPPA and RNAseq measurements, showing three classes of observed relationships: linear, ligand-specific, and no change.

**B)** Heatmaps show genes and proteins with significantly up- or down-regulated expression after ligand treatment. Genes were defined as significantly up- or down-regulated from RNAseq ( $q$ -value  $< .01$ ;  $\log_2FC > 1.5$ ), while proteins were defined from analysis of RPPA ( $q$ -value  $< .01$ ;  $\log_2FC > 0.5$ ).

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

**C)** Euler diagram showing intersections of differentially expressed RPPA proteins and RNAseq genes. The majority of features that were induced in both assays showed concordant responses, defined as both modalities induced in the same direction.

**D)** Distributions of Pearson's correlation coefficients from RNAseq and L1000 gene expression profiles of treatment-matched and treatment-mismatched samples. Both datasets were z-transformed after filtering to overlapping genes and samples. Transcriptional profiles from treatment-matched samples were significantly more correlated than were transcriptional profiles from nonmatched samples matched samples were significantly more correlated than were transcriptional profiles from nonmatched samples.

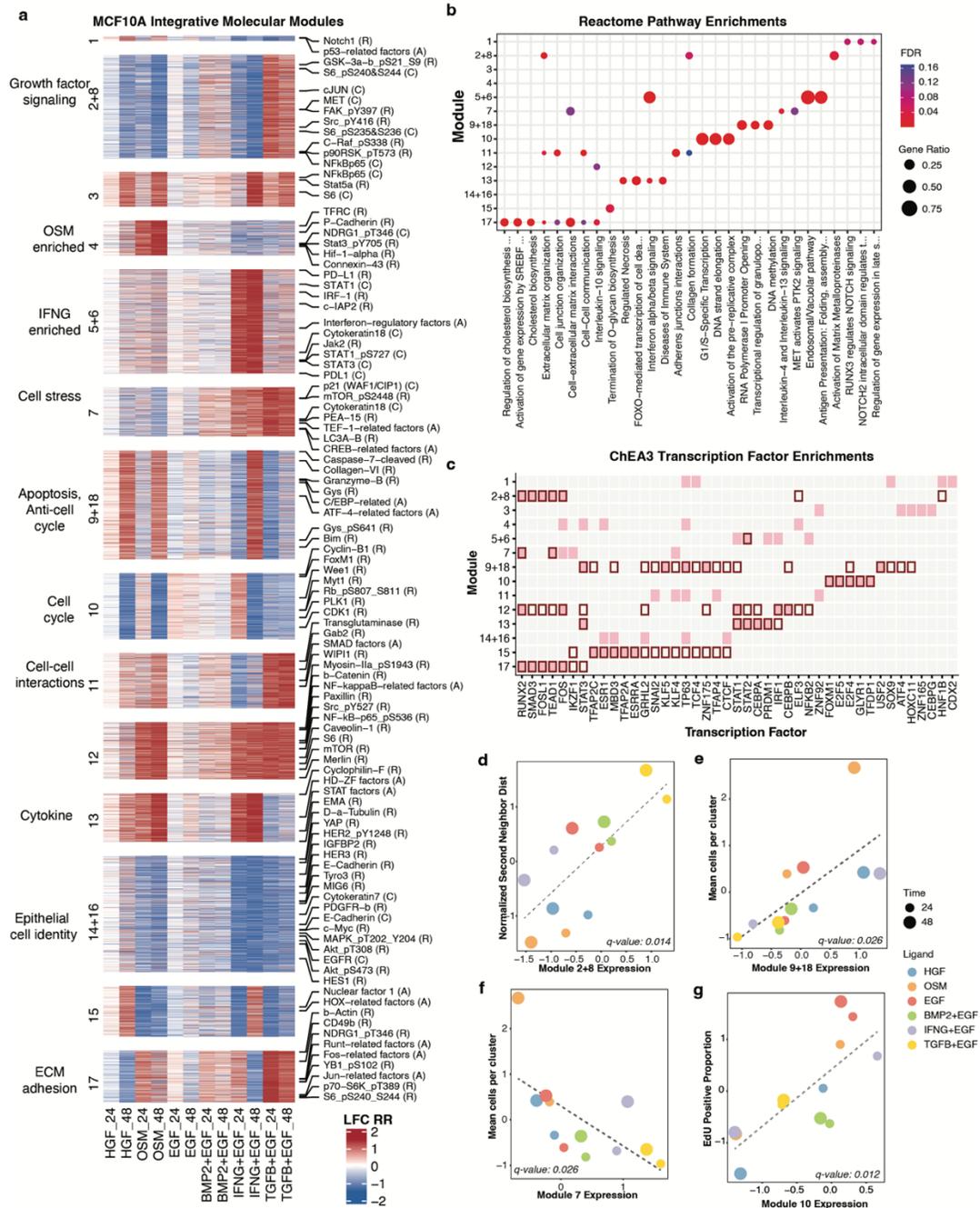
**E)** Dot plot showing the relationship between ATACseq transcriptional start site (TSS) accessibility and gene expression in the EGF 48H samples. Note the switch-like relationship between gene expression and accessibility at the TSS, as has been described previously. The horizontal dotted line indicates the threshold for a gene defined as being expressed.

**F)** Comparison of ATACseq and RNAseq data for the 1- most-variant transcription factor motifs (by standard deviation). Motif enrichment scores (x axis) represents the deviation in motif chromatin accessibility from an expectation based on the average accessibility of the motif across all samples, while ssGSEA scores (y axis) represent the degree of coordinated expression of the TF target gene expression.

We next used a systematic approach to identify modules comprised of coordinately regulated molecular features measured in the different assays (CyCIF, RPPA, GCP, and RNAseq, and ATACseq). Specifically, we examined all molecular features that were induced by at least one ligand (see **Figure 2-4A**) and then scaled each assay dataset with *rrscale*, which is a transformation that normalizes feature distributions, removes outliers, and z-scales feature values [242]. We used gap analysis [259] to identify the optimal number of clusters, and then used consensus clustering with partitioning around medoids (PAM) to identify stable clusters. To further ensure that the clusters represented unique expression patterns, we calculated their pairwise correlations and combined highly correlated pairs, which yielded a final set of 14 molecular modules for interpretation (**Supplementary Figure 2-5A-C**).

Each module represents a unique complement of co-regulated proteomic, transcriptional, and chromatin features (**Figure 2-5A**). Features from each assay were distributed across modules, indicating that our analytical approach enabled integration of features measured in diverse assays (**Supplementary Figure 2-5D**). Each module showed distinct modulation patterns across the ligands; most modules were induced by more than one ligand while a few were ligand-specific, consistent with the findings in **Figure 2-4**. Reactome pathway enrichment analysis demonstrated that each module induced an array of transcriptional programs (**Figure 2-5B**). Transcription Factor enrichment via ChEA3 [229] identified key molecular drivers associated with these modules (**Figure 2-5C**). To explore how our clustering method compared against other published multiomics approaches [243], we performed a Consensus Principal Component Analysis (CPCA) using the R package *MoCluster* [244], which showed similar ligand-specific expression patterns (**Supplementary Figure 2-5E-I**).

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses



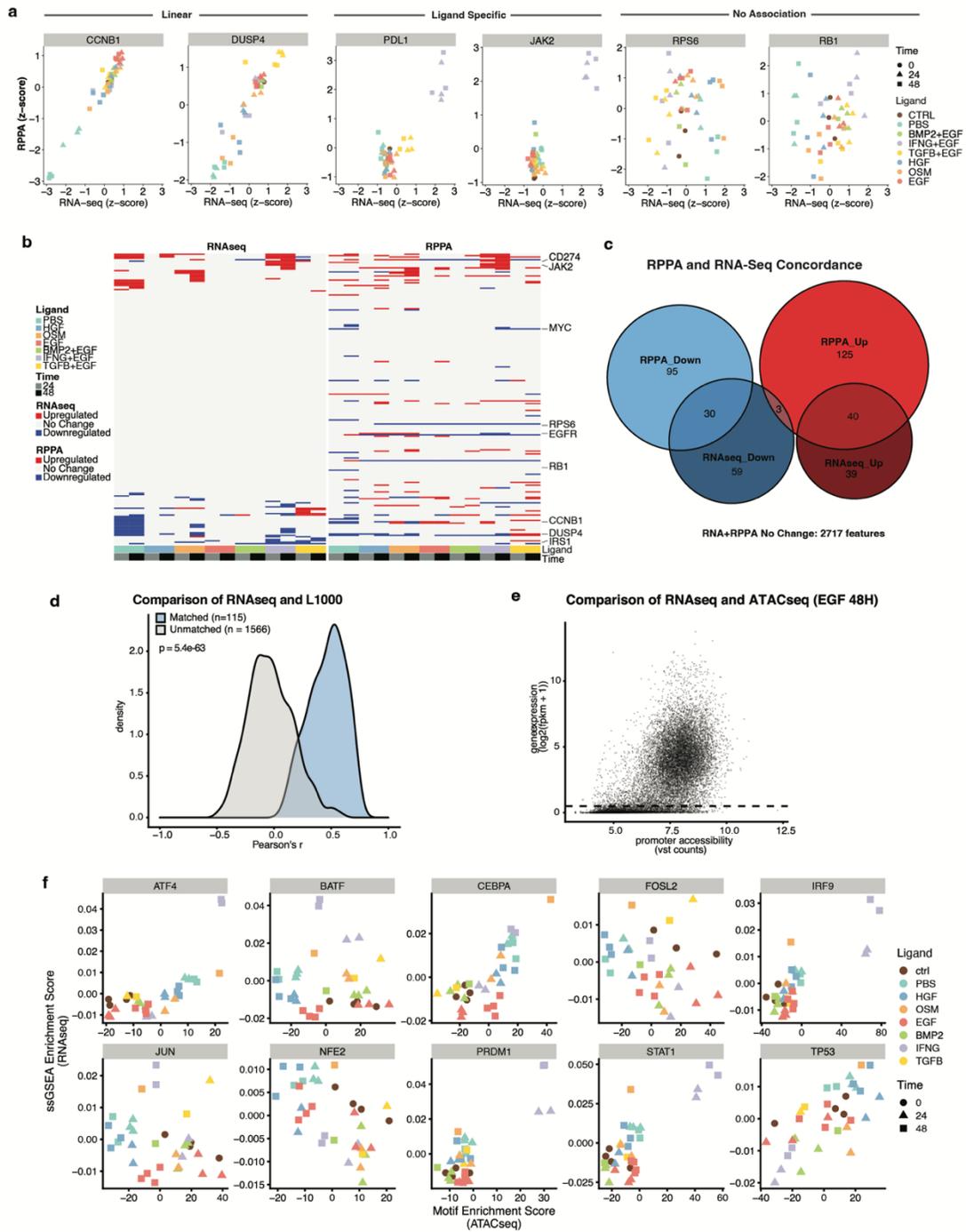
**A)** Heatmap showing the 14 integrative molecular modules for each ligand at 24H and 48H. Features are grouped by cluster. Biological interpretation for modules is indicated on the left; feature callouts for RPPA (R), CyCIF (C), ATACseq (A) are shown to the right.

**B)** Bubble plot shows the top enriched Reactome pathways in each module, computed from RNAseq features. Dot size indicates the gene ratio; dot color indicates FDR value.

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

**C)** Heatmap showing the five top-ranked ChEA3 transcription factor enrichments computed from the RNAseq features in each module (pink). Red border indicates transcription factor enrichments with a q-value below 0.2 (FDR-adjusted Fisher's exact test).

**D-G)** Scatterplots show the relationships between module activity and quantitative phenotypic responses for selected pairs. Dot color indicates the ligand treatment and dot size indicates the time point. The black dotted line shows the linear fit, and the q-value of the fit is shown at the bottom of the plot.



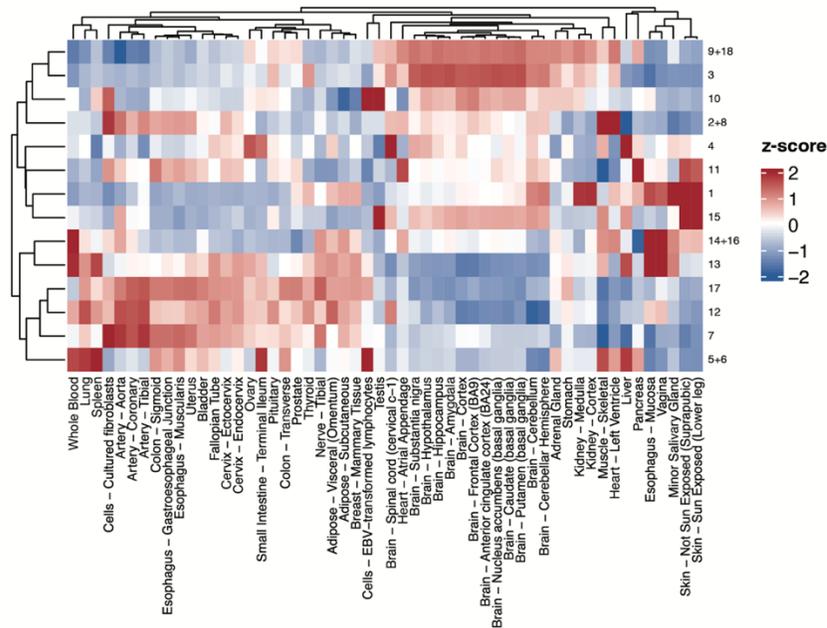
Supplemental Figure 2-5: Identification and characterization of integrative molecular modules

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

- A) Bar plot showing the number of features for each assay included in the integrative modules; note log 10 scale.
- B) Gap analysis used to identify the optimal number of modules. Error bars show +/- the standard error from the simulations.
- C) Module correlation matrix showing Pearson correlation values. Highly correlated cluster pairs 2+8, 5+6, 9+18 and 14+16 were combined to yield 14 clusters.
- D) Bar plot showing the distribution of features for each assay across modules.
- E) Bar plot showing the mean module expression for each of the ligand treatments.
- F) Consensus Principal Component Analysis identifies multiomic modules, and analysis of the resultant eigenvalues for the joint latent variables (JLVs) shows a knee at 8 modules.
- G) Heatmap showing expression of the top 8 JLVs for each treatment condition. Heat color indicates CPCA score.
- H) Heatmap showing the mean expression of the optimized modules across the treatment conditions.
- I) Heatmap showing the correlation in expression patterns for JLVs and module scores shown in g and h. There is high correlation between the first 4 JLVs and the consensus module scores.

*Assessment of molecular modules across diverse tissues.* Elucidating the molecular programs operable across different tissue types is critical for understanding normal organ development and function and for identifying molecular programs that may go awry in the case of disease. We assessed RNA expression of the 14 integrated modules in the GTEx normal tissue dataset [210] to identify molecular programs that may be most active in particular tissue types (**Supplementary Figure 2-6**). We observed tissue-specific activation of the modules. For example, Module 14+16 included features associated with epithelial cell identity such as cytokeratin-7, E-cadherin, claudin-7, and EGFR, and was upregulated in vagina, esophagus, and skin. These tissues are comprised principally of stratified squamous epithelial cells which undergo rapid terminal differentiation as they migrate from a basal zone to cornified surfaces [260], [261], [262]. This suggests that deeper analysis of the molecular features coordinately regulated by module 14+16 may shed light on key molecular programs important for differentiation and maintenance of epithelial cell state across diverse tissues. Module 2+8 was enriched in extracellular matrix organization and collagen formation pathways. This module was highly expressed in artery samples, consistent with the observation that the arterial wall produces a rich and complex extracellular matrix that defines the mechanical properties of the vessel [260], [261]. Additional features included in each of these modules may provide additional insights into their roles in normal and diseased processes in different tissues.

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses



Supplemental Figure 2-6: *GTEx expression analysis*

Heatmap showing GTEx tissue expression of the 14 integrative molecular modules reveals tissue-specific expression, suggesting molecular programs that may be particularly important for mediating normal and diseased functions across tissues.

### 2.4.5 Investigation of the relationship between molecular modules and cellular phenotype

Elucidation of the molecular mechanisms that control cellular phenotype remains a difficult problem in systems biology. We illustrate here how the LINCS ME perturbation dataset can be analyzed to gain insights into mechanisms of phenotype control by linking cellular and molecular responses. We present two examples: a data-driven discovery of associations between phenotypic responses and module activity, followed by a detailed analysis of Module 4 to uncover molecular features associated with the cell clustering and collective motility phenotype induced by OSM.

*Data-driven discovery of phenotype-module associations.* We performed correlation analysis to identify molecular modules that were significantly associated with cellular phenotypes measured by imaging (**Figure 2-5D-G**). For example, Module 2+8 was positively correlated with ‘Normalized Second Neighbor Distance’, a metric that reflects both cell size and cell-cell spatial organization (**Figure 2-5D**,  $p$ -value = 0.014). Several features of this module suggest molecular correlates of this phenotypic response, including pathway enrichments in Extracellular matrix organization and Collagen formation. Additionally, the transcription factor *RUNX2*, which was enriched in this module, has been implicated in modulating cell morphology and cell spreading [263].

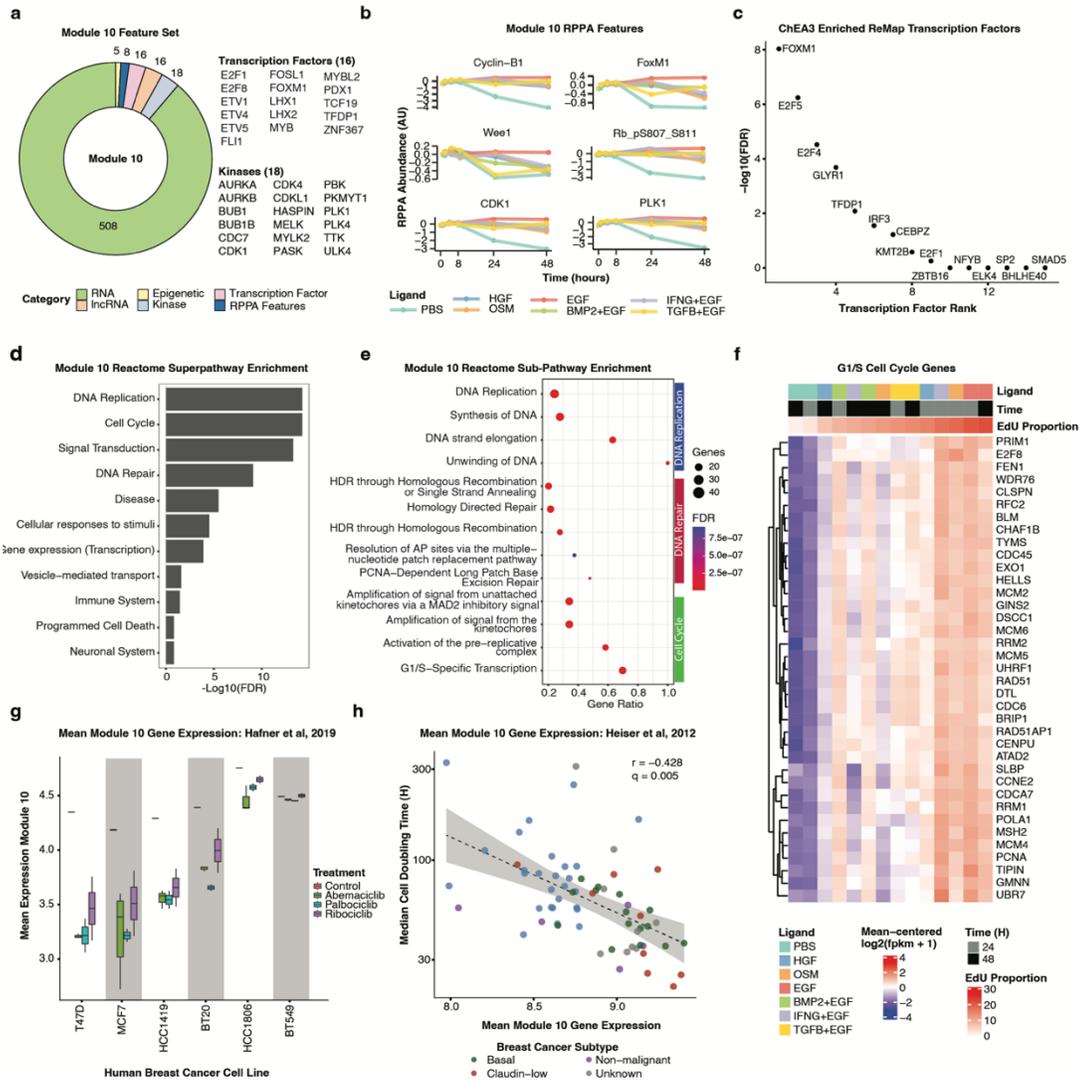
We also identified a specific and robust correlation between Module 10 expression and the fraction of EdU positive cells, a measure of cell proliferation (**Figure 2-5G**,  $p$ -value = 0.012). To explore the putative regulatory components of Module 10, we annotated genes that code for transcription factors, kinases, non-coding RNA, and epigenetic regulators (**Figure 2-6A**). This analysis revealed a suite of factors previously shown to play key roles in regulating cell cycle progression, including the transcription factors:

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

*E2F1*, *FOXM1*, *MYB*, and *TFDP1*; and the kinases: *AURKA*, *CDK1*, *PLK1*, and *BUB1*. Module 10 RPPA features cyclin B, Wee1, and phosphorylated RB are canonical cell cycle proteins that showed temporal dynamics consistent with changes in proliferation, as well as lesser linked features including FOSL1 [264], [265], [266] and PASK [267], [268] (**Figure 2-6B**). ChEA3 transcription factor enrichment [229] identified multiple cell cycle-associated transcription factors including *FOXM1*, *TFDP1* and E2F isoforms (**Figure 2-6C**). Among the most significantly enriched Reactome pathways were Cell Cycle, DNA replication, and DNA repair (**Figure 2-6D**). We analyzed the top 5 sub-pathways within each of these Reactome pathways and found the highest enrichment for G1/S specific transcription, PCNA-dependent base excision repair, and unwinding of DNA (**Figure 2-6E**). Additionally, Module 10 included 86% (37/43) of the genes in a functionally-annotated G1/S gene set, with expression patterns consistent with changes in EdU incorporation (**Figure 2-6F**). There is also evidence for DNA damage and potentially for replication stress in the induction base-excision repair, the G2M checkpoint and activation of DNA damage checkpoint associated kinases. In sum, Module 10 contains cell cycle-associated molecular features from multiple modalities.

To test if the link between Module 10 and cell cycle control generalized beyond MCF10A cells, we analyzed two publicly available independently generated breast cancer cell line data sets. First, we quantified mean Module 10 gene expression scores from 7 breast cancer cell lines treated for 24 hours with a panel of CDK4/6 inhibitors [269]. As expected, this showed robust down-regulation of Module 10 in response to each of the three CDK4/6 inhibitors in the five sensitive cell lines, while the two resistant cell lines showed only modest changes in Module 10 expression (Mann-Whitney U test, p-value = 0.028, **Figure 2-6G**). In a second analysis, we compared Module 10 expression for a panel of 65 breast cancer cell lines<sup>10</sup> against cell doubling time, which revealed a significant correlation, consistent with the interpretation that Module 10 is functionally associated with the cell cycle (**Figure 2-6H**, Pearson R = -0.428). All together, these analyses indicate that our data-driven approach to module detection can identify coordinately regulated molecular features associated with quantitative phenotypic responses and that these findings generalize to independent data sets.

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses



**Figure 2-6: Module 10 is associated with cell cycle progression**

**A)** Donut plot showing distribution of Module 10 features across assays. Transcription factors and kinases in the RNA gene set are called out to the right of the plot.

**B)** Line plot showing 6 of the Module 10 RPPA features.

**C)** Plot of the top 10 most significantly enriched transcription factors inferred from the Module 10 RNAseq gene set.

**D)** Bar plot shows the enrichment of Reactome superpathways from the Module 10 RNA gene set.

**E)** Bubble plot showing the top 5 enriched Reactome subpathways from the Reactome Cell Cycle, DNA Repair, and DNA Replication superpathways. Dot color indicates q-value; dot size indicates the number of genes in Module 10 that are found in each gene set.

**F)** Heat map showing expression of Seurat G1/S cell cycle genes in Module 10 (37 of 43 genes shared), sorted based on the EdU positive proportion.

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

**G)** Boxplot of mean Module 10 gene expression for a panel of breast cancer cell lines treated with three CDK4/6 inhibitors for 24H or an untreated control. Cell lines are ordered by abemaciclib GR50 (increasing). The interquartile range is indicated by the box, with whiskers extending to the minimum and maximum values. Data from Hafner, et al.

**H)** Dot plot of mean Module 10 gene expression from 65 human breast cancer cell lines graphed against their mean doubling time. Cell lines are colored based on their breast cancer subtype classification. The line indicates the linear fit across all cell lines, with the 95% confidence interval represented by the gray shaded area. Data from Heiser et al.

#### 2.4.6 Examination of module activity to elucidate the molecular basis of ligand-induced phenotypic responses

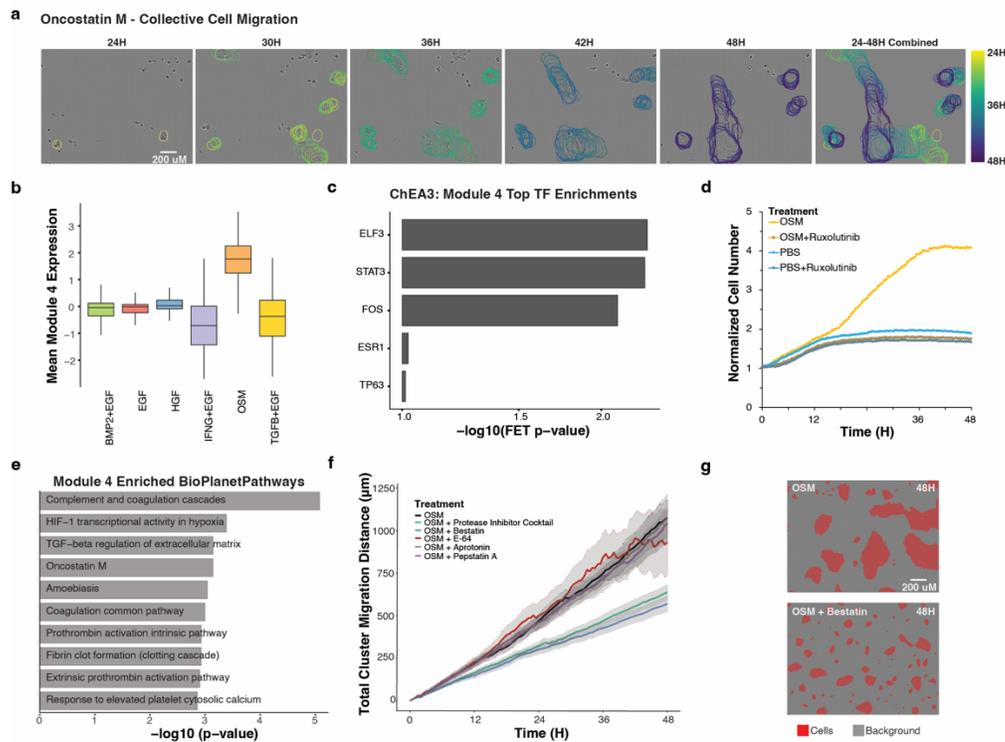
In our final analysis, we illustrate how the modules can be examined to provide insights into the molecular basis of complex phenotypic responses. Here, we focused on OSM, a member of the IL6 cytokine family implicated in immune function, developmental processes, and tissue remodeling [270]. OSM stimulated proliferation and was the only ligand in our panel that induced collective migration, a complex phenotype in which individual cells form tight clusters that undergo migration (**Figure 2-7A**). To date, the molecular correlates of collective cell migration are not well understood, and our dataset provides a unique opportunity to study this behavior.

To gain insight into the molecular features underlying this unique phenotype, we focused on modules that were strongly induced by OSM, including Modules 4, 12 and 13 (**Supplementary Figure 2-5E**). Features in Module 4 were of particular interest, as this module was selectively induced by OSM (**Figure 2-7B**). Module 4 includes RPPA features pSTAT3, P-Cadherin, Connexin-43, and Hif-1-alpha as well as top-ranked transcription factors *ELF3*, *STAT3*, *TP63*, and *FOS* from ChEA3 analysis (**Figure 2-7C**). P-Cadherin and Connexin-43 are intriguing, as they are implicated in the cell adhesion contacts required for mediating the observed clustering phenotype [271], [272]. Based on the coordinated changes in STAT3 across modalities, we tested the functional importance of this axis with Ruxolitinib, a JAK/STAT inhibitor. We found that addition of Ruxolitinib in the presence of OSM strongly inhibited both the growth of cells and cell migration, confirming the importance of JAK/STAT signaling in mediating responses to OSM (**Figure 2-7D**).

To probe more deeply into the Module 4 RNAseq features and augment our Reactome enrichment findings, we tested for enriched pathways using BioPlanet [273] (**Figure 2-7E**). One of the top pathway hits in this analysis was 'OSM', which serves as a validation of the module approach. The most enriched pathway was 'complement and coagulation cascades', two linked processes driven by a series of proteases to stimulate innate immunity and blood clotting [274]. This suggested that protease activity may be critical for mediating OSM-induced cluster migration. To examine the role that proteases play in cluster migration, we treated MCF10A cells with OSM in the presence of a cocktail of five protease inhibitors and found reduced cluster migration, indicating the importance of protease activity in mediating this phenotype (**Figure 2-7F**). We next tested individual components of the protease cocktail and found limited effects of aprotinin, E-64, and pepstatin A. However, with bestatin, an aminopeptidase inhibitor, we observed formation of cell clusters but a failure of these clusters to migrate and merge (**Figure 2-7G**). Thus, these functional studies developed from the module analysis implicate aminopeptidase activity as a critical mediator of OSM-induced collective cell motility in MCF10A cells. Overall, our approach to leverage responses to multiple perturbations enabled identification of

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

molecular programs associated with complex phenotypic responses including cluster migration and cell proliferation.



**Figure 2-7: Analysis of molecular modules identifies functional relationships between molecular and phenotypic responses to OSM**

**A)** OSM induces the formation of cell clusters that undergo collective migration and merge to form large clusters. Representative tracks of OSM-induced cluster migration are shown from 24H to 48H after OSM treatment. Cluster outlines are colored by experimental time point. All images are set to the same scale.

**B)** Boxplot shows the mean expression of molecular features in Module 4 for each of the six ligand treatments. The boxplots' lower and upper hinges correspond to the first and third quartiles. The median is shown as the center line. The upper whisker extends from the hinge to the largest value no further than  $1.5 * \text{IQR}$  from the hinge (where IQR is the inter-quartile range, or distance between the first and third quartiles). The lower whisker extends from the hinge to the smallest value at most  $1.5 * \text{IQR}$  of the hinge.

**C)** Barplot showing the top 5 enriched transcription factors inferred for the Module 2 genes in ChEA3.

**D)** The JAK/STAT inhibitor Ruxolitinib inhibits cell growth in the presence of OSM. Line graph shows the relative number of cells across time. PBS (phosphate buffered saline) treatment serves as a control.

**E)** Barplot of the top 10 enriched pathways in Bioplanet using the module 4 RNAseq gene set.

**F)** OSM-induced collective migration is mediated by protease activity. Line graph shows the accumulated cluster migration distance after OSM + /- a protease inhibitor cocktail and its individual components including bestatin, E-64, aprotonin, and pepstatin A. Solid lines show the population average and gray shaded regions indicate 95% confidence intervals of the mean distance travelled at each time point.

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

**G)** False color phase contrast images at 48H show that bestatin inhibits the formation of large cell clusters when given in conjunction with OSM. Cells are colored red and the background is colored gray.

## 2.5 Discussion

We leveraged the LINCS Consortium framework to systematically quantify the phenotypic and molecular responses of MCF10A mammary epithelial cells after treatment with a diverse panel of ligands. Analysis of this dataset revealed robust molecular and phenotypic responses and enabled identification of ligand-specific signatures, integrated molecular modules, and linkage of phenotypic and molecular responses. These data support the idea that deeply examining a single model system subjected to a range of perturbations with measurements across multiple modalities is crucial to understanding complex biological phenomena.

The robust, multimodal dataset enabled a range of computational analyses. For instance, the coordinated use of a diverse panel of molecular assays facilitated comparisons of the information carried by each assay and revealed that RNAseq and ATACseq assays had the greatest ligand-associated signal. Differences in information content between assays may be due to: intrinsic differences in molecular modalities, the signal available in a particular assay, or differences in the number and diversity of biologically meaningful features in each assay. These findings suggest that comprehensive assays such as RNAseq are well-suited for discovery-based screens or experiments that examine large panels of perturbagens, whereas targeted assays such as CyCIF—which can be adapted through inclusion of different biomarkers—would be expected to excel in more focused hypothesis-driven studies [150], [239].

In our integrated analysis, we joined epigenomic, transcriptional and proteomic changes into co-regulated modules. Critical for this analysis was the use of ligands that stimulate diverse and partially overlapping pathways, as this enabled identification of molecular features that were subtly and variably induced by multiple ligands. We analyzed the modules to identify linkages between molecular features and phenotypic responses. For instance, we identified a set of co-regulated molecular features strongly associated with cell cycle, including both canonical transcriptional factors, pathways, and proteins as well as features that have been implicated but not confirmed in cell cycle regulation, such as PASK [267], [268]. Importantly, we showed that this cell cycle module, which was derived from integrating all 6 ligand perturbations, could generalize to independent datasets comprised of multiple cell lines. Some modules were semi-correlated and contained similar biological programs, as indicated by enrichment of shared pathways and TF programs. Alternate methods to identify modules that permit partial membership of individual features may allow a more nuanced identification of the relationship between molecular features and phenotypic responses [275].

Our findings support the idea that systematically testing multiple perturbations of a single model system can identify molecular programs that are operable in distinct cellular contexts. We assert that identification of these generalizable programs was possible precisely because we used multiple perturbations in a single model system. However, there are also limitations to this approach. For example, a molecular or phenotypic response to a perturbation could be context dependent and may not be observed in other cell lines or model systems. Further exploration of additional cell lines using a panel of perturbations could facilitate identification of the context dependence of the responses we observed and also would enable refinement of the underlying regulatory networks. Indeed, in the disease setting, the assessment of molecular and functional responses in panels of cell lines has proven a

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

powerful approach to identify biological mechanisms common to different disease states. Additionally, an expanded set of perturbations, including ligands, small molecule inhibitors or siRNAs that target other signaling pathways could help to refine the modules we identified here and could also lead to identification of additional functional modules and molecular networks.

Our live-cell imaging studies revealed the induction of phenotypic responses in response to ligand perturbation. OSM uniquely induced MCF10A cells to form tight cell clusters that underwent collective migration. We used our module analysis to explore the molecular basis of this complex phenotypic response and examined modules that were uniquely induced by OSM. Experimental validation identified functional links between OSM-induced molecular and phenotypic responses: protease activity was required for collective cell migration while STAT activation was required for proliferation. Our findings add to the substantial literature that implicates proteases in modulating interactions between cellular and extracellular signals [143]. Future studies that examine the role of other Module 4 features will be needed for a complete understanding of the molecular basis of OSM-induced collective migration. For example, additional complex phenotypic responses could be investigated by growing MCF10A cells as 3D organoids [133].

Together, our findings indicate that this LINCS ME perturbation dataset will serve as a robust and valuable resource for community-wide analysis and exploration. This resource can be utilized by the broader community to gain deeper insights into biological processes such as the molecular basis of different phenotypes, the molecular and phenotypic impact of particular ligands, and how specific molecular features are modulated by perturbation. Additionally, these data can serve as a resource for computational scientists to examine relationships between different molecular modalities, to develop methods for identifying molecular networks, or to elucidate the temporal relationships between different types of molecular changes. We also envision expansion of the dataset to include additional molecular measurements (*e.g.* single-cell RNAseq, single-cell ATACseq, and single-cell proteomics) and perturbation with different ligand combinations. Finally, while MCF10A represents a robust model of epithelial cell biology, analysis of the phenotypic and molecular responses observed in other cell models will be important for establishing broad generalizability of different findings. Our study provides a blueprint of the considerations for generating large-scale, high-quality multi-omic perturbation data, and serves as a reference set against which other cell types could be compared. In addition, our results could be used to help guide future studies by informing the optimal assay, perturbation or time point for more hypothesis-driven studies.

## 2.6 Additional Information

### 2.6.1 Data Availability

Data, metadata and additional analysis reports are available at: [synapse.org/LINCS\\_MCF10A](https://synapse.org/LINCS_MCF10A).

Raw RNAseq and ATACseq data generated for this study can be accessed from the Gene Expression Omnibus (GSE152410). Primary source data for **Figure 2-6G** from GSE99116. Primary source data for Figure 6h is hosted on Synapse.org with Synapse ID: syn2346643 (<https://www.synapse.org/#!/Synapse:syn2346643/wiki/232048>). Supplementary Data 23 contains metadata for the experimental samples and can be merged with each assay's level 3 data. All other data are available from the corresponding author on reasonable request.

provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses

### 2.6.2 Code Availability

Unless otherwise stated, analyses were performed in R (<https://www.R-project.org>). R packages used in analyses included: tidyverse<sup>116</sup> (version 1.3.1), ComplexHeatmap (version 2.8.0), httr (version 1.4.2) and rmarkdown (version 2.9). A complete list of packages and their versions can be found in analysis scripts available at <https://github.com/MEP-LINCS/MDD>. The DOI is <https://zenodo.org/badge/latestdoi/189112490>. Supplementary Data 24 contains a mapping of figures and tables in this paper to the scripts that created them.

### 2.6.3 Contributions

Conceptualization: L.M.H., J.W.G., A.P., and A.L. Study coordination and supervision: L.M.H. Cell culture: S.M.G., K.L.D., R.L.S., T.A.L., M.L., J.W.G., L.M.H., and J.E.K. Immunofluorescence: S.M.G., K.L.D., R.L.S., I.C.M., M.A.D., and L.M.H. Live-cell imaging: S.M.G., I.C.M., C.S.-A., M.A.D., and L.M.H. CyCIF: C.E.M., K.S., Y.W., C.J., C.Y., M.C., and P.K.S. MEMA: K.L.D., R.L.S., D.F.K., M.A.D., and J.E.K. RPPA: Y.L., M.A.D., and G.B.M. RNAseq: S.M.G., D.S.D., D.T., A.M., J.W.G., and L.M.H. ATACseq: S.M.G., D.S.D., J.L., M.A., B.W., L.M.H., and E.F. L1000: N.L., T.N., S.P., X.L., and A.S. GCP: J.M., M.P., and J.J. Integrative analyses: M.A.D., S.M.G., D.S.D., J.E.E., D.J.B.C., C.E., A.B.L., D.T., M.R.B., A.M., and L.M.H. Data curation: M.A.D., D.S.D., E.B., K.D., Z.X., D.V., L.O., and S.S. Project Management: H.S.F. Writing: S.M.G., M.A.D., D.S.D., I.C.M., J.W.G., and L.M.H. All authors reviewed and edited the manuscript.

### 2.6.4 Acknowledgements and funding sources

This work was funded by grants U54-HG008100 to JWG, LMH, and JEK; U54HL127365 to PKS; U54HG008098 and R01-GM104184 to MRB; U54HL127624 to SS and AM; U54-HG008097 to JDJ; U54HL127366 to AS. CE was a LINCS Consortium Postdoctoral Fellow. NCI Cancer Center Support Grant P50CA16672 supported generation of RPPA data. OHSU GPSR and MPSSR receive support from the OHSU Knight Cancer Institute NCI Cancer Center Support Grant P30CA069533. Cartoons created with Biorender.com. We thank Lauren Kronebusch for assistance with scientific editing.

### 2.6.5 Competing Interests

The authors declare the following competing interests: G.B.M. SAB/Consultant: Abbvie, AstraZeneca, Chrysallis Biotechnology, GSK, Ellipses Pharma, ImmunoMET, Infinity, Ionis, Lilly, Medacorp, Nanonstring, PDX Pharmaceuticals, Signalchem Lifesciences, Symphogen, Tarveda, Turbine, Zentalis Pharmaceuticals. Stock/Options/Financial: Catena Pharmaceuticals, ImmunoMet, SignalChem, Tarveda, Turbine. Licensed Technology: HRD assay to Myriad Genetics, DSP patents with Nanostring. J.W.G. has licensed technologies to Abbott Diagnostics and PDX Pharmaceuticals; has ownership positions in Convergent Genomics, Health Technology Innovations, and PDX Pharmaceuticals; serves as a paid consultant to New Leaf Ventures; has received research support from Thermo Fisher Scientific (formerly FEI), Zeiss, Miltenyi Biotech, Quantitative Imaging, Health Technology Innovations and Micron Technologies; and owns stock in Abbott Diagnostics, AbbVie, Alphabet, Amazon, AMD, Amgen, Apple, Berkshire, Cisco systems, Clorox, Colgate Palmolive, Crown Castle Int., Humana, Keysight, Linde, Proctor and Gamble, Qualcomm, Unilever, Gilead, Intel, Johnson & Johnson, Microsoft, Nvidia, Taiwan Semiconductor, and Zimmer Biomet.

### 2.6.6 Supplemental Data

Supplemental data can be accessed through Communications Biology at: <https://www.nature.com/articles/s42003-022-03975-9#Sec59>.

### 3. Chapter III: The cytokine Oncostatin M induces HIF1A dependent collective cell migration in mammary epithelial cells

The following chapter has been adapted from McLean et al (2025) (in preparation)

**Authors:** Ian C. McLean<sup>1</sup>, Sean M. Gross<sup>1,2</sup>, Mark A. Dane<sup>1</sup>, Daniel S. Derrick<sup>1</sup>, Laura M/ Heiser<sup>1,3\*</sup>

<sup>1</sup>Department of Biomedical Engineering, OHSU, Portland, OR, USA

<sup>2</sup>Capsida Biotherapeutics Inc., Thousand Oaks, CA USA <sup>3</sup>Department of Pharmacological Sciences, Mount Sinai Center for Bioinformatics, Icahn School of Medicine at Mount Sinai, New York, NY, USA

<sup>3</sup>Knight Cancer Institute, OHSU, Portland, OR USA

\*Correspondence: [heiserl@ohsu.edu](mailto:heiserl@ohsu.edu) ORCID: 0000-0003-3330-0950

#### 3.1 Abstract

The cellular microenvironment profoundly influences epithelial cell behavior, yet the mechanisms by which specific ligands orchestrate phenotypic transitions remain incompletely understood. Oncostatin M (OSM), a cytokine in the interleukin-6 family, plays pivotal roles in both normal tissue homeostasis and pathological processes, including wound healing, inflammation, and cancer progression. We previously discovered that OSM induces collective cell migration (CCM) in MCF10A cells, a coordinated movement of cell clusters that contributes to tissue repair but also facilitates cancer metastasis. Here, we investigated how OSM drives CCM and phenotypic reprogramming in mammary epithelial MCF10A cells. By comparing OSM-induced responses to those elicited by epidermal growth factor (EGF) and interferon gamma (IFNG), we characterized ligand-specific phenotypes and dissected the molecular networks underpinning OSM-induced CCM. OSM treatment uniquely induced cohesive cell clustering and CCM, distinct from the behaviors observed under EGF or IFNG conditions. Integrative transcriptomic and proteomic analyses identified HIF1A and STAT3 as central regulators of OSM-specific responses. Functional validation revealed that HIF1A drives transcriptional programs associated with hypoxia signaling, metabolic reprogramming, and immune pathways. Complement signaling was identified as a downstream effector of HIF1A, with its inhibition disrupting OSM-induced clustering and migration. These findings establish a novel mechanistic link between OSM signaling, HIF1A activation, and collective migration, offering insights into the regulation of epithelial cell behavior by the microenvironment.

#### 3.2 Introduction

Collective cell migration (CCM) is a highly coordinated process in which groups of cells migrate together while maintaining cell-cell junctions and collective polarity [276]. Unlike single-cell migration, which occurs independently, CCM depends on sustained intercellular communication and mechanical coupling between cells to maintain collective integrity and directionality [277]. This process enables cohesive cellular groups to respond to environmental cues as a unit, facilitating complex behaviors essential for tissue morphogenesis, wound healing, and immune responses.

In developmental contexts, CCM plays a pivotal role in shaping organ architecture. For example, during the formation of neuronal streams in the developing brain, chains of neuroblasts migrate collectively from the subventricular zone to the olfactory bulb [278]. Similarly, during mammary gland morphogenesis, epithelial cells undergo branching morphogenesis, a process driven by collective

migration that ensures the proper organization of ducts and alveoli [93]. In wound healing, keratinocytes at the wound margin migrate collectively, closing the wound while preserving the integrity of the epithelial sheet [279].

The molecular underpinnings of CCM involve diverse signaling networks and cytoskeletal dynamics [277]. CCM often involves the maintenance of adherens junctions and coordinated cytoskeletal remodeling [280]. Cell-cell adhesion molecules, such as E-cadherin, ensure physical cohesion among migrating cells, while integrins mediate interactions with the extracellular matrix, providing traction and directional guidance [281]. Rho family GTPases regulate cytoskeletal rearrangements, enabling cells to generate the mechanical forces required for migration [282]. Furthermore, growth factors and cytokines in the microenvironment, such as transforming growth factor-beta (TGF- $\beta$ ) and epidermal growth factor (EGF), have been shown to provide spatial and temporal cues that guide CCM [283], [284]. These signaling molecules activate downstream pathways, including MAPK, PI3K-AKT, and JAK-STAT, to modulate cellular polarity, adhesion, and motility [82], [285].

Beyond its roles in normal physiology, CCM has been implicated in cancer progression, particularly in metastasis. In breast cancer, tumor cells can adopt a collective migratory mode to invade surrounding tissues and disseminate to distant sites [286]. Unlike single-cell migration, which requires a complete loss of cell-cell junctions, collectively migrating tumor cells retain partial junctional integrity, facilitating coordinated movement through the stroma [287], [288]. Clusters of cancer cells exhibit improved survival during intravasation and extravasation into and out of blood vessels, potentially due to the retention of intercellular signaling and adhesion molecules that protect against anoikis [96]. Additionally, these clusters may carry supportive stromal or immune cells that facilitate colonization at distant sites [97], [98], [99]. Molecularly, collective migration in cancer is associated with partial epithelial-to-mesenchymal transition (EMT), characterized by the partial downregulation of epithelial markers (e.g., E-cadherin) and the upregulation of mesenchymal markers (e.g., vimentin) [118]. Environmental cues such as the cytokine interleukin-6 (IL-6) and hypoxia further promote collective invasion by enhancing tumor cell motility and stromal remodeling [289], [290].

CCM is modulated by multiple extracellular signals, including hypoxia, SDF-1, PDGF, and Oncostatin M (OSM) [147], [291], [292], [293]. OSM is an interleukin-6 family cytokine that is secreted by various cell types, including macrophages, neutrophils, and T cells [54], [294]. OSM signals through a heterodimeric receptor composed of OSMR $\beta$  and gp130, activating downstream pathways such as JAK-STAT, MAPK, and PI3K-AKT [74]. In normal physiology, OSM regulates diverse processes, including hematopoiesis, inflammation, and tissue remodeling [294]. For instance, OSM has been shown to modulate mammary gland development by influencing involution and epithelial differentiation [63].

In cancer, OSM plays a dual role, promoting tumor progression in some contexts while exhibiting anti-tumor effects in others [295]. Alterations in the OSM signaling pathway have been observed in various cancers, including breast, lung, and ovarian cancers, and are associated with increased tumor cell proliferation, invasion, and immune evasion [296], [297], [298]. Notably, OSM induces striking phenotypic changes in epithelial cells, including the alteration of junctional stability and enhanced motility, suggesting a potential role in CCM [147]. The JAK-STAT signaling pathway, activated by OSM, is a key mediator of these effects, driving transcriptional programs that regulate cell proliferation, survival, and migration [299].

In this study, we investigate the role of OSM in inducing CCM in mammary epithelial cells. To model this phenomenon, we used MCF10A cells, a non-tumorigenic human mammary epithelial cell line that recapitulates many features of normal mammary gland biology, including exhibiting typical epithelial cell morphology, polarity, and genetic stability [132], [134]. MCF10A cells are widely used as a model system to study epithelial cell behavior due to their ability to form organized structures in 3D culture and respond dynamically to various ligands and environmental cues [134].

Our study integrates live-cell imaging, transcriptomics, and functional assays to uncover the molecular mechanisms underlying OSM-induced CCM. We show that OSM activates a distinct transcriptional program in MCF10A cells, characterized by the upregulation of pathways associated with hypoxia and immune signaling. Among the key regulators identified, hypoxia-inducible factor 1-alpha (HIF1A) emerged as a central mediator of OSM-induced transcriptional changes. Functional validation revealed that HIF1A plays a critical role in driving the clustering and coordinated migration of MCF10A cells in response to OSM. These findings provide insight into the intersection of HIF1A signaling, immune-related pathways, and CCM, and delineate a novel molecular mechanism underlying cell migration.

### 3.3 Methods

#### MCF10A Cell Culture

Cell culture and ligand perturbation experiments were carried out according to the methods described by Gross et al. 2022 [147]. Briefly, for routine cell growth and passaging, MCF10A cells were maintained in a growth medium consisting of DMEM/F12 (Invitrogen #11330-032), 5% horse serum (Sigma #H1138), 20 ng/ml EGF (R&D Systems #236-EG), 0.5 µg/ml hydrocortisone (Sigma #H-4001), 100 ng/ml cholera toxin (Sigma #C8052), 10 µg/ml insulin (Sigma #I9278), and 1% Pen/Strep (Invitrogen #15070-063). For perturbation studies, a growth factor-free medium was used, which included DMEM/F12, 5% horse serum, 0.5 µg/ml hydrocortisone, 100 ng/ml cholera toxin, and 1% Pen/Strep.

MCF10A cells were cultured until they reached 50-80% confluence in growth medium, followed by detachment using 0.05% trypsin-EDTA (Thermo Fisher Scientific #25300-054). After detachment, 6,000 cells were seeded into collagen-1 (Cultrex #3442-050-01) coated 24-well plates (Thermo Fisher Scientific #267062) with growth medium. Six hours later, the cells were washed with PBS, and growth factor-free medium was added. The cells were then incubated for 18 hours in the new medium. Afterward, cells were treated with either 10 ng/ml EGF (R&D Systems #236-EG), 10 ng/ml OSM (R&D Systems #8475-OM), or 20 ng/ml IFNG (R&D Systems #258-IF) + 10 ng/mL EGF.

#### Live Cell Imaging

Live-cell imaging was conducted using the Incucyte S3 microscope (Essen BioScience, #4647), with images captured every 30 minutes for a duration of up to 48 hours. The resulting live-cell image stacks were processed by first registering them with a custom Fiji script and then segmented using CellPose v3.01 [152], [224]. Image tracking was performed using the Baxter Algorithms pipeline [151].

Subsequent analysis of the cell tracking data was conducted within RStudio [300]. Cell counts were calculated by determining the number of cells per field, normalized by the initial T0 count for each field. Nearest neighbor distances were computed by measuring the Euclidean distances in pixels from the centroid of each cell to the centroid of the second nearest cell within the imaging field. To adjust for variations in cell count, the average nearest neighbor distances for each image were normalized by the

expected mean distance to the nearest neighboring cell under a random cell distribution model [226]. Cytoplasmic size was determined as the average size of the cytoplasm 24 hours post-ligand addition. Cell motility was assessed by excluding tracks with jumps greater than 200 pixels in a 30-minute interval. Motility was quantified as the slope of the mean squared displacement (MSD) over time intervals ranging from 30 minutes to 6 hours [301]. This slope was determined by constructing a linear model comparing MSD to time intervals, and it serves as a proxy for the diffusion coefficient associated with Brownian motion [301]. Finally, CCM was estimated by calculating the Cosine Similarity between displacement vectors of the 10 nearest cells every 30 minutes. Statistical comparisons of phenotypic scores between ligand conditions was performed using Dunnett's test, with EGF serving as the control [302].

### **Immunofluorescent Imaging**

For all immunofluorescent experiments, cells were fixed in 4% formaldehyde after 48 hours of ligand treatment. Cells were blocked for 1 hour in a PBS solution (5% Normal Goat Serum (Cell Signaling #5425), 0.3% Triton X (Thermo Fisher Scientific #85111)), then incubated overnight with the primary conjugated antibody (Cell Signaling #2677) at a 1:500 dilution. Cells were washed thorough then counterstained with 0.5  $\mu\text{g} / \mu\text{L}$  DAPI (PromoKine PD-CA707-40043). Imaging was performed using the InCell 6000 (GE Healthcare).

### **RNAseq, RPPA Data, and Cyclic IF**

The RNA sequencing (RNAseq), reverse phase protein array (RPPA) data, and cyclic immunofluorescence (cyclicIF) data were generated in a prior investigation [147]. For detailed information on the methodologies used for data generation, normalization, and integration, please refer to the original publication.

### **Causal Pathway Analysis and Network Analysis**

Causal Pathway was utilized to construct gene and protein interaction networks [176]. The algorithm received integrated features from RNAseq, RPPA, and cyclic IF datasets, focusing on those from the OSM and IFNG conditions. Data from proteomic assays were favored for features shared between assays. Only features with a log-fold change (LFC) of 1 or greater compared to the time zero control were included in the analysis. Default parameters were applied for the causal pathway analysis. The generated networks were subsequently integrated and analyzed in Cytoscape [303].

To identify gene/protein nodes for experimental testing, a comparative analysis was performed between the IFNG and OSM networks. Each node was assigned a score based on network rewiring, which compared the OSM network to the IFNG network using Dynet [304]. This score was then scaled according to the LFC in the OSM condition. The fourteen nodes with the highest comparative network importance scores were selected for experimental validation.

### **siRNA Screen and Drug Experiments**

Reverse transfection of MCF10A cells was performed according to DharmaFECT Transfection vendor protocols. Cells were seeded in collagen coated 24-well plates with growth media as previously described (see methods: MCF10A cell culture). After 6 hours in culture, media was replaced with 25 nM siRNA and 0.5  $\mu\text{L}$  Lipofectamine™ RNAiMAX Transfection Reagent (Thermo Fisher Scientific #13778075) per well in antibiotic free media. After 48 hours of transfection, cells were washed thoroughly with PBS

and replaced with assay media and indicated ligand treatments. The siRNA screen was run in triplicate. A full list of siRNAs used in the study is included (**Supplemental Table 1**). Immediately after ligand treatment, live-cell imaging and phenotypic quantification was performed as described previously (see methods: Live Cell Imaging). Statistical comparisons between phenotypic scores were performed by pairwise student's t-tests between knockdown conditions and the siSCR condition for each ligand treatment [305].

Cells were cultured and plated for drug experiments as previously described (see methods: MCF10A cell culture). Concurrent to ligand treatment, cells were incubated with either vehicle, or 25  $\mu$ Mol - 75  $\mu$ Mol Compstatin (Selleckchem #S8522).

### shRNA Integration

Plasmic constructs for non-targeting SMARTvector Lentiviral shRNA and shRNA sequences targeting HIF1A were purchased from Horizon Discovery (Horizon Discovery # VSC11287, #3SH11243-02EG3091). Expression of the shRNA was driven by the EF1A promoter and contained Puromycin resistance and TurboGFP elements to ensure proper integration. *E. Coli* containing the plasmid constructs were inoculated into LB broth medium containing 100  $\mu$ L/mL carbenicillin and incubated at 37 C for 18 hours with shaking. Plasmids were extracted using the HiSpeed Plasmid Midi Kit (Qiagen #12643).

Plasmids were then packaged into lentiviral particles by transfecting 293T human embryonic kidney cells (ATCC # CRL-3216) with the shRNA plasmid construct and lentiviral package plasmids (Addgene #8454, #12263) using LF2000 transfection reagent (Invitrogen #11668019). Virus containing media was collected through a 0.45  $\mu$ m low-protein binding filter (Sterlitech #PES4547100). MCF10A cells were then transduced with virus containing media and 10  $\mu$ g / mL polybrene. After a four-hour transduction, media was replaced with growth media, and shRNA containing cells were selected with 0.5  $\mu$ g/mL puromycin. Expression of shRNA was confirmed using fluorescent imaging of GFP on the Incucyte S3 microscope (Essen BioScience, #4647).

### Immunoblotting

HIF1A was detected in cells transduced with shRNA through immunoblotting. After 24 hours of treatment with OSM EGF, or CoCl<sub>2</sub> cells were lysed in hot Laemmli buffer (Santa Cruz Biotechnology #sc-286962). Cell lysates were boiled at 95C for 10 minutes, then total protein concentration was determined and normalized through BCA assay (Thermo Fisher Scientific #23225). 25  $\mu$ L of cell lysate and HIF1A lysate control (Novus Biologicals #NBP2-04440) was loaded onto gels (Invitrogen #NP0321BOX) and proteins were separated at 200V for 45 min then transferred at 30V overnight at 4C. After transfer the membrane was blotted with TBST and 5% dry milk for 1 hr., then probed for HIF1A at a primary antibody concentration of 1:1000 (BD Biosciences #610959) and secondary antibody 1:10000 (Jackson Immunoresearch #715-035-150). Chemiluminescence was imaged on the Alpha Innotech FluorChem Imaging System (#22424).

### scRNA-seq Library Preparation and Sequencing

All cell lines and ligand conditions were multiplexed using Hashtag Oligonucleotide barcoding (TotalSeq-B, Biolegend #399904) according to the vendor's recommendations. The mRNA library was created using

the Chromium Single Cell 3' GEM, Library & Gel Bead Kit v3 (10X Genomics #1000092) and then sequenced using an Illumina NovaSeq for 800 million reads.

### scRNA-seq Data Processing and Analysis

Raw files were converted to FASTQ format with bcl2fastq (version 2.20.0.445). Cellranger count was used to align reads to the GRCh38 transcriptome (GCF\_000001405.40). The R package deMULTIplex was used to demultiplex the hash-tagged samples and assign cell lines and treatments to cells [306]. Seurat (4.0.5) was used to perform variable feature identification, dimensionality reduction, unsupervised clustering, visualization, and differential gene expression [307]. Differential expression analysis was performed using the FindMarkers function of Seurat with default parameters. GSEA was performed with the R package clusterProfiler on Gene Ontology Biological Process gene sets using significantly upregulated genes compared to the shSCR control (LFC > 0.5, p-value < 0.05) [308].

## 3.4 Results

### 3.4.1 OSM treatment initiates MCF10A epithelial collective cell migration

In this study, we examined the phenotypic effects and molecular mechanisms activated by Oncostatin M (OSM) treatment on MCF10A cells and compared them to those elicited by epidermal growth factor (EGF) and interferon gamma (IFNG). Prior studies demonstrated that OSM treatment induces striking changes in cell morphology and behavior, distinct from those observed under typical EGF conditions or with IFNG, another activator of JAK-STAT signaling [249].

MCF10A cells were treated with OSM, EGF, or EGF + IFNG for 48 hours, and their phenotypic responses were assessed through live-cell imaging and immunofluorescent microscopy. As expected, EGF-treated cells displayed the characteristic cobblestone pattern typical of MCF10A epithelial sheets (**Figure 3-1A**) [133]. In contrast, OSM-treated cells exhibited a distinct phenotypic shift, forming tight clusters that were significantly more compact than those observed in the EGF condition (**Figure 3-1A**). Notably, these clusters retained robust cell-cell junctions while displaying collective migratory behavior. Time-lapse imaging revealed that OSM-induced migration occurred as cohesive clusters, a hallmark of collective cell migration (CCM) (**Figure 3-1B**). This phenotype was absent in both EGF and IFNG conditions.

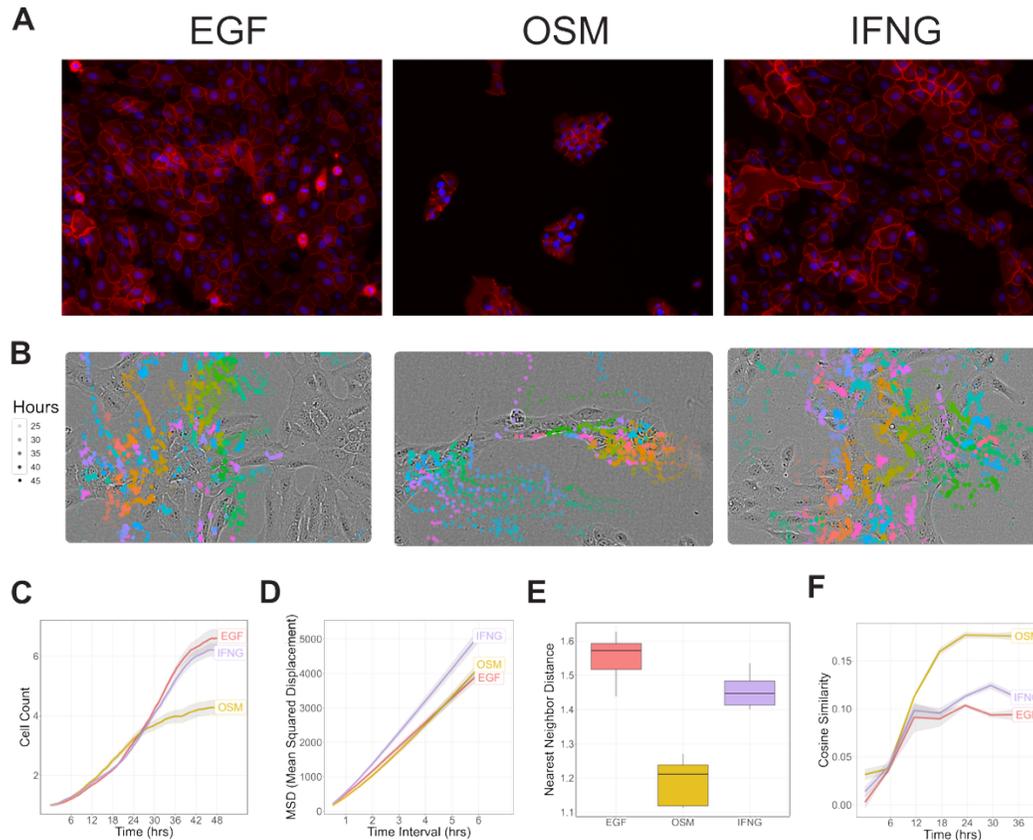
The live-cell imaging data was quantified to compare cell phenotype across conditions: Cell Count to estimate proliferation, Mean Squared Displacement to estimate motility, Nearest Neighbor Distance to assess cell clustering, and Cosine Similarity of displacement vectors to assess collective migration (**Figure 3-1C-F**). Statistical comparisons of OSM and IFNG treatments to EGF using Dunnett's test ( $p < 0.05$ ) demonstrated that neither ligand significantly influenced proliferation compared to EGF (**Figure 3-1C**, **Supplemental Figure 3-1**) [302]. It is noteworthy that OSM-treated cells exhibited comparable proliferation rates to those observed under EGF treatment, highlighting its ability to drive mitogenic effects through non-RTK mechanisms [147].

IFNG-treated cells did not form clusters or exhibit an increased CCM score, despite activating the JAK-STAT pathway, as confirmed in prior studies [147]. Quantitatively, IFNG induced greater motility, as evidenced by a higher mean squared displacement compared to EGF-treated cells (**Figure 3-1D**).

IFNG slightly reduced Nearest Neighbor Distance, suggesting mild clustering, while OSM significantly decreased this metric, indicating strong clustering (**Figure 3-1E**, **Supplemental Figure 3-1**). Most notably, Cosine Similarity increased significantly in the OSM condition, indicating collective cell migration, but not

## HIF1A dependent collective cell migration in mammary epithelial cells

in the IFNG condition (**Figure 3-1F, Supplemental Figure 3-1**). The distinction between IFNG-induced motility and OSM-induced CCM demonstrate the differential downstream effects of JAK-STAT activation by these two ligands: IFNG drives a more dispersed, single-cell migratory phenotype, whereas OSM promotes tightly clustered collective cell migration. This phenotypic contrast motivated us to further explore the potential of OSM as a tool for dissecting the molecular mechanisms underlying CCM in epithelial cells.

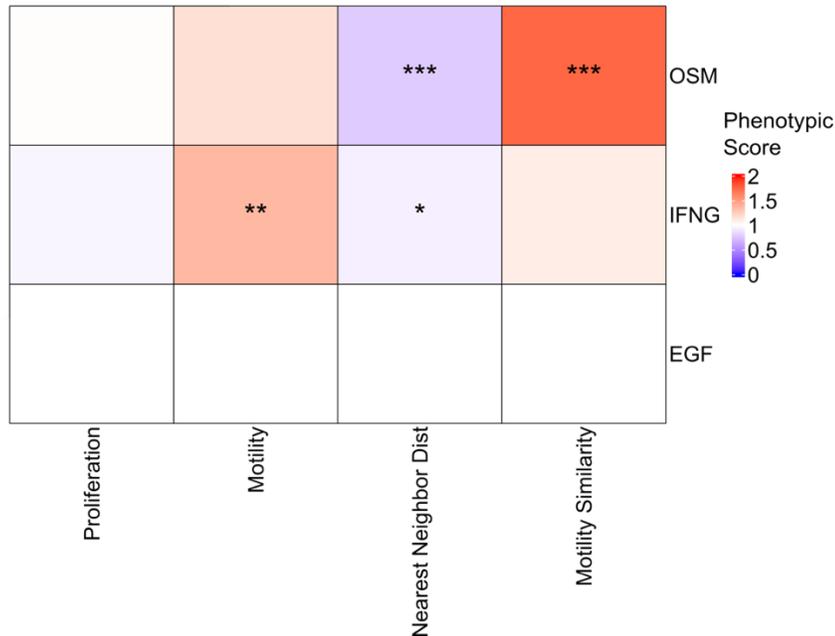


**Figure 3-1: Treatment of MCF10A cells with OSM but not IFNG initiates cell clustering and collective cell migration**

**A)** MCF10A cells were treated with EGF, OSM, or IFNG for 48 hours and stained for DAPI (blue) and B-Catenin (red).

**B)** Representative migratory tracks derived from cell tracking data during 24-48 hours of ligand treatment. Individual tracked cells are represented by different colors. The opacity of each circle represents increase in experimental duration.

**C-F)** Quantification of cell phenotype from live-cell imaging.



**Supplemental Figure 3-1: Statistical comparison of phenotypic scores**

Statistical comparison of the quantifications of cellular proliferation, motility, cell clustering / spreading, and collective migration. Phenotypic scores have been normalized to the EGF condition. Dunnett's test was employed to compare phenotypic scores to the EGF control. Significance is shown as follows: (\* - p-value < .05, \*\* - p-value < .01, \*\*\* - p-value < .001).

### 3.4.2 Comparative network analysis identifies OSM-specific molecular regulators

To elucidate the molecular mechanisms driving OSM-induced collective cell migration (CCM), we analyzed publicly available data from the Library of Integrated Network-based Cellular Signatures (LINCS) [147]. Bulk RNA sequencing (RNA-seq) and two proteomic assays—cyclic immunofluorescence (cyclic IF) and reverse-phase protein array (RPPA)—were performed on MCF10A cells treated with either OSM or IFNG for 48 hours [227], [309]. Our approach was specifically designed to identify molecular signatures unique to OSM-induced phenotypes, particularly in contrast to the IFNG condition, which also activates the JAK-STAT pathway but does not elicit similar phenotypic responses. By examining the differences between the two conditions, we aimed to tease apart the molecular mechanisms underlying OSM-induced CCM, offering new insights into its distinct regulatory landscape.

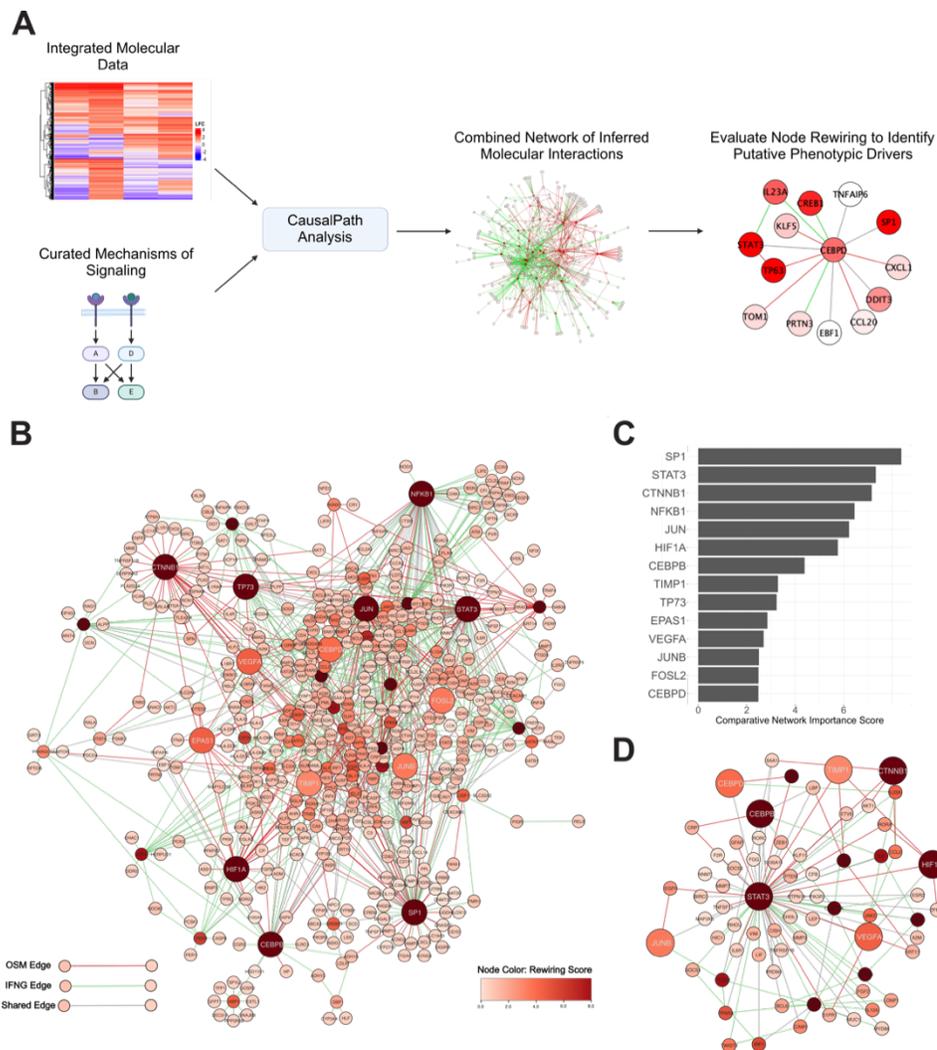
To achieve a systems-level understanding of CCM, transcriptional and proteomic datasets were integrated for each condition (**Figure 3-2A, Supplemental Figure 3-2**). CausalPath analysis was applied to infer protein-protein interaction networks supported by the omics data, and the resulting networks were visualized using Cytoscape (**Figure 3-2B**) [176], [303]. We reasoned that molecular nodes that exhibited the highest degree of deviation between the OSM and IFNG conditions could represent potential key drivers of the OSM-induced CCM phenotype. To identify these nodes, a rewiring metric was calculated for each node to quantify the number of edges that must be added or removed to transition a node's connectivity from one condition to the other [304]. This metric was weighted by the log-fold change (LFC) induced by OSM for each node, prioritizing nodes that were both highly rewired and upregulated under OSM treatment. By comparing the networks derived from OSM and IFNG conditions, we sought to

identify regulators uniquely associated with OSM-induced CCM and distinguish them from general JAK-STAT pathway components.

STAT3 ranked second on the ranking of nodes with the highest combined rewiring and OSM-specific upregulation scores, despite its activation under both OSM and IFNG conditions (**Figure 3-2C**). This highlights the value of a network-based approach: while STAT3 activation occurs under both treatments, the effect of its upregulation differs significantly. This approach helps pinpoint downstream effectors uniquely modulated under OSM-induced conditions, offering insights into the distinct phenotypic divergence observed, where OSM induced clustering and CCM while IFNG did not, despite both activating STAT3 (**Figure 3-1B**). Eight of the fourteen top-rewired nodes were centered around the STAT3 subnetwork, consistent with the pivotal role of STAT3 in cytokine signaling and cellular phenotypic transitions (**Figure 3-3D**).

This analysis also identified several general transcriptional regulators, including SP1, JUN, JUNB, FOSL2, and CTNNB1. These nodes are well-established mediators of cytokine signaling and have been implicated in diverse cellular processes such as proliferation, differentiation, and migration [310], [311], [312], [313]. TP73, another highly rewired node, has been linked to epithelial cell plasticity and tumor suppression [314], [315]. We also identified CEBPB and CEBPD, two transcription factors known to regulate epithelial lineage commitment [316], [317].

Nodes associated with hypoxia signaling, such as HIF1A, VEGFA, TIMP1, and EPAS1, were highly rewired. HIF1A, a master regulator of hypoxia, is particularly notable as prior research has demonstrated its activation by OSM in normal and cancerous cellular contexts [318], [319]. Hypoxia-related pathways are well-documented to contribute to modulating migratory phenotypes, including collective cell migration [118], [120], [290]. The presence of hypoxia-associated nodes in the OSM network suggests a potential link between OSM signaling and partial epithelial-to-mesenchymal transition (EMT), a process often implicated in enhanced migratory and invasive capacities of epithelial cells. Partial EMT is characterized by retained cell-cell junctions alongside increased motility, aligning with the observed CCM phenotype in OSM-treated MCF10A cells [118]. These findings provide a potential mechanistic link between OSM signaling and the hypoxia-associated promotion of migratory phenotypes.



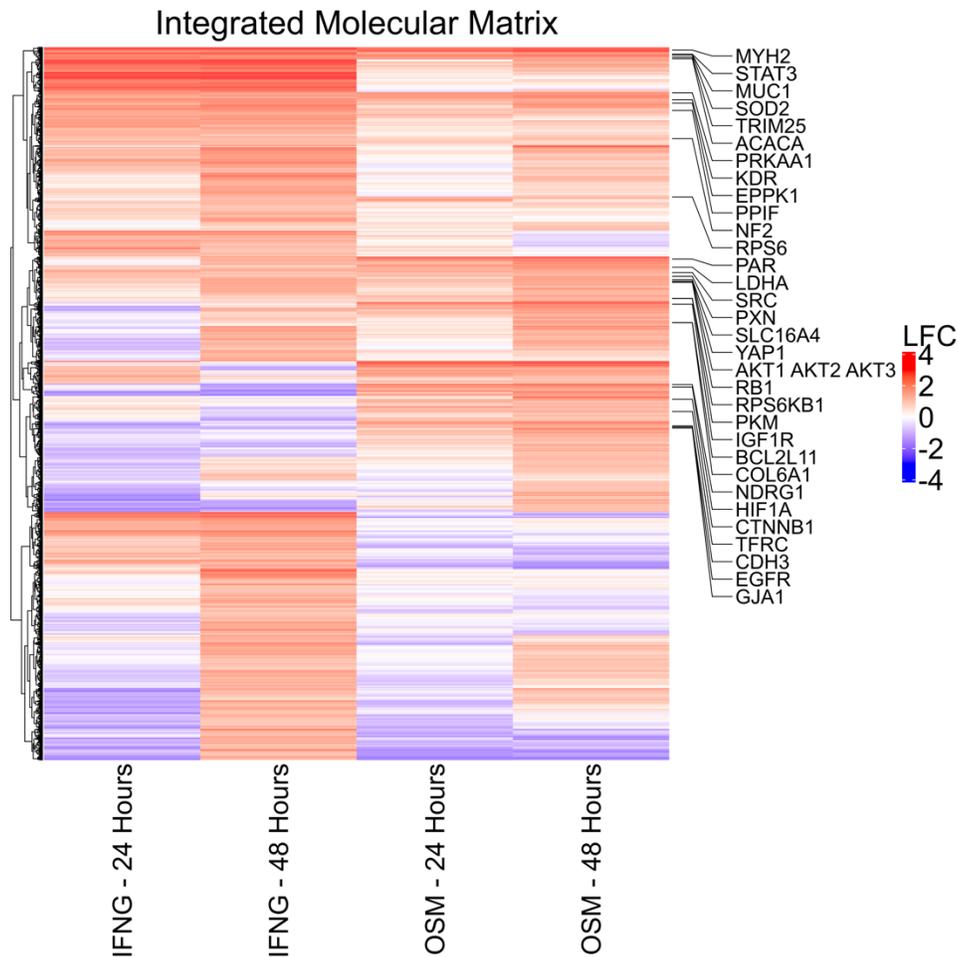
**Figure 3-2: Comparative network analysis of omics data reveals putative drivers of OSM-induced CCM**

**A)** Workflow for the comparative network analysis to identify molecular subnetworks and nodes perturbed by OSM. Integrated molecular data collected from OSM and IFNG treated cells was analyzed using CausalPath.

**B)** Combined molecular network of OSM and IFNG activated nodes. Each node in the combined network was evaluated for rewiring between conditions.

**C)** The top scoring 14 nodes in the network rewired by OSM treatment.

**D)** The majority of top rewired nodes are centered around the STAT3 subnetwork.



**Supplemental Figure 3-2: Integrated and filtered molecular data collected from OSM and IFNG treated cells**

Proteomic (RPPA, CycIF) and transcriptional (RNAseq) data was collected from MCF10A cells treated with OSM or IFNG for 24 and 48 hours. The data from each assay was normalized to the Time 0 control, filtered for differentially expressed features (LFC > 1, p-value < .05) and then integrated into a combined dataset. Features from the RPPA assay are called out in the heatmap.

### 3.4.3 Functional Validation Reveals OSM-Specific Phenotypic Regulators

We hypothesized that knockdowns yielding OSM-specific phenotypic effects would help identify nodes critical for mediating OSM-induced CCM, distinguishing them from general MCF10A cellular responses. To test this, we validated the functional and phenotypic significance of the rewired nodes identified in our network analysis using siRNA knockdown. MCF10A cells were subjected to siRNA-mediated knockdown of the 14 nominated nodes under three experimental conditions: OSM, EGF, and IFNG treatments. The inclusion of EGF and IFNG conditions allowed us to evaluate the specificity of phenotypic changes associated with OSM-induced responses. The siRNA screen was conducted in triplicate to ensure robust and reproducible findings. Dynamic phenotypic changes were monitored through live-cell imaging over 48 hours. We calculated the same phenotypic metrics described in Section 1. These metrics were normalized to the scramble condition for each ligand treatment, enabling assessment of how each knockdown altered the phenotype under OSM, EGF, and IFNG conditions

(**Figure 3-3A-D**). To evaluate the statistical significance of these changes, we applied Student's t-test to compare raw phenotypic metrics from knockdown conditions to the scramble control for each ligand treatment [305]. Additionally, fixed-cell immunofluorescent (IF) imaging was employed to capture detailed phenotypic alterations (**Figure 3-3E**).

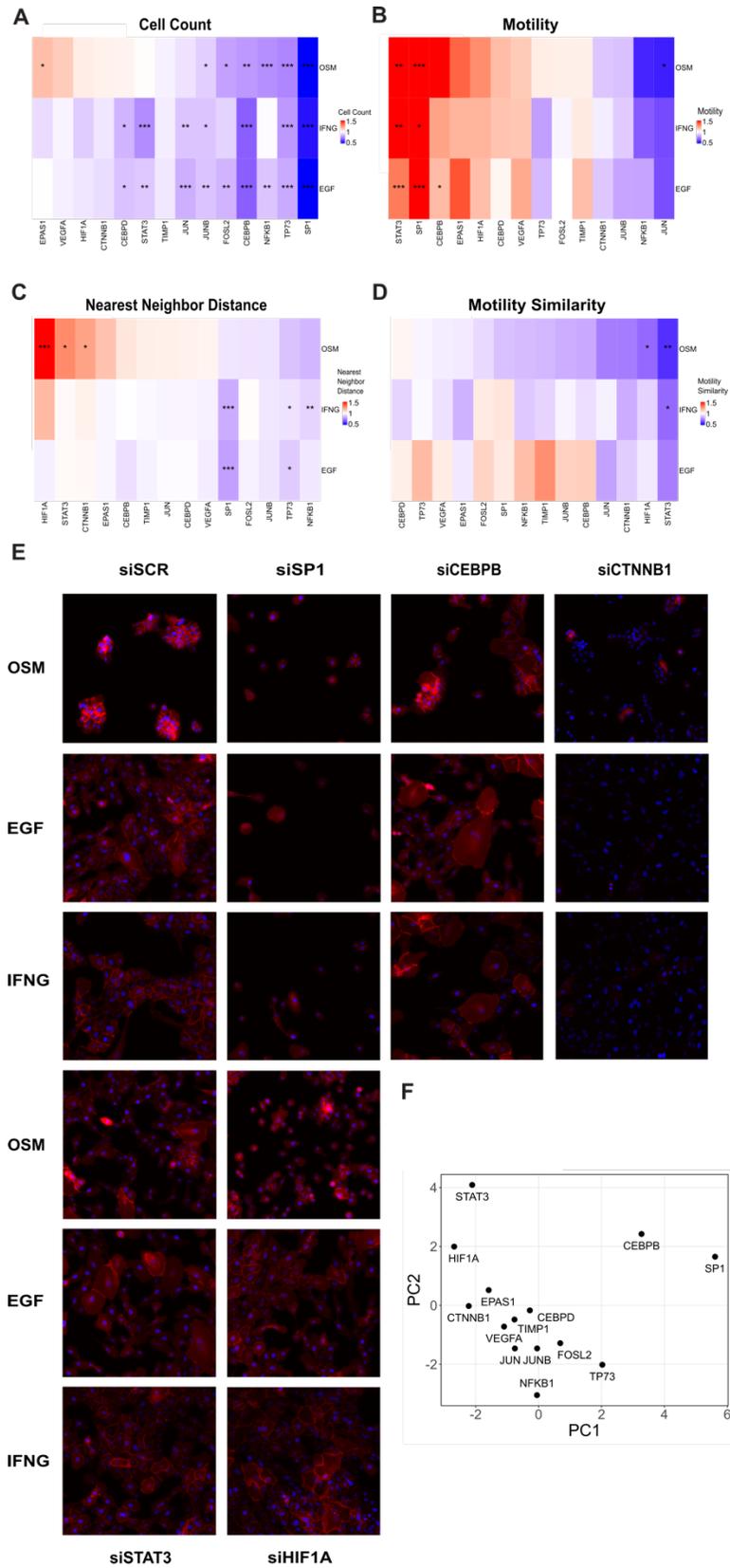
The siRNA experiment revealed a diverse array of phenotypic effects, including both condition-specific changes and those shared across conditions. This diversity reflects the broad biological roles of the nodes targeted by the knockdowns. Shared effects across all conditions reflect MCF10A intrinsic molecular functions and provided validation of knockdown. For example, knockdown of SP1 and JUNB inhibited cell count across all treatments, consistent with their known roles in promoting cell proliferation [320], [321] (**Figure 3-3A, 3-3E**). Interestingly, STAT3 knockdown increased cell motility in all conditions, an unexpected finding that contrasts with reports in the literature that primarily suggest that STAT3 activation can enhance migratory and invasive cell behavior [322] (**Figure 3-3B, 3-3E**). Knockdown of CEBPB decreased proliferation across all conditions and disrupted epithelial cell morphology, a finding that is particularly intriguing given CEBPB's role as a regulator of epithelial differentiation [317] (**Figure 3-3A, 3-3E**). These effects further emphasize its potential role in maintaining epithelial integrity.

When focusing on OSM-specific regulators, several nodes emerged as unique phenotypic mediators. Knockdown of EPAS1 (HIF2A) uniquely increased proliferation in OSM-treated cells (**Figure 3-3A**). This observation contrasts with prior reports suggesting that HIF2A promotes cell cycle progression and proliferation in hypoxic environments and may indicate a context dependent relationship between HIF2A and cellular proliferation [323]. JUN knockdown specifically decreased motility under OSM treatment, suggesting a role in modulating OSM-driven migratory behavior (**Figure 3-3B**).

STAT3, HIF1A, and CTNNB1 knockdowns decreased nearest-neighbor distances and disrupted clustering under OSM conditions (**Figure 3-3C**). CTNNB1, a core junctional protein, plays a well-established role in maintaining cell-cell adhesion, making this finding particularly relevant to the observed clustering phenotype [324] (**Figure 3-3E**). The regulation of clustering and motility similarity by STAT3 was expected, given its central role in OSM signaling, and confirms STAT3 as a key driver of OSM-induced cell clustering and CCM (**Figure 3-3D**). Interestingly, HIF1A knockdown specifically reduced cosine similarity and disrupted cluster formation under OSM treatment, suggesting a critical role for HIF1A signaling in OSM-induced CCM. HIF1A has been linked to migratory and invasive phenotypes, particularly through its regulation of genes involved in cytoskeletal dynamics and extracellular matrix remodeling [325], [326]. The co-regulation of clustering and motility similarity by HIF1A knock-down highlights its potential as a key regulator of OSM-induced CCM.

We performed principal component analysis (PCA) of phenotypic scores across all treatments to identify siRNA perturbations that most strongly influenced changes in OSM-induced cellular phenotype (**Figure 3-3F**). The PCA plot revealed distinct clustering of knockdowns based on their effects, with nodes such as SP1 and CEBPB separating from others due to their pronounced impact on cell count and morphology disruption. Notably, STAT3 and HIF1A formed a separate group, highlighting their roles as OSM-specific regulators. Supplementary scatterplots comparing phenotypic scores further supported these findings, identifying nodes with unique effects on OSM-induced phenotypes (**Supplemental Figure 3-3A-B**).

HIF1A dependent collective cell migration in mammary epithelial cells

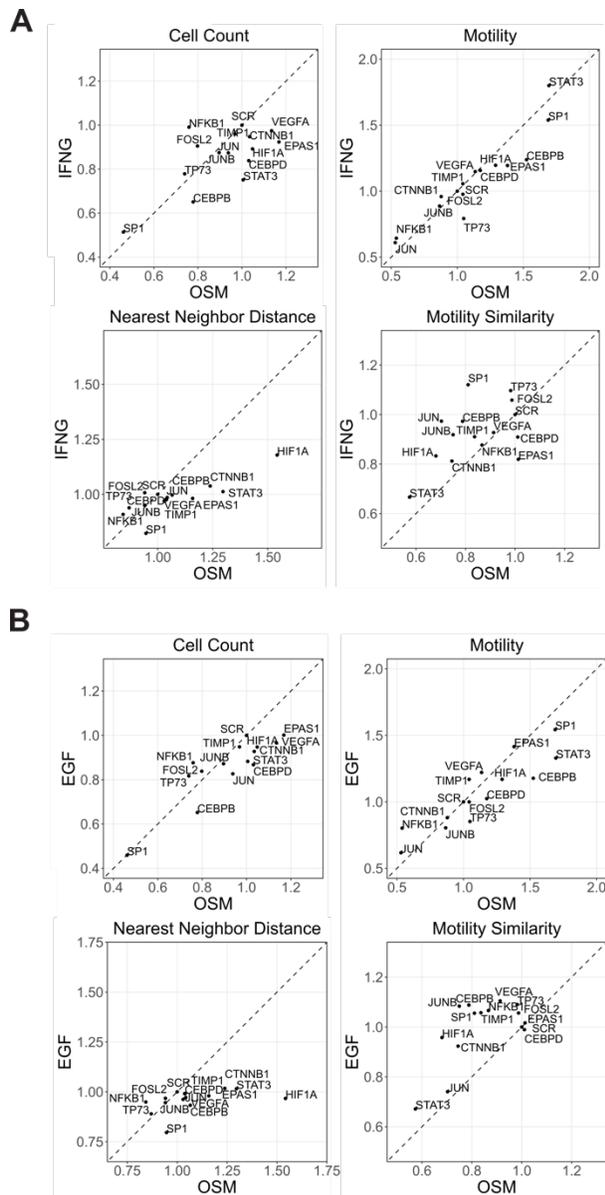


**Figure 3-3: Quantification and visualization of siRNA screen to identify drivers of OSM-induced CCM**

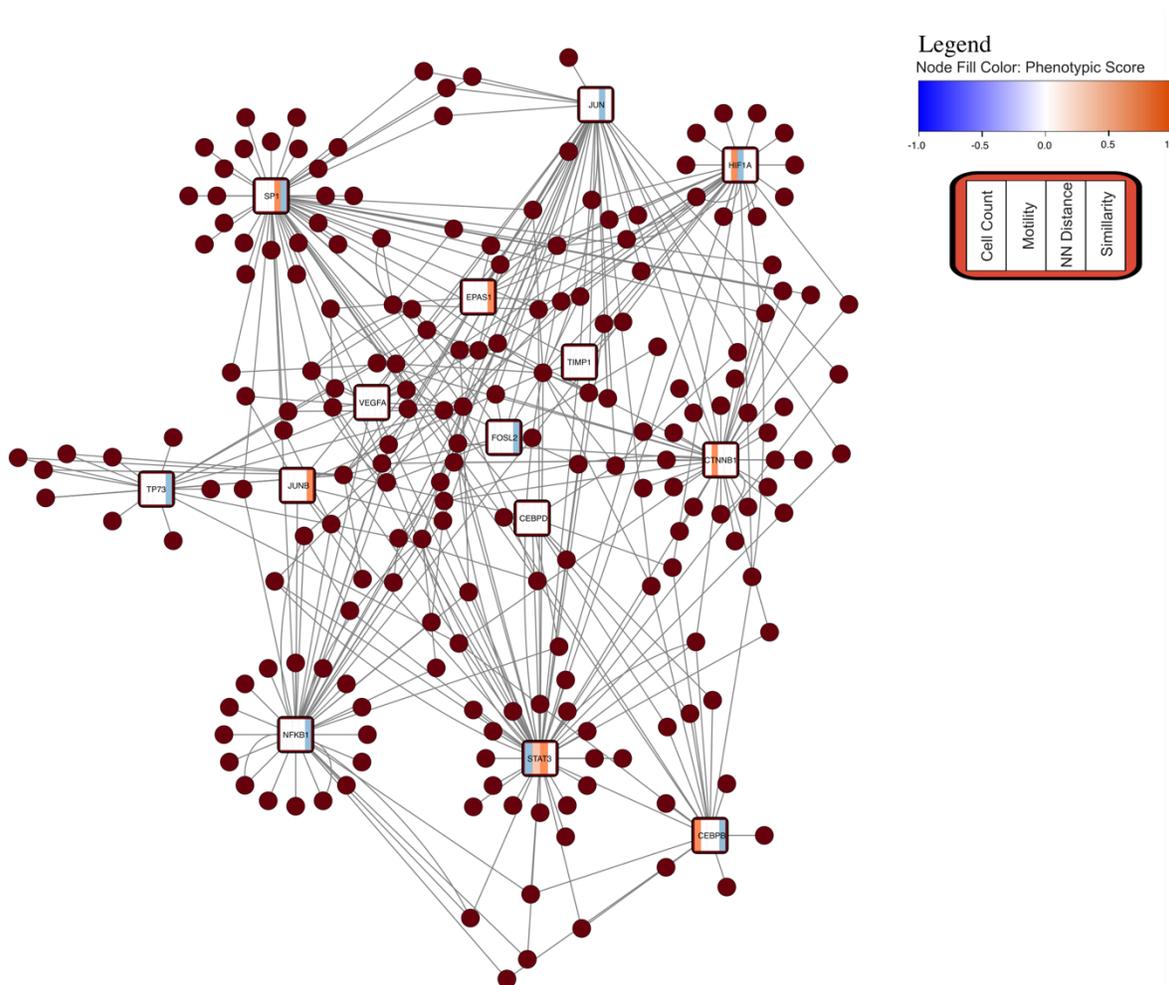
**A-D)** Quantification of cell phenotype induced by knockdown of nominated nodes. Phenotypic scores have been normalized to the siSCR condition for each ligand treatment. Statistical comparisons were performed using Students T-Test. Significance is shown as follows: (\* - p-value < .05, \*\* - p-value < .01, \*\*\* - p-value < .001).

**E)** Representative immunofluorescent images from select knockdowns taken 48 hours after ligand and siRNA treatment.

**F)** Principal component analysis of phenotypic metrics across all ligand treatments.

**Supplemental Figure 3-3: Integrated and filtered molecular data collected from OSM and IFNG treated cells**

**A-B)** Comparison of the phenotypic effects of knockdown on OSM and IFNG (A) or EGF (B) treated cells.



Supplemental Figure 3-4: *OSM molecular network of phenotypic perturbations*

Integrated molecular network of OSM treated cells. Nodes nominated for siRNA screening are labelled by the phenotypic effect of knockdown. Non-significant changes in phenotype are colored white.

### 3.4.4 HIF1A as a Key Mediator of OSM-Induced CCM

Given its known role in partial epithelial-to-mesenchymal transition (EMT) and collective cell migration (CCM), HIF1A was prioritized for further investigation [118]. Analysis of RPPA data and RNA-seq data revealed a sustained upregulation of HIF1A protein and gene expression during the 48 hours of OSM treatment (**Figure 3-4A, Supplemental Figure 3-5A**). This upregulation was unique to the OSM condition compared to EGF and IFNG treatments. These findings concur with previous literature demonstrating that OSM upregulates HIF1A in a transcriptional manner, independent of oxygen concentration [318]. While previous studies have linked HIF1A to CCM, the molecular mechanisms by which it induces CCM in epithelial cells remain unclear, presenting an opportunity to explore how HIF1A orchestrates these complex behaviors.

HIF1A is a transcription factor, so we hypothesized that its downstream transcriptional effectors may be critical for mediating CCM. To identify transcriptional programs that HIF1A is activating in OSM-treated cells, we engineered shRNA MCF10A cells with HIF1A knocked down. Knockdown was confirmed by

Western blot (**Supplemental Figure 3-5B**). We then subjected shHIF1A and shSCR cells treated with OSM, EGF, or IFNG to scRNA-seq. UMAP projection of the resulting data showed that ligand treatment predominately separated the conditions (**Figure 3-4B**). Assessment of HIF1A transcript also confirmed that HIF1A was effectively knocked down in shHIF1A cells (**Figure 3-4C**).

To investigate the mechanisms underlying OSM-induced CCM, we sought to compare the transcriptional signatures induced by OSM activation of HIF1A to those elicited by other ligands, hypothesizing that OSM-specific changes would identify candidate drivers of CCM. Differential gene expression analysis comparing shSCR and shHIF1A lines for each treatment revealed significant overlap (27 genes) between EGF and IFNG HIF1A-regulated genes, which may represent constitutive low-level transcriptional activity of HIF1A (**Figure 3-4D**). In contrast, the majority of HIF1A-regulated genes in the OSM condition (79 genes) were unique to OSM, indicating a distinct molecular program activated by HIF1A under these conditions. Because CCM is exclusive to the OSM condition, we focused our subsequent analysis on these unique OSM-HIF1A-regulated genes. Comparison of these unique genes to literature-derived Hallmarks and ChIP-Seq-derived (ChEA3) gene sets of hypoxia or HIF1A transcriptional targets revealed overlap with canonical hypoxia-related genes, such as NDRG1 and CA9 [229], [327], [328], [329] (**Figure 3-4E, Supplemental Figure 3-5C-D**). However, OSM-HIF1A also uniquely upregulated additional genes not typically associated with hypoxia or canonical HIF1A transcriptional regulation, suggesting the activation of a specialized transcriptional program driving OSM-induced CCM.

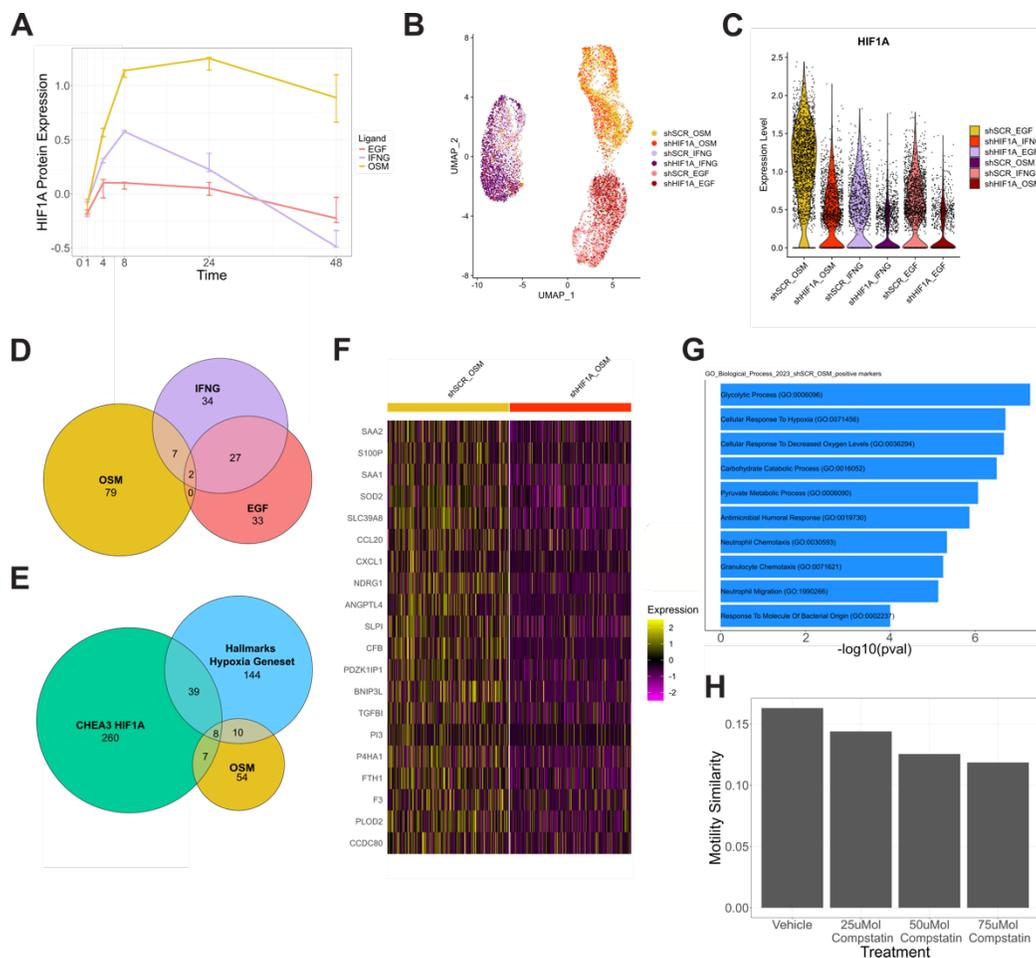
We next sought to examine the biological processes represented in the unique OSM-regulated HIF1A gene set (**Figure 3-4F**). Gene set enrichment analysis (GSEA) using Gene Ontology biological process annotations revealed that HIF1A, in OSM-treated cells, activates transcriptional programs associated with the Gene Ontology terms glycolysis and metabolic reprogramming, including terms such as "glycolytic process", "carbohydrate catabolic process", and "pyruvate metabolic process" (**Figure 3-4G**) [330]. Additionally, we observed enrichment of hypoxia-related gene sets, including "cellular response to hypoxia" and "cellular response to decreased oxygen levels." These findings align with HIF1A's well-documented role in promoting glycolysis and mediating the cellular response to hypoxia, confirming that HIF1A retains its canonical functionality in the OSM condition [331]. Furthermore, the glycolysis-related pathways suggest a metabolic shift that may be uniquely tied to OSM-induced phenotypes, providing additional insight into how HIF1A contributes to the molecular program underlying the OSM-induced phenotype.

Interestingly, we also identified gene sets associated with immune processes, including "antimicrobial humoral response," "neutrophil chemotaxis," and "neutrophil migration." The enrichment of terms related to neutrophil migration and chemotaxis was particularly intriguing, as neutrophils undergo collective cell migration to reach sites of infection [332], [333], [334]. This led us to hypothesize that some of the genes involved in this pathway might also contribute causally to the CCM phenotype observed in OSM-treated cells. To further investigate, we examined the neutrophil chemotaxis gene set and identified HIF1A-regulated genes in this context (**Supplemental Figure 3-5E-F**). Many of these genes were related to complement pathway activation, a finding of great interest given that complement signaling has been shown to be essential for cell adhesion and various forms of CCM [335], [336], [337]. This overlap between HIF1A-mediated hypoxic signaling, neutrophil migration processes, and complement activation suggests a potential mechanistic link by which HIF1A orchestrates OSM-induced CCM in epithelial cells.

## HIF1A dependent collective cell migration in mammary epithelial cells

We then sought to experimentally determine if complement signaling contributes to OSM-induced CCM. We treated cells concurrently with OSM and Compstatin, an inhibitor of C3, the convergence point of all modes of complement activation and performed live-cell imaging to assess any changes in phenotype [338]. Cells treated with Compstatin exhibited reduced clustering and motility similarity, resembling phenotypes observed in HIF1A knockdown conditions (**Figure 3-4H**). This strongly suggests that complement signaling plays a functional role in mediating OSM-induced CCM.

These findings establish a novel mechanistic link between OSM signaling, HIF1A activation, and complement pathway involvement in driving collective cell migration. By connecting the well-documented role of HIF1A in hypoxia responses and metabolic reprogramming with the unique transcriptional programs induced by OSM, we reveal how HIF1A mediates phenotypic outcomes critical for CCM. Furthermore, the identification of complement signaling as a downstream effector may indicate a novel and non-canonical role of HIF1A in influencing cell behavior.



**Figure 3-4: scRNAseq reveals HIF1A transcriptional program driving OSM-induced CCM**

**A)** OSM uniquely upregulates the HIF1A protein over 48 hours of ligand treatment. Error bars represent standard deviation.

**B)** UMAP projection of scRNAseq data indicates that transcriptional signals separate primarily by ligand treatment.

HIF1A dependent collective cell migration in mammary epithelial cells

C) HIF1A expression is significantly reduced in shHIF1A cell lines compared to shSCR.

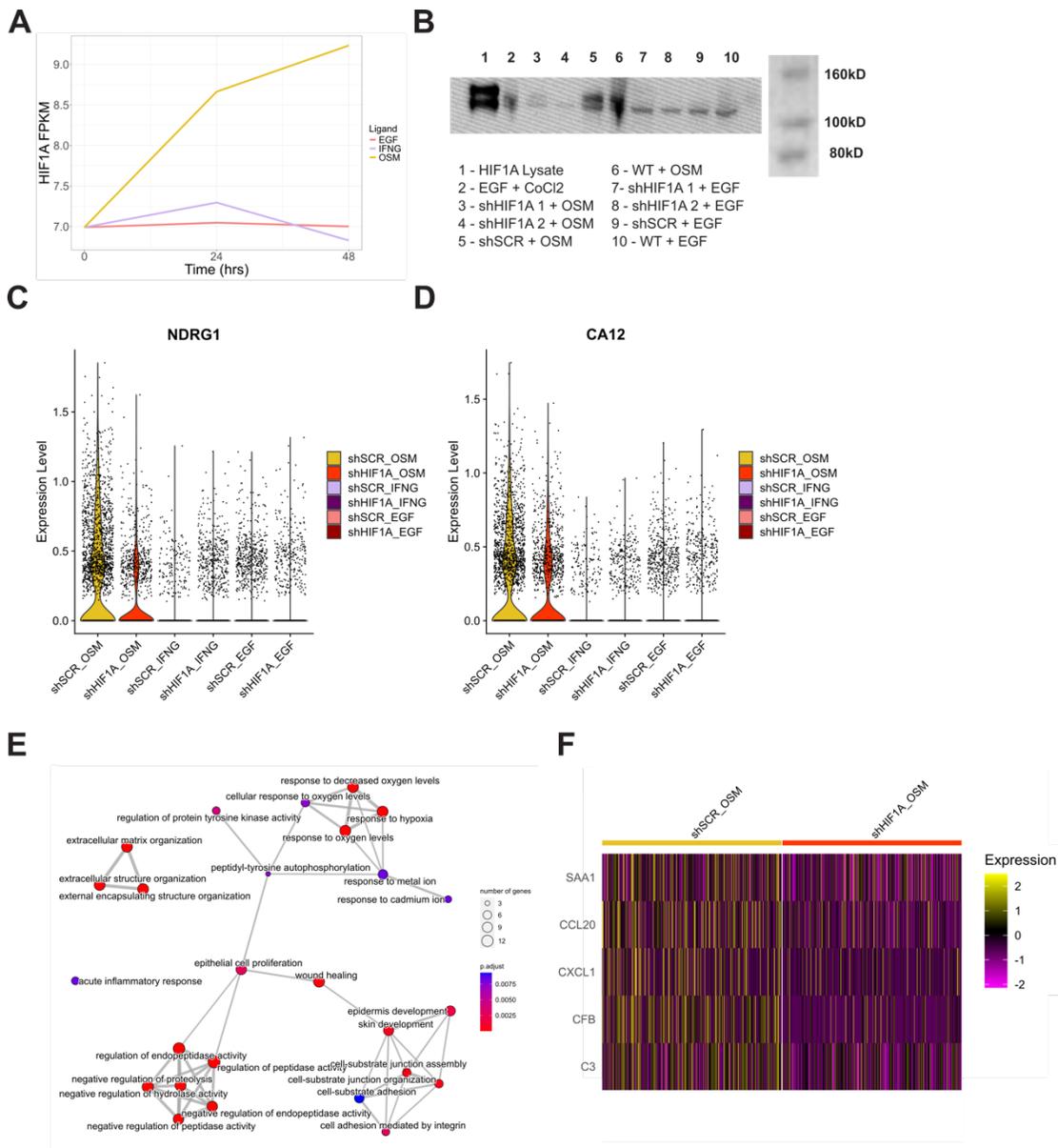
D) We identified HIF1A-regulated genes for each ligand condition by performing differential gene expression analysis between shSCR vs shHIF1A. Venn diagram represents a comparison of HIF1A-regulated gene sets for each treatment.

E) Comparison of unique OSM-HIF1A regulated genes to literature derived gene sets representing hypoxia and HIF1A transcription factor activity.

F) Relative expression of the top 20 differentially expressed unique OSM-HIF1A regulated genes.

G) We performed gene set enrichment analysis to identify Gene Ontology terms enriched in OSM treated shSCR cell vs shHIF1A cells. Top ten most enriched gene sets are shown.

H) Complement inhibition reduces CCM in OSM treated cells.



**Supplemental Figure 3-5: scRNAseq supplemental figures**

- A) Bulk RNAseq demonstrates that OSM uniquely upregulates HIF1A transcriptionally.
- B) Western blot confirming HIF1A knockdown in shHIF1A cell lines.
- C-D) Relative gene expression of NDRG1 and CA9: genes that are canonically activated by HIF1A.
- E) Gene set enrichment analysis of OSM treated shSCR vs. shHIF1A cells.
- F) Genes in the neutrophil chemotaxis Gene Ontology gene set that are upregulated in OSM treated shSCR cells.

### 3.5 Discussion

Cell migration is a key process in various physiological and pathological contexts, including tissue development, wound healing, and cancer metastasis. Collective cell migration (CCM), where groups of cells move together as a cohesive unit, plays a critical role in maintaining tissue architecture and facilitating invasion during cancer metastasis [276]. The mechanisms driving CCM are complex, involving complex interactions between signaling pathways, cytoskeletal dynamics, and cell-cell communication. In this study, we sought to explore how the cytokine OSM drives CCM in MCF10A cells. Through high-throughput screening and transcriptomic analysis, we identified HIF1A and STAT3 as central regulators of OSM-induced CCM, shedding light on the molecular pathways that drive this complex cellular behavior.

Our results demonstrate that OSM induces a unique transcriptional program that drives CCM, and we pinpoint HIF1A and STAT3 as key transcription factors that regulate this program. By using a siRNA-based screen, we mapped the functional role of these nodes in both general cellular phenotypes and OSM-specific phenotypes. Knockdown of HIF1A significantly impaired OSM-induced CCM, confirming its critical role in this process. Similarly, STAT3, which is known to mediate cellular responses to cytokine signaling, was also found to be essential for the promotion of CCM in response to OSM [322]. These findings suggest that the interplay between OSM, HIF1A, and STAT3 drives a unique program that governs CCM, which could have significant implications for understanding the mechanisms of metastasis and tissue remodeling in cancer.

HIF1A has long been recognized as a key regulator of cellular responses to hypoxia, where it mediates processes such as metabolic reprogramming, cell survival, and motility [326]. More recently, HIF1A has been implicated in partial epithelial-to-mesenchymal transition (EMT), a process that is often associated with cancer metastasis [118], [119]. In our study, we found that HIF1A was upregulated in OSM-treated cells and was crucial for the induction of CCM. This is consistent with prior work suggesting that HIF1A is involved in regulating the migratory behavior of cancer cells under various conditions [339], [340]. To further investigate the transcriptional program regulated by HIF1A in OSM-treated cells, we performed single-cell RNA sequencing, which revealed distinct changes in gene expression compared to cells treated with other ligands such as EGF or IFNG. Specifically, gene set enrichment analysis (GSEA) showed that HIF1A-regulated genes were enriched for pathways related to glycolysis, hypoxia, and immune responses. These findings suggest that HIF1A plays a pivotal role not only in promoting cellular motility but also in modulating immune-related signaling pathways, which may contribute to the collective migration observed in OSM-treated cells.

An unexpected finding in our study was the enrichment of complement-related immune pathways in HIF1A-regulated genes. Given that complement signaling is involved in a variety of cellular behaviors,

including migration, we hypothesized that complement activation could contribute to OSM-induced CCM. To test this, we used Compstatin, a complement inhibitor, and observed a reduction in both clustering and motility of OSM-treated cells, similar to the effects seen with HIF1A knockdown. This suggests that complement signaling is a downstream effector of HIF1A in driving CCM. This novel connection between OSM signaling, HIF1A activation, and complement-mediated motility provides new insights into the molecular mechanisms that regulate CCM and could have important implications for understanding cancer invasion and metastasis.

While our results provide a robust characterization of OSM-induced CCM in MCF10A cells, several limitations warrant further exploration. First, our study was conducted exclusively in MCF10A cells, a non-cancerous human mammary epithelial cell line [132]. To better understand the broader relevance of our findings, future studies should investigate the role of CCM in other epithelial cell lines and cancer-derived models. These include primary cells and patient-derived organoids, which more closely recapitulate the *in vivo* microenvironment and may provide insights into how CCM mechanisms are altered in cancerous conditions. Expanding the cell line repertoire will help determine the generalizability of our results across different tissue types and cancer models.

Additionally, while we performed single-cell RNA sequencing to capture the transcriptional response to OSM-induced signaling, our analysis focused primarily on the transcriptional consequences of HIF1A activation. It is important to note that HIF1A modulates various cellular pathways that could influence CCM. Pathways like mTOR signaling, integrin-mediated adhesion, and autophagy, which are also regulated by HIF1A, may contribute to CCM by affecting cellular motility, adhesion, and metabolic support for migration [341], [342], [343]. Further investigation into how HIF1A coordinates these signaling pathways will provide a more comprehensive understanding of its role in CCM and potentially uncover additional therapeutic targets.

Finally, while our study focused on transcriptional changes, other factors such as post-translational modifications, cellular localization, and crosstalk with other signaling networks could play critical roles in regulating CCM. Future research could address these aspects by incorporating proteomic analysis and real-time imaging techniques to track signaling events in live cells.

In conclusion, this study provides a detailed analysis of the signaling pathways and transcriptional networks involved in OSM-induced CCM. By expanding our understanding of how OSM and its downstream effectors regulate cellular migration, we lay the groundwork for future investigations into how CCM can be modulated for therapeutic purposes in cancer and tissue regeneration.

## 3.6 Additional Information

### 4.6.1 Data Availability

Raw live-cell images of MCF10A cells treated with ligands, siRNA, and drug are deposited on ....

All molecular data analyzed in this study and a detailed description of experimental and analytical methods are available at [synapse.org/LINCS\\_MCF10A](https://synapse.org/LINCS_MCF10A).

### 4.6.2 Contributions

Conceptualization: L.M.H., and I.C.M. Study coordination and supervision: L.M.H. Cell culture: I.C.M. Live-cell imaging: I.C.M. IF-imaging: I.C.M. scRNAseq: I.C.M. Data analyses: I.C.M. Writing: I.C.M and L.M.H. All authors reviewed and edited the manuscript.

#### 2.6.4 Acknowledgements and funding sources

This work was supported by NIH research grants U54-CA209988 and U54 HG008100, as well as the Anna Fuller Fund (L.M.H.). OHSU Massively Parallel Sequencing Shared Resource receives support from the OHSU Knight Cancer Institute NCI Cancer Center Support Grant P30CA069533. Cartoons created with Biorender.com.

#### 2.6.5 Competing Interests

The authors declare no competing interests.

#### 2.6.6 Supplemental Data

Supplemental data can be accessed through BioArchive at:

## 4. Chapter IV: Integrative Assessment of Ligand Combinations Reveals Synergistic Molecular Mechanisms Driving Complex Phenotypic Responses

The following chapter has been adapted from McLean et al (2025),  
**doi:** <https://doi.org/10.1101/2025.04.03.647095>

**Authors:** Ian C. McLean<sup>1</sup>, Sean Gross<sup>1,2</sup>, Jeremy Copperman<sup>3,4</sup>, Daniel Derrick<sup>1</sup>, Laura Heiser<sup>1,4\*</sup>

<sup>1</sup>Department of Biomedical Engineering, OHSU, Portland, OR, USA

<sup>2</sup>Capsida Biotherapeutics Inc., Thousand Oaks, CA USA <sup>3</sup>Department of Pharmacological Sciences, Mount Sinai Center for Bioinformatics, Icahn School of Medicine at Mount Sinai, New York, NY, USA

<sup>3</sup>Cancer Early Detection Advanced Research Center, OHSU, Portland, OR USA <sup>5</sup>Broad Institute of MIT and Harvard, Cambridge, MA, USA

<sup>4</sup>Knight Cancer Institute, OHSU, Portland, OR USA

\*Correspondence: [heiserl@ohsu.edu](mailto:heiserl@ohsu.edu) ORCID: 0000-0003-3330-0950

## 4.1 Abstract

The microenvironment surrounding cells plays a critical role in determining cellular phenotype. Key components of the microenvironment include the diverse milieu of ligands and cytokines that bind cell surface receptors to initiate changes in molecular programs. While the responses to extracellular signals have been extensively studied in isolation, little is known about the effects of combinations of signals on phenotypic and transcriptional responses. In this study, we used a coordinated approach to systematically investigate the combinatorial effects of the cytokines Oncostatin M (OSM) and Transforming Growth Factor Beta 1 (TGFB), and the growth factor Epidermal Growth Factor (EGF) on MCF10A mammary epithelial cells. Quantitative analysis of live-cell imaging data revealed a complex array of phenotypic responses after ligand treatment, including changes in proliferation, motility, cell clustering, and cytoplasmic size. We observed that all ligand combinations produce emergent phenotypic responses distinct from the maximal effects of individual ligands, suggesting induction of new molecular programs. Companion RNA sequencing studies revealed synergistic upregulation of genes involved in cell migration, epithelial differentiation, and chemotactic signaling. Notably, these included the chemokines CXCL3, CXCL5, and PPBP, which are known drivers of epithelial proliferation and migration. Additionally, transcription factor enrichment analyses and Reverse Phase Protein Array (RPPA) studies highlighted distinct changes in transcription factor activity and pathway utilization following combination treatment, including enhanced activation of MAP kinase and CREB signaling. Using partial least squares regression, we identified robust transcriptional signatures associated with quantitative cellular phenotypes. We validated these signatures in independent datasets, confirming that they generalize across cellular contexts. Finally, an in-depth functional analysis of cell motility with RNA interference and pathway inhibition revealed that synergistic upregulation of CXCR2 signaling, mediated by CREB transcription factor activation, contributes to increases in cell motility across ligand conditions. These findings demonstrate the importance of combinatorial signaling in reprogramming epithelial phenotypes and reveal potential therapeutic targets for disrupting synergistic pathways in disease contexts such as cancer progression. Together, this study provides a framework for understanding how complex ligand interactions shape phenotypic and molecular landscapes.

## 4.2 Introduction

The intricate interplay between cells and their microenvironment is a fundamental determinant of cellular behavior and function. Central to this dynamic relationship is the diverse array of ligands, cytokines, and extracellular matrix proteins that can initiate myriad intracellular responses that ultimately shape cellular phenotype. While the impacts of individual extracellular signals have been extensively studied, our understanding of how cells integrate and respond to combinations of extracellular signals remains limited.

In this study, we investigate the combinatorial effects of three ligands: Oncostatin M (OSM), Transforming Growth Factor Beta 1 (TGFB), and Epidermal Growth Factor (EGF), on the phenotypic and molecular responses of MCF10A mammary epithelial cells. These ligands hold pivotal roles in the normal development and function of mammary tissue, and their dysregulation is associated with disease [63], [64], [344]. EGF canonically activates the MEK/ERK and PI3K signaling pathways [67], OSM signals through JAK/STAT pathways [74], and TGFB orchestrates SMAD-mediated processes [70]. The signaling pathways activated by EGF regulate the motility, proliferation, and invasion of normal and malignant breast epithelial cells [67]. TGFB induces epithelial-to-mesenchymal transition (EMT) in breast epithelial

cells, achieved through the induction of EMT-associated transcription factors including SNAI1 and SNAI2, resulting in stereotyped changes in cell morphology and motility [285]. TGF $\beta$  also influences breast epithelial cell proliferation by activating p21 and suppressing key cell cycle transcription factors such as MYC, leading to cell cycle arrest [345]. Activation of JAK/STAT signaling downstream of OSM and other IL6 family cytokines modulates invasive properties and induces changes in modes of migration [147], [346]. Despite their well-defined molecular consequences, the interplay between these signaling pathways and their combined impact on cellular behavior remains poorly understood.

Prior investigations into the combinatorial effects of perturbations have primarily focused on exploring and predicting the interplay of therapeutic inhibitors [347]. Computational modeling approaches have demonstrated that synergistic drug interactions can be partially predicted from the transcriptional profiles of cells treated with individual agents [178], [348], [349], [350]. In addition, there is evidence indicating a strong correlation between synergistic gene expression patterns and the degree of drug synergy, demonstrating a robust association between combinatorial transcriptional dynamics and resultant phenotypic responses [182]. However, the exploration of combinatorial effects and relationships between transcription and phenotype have typically been confined to a single phenotypic response, notably viability, without extending to more complex phenotypes such as cell motility, morphology or spatial arrangement. Expanding our understanding of combinatorial effects beyond drug perturbations and viability holds great promise for uncovering mechanisms of signal integration at molecular and phenotypic levels.

Employing live-cell imaging and RNA-seq, here we systematically investigate the phenotypic and transcriptional responses of MCF10A mammary epithelial cells to combinations of EGF, TGF $\beta$ , and OSM. MCF10A is a well-characterized model system that has been utilized extensively to investigate tissue development, migration, and proliferation [133], [351], [352], [353], [354]. We found that all combinations of ligands induce phenotypic responses that differ significantly from their respective single ligand phenotypes, suggesting activation of additional molecular programs. Motivated by this finding, we performed comprehensive transcriptomic analysis to identify synergistic transcriptional programs in each combination condition, which revealed specific transcriptional programs modulated in response to combination treatments. We used Partial Least Squares Regression (PLSR) [355] to decipher the complex relationship between transcriptional programs and cellular phenotype. Our comprehensive analysis revealed that when combined, EGF and OSM synergistically amplify molecular programs associated with leukocyte chemotaxis and CXCR2 activation, resulting in increased cell motility. Functional validation demonstrated that synergistic upregulation of CXCR2-associated chemotactic factors is mediated by CREB transcription factor activation.

## 4.3 Methods

### MCF10A Cell Culture

Cell culture and ligand perturbation experiments were conducted as previously detailed [147]. Briefly, for routine growth and passaging, cells were cultured in growth media containing DMEM/F12 (Invitrogen #11330-032), 5% horse serum (Sigma #H1138), 20 ng/ml EGF (R&D Systems #236-EG), 0.5  $\mu$ g/ml hydrocortisone (Sigma #H-4001), 100 ng/ml cholera toxin (Sigma #C8052), 10  $\mu$ g/ml insulin (Sigma #I9278), and 1% Pen/Strep (Invitrogen #15070-063). For perturbation experiments, growth factor-free media was used, composed of DMEM/F12, 5% horse serum, 0.5  $\mu$ g/ml hydrocortisone, 100 ng/ml cholera toxin, and 1% Pen/Strep.

MCF10A cells were grown to 50-80% confluence in GM and detached using 0.05% trypsin-EDTA (Thermo Fisher Scientific #25300-054). Post-detachment, 5,000 cells were seeded into collagen-1 (Cultrex #3442-050-01) coated 24-well plates (Thermo Fisher Scientific #267062) in growth media. Six hours after seeding, cells were washed with PBS and growth factor-free media was added. After an 18-hour incubation in the new media, cells were treated with single ligand or combinations of ligands in fresh growth factor-free media: 10 ng/ml EGF (R&D Systems #236-EG), 10 ng/ml OSM (R&D Systems #8475-OM), and 10 ng/ml TGF $\beta$  (R&D Systems #240-B).

### **Live-Cell Imaging and Image Analysis Pipeline**

Live-cell imaging was performed using the Incucyte S3 microscope (Essen BioScience, #4647). Images were captured every 30 minutes for up to 48 hours. Live-cell image stacks were then registered using a custom Fiji script [224] and segmented with CellPose [152]. Image tracking was carried out using the Baxter Algorithms pipeline [151].

All analysis of cell tracking data was performed in RStudio [300]. The cell count metric was determined by counting the number of cells per field and normalizing this count by the T0 count for that field. Nearest neighbor distances were measured by calculating the pixel Euclidean distances from each cell centroid to the centroid of the second nearest cell in the imaging field. To account for variations in cell count, the mean nearest neighbor distances for each image were normalized by the expected mean distance to the nearest neighboring cell if the cells were distributed randomly [226]. Cytoplasmic size was calculated as the average cytoplasmic size 24 hours after ligand addition. Cell motility was quantified by first removing tracks with distance jumps greater than 200 pixels in 30 minutes. Motility was estimated as the slope of the mean squared displacement (MSD) [301] over time intervals ranging from 30 minutes to 6 hours. The slope of the MSD for each treatment was derived by constructing a linear model comparing MSD to the time interval. This value is proportional to the diffusion coefficient for Brownian motion [301]

To assess the statistical significance of the deviation in combination ligand phenotypes from single ligand conditions, ANOVA was used followed by Tukey's Honest Significant Differences test for post-hoc comparisons. For phenotype comparisons with EGF and PBS, ANOVA followed by Dunnett's test for post-hoc comparisons was performed using the DescTools package in RStudio [356]. A p-value of less than 0.05 was considered significant for all tests.

### **RNAseq Generation and Analysis**

MCF10A cells were transferred into RLT Plus buffer (Qiagen) containing 1%  $\beta$ -ME, flash-frozen in liquid nitrogen, and stored at  $-80^{\circ}\text{C}$  until RNA extraction. Total RNA was isolated from the frozen samples using the Qiagen RNeasy Mini Kit. cDNA libraries were prepared from poly(A)-selected RNA using the Illumina TruSeq Stranded mRNA Library Preparation kit. The Illumina HiSeq 2500 platform to generate 100-bp paired-end reads. Short read sequencing assays were performed by the OHSU Massively Parallel Sequencing Shared Resource.

RNAseq data was preprocessed and aligned using a pipeline adapted from Tatlow, et al [245]. TrimGalore (v. 0.4.3) was used to trim adapter sequences and low-quality bases using CutAdapt (v. 1.10) and to generate FASTQ quality reports using FastQC (v. 0.11.5). After adapter trimming, reads were filtered to have a minimum length of 30 bp. Trimmed reads were pseudo-aligned to the GENCODE V24

transcriptome (GRCh38.p5) using Kallisto (v. 0.46.2). Gene-level quantifications were obtained from transcript-level abundance estimates using the R package tximport (v. 1.8.0) in R (v. 3.5.0).

We performed multiple differential gene expression analyses, comparing each ligand condition to time zero controls (**Figure 4-2A, B, D, E, F**) and comparing each two-ligand combination condition to the respective single ligand conditions (**Figure 4-3**). Both sets of analyses were conducted using RNA-seq gene-level summaries with the R package DESeq2 (version 1.24.0) [228]. For all analyses, significantly induced genes were defined as genes with a LFC > 1.5 or LFC < -1.5 and a p-value < .05, adjusted using the Benjamini and Hochberg method [357].

Transcription factor enrichment scores were calculated using Priori with default settings, using TPM values as input [175]. Gene modules were identified through K-means clustering of the top 200 most up-regulated and top 200 most down-regulated genes, employing the ComplexHeatmap package [358]. The number of clusters was determined by gap analysis. Statistical comparisons of differentially expressed genes within each module were performed using chi-squared analysis, followed by examination of standardized residuals. A p-value of less than 0.05 was considered significant.

Gene set enrichment analyses were conducted using the Gprofiler package, focusing on Gene Ontology categories for Biological Process, Molecular Function, and Cellular Component [359]. Gene sets were considered statistically significant if the adjusted p-value was below 0.01. The most enriched gene sets for each analysis were selected for visualization by ranking the gene sets first by the smallest p-value and subsequently by the highest odds ratio. P-value adjustments for multiple comparisons were made using the g:SCS method from the Gprofiler package [360].

### **Partial Least Squares Regression**

PLSR models were built using replicate-level gene expression data as the independent variable to predict the replicate phenotypic metrics as the dependent variable. This pairing of replicates allowed us to utilize biological variation in both phenotypic and transcriptional responses. To exclude low-variance genes from the model, we used the VST method in the Seurat RStudio package (Version 3) to select the top 2,500 most variable genes for input [361]. Gene expression for each condition was normalized to T0 and scaled, and phenotypic metrics were scaled as well. The PLSR models were constructed using the PLS package in RStudio [362]. The number of components in each model was determined by identifying the elbows in the relative root mean squared error plots. Leave-one-out analysis was conducted to assess robustness. Genes were ranked by importance in the model by calculating the Variable Importance in Projection (VIP) scores [363]. The top 100 VIP scoring genes, either positively or negatively correlated with the first component of the model, were used as input for gene set enrichment analyses. The enrichment was performed using the Gprofiler package, focusing on Gene Ontology categories for Biological Process, Molecular Function, and Cellular Component [364], [365].

### **Partial Least Squares Regression Model Validation**

Orthogonal validation was conducted for the PLSR models predicting cell count and motility. This involved analyzing publicly available datasets containing RNAseq data paired with proliferation rates from 34 breast cancer cell lines [366] and motility estimates from 28 breast cancer cell lines [367]. We input the log<sub>2</sub>-normalized gene abundance data from these cell lines into our models to predict cell

count or motility for each line. Pearson correlation was then calculated to compare the predicted phenotypes to the experimentally determined metrics.

Additionally, we validated the PLSR model predicting cell count using the Project Achilles dataset for breast cancer cell lines from the Cancer Dependency Map Portal (DEPMAP) [368], [369]. We compared VIP scores to Gene Effect scores calculated by DEPMAP from CRISPR screens and investigated the statistical significance of the relationship between VIP and Gene Effect scores using chi-squared analysis.

### **CXCR2 Inhibition**

Cells were cultured in 24-well plates following previously established protocols. After cell attachment and subsequent culture in assay media, 5 nM AZD5069 (MedChemExpress, #19855) or DMSO was added along with the ligand treatments. Cell imaging and motility assessments were performed as previously described. To determine the statistical significance of changes in motility, we first fitted an ordinary least squares linear model to the data using the `estimatr` package in RStudio [370]. Then, we estimated marginal means with the `emmeans` package and computed pairwise contrasts to compare motility across all ligand conditions for inhibitor-treated versus DMSO-treated cells [371]. P-values were adjusted using Tukey's method for multiple comparisons, and a p-value < 0.05 was considered significant.

### **Reverse Phase Protein Array Sample Preparation**

Cells were lysed and collected by manual scraping into 50-100  $\mu$ L of lysis buffer (1% Triton X-100, 50 mM HEPES pH 7.4, 150 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 1 mM EGTA, 100 mM Na pyrophosphate, 1 mM Na<sub>3</sub>VO<sub>4</sub>, 10% glycerol, 1x complete EDTA-free protease inhibitor cocktail (Roche #11873580001), 1x PhosSTOP phosphatase inhibitor cocktail (Roche #4906837001)). The lysates were incubated on ice for 20 minutes, followed by centrifugation at 14,000 rpm for 10 minutes at 4°C. The supernatant was collected, quantified using a BCA assay, and then mixed with 4X SDS sample buffer (40% glycerol, 8% SDS, 0.25 M Tris-HCl, 10%  $\beta$ -mercaptoethanol, pH 6.8). The mixture was boiled for 5 minutes and stored at -80°C. Three sets of biological replicates were submitted for RPPA testing. The samples underwent standard pre-processing using protocols established at the MD Anderson Cancer Center RPPA core [372]. Statistical significance in antibody intensity between the EGF+OSM condition and both single ligand conditions was determined using ANOVA followed by Dunnett's test using the `DescTools` package [356]. The RPPA data used in this study were part of a larger panel of conditions and perturbations assessed in MCF10A cells. The full dataset is available in **Supplementary Table 1**.

### **SiRNA Assay and QPCR**

Cells were seeded in growth media at a density of 25,000 cells per well in a 6-well plate. Seven hours after seeding, the cells were transfected with either a commercially validated siRNA pool targeting CREB1 (Horizon Discovery #L-003619-00-0005) or a negative control siRNA (Horizon Discovery #L-003619-00-0005) using Lipofectamine RNAiMAX Transfection Reagent (Invitrogen #13778100) at a concentration of 25 nM siRNA. After 48 hours, the cells were treated with EGF, OSM, or a combination of EGF and OSM for an additional 24 hours. Each treatment condition was performed in triplicate.

To evaluate the changes in mRNA levels following siRNA-mediated CREB knockdown, we extracted total RNA from treated cells using the RNeasy Mini Kit (Qiagen # 74104) as per the manufacturer's protocol. We synthesized cDNA using the iScript cDNA Synthesis Kit (Bio-Rad #1708891). The mRNA expression levels were quantified through real-time qPCR with SYBR green chemistry on the Bio-Rad CFX Opus 384

Real-Time PCR System (Bio-Rad #12011452). The results were normalized to Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) levels using the  $2^{-\Delta\Delta Ct}$  method [373]. Statistical significance between expression levels was determined by student's T-test performed on  $\Delta Ct$  values [305]. The primer sequences for the target mRNAs are provided in **Supplementary Table 2**.

## 4.4 Results

### 4.4.1 EGF, OSM and TGFB treatment combinations induce emergent phenotypic responses

We nominated for study a panel of three ligands (EGF, OSM and TGFB) that canonically activate distinct signaling pathways and that have been shown to induce strong phenotypic responses in MCF10A mammary epithelial cells [147] (**Figure 4-1A**). Although the effects of these individual ligands have been examined in various cellular contexts including MCF10A [147], [374], [375], [376], [377], predicting how cells will respond to dual treatment at the molecular or phenotypic level is challenging due to the complex responses elicited by single ligands.

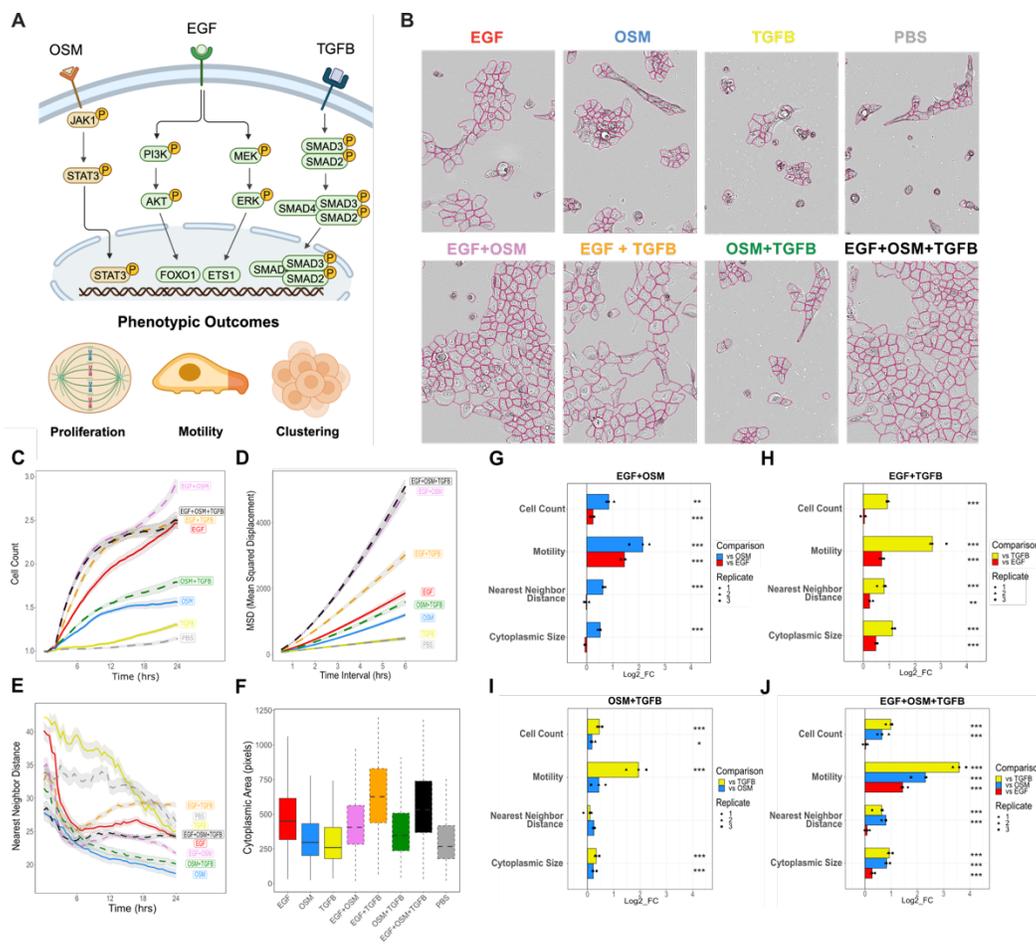
MCF10A cells cultured in growth factor free media were exposed to EGF, OSM, and TGFB independently, or in pairwise or triple combination, and subjected to live-cell imaging every 30 minutes for 24 hours. EGF is typically included as a supplement in MCF10A culture media, so this condition serves as a positive control for our study [132]; PBS serves as a growth-factor-starved negative control. We employed quantitative image analysis to assess changes in cell count, motility, cell clustering and cell spreading, and cytoplasmic size (**Figure 4-1B**). In our initial analyses, we describe the quantitative responses to each single ligand compared to PBS control. Consistent with previous studies (Gross et al., 2022), we found that independent treatment with EGF strongly induced proliferation, whereas TGFB and OSM induced only modest increases in cell count compared to PBS (**Figure 4-1C, Supp Fig 4-1A**). Cell motility was evaluated by calculating the mean squared displacement (MSD) over all cells for each treatment and deriving the diffusion coefficient from the MSD curves (**Figure 4-1D**). Like cell proliferation, EGF treatment significantly increased cell motility compared to PBS control, while OSM induced a modest increase in motility, and TGFB showed no change in motility compared to PBS (**Figure 4-1D, Supp 4-1A**). The calculation of the distance to the nearest neighboring cell revealed changes in cell clustering (decreased neighbor distance) and cell spreading (increased neighbor distance). EGF and TGFB treatment did not significantly affect the nearest neighbor distance compared to PBS control; however, OSM led to a decrease in this metric, indicating tight cell clustering (**Figure 1E, Supp Fig 4-1A**). Lastly, we observed that EGF treatment led to an increase in cytoplasmic size compared to the PBS control, whereas TGFB or OSM alone had no effect (**Figure 4-1F, Supp Fig 4-1A**). Notably, the addition of TGFB to EGF further amplified the increase in cytoplasmic size beyond that induced by EGF alone (**Supp Fig 4-1B**). In total, these analyses demonstrate that each treatment regime induced a distinct constellation of phenotypic changes.

Given the distinct phenotypic responses associated with each ligand, we hypothesized that the phenotypic responses to combination treatments would adhere to the Highest Single Agent (HSA) model [179]. The HSA model, used to study drug combination synergy, posits that the expected phenotypic response for a combination treatment is equal to the maximal effect of any single agent. Using this framework, we anticipated that the phenotypic responses to combination ligand conditions would primarily reflect the influence of the single ligand with the most significant impact on each aspect of cell phenotype. To test this hypothesis, we compared the quantified phenotypic metrics for each combination condition to those of each respective single ligand condition. A phenotype was considered

## Combinations Reveals Synergistic Molecular Mechanisms Driving Complex Phenotypic Responses

emergent (i.e., it did not mirror either single ligand response) if we detected a significant change that deviated from each single ligand condition, assessed using a one-way analysis of variance (ANOVA) followed by post-hoc Tukey's honest significant difference test (p-value < 0.05).

Consistent with the HSA model, cell count observed under EGF+TGFB and EGF+OSM+TGFB treatments remained unchanged compared to EGF alone. However, EGF+OSM and OSM+TGFB treatments resulted in a cell count greater than their individual ligand treatments (**Figure 4-1G, H, I, J**). This finding suggests an interaction between OSM and the other ligands that influences changes in cell proliferation. We also observed emergent phenotypic changes in cell motility, distance to nearest neighboring cells, and cytoplasmic size when ligands were applied in combination. Treatment with OSM and TGFB individually resulted in decreased cell motility when compared to EGF treated cells (**Supp Fig 4-1B**). However, combination treatment with EGF+OSM, EGF+TGFB, and EGF+OSM+TGFB resulted in an emergent increase in cell motility compared to all single ligand conditions (**Figure 4-1G, H, J**). Finally, combination treatment with EGF+TGFB resulted in an emergent phenotypic increase in both nearest neighbor distance and cytoplasmic size, indicating cell spreading (**Figure 4-1H**). Our findings indicate that across various ligand combination treatments and multiple phenotypic responses, combination treatment induces emergent behavior that deviates from that of individual ligand treatments. This suggests that combination treatments induce molecular programs not induced by single ligand treatments.



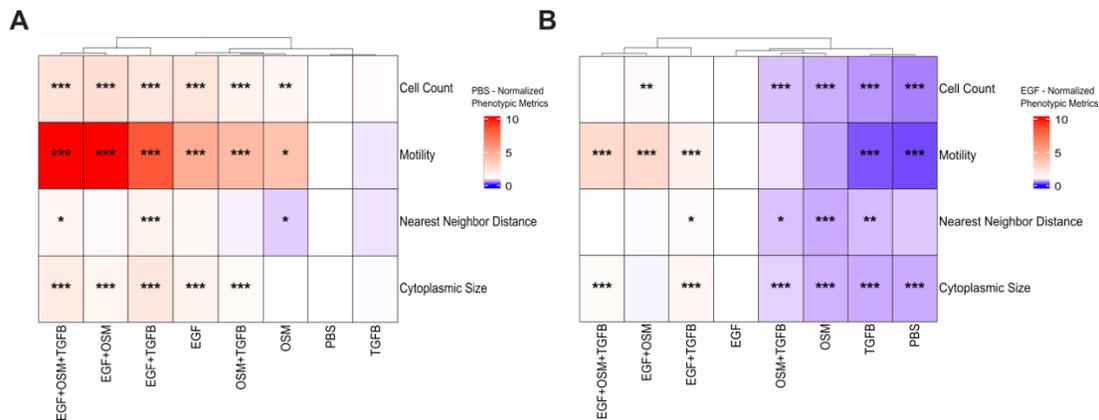
**Figure 4-1: Combination treatments induce diverse and emergent phenotypic behavior**

**A)** Overview of the signaling pathways activated by each treatment, highlighting the canonical pathways associated with each ligand. Schematic representations depict the potential phenotypic outcomes (e.g., proliferation, motility, and clustering/spreading) influenced by these pathways.

**B)** Representative images of MCF10A cells under different ligand treatments at 24 hours, demonstrating changes in cell phenotype.

**C-F)** Quantification of cell phenotype from 0-24 hours (cell count normalized to T0, MSD, nearest neighbor distance) or at 24 hours (cytoplasmic area).

**G-J)** Quantified phenotypic responses for each combination condition were compared to each single ligand condition comprising that combination. ANOVA followed by post-hoc Tukey's honest significant difference test was used to assess significance, with p-value < 0.05 considered significant.



**Supplemental Figure 4-1: Statistical comparison of phenotypic responses to control conditions**

**A)** Quantified phenotypic metrics for all ligand treatments normalized by and compared to PBS. Statistical significance was determined using Dunnett's test with a p-value < .05 considered significant.

**B)** Quantified phenotypic metrics for all ligand treatments normalized by and compared to EGF. Statistical significance was determined using Dunnett's test with a p-value < .05 considered significant.

#### 4.4.2 Identification of molecular programs induced by single and combination ligand treatments

We next used RNA sequencing to examine the molecular mechanisms driving response to ligand combinations. Cells were harvested after 24 hours of ligand treatment and subjected to bulk RNA sequencing. We posited that treatments leading to the strongest changes in cell phenotype compared to the PBS control would likewise display the greatest change in transcriptional responses. To test this hypothesis, we first assessed the overall transcriptional perturbation for each ligand condition by quantifying the number of differentially expressed genes as compared to PBS control (LFC > 1.5 or LFC < -1.5 and q-value < .05).

Comparison of differentially expressed genes revealed that treatments modulate gene expression programs to different extents; the EGF+OSM treatment induced the greatest number of differentially expressed genes, while independent treatment with TGFB had only modest impact on transcription (**Figure 4-2A, Supp Fig 4-2A, B**). To assess the overall change in cell phenotype for each condition, we calculated the total magnitude of the four phenotypic metrics relative to the PBS control. This was done

by representing the phenotypic metrics as a four-dimensional vector, where the origin indicates no change from PBS control, and then computing overall change in cell phenotype as the vector magnitude across these four dimensions. Then, for each treatment, we directly compared the overall change in cell phenotype to the number of differentially expressed genes (**Figure 4-2B**). We found a strong correlation between number of differentially expressed genes and overall change in phenotype ( $R^2 = 0.75$ ,  $p$ -value = .012). The treatments that provoked the most significant changes in gene expression also corresponded to the most substantial overall changes in cell phenotype (EGF+OSM+TGFB, EGF+OSM, and EGF+TGFB). Conversely, treatments that exerted a minimal change on cell phenotype led to the fewest changes in gene expression, highlighting a clear relationship between gene expression alterations and phenotypic outcomes across different treatment conditions.

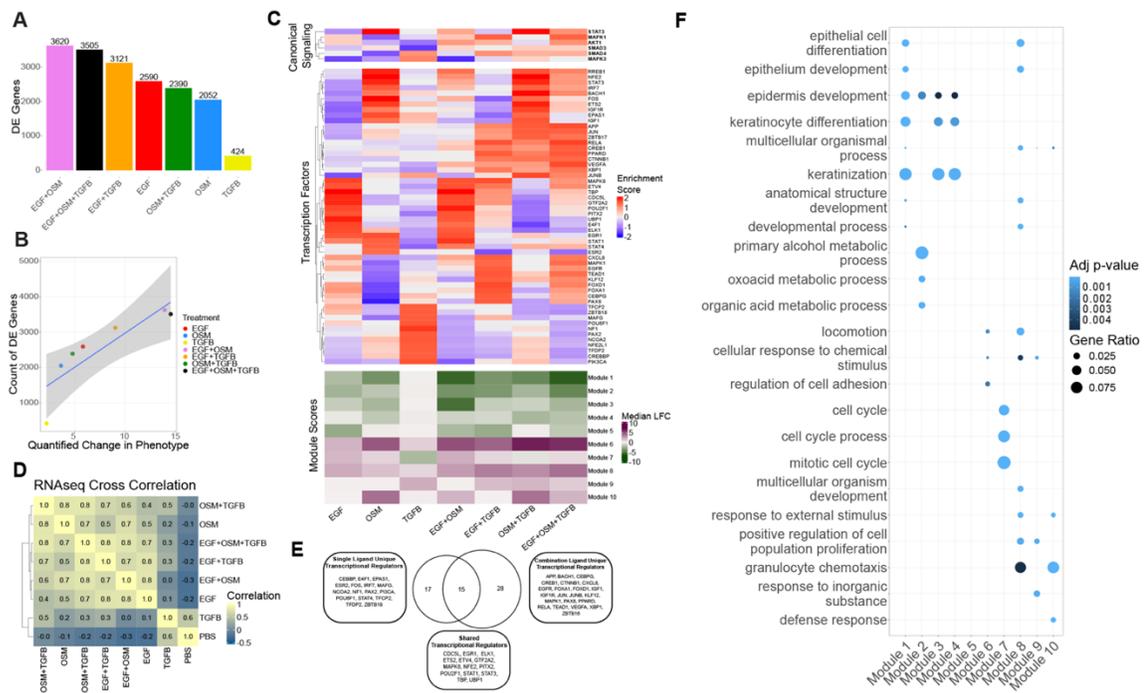
To further investigate the transcriptional programs modulated by single and combination treatments, we next performed transcription factor enrichment on the RNAseq dataset using *Priori*, an approach that leverages prior biological information to infer the activity of 223 transcriptional regulators [175]. *Priori* evaluates the activation levels of canonical transcription factors and enables data-driven identification of key signaling proteins that may play a crucial role in mediating phenotypic and combination responses. Analysis of the relative enrichment for canonical transcriptional regulators associated with EGF, OSM, and TGFB signaling with single-ligand treatments recapitulates established signaling pathways. Namely, EGF enriches AKT and MAPK transcriptional processes, OSM induces STAT3, while TGFB induces SMAD4 programs, respectively (**Figure 4-2C, Canonical Signaling**). Relative to EGF alone, EGF+TGFB increases MAPK enrichment, EGF+OSM increases AKT enrichment, and the three-ligand combination condition increases enrichment for both pathways. This suggests that activation of the MAPK and PI3K signaling cascade may contribute to the high level of proliferation observed in these conditions. Conversely, EGF+OSM diminishes STAT3 enrichment relative to the OSM condition (**Figure 4-2C, Canonical Signaling**). We also compared the top decile of transcription factors most strongly induced by each treatment condition (**Figure 4-2C, Transcription Factors**), which revealed that approximately half of the transcriptional regulators enriched in single ligand treatments are also active in their respective combination conditions (16/33) (**Figure 4-2E, Supp Fig 4-2C**). However, 28 (28/44) transcriptional regulators show enrichment only in combination treatments and not in single ligand conditions. These factors comprise a diverse array of transcriptional regulators including growth factor signaling molecules (EGFR, IGF1, IGF1R, MAPK1, VEGFA) and regulators of epithelial and immune cell differentiation (APP, BACH1, CEBPG, CREB1, CTNNB1, RELA). This may indicate that a common molecular program is induced in response to treatment with multiple ligands.

We next sought to identify coordinated molecular programs associated with the observed phenotypic responses. Focusing on the subset of genes most strongly induced relative to the T0 control, we filtered for the top 200 most up-regulated and top 200 most down-regulated genes for each treatment ( $p$ -value < 0.05, LFC > 1.5). Pairwise correlation of the log fold changes revealed that the combination treatments share transcriptional similarities with at least one of the individual ligands in each pair (Pearson correlation > 0.7) (**Figure 4-2D**). To further explore what may be driving the distinct phenotypic responses in the combination conditions, we applied K-means clustering [378] and gap analysis [259], which identified 10 coordinated gene modules (**Supp Fig 4-2D**). Gene module scores, calculated as the median expression for each module, were not exclusively up- or downregulated in any combination condition suggesting that combination responses do not activate unique transcriptional gene modules

distinct from single ligand conditions, but rather modulate existing transcriptional programs (**Figure 4-2C, Module Scores**).

We further explored the molecular programs in each module through gene set enrichment analysis, focusing on curated Gene Ontology categories (**Figure 4-2F**) [360], [364], [365]. Gene modules 1, 2, 3, and 4 are enriched for terms related to epithelial cell and keratinocyte differentiation and development. These modules are downregulated across all treatments except for the TGF $\beta$  condition. Additionally, Modules 1 and 2 are more strongly downregulated by the EGF+OSM and EGF+OSM+TGF $\beta$  conditions as compared to their respective single ligand treatments (**Supp Fig 4-2E, H**). These two treatments also show the most significant increase in motility compared to the PBS control, suggesting that downregulation of these gene modules corresponds to dedifferentiation associated with cell motility (**Figure 4-1C, D**).

Module 7 is enriched for cell cycle-related terms, with its expression highest under the EGF and EGF+OSM conditions, consistent with the increased cell count observed in these conditions. Gene Ontology terms associated with general cell motility (locomotion) are significantly enriched in modules 6 and 8, while specific mechanisms of cell motility (granulocyte chemotaxis) are enriched in modules 8 and 10. Module 6 is upregulated by EGF+TGF $\beta$ , OSM+TGF $\beta$ , and EGF+OSM+TGF $\beta$  compared to all single ligand responses, suggesting an underlying amplification of general cell motility transcriptional programs when these signals are combined (**Supp Fig 4-2F, G, H**). Conversely, module 10 is only upregulated by conditions containing OSM, indicating that a unique chemotaxis program is activated only when OSM is included in the treatment.



**Figure 4-2: Transcriptional programs induced by single and combination ligand treatments**

**A)** Number of differentially expressed genes (LFC > 1.5 or < -1.5, q-value < 0.05) across treatments relative to PBS control, showing the greatest response in EGF+OSM treatment.

## Combinations Reveals Synergistic Molecular Mechanisms Driving Complex Phenotypic Responses

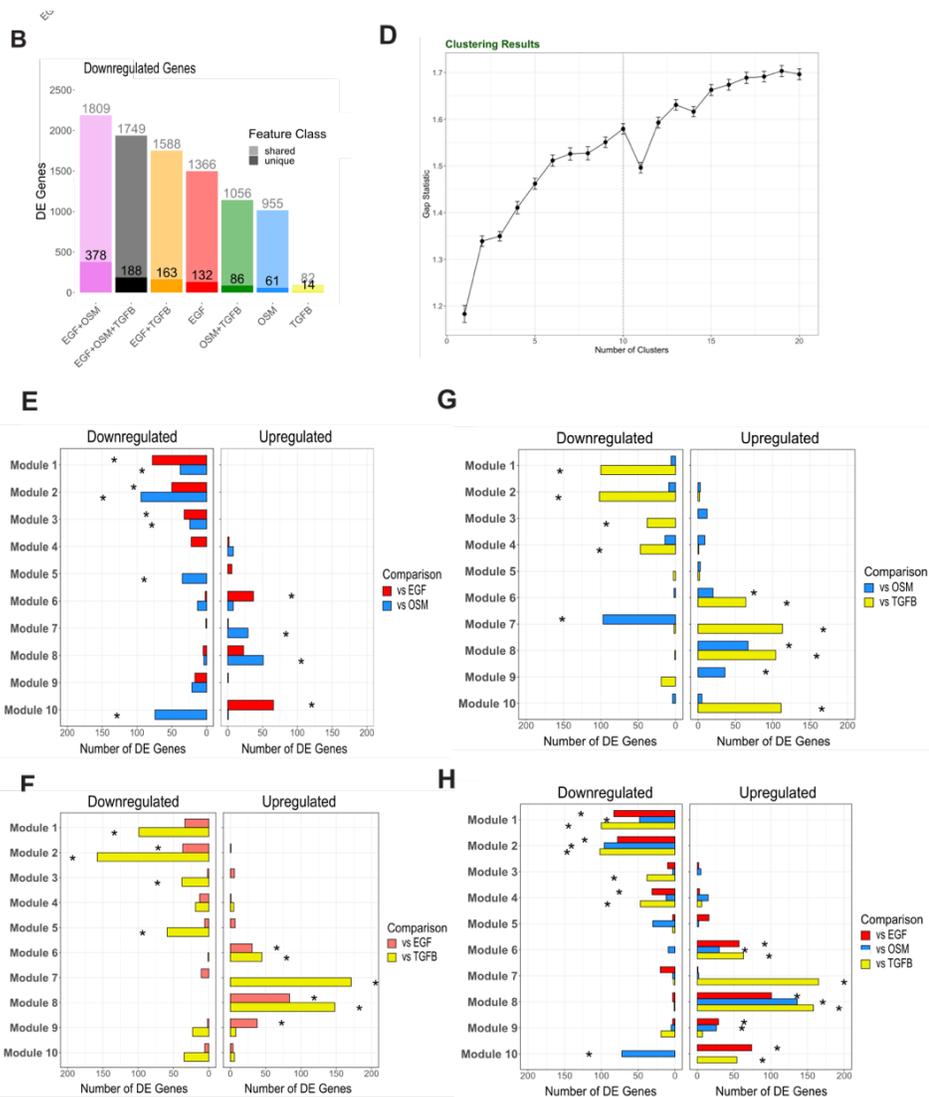
**B)** Correlation between the number of differentially expressed genes and overall changes in cell phenotype ( $R^2 = 0.75$ ,  $p = 0.012$ ).

**C)** Enrichment analysis of canonical transcriptional regulators (Canonical Signaling, top heatmap), the top decile of transcription factors (Transcription Factors, middle heatmap), and median LFC for gene modules (Module Score, bottom heatmap) activated in response to each treatment condition.

**D)** Pairwise Pearson correlations of gene expression log fold changes compared to T0 between all treatments.

**E)** Venn diagram comparing transcriptional regulators enriched in single versus combination treatments.

**F)** Enrichment of Gene Ontology terms for each gene module identified.



**Supplemental Figure 4-1: Detailed transcriptional analysis of single and combination ligand treatments**

**A)** Number of upregulated differentially expressed genes (LFC > 1.5,  $q$ -value < 0.05) for each treatment relative to PBS control. Genes unique to each treatment are shown with solid transparency, while those shared with at least one other condition are shown with lighter transparency.

**B)** Number of downregulated differentially expressed genes (LFC < -1.5, q-value < 0.05) for each treatment relative to PBS control, with transparency representing unique versus shared genes, as in panel A.

**C)** Upset plot showing the overlap of transcriptional regulators activated by single and combination treatments, highlighting the shared and unique regulatory programs across conditions.

**D)** Gap analysis identifies 10 optimal gene modules across treatments, based on clustering of transcriptional data.

**E-H)** Comparisons of gene module scores for each combination treatment versus the respective single ligand conditions. Statistical significance was determined by Chi-squared analysis, comparing the number of differentially expressed genes in each treatment relative to T0 for each module. P-value < .05 was considered significant.

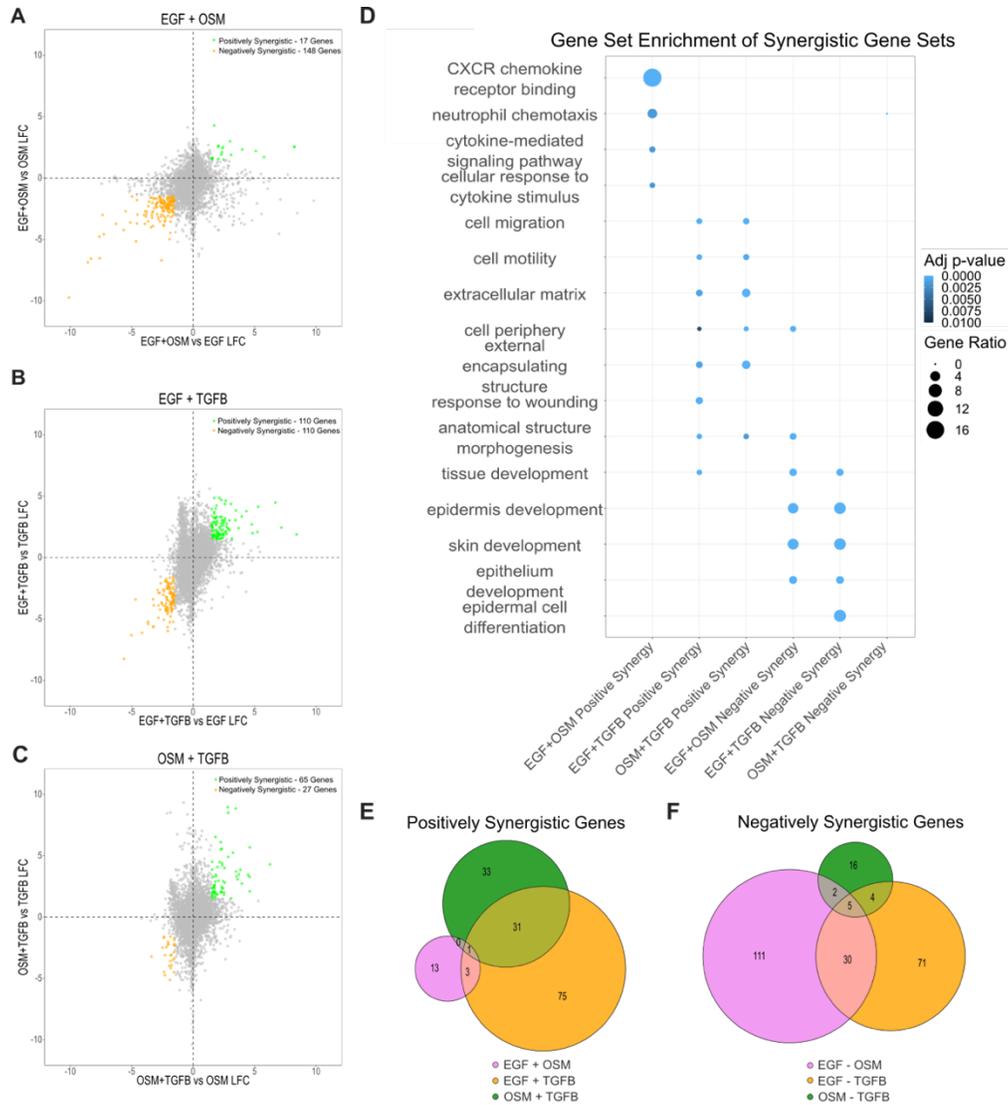
#### 4.4.3 Combination treatments result in specific synergistic transcriptional programs

The previous analyses suggest that combination treatments primarily amplify the transcriptional programs induced by single ligand responses. To further investigate this observation, we compared the transcriptional responses in combination treatments to predictions made using a simple additive model. Previous studies have shown that upregulated gene responses to ligand combinations typically follow either additive or multiplicative patterns [379]. To determine if this applies to our dataset, we developed a simple model by summing the fold changes in expression relative to the T0 control for upregulated genes (LFC > 0.5) from single ligand conditions and compared these predicted values to the observed expression levels in the combination treatments. Using this model, our analysis revealed a strong correlation between the predicted and actual responses for all ligand combination conditions (EGF+OSM -  $R^2 = 0.73$ , p-value < .001; EGF+TGFB -  $R^2 = 0.70$ , p-value < .001; OSM+TGFB -  $R^2 = 0.76$ , p-value < .001; EGF+OSM+TGFB -  $R^2 = 0.66$ , p-value < .001), suggesting that the combination responses can largely be explained by a simple additive model (**Supp Fig 4-3A-D**). These findings indicate that ligand combination responses predominantly recapitulate the transcriptional responses of the individual ligands.

The observation that the transcriptional responses to combination treatments predominantly reflect the effects of individual ligands was unexpected, considering the complex phenotypic changes we observed in each combination condition. We hypothesized that these emergent phenotypic effects might arise from specific synergistic transcriptional patterns not captured in our previous analysis. Here we more deeply explore that observation by borrowing frameworks developed in drug combination studies to identify synergistic gene expression programs by applying HSA modeling to molecular quantification [182]. For each combination treatment, we calculated the log fold change in expression for all genes compared to each respective single ligand condition, then visualized the changes in expression in x-y scatterplots (**Figure 4-3A-C**). We designated a gene as positively synergistic if the log fold change in expression after combination treatment exceeded 1.5 compared to each of the single ligand conditions (adjusted p-value < 0.05) (**Figure 4-3A-C**). Similarly, we defined negatively synergistic genes as those with a log fold change of at least -1.5 compared to both respective single ligand conditions. This comprehensive analysis revealed that for each pairwise combination, there exist subtle yet significant synergistic transcriptional programs. The number of positively synergistic genes per combination condition ranged from 17 to 110, while negatively synergistic genes ranged from 27 to 148. To evaluate how the range of synergistic modulation in our ligand combinations compares to previous studies using drug combinations, we performed the same analysis on MCF7 malignant breast epithelial cells treated with Tamoxifen, Mefloquine, and Withaferin individually and in combination (**Supp Fig 4-3E-G**) [182]. These drugs were selected based on their phenotypic synergy in reducing cell viability when combined. When comparing our results, we found that although the degree of transcriptional synergy in our ligand

combinations was far surpassed by the level of synergy observed in the Mefloquine and Tamoxifen combination (2235 genes), it was comparable to the two other combination conditions (50 genes, 137 genes). Interestingly, even in the Mefloquine and Withaferin combination, which showed relatively small transcriptional synergy, there was still substantial phenotypic synergy. This suggests that while a limited number of genes may be synergistically expressed in our ligand combination treatments, their strong modulation may have functional importance.

We performed gene set enrichment analysis of the positive and negative synergistic gene sets identified for each combination treatment using Gene Ontology Biological Process, Molecular Function, and Cellular Component gene sets [360], [364], [365]. Positively synergistic genes induced by the combination of EGF and OSM resulted in the upregulation of chemotactic transcriptional programs (**Figure 4-3D**). Combinations containing TGFB exhibited a large overlap of positively synergistic genes (EGF+TGFB = 32/110, OSM+TGFB = 32/65) (**Figure 4-3E**). These combinations also synergistically induced transcriptional programs related to ECM remodeling and cell motility, consistent with TGFB's known role in promoting epithelial-to-mesenchymal transition and the emergent enlargement of cell cytoplasm in this condition [72]. Notably, these EMT-related programs were not strongly induced by TGFB treatment alone and required either OSM or EGF in combination. This aligns with the observation from live-cell imaging that TGFB only induces migration and alterations in cell morphology in the presence of OSM or EGF (**Figure 4-1D, E**). Consistent with the live-cell image data, EGF+OSM and EGF+TGFB both downregulate epithelial differentiation processes (epithelium development, skin development). Epithelial cells that undergo transdifferentiation and loss of epithelial identity, notably through epithelial-to-mesenchymal transition, exhibit increased motility and changes in cell morphology, which we observed in both treatment combinations [380]. Negatively synergistic gene sets for EGF+OSM and EGF+TGFB had a large overlap, with 35 genes in common (**Figure 4-3F**), suggesting that these shared repressive transcriptional programs contribute to these processes (**Figure 4-1D, E**).



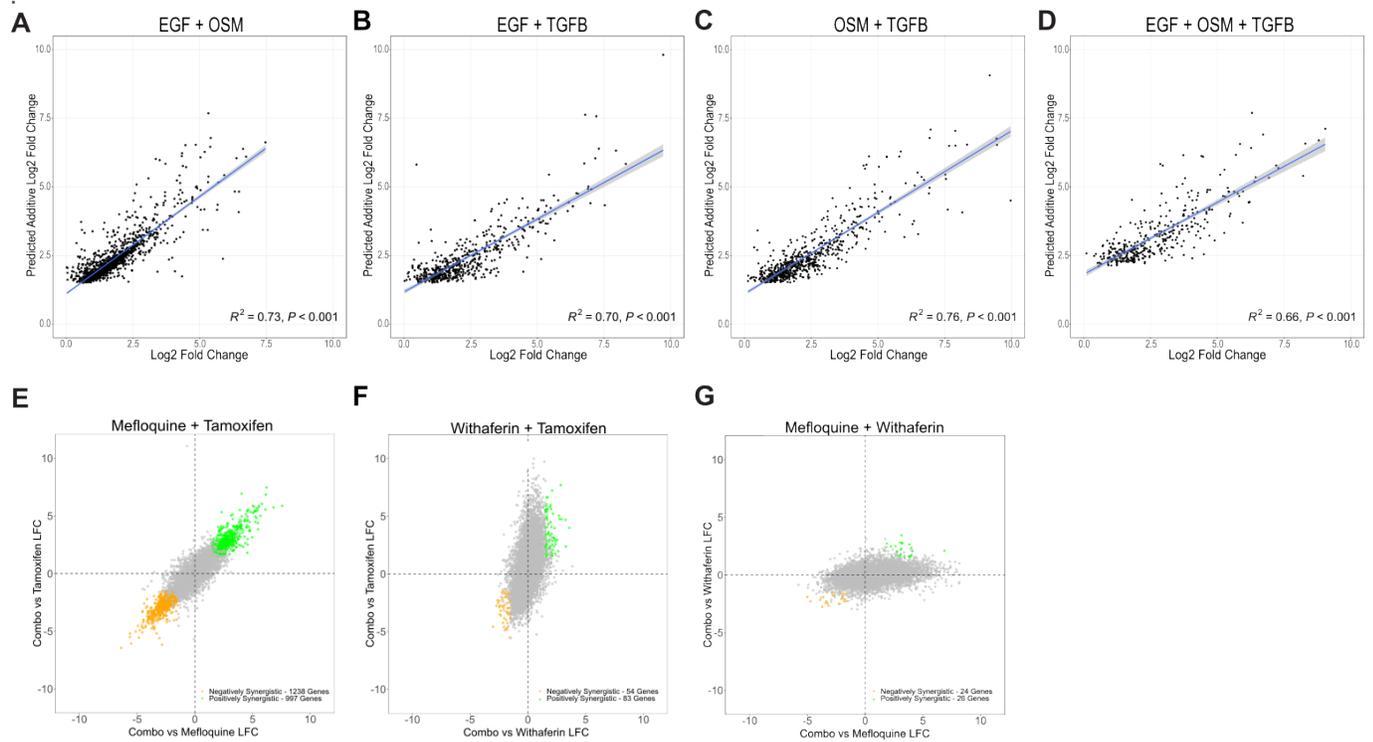
**Figure 4-3: Synergistic transcriptional programs induced by ligand combination treatments**

**A-C)** Scatterplots showing log fold change in gene expression for combination treatments versus each respective single ligand condition (EGF+OSM, EGF+TGFB, OSM+TGFB). Genes exhibiting positive synergy (LFC > 1.5, adjusted p-value < 0.05 compared to both treatments) are shown in green, and genes with negative synergy (LFC < -1.5, adjusted p-value < 0.05 compared to both treatments) are shown in orange.

**D)** Gene Ontology enrichment analysis of positively and negatively synergistic gene sets.

**E)** Venn diagram showing overlap of positively synergistic genes between combination conditions.

**F)** Venn diagram depicting negatively synergistic genes between combination conditions



**Supplemental Figure 4-3: Additive modeling of ligand combinations and transcriptional synergy analysis of drug treatments**

**A-D)** Scatterplots depicting the correlation between the observed log<sub>2</sub> fold change (LFC) for upregulated genes (LFC > 0.5) in each ligand combination condition (EGF+OSM, EGF+TGFB, OSM+TGFB, and EGF+OSM+TGFB) and the predicted LFC based on a purely additive model summing the LFC of the individual ligands. Pearson correlation coefficients ( $R^2$ ) are reported for each comparison, demonstrating strong correlations ( $R^2$  values ranging from 0.66 to 0.76,  $p$ -value < .001), indicating that ligand combination responses are predominantly additive.

**E-G)** Transcriptional synergy analysis applied to a published dataset of MCF7 cells treated with Tamoxifen, Mefloquine, and Withaferin individually and in combination. Scatterplots highlight the number of synergistic genes (positive and negative) induced by each drug combination, providing a comparison to ligand-induced transcriptional synergy.

#### 4.4.4 Partial Least Squares Regression uncovers transcriptional signatures driving cellular phenotype

We next sought to more directly link image and RNAseq data to uncover molecular drivers of complex cellular phenotypes. To achieve this goal, we utilized Partial Least Squares Regression (PLSR), an efficient statistical prediction tool that is especially appropriate for small sample data with many possibly correlated variables [355]. We constructed PLSR models to identify gene signatures to predict each of the four phenotypic metrics derived from live-cell imaging data (**Figure 4-1**). The three biological replicates of imaged cells shown in Figure 1 were harvested for the RNA sequencing described above, enabling direct linkage of image and RNAseq data for each experimental sample. This design leveraged biological variation in both phenotypic and transcriptional responses across replicates. The inputs for each model consisted of replicate-level phenotypic metrics and replicate-level Log Fold Change (LFC) gene expression compared to T0 control. Leave-one-out analysis was used to determine the optimal

number of latent variables in each model and to evaluate model robustness [381]. The Relative Root Mean Squared Error of Prediction (RRMSEP) for the four models ranged from 0.205 to 0.563, indicating good fit.

For each PLSR model, we identified the gene signature most strongly associated with phenotypic changes by calculating Variable Importance in Projection (VIP) scores [363]. VIP scoring estimates the overall significance of each feature in the model without distinguishing between positively and negatively correlated features. To further refine the top-scoring genes in each model, we categorized them based on their correlation with the phenotype of interest. For each model, we selected for further analysis the top 100 genes with the highest VIP scores and a positive correlation to the model's first component and the top 100 genes with a negative correlation to the first component. There were varying degrees of overlap among the highest-scoring VIPs from different phenotypic signatures. The largest overlap was observed between the Nearest Neighbor Distance and Cytoplasmic Size signatures, likely due to the inverse relationship between cell spreading and cell clustering (**Figure 4-4A, B**). Similarly, there was overlap between the Cell Count and Motility signatures, which is expected as these biological processes have been shown to be driven by similar molecular mechanisms [367]. Despite these overlaps, a significant number of genes were uniquely identified as VIPs for each phenotype, indicating that the PLSR regression identifies distinct biological mechanisms associated with each phenotype. We conducted gene set enrichment analysis for each signature to investigate the underlying cellular processes [360], [364], [365]. Enrichment for the top negatively correlated VIPs for Nearest Neighbor Distance include terms related to inflammatory response (acute-phase response, complement activation) (**Supp Fig 4-4A, B**). The top positively correlated VIPs associated with cytoplasmic size are enriched in terms related to synthesis processes for cell membrane components (isoprenoid biosynthetic process, cholesterol biosynthetic process, phospholipid biosynthetic process) and cell-substrate interactions (cell-substrate junction, focal adhesion), indicating a shared transcriptional program involved in membrane and cytoskeletal remodeling that regulates cytoplasmic size (**Supp Fig 4-4C, D**).

We next focused on exploring and validating the Cell Count PLSR model, a well-studied phenotype that serves as an excellent test case of our approach to link image and molecular data. Genes positively correlated with cell count exhibited enrichment for gene sets canonically associated with proliferation, including mitotic cell cycle, DNA replication, and cyclin E2-CDK2 complex (**Figure 4-4C, D**). Among the positively correlated VIPs are essential components of mitosis, including several cell cycle checkpoint genes (CDK2, CDC6, CDCA4) [382], [383], [384], microtubule regulation genes (TUBB, TUBB4B, TUBG1) [385], and growth-regulating secreted factors (AREG, IL1A, CSF3) [386], [387], [388]. Conversely, the top negatively correlated VIPs associated with cell count were enriched for cellular component terms such as vacuolar lumen and lysosomal lumen, implying an upregulation of autophagy components in the absence of proliferation, potentially due to cellular stress [389]. These findings suggest that the Cell Count PLSR model captures expected biologically relevant pathways and processes associated with proliferation.

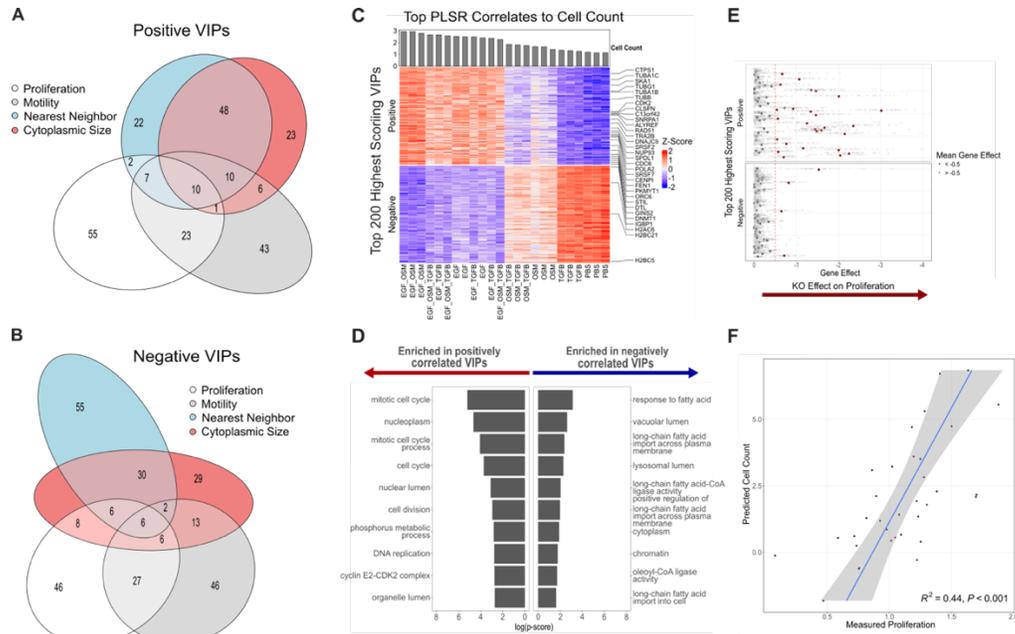
Assessing computational models using orthogonal approaches is crucial to ensure their robustness and reliability. Here, we leveraged independent, publicly available datasets to assess model generalizability and to validate our Cell Count model. First, we investigated whether the cell count signature identified genes essential for viability across a panel of diverse cancer cell lines. To achieve this, we utilized the Project Achilles dataset from the Cancer Dependency Map Portal (DEPMAP), which used high-

throughput CRISPR Cas9 to experimentally determine gene essentiality across thousands of cancer cell lines [368], [369]. The Gene Effect score quantifies the impact of gene knockout on cell viability. The more negative the experimentally determined Gene Effect score, the greater its impact on viability. A Gene Effect score less than -0.5 indicates cell depletion, while a score less than -1 indicates strong cell killing. In contrast, a Gene Effect score of 0 signifies that a gene is not implicated in viability.

We examined the essentiality of the highest scoring VIPs included in our PLSR signature by analyzing their experimentally determined Gene Effect scores for the 94 breast cancer cell lines included in the DEPMAP database (**Figure 4-4E**). VIP genes that showed the highest positive correlation with cell count were significantly more likely to have a Gene Effect score  $\leq -0.5$  as compared to genes with negatively correlated VIP scores ( $\chi^2 = 902.41$ , p-value  $< 2.2e-16$ ). Moreover, across all genes included in our Cell Count PLSR model, those identified as high-scoring VIPs were more frequently associated with a Gene Effect score  $\leq -0.5$  than were genes with low-scoring VIPs ( $\chi^2 = 3841.6$ , p-value  $< 2.2e-16$ ), demonstrating the predictive capability of our model in assessing gene essentiality across diverse breast cancer cell lines (**Supp Fig 4-4E**). Notably, genes with high VIP scores in the PLSR model that also had large Gene Effect scores include canonical cell cycle components (CDC20, PLK1, GINS1) [385] as well as recently discovered regulators (ALYREF) [390] and prognostic markers of breast carcinogenesis (GINS2) [391]. This underscores that analyzing transcription and phenotype in a single cell line across multiple perturbations offers insights into mechanisms governing cell viability across diverse biological, while also revealing recently discovered genes involved in cell proliferation and potentially identifying novel, undiscovered ones.

We also assessed the generalizability of the Cell Count PLSR model to predict proliferation in other cell line models. We analyzed publicly available datasets comprised of paired RNAseq and proliferation rates from 34 breast cancer cell lines [366]. For each cell line, we input the log<sub>2</sub>-normalized gene abundance data into our model to predict cell count. The model's predicted cell count was significantly correlated with experimental measures of proliferation ( $R^2 = 0.44$ , p-value  $< .001$ ) (**Figure 4-4F**). These results demonstrate that the Cell Count model generalizes across diverse cellular contexts beyond MCF10A and showcases the power of our approach to link molecular and image-based measurements. Taken together, these findings demonstrate the robustness of our approach in identifying functional molecular programs that govern complex phenotypic responses

## Combinations Reveals Synergistic Molecular Mechanisms Driving Complex Phenotypic Responses



**Figure 4-4: Partial Least Squares Regression (PLSR) links transcriptional signatures to phenotypic metrics**

**A-B)** PLSR models were used to link RNA sequencing data to phenotypic metrics derived from live-cell imaging (Cytoplasmic Size, Nearest Neighbor Distance, Motility, and Cell Count). Overlap in top positive VIP genes for each PLSR model is visualized in (A), while overlap for top negative VIPs is shown in (B).

**C)** Z-scored gene expression of the top 100 positive VIPs and top 100 negative VIPs for the cell count PLSR model.

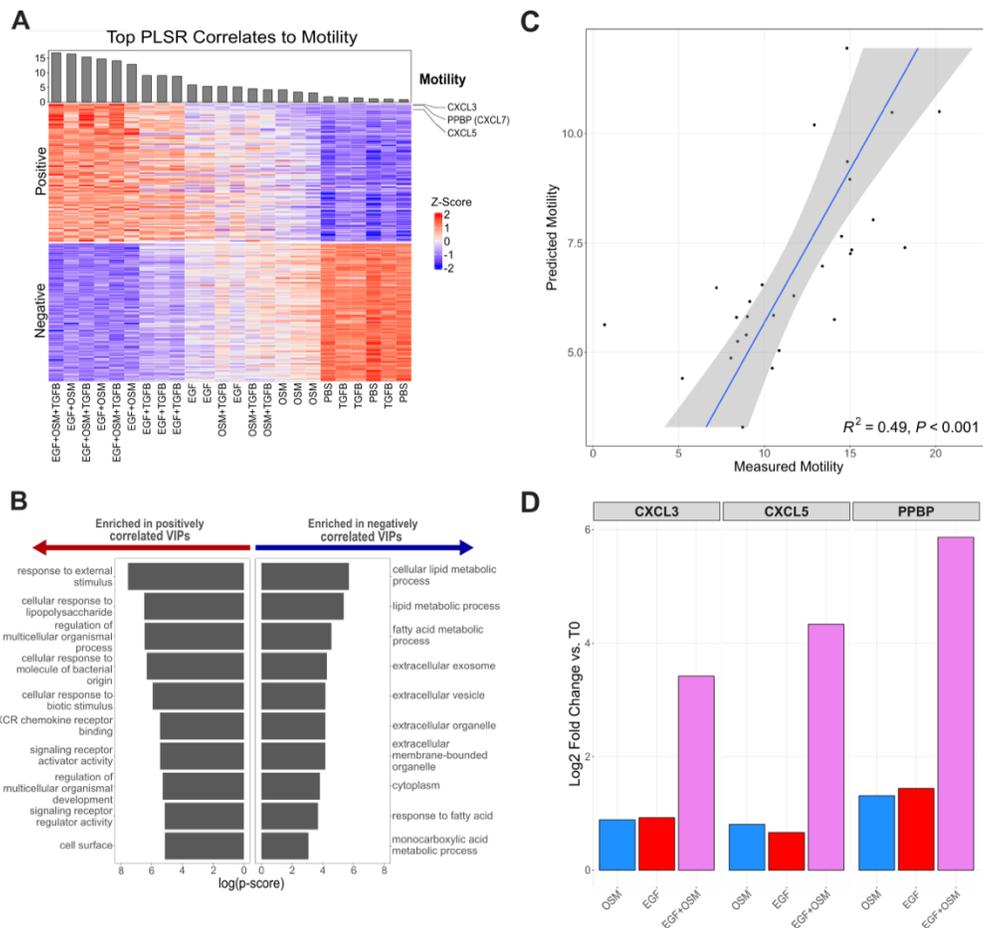
**D)** Gene set enrichment analysis of the top VIPs for the Cell Count model.

**E)** Validation of Cell Count PLSR model using the DEPMap Cancer Dependency Map (Project Achilles). Genes positively correlated with Cell Count show significant enrichment for essentiality in 94 breast cancer cell lines ( $\chi^2 = 902.41$ , p-value <  $2.2e-16$ ).

**F)** Generalization of the Cell Count PLSR model to other cellular contexts. The model was applied to RNAseq data from 34 breast cancer cell lines and showed significant correlation ( $R^2 = 0.44$ , p-value < .001) with experimental proliferation rates.



Having established the validity of our Cell Motility model, we next more deeply explored it to uncover novel biological mechanisms driving this phenotypic response. Among the top most important genes are CXCL3, PPBP (CXCL7), and CXCL5, ranking first, second, and fifth in VIP scores (**Figure 4-5D**). These genes encode chemotactic ligands that signal through the CXCR2 chemokine receptor, a pathway known to enhance mammary epithelial cell migration [257]. Furthermore, our previous RNAseq analysis showed that CXCL3, CXCL5, and PPBP are positively synergistically upregulated under the EGF+OSM condition, which produced the most significant increase in cell motility among all conditions tested (**Figure 4-1D**, **Figure 4-5D**). Motivated by this, we focused on the EGF+OSM combination condition to functionally investigate the mechanism by which this combined treatment synergistically enhances motility as compared to the individual effects of EGF and OSM.



**Figure 4-5: Expression of CXCR2 Agonists Correlates with Cell Motility**

**A)** Z-scored expression for the top 100 VIP genes positively and negatively with Cell Motility.

**B)** Gene set enrichment analysis of top VIPs, showing significant enrichment in pathways related to CXCR chemokine receptor binding.

**C)** Correlation between predicted cell motility values from the Cell Motility PLSR model and experimentally measured migration rates from 28 breast cancer cell lines. A significant correlation ( $R^2 = 0.49$ ,  $p$ -value  $< .001$ ) between predicted motility and experimental migration rates supports the relevance of the model in capturing motility-associated biological pathways.

D) LFC values compared to T0 control for the EGF+OSM combination condition and respective single ligand conditions show the synergistic upregulation of CXCR2 agonists (CXCL3, CXCL5, and PPBP).

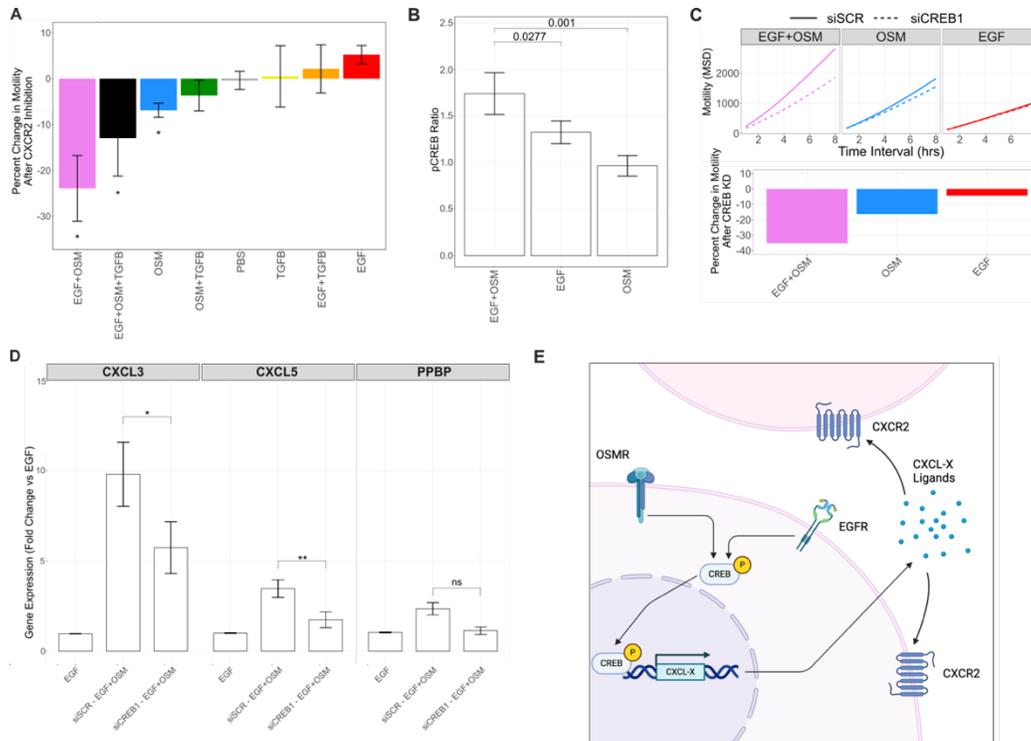
We hypothesized that the upregulation of CXCL3, CXCL5 and PPBP (CXCL7) contributes to the increased cell motility observed in the EGF+OSM condition compared to EGF and OSM single ligand conditions. To test this hypothesis, we first experimentally tested whether CXCR2 activation (the common receptor for CXCL3, CXCL5, and PPBP) influences cell motility in the EGF+OSM condition. We treated MCF10A cells with the ligand panel in the presence of AZD5069, a small molecule inhibitor of CXCR2 receptor activation, and then assessed cell motility. CXCR2 inhibition significantly suppressed cell motility in the EGF+OSM+TGFB, EGF+OSM, and OSM conditions, with the most substantial decrease observed in the EGF+OSM condition (23.9% median decrease across three biological replicates) (**Figure 4-6A, Supp Fig 4-5A**).

We then sought to identify the transcriptional regulators involved in the synergistic upregulation of CXCL3, CXCL5, and PPBP in the EGF+OSM condition. Given that this upregulation is synergistic in the EGF+OSM condition, and considering the emergent increase in cell motility observed with this combination treatment compared to either ligand alone, we hypothesized that the activation of transcriptional regulators would also exhibit a synergistic pattern in this context. To identify transcriptional regulators synergistically activated by EGF+OSM, we assessed proteomic changes with Reverse Phase Protein Array (RPPA) analysis 1 hour after ligand treatment [227]. Statistical analysis of the RPPA data identified 3 proteins with statistically significant change ( $p$ -value  $< .05$ ) in expression between the EGF+OSM condition and both single ligand conditions: P70 S6 Kinase (pThr389), S6 (pSer240/244), and CREB (pS133) (**Figure 4-6B, Supp Fig 4-5B**). CREB is a transcription factor known to enhance CXCL3, CXCL5, and PPBP expression [392], and its upregulation is consistent with our hypothesis that enhancement of chemotactic signaling contributes to the increased motility in the EGF+OSM condition.

We functionally validated CREB's role in motility by performing siRNA knock-down in the presence of EGF, OSM, and EGF+OSM treatments followed by assessment of cell motility. CREB inhibition had minimal impact on EGF treated cells, a modest impact on OSM treated cells, and the greatest impact on EGF+OSM treated cells, mirroring the effects of CXCR2 inhibition (**Figure 4-6C**). Knockdown of CREB was confirmed through qPCR (**Supp Fig 4-5C**). We next sought to confirm the relationship between CREB activation and chemokine expression via qPCR analysis of CREB knockdown cells. Under EGF+OSM treatment, the expression of CXCL3 and CXCL5 was reduced, confirming that CREB indeed regulates the expression of these chemotactic ligands (**Figure 4-6D, Supp Fig 4-5D**).

In summary, our findings reveal that EGF+OSM promotes a synergistic increase in cell motility compared to the single ligand conditions. RNAseq analysis revealed that CXCL3, CXCL5, and PPBP are strongly correlated to cell motility and synergistically upregulated under the EGF+OSM condition. Inhibiting CXCR2 significantly reduced cell motility, and RPPA analysis indicated a synergistic increase in CREB phosphorylation, which enhances the expression of CXCL3, CXCL5, and PPBP. CREB knockdown experiments confirmed its crucial role, demonstrating that CREB activation drives the upregulation of these chemokines. These results collectively suggest that the synergistic activation of CREB, in response to combined EGF and OSM stimulation, drives the transcriptional upregulation of key chemokines, thereby promoting increased cell motility (**Figure 4-6E**).

## Combinations Reveals Synergistic Molecular Mechanisms Driving Complex Phenotypic Responses



**Figure 4-6: Synergistic Transcriptional Upregulation of CXCR2 Chemotactic Signaling Molecules via CREB Activation Promotes Increased Cell Motility**

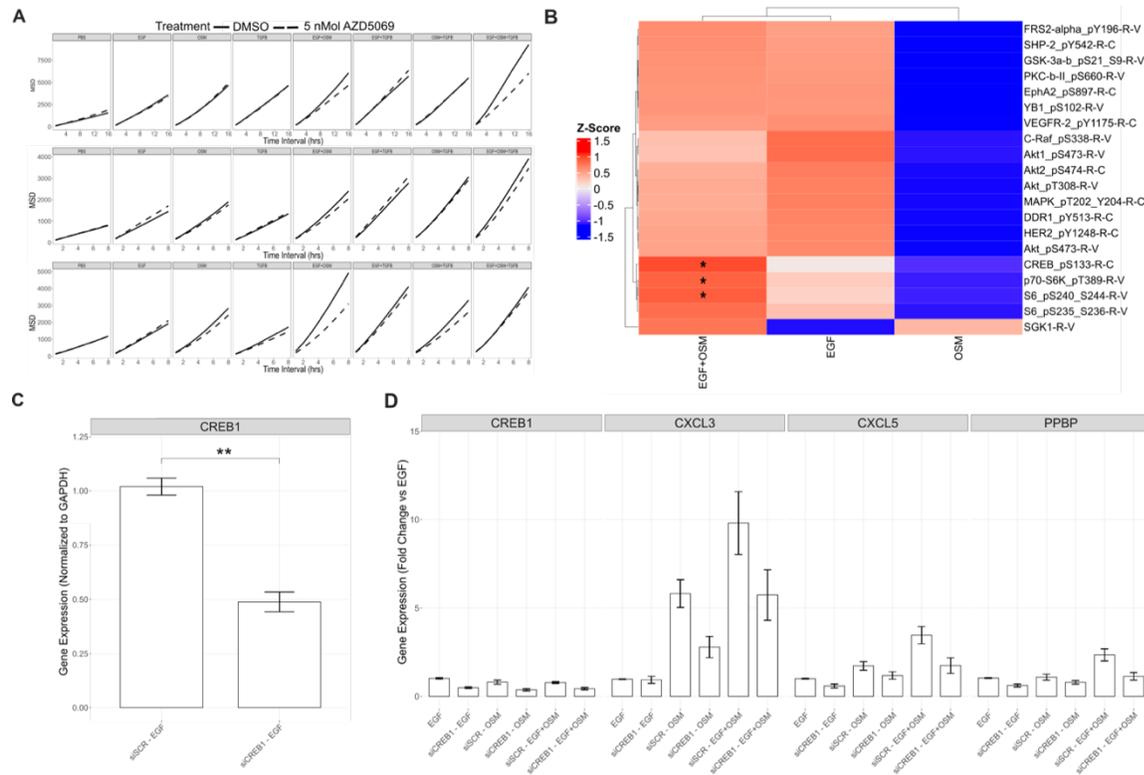
**A)** Cell motility assays following treatment with single and combination ligand panel in the presence or absence of AZD5069, a CXCR2 inhibitor. CXCR2 inhibition significantly decreased cell motility in the EGF+OSM, EGF, and OSM conditions. Data shown as median change in motility with standard deviation from three biological replicates. P-value < .05 was considered significant.

**B)** Reverse Phase Protein Array (RPPA) analysis 1 hour after treatment with EGF, OSM, EGF and EGF+OSM. Statistically significant changes in protein expression (p-value < .05) were assessed using Dunnett's test.

**C)** Effects of CREB knockdown on cell motility of cells treated with EGF, OSM, or EGF+OSM. Mean Squared Displacement (MSD) is shown in the top panel and change in motility is shown in the bottom.

**D)** qPCR analysis of chemokine expression in CREB knockdown cells following EGF+OSM treatment. Barplots depict fold change in expression compared to the EGF control. Student's t-test was used to assess significance (p-value < .05).

**E)** Putative mechanism of synergistic activation of CREB in response to combined EGF and OSM stimulation drives the upregulation of CXCL3, CXCL5, and PPBP, leading to increased cell motility via CXCR2 activation.



**Supplemental Figure 4-5: Cell Motility Assays, CREB Knockdown Efficiency, and RPPA Data**

**A)** Cell motility assay data showing the effects of CXCR2 inhibition (AZD5069) on MCF10A cell motility in the EGF+OSM, EGF, and OSM conditions, across all three biological replicates. The data depict the mean squared displacement, with CXCR2 inhibition significantly reducing motility in the EGF+OSM condition.

**B)** Reverse Phase Protein Array (RPPA) data from cells treated with EGF, OSM, and EGF+OSM. Dunnett's test was used to compare the single ligand conditions to the EGF+OSM combination, revealing statistically significant changes ( $p$ -value  $< .05$ ) in CREB activation in the EGF+OSM condition.

**C)** qPCR confirmation of CREB knockdown in MCF10A cells, with treatment conditions for siRNA targeting CREB.

**D)** Full qPCR analysis of chemokine expression in CREB knockdown MCF10A cells treated with EGF, OSM, and EGF+OSM. The data show fold change of gene expression compared to EGF and normalized to GAPDH.

## 4.5 Discussion

Cells operate within a complex microenvironment in which signals from the local environment are continuously changing. These signals are integrated by cells, impacting transcription and signaling processes and thus significantly affecting cellular functions. While prior research has primarily investigated how individual signals combine to affect phenotypic and transcriptional outcomes, often in the context of drugs and a limited number of phenotypes like cell survival and proliferation, this study seeks to expand these concepts to the more intricate realm of ligand interactions. Here we explored the interactions between the ligands EGF, OSM, and TGF $\beta$  and their effects on molecular and cellular responses. By analyzing the outcomes of both single-ligand and combined treatments, we reveal how these signals drive distinct transcriptional programs that impact cell motility, proliferation, and

differentiation. Our findings show that combining ligand treatments can lead to unexpected phenotypic behaviors. For example, certain ligand combinations resulted in greater cell counts (EGF+OSM, OSM+TGFB), enhanced motility (EGF+OSM, EGF+TGFB, EGF+OSM+TGFB, OSM+TGFB), and altered both cell morphology and spatial arrangement (EGF+TGFB, OSM+TGFB, EGF+OSM+TGFB) as compared to their single-ligand treatments. These results demonstrate that the combined effects of ligand treatments can exceed the sum of their individual parts, emphasizing the challenge of predicting phenotypic responses from single-ligand data alone.

Distinct signaling pathways activated by EGF, OSM, and TGFB each play unique roles in cellular responses [67], [71], [393]. Our RNA sequencing data reveal that combination treatments mimic the gene expression profiles of one or both single-ligand conditions. An additive model of single-ligand gene expression showed a strong correlation with gene expression patterns observed in combination treatments, with the magnitude of transcriptional changes aligning with corresponding phenotypic effects. Notably, gene programs associated with epithelial differentiation are downregulated after nearly all treatment paradigms tested, with the notable exception of TGFB. This finding is intriguing given TGFB's well-studied role in epithelial-to-mesenchymal transition (EMT) [72]. This suggests that EGF might be necessary for TGFB to trigger the conventional EMT in MCF10A cells, highlighting the dependence of well-studied ligand-induced phenotypes (e.g. EMT induced by TGFB), on the presence of other ligands in the microenvironment.

Moreover, we identified unique transcription factors enriched across all combination conditions, including regulators of epithelial differentiation such as APP, BACH1, and CTNNA1. The shared enrichment of these transcription factors suggests that the combination of these ligands converges on similar signaling pathways to influence differentiation state. Building on prior research showing that small degrees of transcriptional synergy can influence phenotypic synergy [182], we utilized an HSA-based modeling approach to define synergistic transcriptional programs. This approach uncovered subtle but significant synergistic gene sets specific to each ligand combination. While most studies have focused on phenotypic synergy, extending these frameworks into molecular synergy provides valuable insights into the underlying mechanisms driving combined ligand effects. Our studies support the adoption of existing frameworks designed for drug-induced changes in viability to gain insights into complex phenotypic responses. With the advent of spatially resolved assays [394], we envision that these frameworks could be applied to a broad array of data types and biological questions.

The use of Partial Least Squares Regression (PLSR) to connect transcriptional data with phenotypic outcomes is a significant strength of this study. Identifying gene signatures linked to cell count, motility, spatial organization and cytoplasmic size offers insights into the molecular drivers of these processes. Validation of the PLSR model with publicly available datasets and the Cancer Dependency Map underscores the robustness and generalizability of our PLSR models, with potential applications for identifying critical regulators of cell proliferation and survival [366], [367], [368], [369]. Furthermore, this approach holds significant promise for uncovering novel regulators of cell phenotype by linking previously uncharacterized genes to specific cellular behaviors. However, alternative machine learning approaches such as random forests or neural networks could be employed in future studies to capture more complex non-linear relationships between transcriptional data and phenotypic outcomes [395], [396], [397], [398]. These methods may offer complementary insights and help further refine our understanding of the complex signaling networks governing cellular behavior. Additionally, expanding

the analysis to include time-course transcriptional data could provide a dynamic view of how these programs evolve over time in response to ligand treatments.

Our analysis of CXCR2 chemotactic signaling and CREB activation provides important mechanistic insights into molecular drivers of cell motility. The synergistic upregulation of CXCL3, CXCL5, and PPBP in the EGF+OSM condition, along with the effects of CXCR2 inhibition and CREB knockdown, highlights a key signaling axis involved in cell motility. CREB's role in promoting chemokine expression further clarifies the molecular mechanisms underlying the observed changes. We demonstrated that EGF+OSM, only when applied in combination, phosphorylates and activates CREB transcription factor, leading to the transcriptional upregulation of CXCR2 receptor agonists. Activation of CXCR2 then contributes to increased cell motility.

Previous studies have established CREB's role in promoting chemokine activity in malignant epithelial cells, transgenic mice, and other cell lines [392], [399], [400]. While EGF is known to activate CREB in various contexts, including in breast tissue [401], we reveal that in MCF10A cells, this pathway is uniquely activated by the combined presence of OSM and EGF. Additionally, prior research has shown that CXCR2 is overexpressed in breast cancer epithelial cells and enhances malignant cell migration [402], [403]. Here, we provide evidence for an autocrine signaling mechanism achieving similar outcomes in cell motility. However, it is important to note that the observed phenotypic changes likely also involve protein and signaling-driven mechanisms. While this study focused on the CREB-CXCR2 axis, future research should investigate other interacting pathways to provide a more comprehensive view of these regulatory networks on changes in motility. For instance, identifying which CXCR2 agonists actively bind to the receptor and exploring the roles of other transcription factors or co-regulators in chemokine expression and motility could offer deeper insights into these processes.

This study has several limitations that should be addressed in future research. First, while we focused on specific quantitative changes in cellular phenotype, many other aspects remain unexplored, such as metabolic activity, apoptosis, and differentiation status. Examining these phenotypes could provide critical insights into how ligand combinations influence cellular behavior. For instance, changes in metabolic activity could elucidate the energetic requirements for motility [404], apoptosis assays might reveal how ligand signaling impacts cell survival [405], and differentiation markers could help determine whether ligand combinations push cells toward specific lineages [406]. Additionally, investigating chromatin remodeling and transcriptional dynamics could uncover upstream regulatory mechanisms driving the observed phenotypic changes [407].

Second, our research was conducted using MCF10A cells, though we validated our results with external cancer cell datasets containing thousands of paired RNA-seq and phenotypic profiles to address this limitation. Future studies could expand the range of cell lines used, incorporating primary cells that more closely mimic the physiological state of cells in vivo [408]. Additionally, using patient-derived organoids, a more complex model system that includes additional cell-cell contacts and extracellular matrix, could help determine the generalizability of our findings across different biological contexts [409].

Lastly, while our study observed various types of cell motility, we did not differentiate between distinct motility behaviors. Understanding the differences between random motility, directed migration, and collective movement could offer deeper insights into the regulation of these processes by ligand combinations [410]. To address this, future experiments could include chemotactic gradient assays to

evaluate directed migration [411], scratch assays to study wound healing-like behavior [412], and 3D model systems to investigate the contributions of the ECM [413]. Such approaches would provide a more nuanced understanding of how specific signaling pathways influence distinct motility types.

Overall, this study offers a comprehensive analysis of how EGF, OSM, and TGF $\beta$  signaling pathways interact to influence cellular responses through complex transcriptional programs. The integration of transcriptomic and phenotypic data using machine learning approaches enhances our understanding of the molecular mechanisms governing cell behavior, with potential implications for developing targeted therapeutic strategies to modulate cell motility and proliferation in cancer and other diseases.

## 4.6 Additional Information

### 4.6.1 Data Availability

Raw live-cell images of MCF10A cells treated with single and combination ligands are deposited on Zenodo [10.5281/zenodo.14261795](https://zenodo.org/doi/10.5281/zenodo.14261795).

RNA sequencing data and processed counts data generated for this study can be accessed from the Gene Expression Omnibus: [GSE282654](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE282654).

### 4.6.2 Contributions

Conceptualization: L.M.H., I.C.M, and J.C. Study coordination and supervision: L.M.H. Cell culture: I.C.M. and S.M.G. Live-cell imaging: I.C.M. and S.M.G. RPPA: I.C.M. and S.M.G. RNAseq: I.C.M., S.M.G., and D.S.D. Data analyses: I.C.M. Writing: I.C.M and L.M.H. All authors reviewed and edited the manuscript.

### 2.6.4 Acknowledgements and funding sources

This work was supported by NIH research grants U54-CA209988 and U54 HG008100, as well as the Anna Fuller Fund (L.M.H.). OHSU Massively Parallel Sequencing Shared Resource receives support from the OHSU Knight Cancer Institute NCI Cancer Center Support Grant P30CA069533. Cartoons created with Biorender.com.

### 2.6.5 Competing Interests

The authors declare no competing interests.

### 2.6.6 Supplemental Data

Supplemental data can be accessed through BioArchive at:

**doi:** <https://doi.org/10.1101/2025.04.03.647095>

## 5. Chapter V: Conclusions and future work

### 5.1 Summary of contributions

#### 5.1.1 Three dimensional organoid experiments

One limitation of the research presented in this dissertation is that it has been primarily restricted to work done in MCF10A cells. While MCF10A cells are a widely used model system for studying normal mammary epithelial cell behavior, they are inherently a simplified representation of the mammary gland *in vivo*. These cells lack several key characteristics of the mammary microenvironment, including extracellular matrix (ECM) components, interactions with immune cells, and the three-dimensional architecture that defines the mammary gland [134]. As such, MCF10A cells serve as a powerful reductionist tool but do not fully capture the complexity of mammary tissue *in vivo*. Recognizing this limitation, part of this graduate research involved establishing more complex model systems in the lab to validate the findings obtained from MCF10A cells, particularly those related to the Oncostatin M (OSM)-induced collective cell migration (CCM) phenotype.

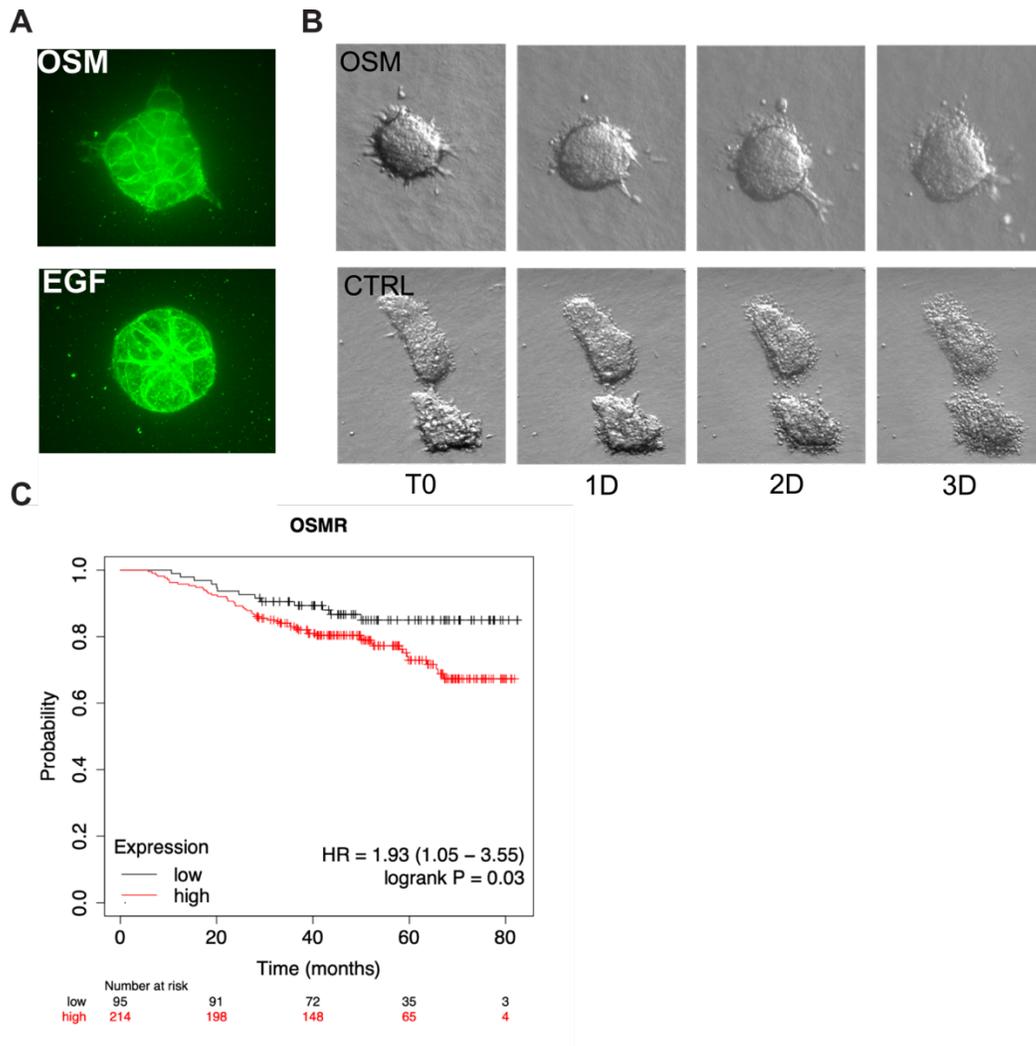
To address this, MCF10A cells were cultured in a three-dimensional (3D) matrix using Matrigel to better mimic *in vivo* tissue architecture. When embedded in Matrigel, MCF10A cells formed organoid structures that recapitulated aspects of mammary gland morphology, including lumen formation and polarized epithelial architecture [133]. Using this system, I treated MCF10A organoids with OSM and control conditions, such as epidermal growth factor (EGF). I observed that OSM treatment induced an invasive phenotype in these organoids, which was imaged using confocal microscopy (**Figure 5-1A**). These results suggest that OSM induces phenotypic changes in 3D organoid systems consistent with CCM observed in 2D culture models, providing further evidence that this phenotype is relevant in more complex tissue contexts.

Thus far, this dissertation has focused on normal mammary epithelial tissues; however, CCM also plays a critical role in cancer metastasis [100]. Previous studies in the literature, such as those from Andy Ewald's group, have demonstrated that breast cancer metastases occurs more efficiently via CCM [96]. One seminal study found that in mouse models of tumors, metastatic sites were exclusively seeded by polyclonal lineages of breast cancer cells [286]. This implies that a collective process occurred during the metastatic cascade. Given the link between CCM and metastases, microenvironmental signaling activating this process could be important to understanding the factors that influence patient prognosis.

To investigate whether the OSM-induced CCM phenotype observed in normal mammary cells extends to cancerous tissues, I used a method to culture tumor-derived organoids from MMTV-PyMT mice. These mice develop mammary tumors driven by the polyoma middle T antigen (PyMT), a model that closely mimics human luminal breast cancer [414]. Tumors were enzymatically digested into small fragments and cultured within collagen matrices to recreate a 3D microenvironment. Previous work has shown that under certain conditions, tumor-derived organoids can invade the surrounding collagen matrix in either a single-cell or collective manner [94]. Using this system, I assessed whether OSM induces an invasive phenotype in tumor-derived organoids. Confocal microscopy revealed that OSM treatment resulted in invasive strands of collectively migrating cells, demonstrating that the CCM phenotype observed in normal epithelial organoids is also present in cancer-derived systems (**Figure 5-1B**).

Altogether, these findings suggest that the OSM-induced CCM phenotype exists in more complex model systems, including those of cancer origin. This is significant because previous studies and data collected

from patient breast cancer tumors have linked OSM receptor expression to breast cancer progression, metastasis, and poor patient prognosis, but the mechanistic basis for these associations remains unclear (**Figure 5-1C**). The results presented here provide a potential mechanistic link between OSM and breast cancer progression, centered around the initiation of CCM processes and the promotion of metastasis. However, it is important to note that these results are preliminary, and additional work is required to confirm and further characterize these findings.



**Figure 5-1: OSM-induced CCM in 3D model systems of normal and malignant mammary tissue**

**A)** MCF10A organoids were grown in Matrigel and collagen for 2 weeks then treated with OSM or EGF for 72 hours. OSM treatment induced invasion into the surrounding matrix. Organoids are imaged using a confocal microscope and a live-cell fluorescent die.

**B)** Organoids derived from MMTV-PyMT mammary tumors were embedded in collagen gel and treated with OSM or PBS. After 1 week of ligand treatment, OSM treated organoids exhibited a proliferative and invasive phenotype, while PBS treated organoids adopt a non-migratory and apoptotic state. Organoids were imaged using light sheet microscopy.

C) Survival analysis of basal breast cancer patients stratified for OSM receptor expression. Data and analysis were derived from Kaplan Meir Plotter [415].

### 5.1.1 Collaborative works not discussed

Beyond the core research projects presented in this dissertation, I have had the opportunity to collaborate on a range of interdisciplinary studies. The publications listed below reflect additional contributions that, although not covered in this work, have played a key role in shaping my development as a researcher

In review or preparation:

- Copperman J, **McLean I**, Heiser L, Gray J, Zuckerman D. Single-cell morphodynamical trajectories enable prediction of gene expression accompanying cell state change. <https://www.biorxiv.org/content/10.1101/2024.01.18.576248v2>. (In Review)
- Johnson, J. A. I., Stein-O'Brien, G. L., Booth, M., Heiland, R., Kurtoglu, F., Bergman, D. R., Bucher, E., Deshpande, A., Forjaz, A., Getz, M., Godet, I., Lyman, M., Metzcar, J., Mitchell, J., Raddatz, A., Rocha, H., Solorzano, J., Sundus, A., Wang, Y., **McLean I**, ... Macklin, P., et al. Digitize your biology! Modeling multicellular systems through interpretable cell behavior. <https://www.biorxiv.org/content/10.1101/2023.09.17.557982v3>. (Accepted with Revisions: Cell)
- Eric B. Berens, Sokchea Khou, Elaine Huang, Amber Hoffman, Briana Johnson, Nell Kirchberger, Sam Sivagnanam, Nicholas Calistri, Daniel Derrick, Tiera Liby, **Ian McLean**, Aryn A. Alanizi, E. Shelley Hwang, Pepper Schedin, Hugo Gonzalez, Zena Werb (deceased), Laura M. Heiser, Lisa M. Coussens. "Neoplastic immune mimicry is a generalizable phenomenon in breast cancer and epithelial CD69 enables early tumor progression". (In Preparation)

## 5.2 Biological conclusions

Understanding how microenvironmental signals influence cell behavior is critical for unraveling the complex dynamics of cellular processes in both normal and pathological states. These signals operate within a vast network of interactions, and their molecular consequences are challenging to decipher due to the sheer diversity and context dependency. However, every piece of new information adds clarity to this puzzle, potentially revealing pathways of therapeutic relevance in pathological processes. By systematically exploring these signals, we can better understand how cells respond and adapt to their environments, advancing both fundamental biology and applied medicine.

This dissertation investigates the interplay between microenvironmental signals and cellular behavior, emphasizing the role of combinatorial signaling, collective cell migration (CCM), and systematic data generation. By focusing on MCF10A mammary epithelial cells as a model system, we have provided novel insights into how distinct ligands and their combinations drive phenotypic and molecular changes. The following sections summarize the major contributions of this work and place them within the broader context of systems biology and current research trends.

### The LINCS Microenvironment Perturbation Dataset

Large-scale genomics initiatives, such as The Cancer Genome Atlas (TCGA), have revolutionized our understanding of cancer biology by providing comprehensive datasets that connect genetic alterations to clinical outcomes [158]. These resources have profoundly impacted both therapeutic strategies and the broader research community, enabling data-driven approaches to precision medicine. Inspired by the

success of such initiatives, the LINCS Microenvironment (ME) perturbation dataset represents a continuation of an effort to catalog cellular responses to extracellular signals. This dataset provides a systematic resource for exploring how diverse ligands drive transcriptional, proteomic, epigenomic, and phenotypic changes.

This resource serves as an invaluable tool for researchers aiming to understand how specific ligands influence cellular behavior on a molecular level. By systematically cataloging cellular responses across a wide range of signals, the dataset provides a foundation for exploring the complex network of interactions that regulate cell behavior. While the LINCS dataset has the potential to guide therapeutic development by pinpointing key pathways involved in cellular responses, its broader contribution lies in providing a model for large-scale, comprehensive efforts to catalog cellular responses in different contexts. Alongside this effort, other ongoing initiatives, such as DEPMAP, are also working toward similar goals, focusing on mapping the genetic and pharmacological dependencies of a variety of cancer models [368]. As more ligands, cell types, and physiological conditions are included in both the LINCS and other collaborative datasets, these efforts will strengthen our ability to understand how cells interpret and integrate microenvironmental signals.

### **Insights into Collective Cell Migration**

The molecular foundations of CCM have been extensively studied, with a growing body of literature elucidating key signaling pathways that govern this process. Studies have identified various growth factors, cytokines, and extracellular matrix components that influence CCM, including TGF $\beta$ , EGF, HGF, and CXCL12 [416], [417], [418], [419]. These molecules regulate several aspects of CCM, such as actin cytoskeleton rearrangement, intercellular junction remodeling, and polarity establishment. The Hippo, Wnt, and Notch pathways, among others, have been implicated in controlling the balance between collective and individual cell movement [420]. Recent work has also highlighted the role of immune signaling in guiding CCM, emphasizing how the inflammatory microenvironment influences both normal and pathological migration [421]. Despite these advances, many details regarding the regulation of CCM, especially in cancerous contexts, remain to be fully understood.

In this dissertation, we demonstrate that Oncostatin M (OSM) is a potent inducer of CCM in mammary epithelial cells, primarily through the activation of HIF1A. OSM-induced HIF1A signaling drives a transcriptional program that includes hypoxia-related genes and immune response pathways, such as complement activation. These findings align with previous research linking hypoxia-driven processes to invasive behaviors in cancer cells, while also introducing a novel connection between OSM signaling and immune responses in the context of migration.

The role of OSM in cancer progression is of particular interest. Growing evidence indicates that OSM, along with other cytokines such as IL-6 and TNF- $\alpha$ , plays a critical role in metastasis [422]. OSM has been associated with poor prognosis in various cancers, including breast and ovarian cancers, where its expression correlates with increased metastatic spread and fibroblast activation [297], [298]. OSM's ability to promote collective migration suggests that its influence on metastasis may involve complex interactions between tumor cells and the tumor microenvironment, making it a promising target for further investigation.

### **Advancing the Understanding of Combinatorial Signaling**

Cells are embedded in a dynamic microenvironment and constantly receive multiple signaling cues simultaneously. While canonical pathways activated by individual ligands have been widely studied, much less is known about how different signals interact and the emergent behaviors they induce. This dissertation explores the significance of combinatorial signaling, particularly through the interactions between EGF, OSM, and TGFB. These combinations resulted in distinct molecular and phenotypic outcomes that were not predictable based on the effects of each individual ligand alone.

The study of combinatorial signaling is increasingly viewed as a critical approach in both basic cellular biology and therapeutic development. In clinical settings, combination therapies, such as those used in chemotherapy, have been shown to enhance treatment efficacy and reduce the risk of resistance [423]. The success of these therapeutic strategies stems from the understanding that multiple molecular pathways must be targeted to effectively treat diseases like cancer. Similarly, in cellular signaling, the interplay between different ligands can significantly amplify or modulate specific cellular responses, potentially revealing novel pathways that are critical for developmental processes, immune responses, and disease progression [424]. The importance of combinations lies in the realization that cells do not respond to signals in isolation. Rather, they function in a milieu of concurrent signals, each influencing the others in a network of complex interactions. These combinatorial inputs shape cellular behavior and are integral to maintaining cellular homeostasis and adapting to environmental changes. By studying these interactions, we uncover new dimensions of signaling biology that extend beyond the well-characterized canonical pathways, shedding light on previously unexplored mechanisms of cellular communication.

Understanding ligand combinations and their integration is essential not only for providing insights into basic cellular processes but also for developing more effective and nuanced therapeutic strategies. Cells *in vivo* are constantly exposed to multiple signals from their microenvironment, and the responses they generate depend on how these signals are integrated. The interplay between ligands can alter the balance between cell survival, differentiation, and migration, with far-reaching implications for disease states such as cancer, fibrosis, and immune disorders. By unraveling the complexities of these interactions, we can gain a deeper understanding of how cells process signals in a collective manner and how this influences physiological outcomes.

### **Toward a Systems Biology Framework for Cellular Communication**

A recurring theme throughout this dissertation is the importance of systems biology approaches in addressing complex biological questions. By integrating high-throughput experimental data with computational analyses, systems biology offers a powerful framework for understanding how biological tissues behave.

Looking toward the future, systems biology aims to provide a holistic understanding of cellular behavior by in part integrating diverse data types. These integrated datasets offer unprecedented insights into how cells respond to environmental cues, revealing the complex interactions between individual signaling pathways and broader cellular networks. As we continue to develop computational tools and algorithms to link data across modalities, we move closer to creating predictive models that can guide experimental research and therapeutic interventions.

In our own work, we have employed these systems biology approaches to study cellular responses to various ligands and to explore the emergent behaviors that arise from combinatorial signaling. By

expanding this approach, we aim to provide a unified understanding of how cells, tissues, and organisms respond to their environments. In the future, systems biology will be critical for deciphering how complex signaling networks govern tissue development, homeostasis, as well as how they malfunction in disease states. Expanding experimental datasets, refining computational models, and integrating data across multiple scales—from molecular interactions to tissue dynamics—will be essential to realize this vision, ultimately providing new opportunities for the development of therapeutic strategies targeting complex signaling networks in disease.

### 5.3 Future directions

The work in Chapter 2 developed the LINCS Microenvironment (ME) perturbation dataset, cataloging the molecular and phenotypic responses of MCF10A cells to diverse ligands. This resource demonstrated the potential of systematic profiling to uncover relationships between microenvironmental signals and cellular phenotypes. Future efforts should expand this approach to deeply profile additional ligands and cellular contexts, including different cell types, such as stromal or immune cells, and physiological environments. Extending these datasets will enhance our ability to model tissue-level processes and predict outcomes of microenvironmental perturbations. Additionally, frameworks for analyzing and integrating these large datasets are needed to synthesize information and build comprehensive models of cellular communication and tissue dynamics.

Chapter 3 focused on the role of OSM in inducing collective cell migration (CCM) in MCF10A cells. This process was mediated by HIF1A, which regulated a transcriptional program consisting of hypoxia-related pathways and immune signaling. Notably, complement activation was linked to OSM-induced CCM, suggesting a novel interplay between immune pathways and epithelial cell migration. Future studies could explore whether OSM-induced CCM is conserved across other model systems, such as MMTV-PyMT organoids or xenograft models [219], [414]. Understanding the role of OSM in cancer metastasis is particularly important, given its association with poor clinical outcomes [297], [298]. OSM's effects on fibroblast activation, ECM degradation, and tumor progression warrant deeper investigation. The involvement of HIF1A in this process raises further questions about whether its activation is driven by general hypoxia-related mechanisms or specific cytokine-induced transcriptional changes. Additionally, the connection between complement signaling and CCM should be dissected, focusing on whether this pathway involves integrin regulation, signaling cascades, or other mechanisms. Finally, studying the contributions of cell junction proteins, such as P-cadherin, will provide insights into how cell adhesion supports CCM.

The work presented in Chapter 4 explored the combinatorial effects of the cytokines OSM and Transforming Growth Factor Beta 1 (TGFB), and the growth factor Epidermal Growth Factor (EGF) on MCF10A mammary epithelial cells. Analyses revealed a signaling axis of synergistic upregulation of genes involved in migration and chemotactic signaling, mediated by CREB activation and CXCR2 signaling. Future research could extend these findings by investigating combinatorial signaling effects in other cellular contexts and model systems. For example, additional ligand combinations or different cell types, such as immune or mesenchymal cells, may reveal broader principles of signal integration. Assays could be performed to probe signaling pathways upstream of CREB, like p38 kinase, to determine how the combination of EGF and OSM promotes synergistic activation [425]. Furthermore, general frameworks for understanding phenotypic synergy, commonly applied in drug studies, are not easily transferable to complex behaviors like migration or molecular synergy. Expanding experimental datasets to include more

ligands and diverse cellular contexts will provide a systematic understanding and quantification of synergy, ultimately leading to a better understanding of combinatorial effects on cellular behavior. The development of high-throughput approaches to study combinatorial effects systematically will enhance our understanding of how diverse signals influence cellular behavior. Applying computational frameworks to integrate transcriptional, proteomic, and phenotypic data will help elucidate the mechanisms underlying emergent behaviors induced by complex signaling interactions.

By advancing these lines of investigation, we can deepen our understanding of how microenvironmental signals orchestrate cellular behavior and identify new therapeutic targets for pathological processes, including cancer progression and metastasis.

## Citations

- [1] H. Macias and L. Hinck, "Mammary gland development," *WIREs Developmental Biology*, vol. 1, no. 4, pp. 533–557, Jul. 2012, doi: 10.1002/wdev.35.
- [2] S. K. Biswas, S. Banerjee, G. W. Baker, C.-Y. Kuo, and I. Chowdhury, "The Mammary Gland: Basic Structure and Molecular Signaling during Development," *Int J Mol Sci*, vol. 23, no. 7, p. 3883, Mar. 2022, doi: 10.3390/ijms23073883.
- [3] J. E. Visvader, "Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis," *Genes Dev*, vol. 23, no. 22, pp. 2563–2577, Nov. 2009, doi: 10.1101/gad.1849509.
- [4] J. M. Williams and C. W. Daniel, "Mammary ductal elongation: Differentiation of myoepithelium and basal lamina during branching morphogenesis," *Dev Biol*, vol. 97, no. 2, pp. 274–290, Jun. 1983, doi: 10.1016/0012-1606(83)90086-6.
- [5] N. Gjorevski and C. M. Nelson, "Integrated morphodynamic signalling of the mammary gland," *Nat Rev Mol Cell Biol*, vol. 12, no. 9, pp. 581–593, Sep. 2011, doi: 10.1038/nrm3168.
- [6] P. Rainard, F. B. Gilbert, and P. Germon, "Immune defenses of the mammary gland epithelium of dairy ruminants," *Front Immunol*, vol. 13, Oct. 2022, doi: 10.3389/fimmu.2022.1031785.
- [7] J. N. Lilla and Z. Werb, "Mast cells contribute to the stromal microenvironment in mammary gland branching morphogenesis," *Dev Biol*, vol. 337, no. 1, pp. 124–133, Jan. 2010, doi: 10.1016/j.ydbio.2009.10.021.
- [8] M. D. Sternlicht, "Key stages in mammary gland development: The cues that regulate ductal branching morphogenesis," *Breast Cancer Research*, vol. 8, no. 1, p. 201, Feb. 2005, doi: 10.1186/bcr1368.
- [9] T. Stein, N. Salomonis, and B. A. Gusterson, "Mammary Gland Involution as a Multi-step Process," *J Mammary Gland Biol Neoplasia*, vol. 12, no. 1, pp. 25–35, Mar. 2007, doi: 10.1007/s10911-007-9035-7.
- [10] J. L. Inman, C. Robertson, J. D. Mott, and M. J. Bissell, "Mammary gland development: cell fate specification, stem cells and the microenvironment," *Development*, vol. 142, no. 6, pp. 1028–1042, Mar. 2015, doi: 10.1242/dev.087643.

- [11] M. Wicker and K.-U. Wagner, "Cellular Plasticity in Mammary Gland Development and Breast Cancer," *Cancers (Basel)*, vol. 15, no. 23, p. 5605, Nov. 2023, doi: 10.3390/cancers15235605.
- [12] N. Harbeck *et al.*, "Breast cancer," *Nat Rev Dis Primers*, vol. 5, no. 1, p. 66, Sep. 2019, doi: 10.1038/s41572-019-0111-2.
- [13] E. Orrantia-Borunda, P. Anchondo-Nuñez, L. E. Acuña-Aguilar, F. O. Gómez-Valles, and C. A. Ramírez-Valdespino, "Subtypes of Breast Cancer," in *Breast Cancer*, Exon Publications, 2022, pp. 31–42. doi: 10.36255/exon-publications-breast-cancer-subtypes.
- [14] K. A. Akshata Desai, "Triple Negative Breast Cancer – An Overview," *Hereditary Genetics*, 2012, doi: 10.4172/2161-1041.S2-001.
- [15] E. Boakes, A. Woods, N. Johnson, and N. Kadoglou, "Breast Infection: A Review of Diagnosis and Management Practices," *Eur J Breast Health*, vol. 14, no. 3, pp. 136–143, Jul. 2018, doi: 10.5152/ejbh.2018.3871.
- [16] D. Nagarajan and S. McArdle, "Immune Landscape of Breast Cancers," *Biomedicines*, vol. 6, no. 1, p. 20, Feb. 2018, doi: 10.3390/biomedicines6010020.
- [17] K. Polyak and R. Kalluri, "The Role of the Microenvironment in Mammary Gland Development and Cancer," *Cold Spring Harb Perspect Biol*, vol. 2, no. 11, pp. a003244–a003244, Nov. 2010, doi: 10.1101/cshperspect.a003244.
- [18] M. C. Neville, T. B. McFadden, and I. Forsyth, "Hormonal regulation of mammary differentiation and milk secretion.," *J Mammary Gland Biol Neoplasia*, vol. 7, no. 1, pp. 49–66, 2002, doi: 10.1023/A:1015770423167.
- [19] E. A. Peterson *et al.*, "Amphiregulin Is a Critical Downstream Effector of Estrogen Signaling in ER $\alpha$ -Positive Breast Cancer," *Cancer Res*, vol. 75, no. 22, pp. 4830–4838, Nov. 2015, doi: 10.1158/0008-5472.CAN-15-0709.
- [20] N. F. Boyd *et al.*, "Evidence That Breast Tissue Stiffness Is Associated with Risk of Breast Cancer," *PLoS One*, vol. 9, no. 7, p. e100937, Jul. 2014, doi: 10.1371/journal.pone.0100937.
- [21] J. Luo *et al.*, "The oncogenic roles and clinical implications of YAP/TAZ in breast cancer," *Br J Cancer*, Feb. 2023, doi: 10.1038/s41416-023-02182-5.
- [22] D. N. Danforth, "The Role of Chronic Inflammation in the Development of Breast Cancer," *Cancers (Basel)*, vol. 13, no. 15, p. 3918, Aug. 2021, doi: 10.3390/cancers13153918.
- [23] J. N. Amens, G. Bahçecioglu, and P. Zorlutuna, "Immune System Effects on Breast Cancer," *Cell Mol Bioeng*, vol. 14, no. 4, pp. 279–292, Aug. 2021, doi: 10.1007/s12195-021-00679-8.
- [24] S. Satpathi, S. S. Gaurkar, A. Potdukhe, and M. B. Wanjari, "Unveiling the Role of Hormonal Imbalance in Breast Cancer Development: A Comprehensive Review," *Cureus*, Jul. 2023, doi: 10.7759/cureus.41737.
- [25] M. Ajmal, M. Khan, and K. Van Fossen, *Breast Fibroadenoma*. 2024.

- [26] D. Hu *et al.*, “Cancer-associated fibroblasts in breast cancer: Challenges and opportunities,” *Cancer Commun*, vol. 42, no. 5, pp. 401–434, May 2022, doi: 10.1002/cac2.12291.
- [27] Y. Zhang *et al.*, “Hypoxia in Breast Cancer—Scientific Translation to Therapeutic and Diagnostic Clinical Applications,” *Front Oncol*, vol. 11, Mar. 2021, doi: 10.3389/fonc.2021.652266.
- [28] Y. Jia *et al.*, “Exosome: emerging biomarker in breast cancer,” *Oncotarget*, vol. 8, no. 25, pp. 41717–41733, Jun. 2017, doi: 10.18632/oncotarget.16684.
- [29] Q. Feng *et al.*, “A class of extracellular vesicles from breast cancer cells activates VEGF receptors and tumour angiogenesis,” *Nat Commun*, vol. 8, no. 1, p. 14450, Feb. 2017, doi: 10.1038/ncomms14450.
- [30] X. Yuan *et al.*, “Breast cancer exosomes contribute to pre-metastatic niche formation and promote bone metastasis of tumor cells,” *Theranostics*, vol. 11, no. 3, pp. 1429–1445, 2021, doi: 10.7150/thno.45351.
- [31] H. Hashizume *et al.*, “Openings between defective endothelial cells explain tumor vessel leakiness,” *Am J Pathol*, vol. 156, no. 4, pp. 1363–80, Apr. 2000, doi: 10.1016/S0002-9440(10)65006-7.
- [32] C. O. Madu, S. Wang, C. O. Madu, and Y. Lu, “Angiogenesis in Breast Cancer Progression, Diagnosis, and Treatment,” *J Cancer*, vol. 11, no. 15, pp. 4474–4494, 2020, doi: 10.7150/jca.44313.
- [33] B. Ruffell and L. M. Coussens, “Macrophages and Therapeutic Resistance in Cancer,” *Cancer Cell*, vol. 27, no. 4, pp. 462–472, Apr. 2015, doi: 10.1016/j.ccell.2015.02.015.
- [34] J. Park *et al.*, “Cancer cells induce metastasis-supporting neutrophil extracellular DNA traps,” *Sci Transl Med*, vol. 8, no. 361, p. 361ra138, Oct. 2016, doi: 10.1126/scitranslmed.aag1711.
- [35] P. Wu *et al.*, “Adaptive Mechanisms of Tumor Therapy Resistance Driven by Tumor Microenvironment,” *Front Cell Dev Biol*, vol. 9, Mar. 2021, doi: 10.3389/fcell.2021.641469.
- [36] V. Debien *et al.*, “Immunotherapy in breast cancer: an overview of current strategies and perspectives,” *NPJ Breast Cancer*, vol. 9, no. 1, p. 7, Feb. 2023, doi: 10.1038/s41523-023-00508-3.
- [37] O. Oey, A. F. Sunjaya, Y. Khan, and A. Redfern, “Stromal inflammation, fibrosis and cancer: An old intuition with promising potential,” *World J Clin Oncol*, vol. 14, no. 7, pp. 230–246, Jul. 2023, doi: 10.5306/wjco.v14.i7.230.
- [38] J. Liu, X. Wang, Y. Deng, X. Yu, H. Wang, and Z. Li, “Research Progress on the Role of Regulatory T Cell in Tumor Microenvironment in the Treatment of Breast Cancer,” *Front Oncol*, vol. 11, Nov. 2021, doi: 10.3389/fonc.2021.766248.
- [39] N. McNamee, M. Catalano, A. Mukhopadhyaya, and L. O’Driscoll, “An extensive study of potential inhibitors of extracellular vesicles release in triple-negative breast cancer,” *BMC Cancer*, vol. 23, no. 1, p. 654, Jul. 2023, doi: 10.1186/s12885-023-11160-2.

- [40] H. J. Burstein *et al.*, “VEGF as a Marker for Outcome Among Advanced Breast Cancer Patients Receiving anti-VEGF Therapy with Bevacizumab and Vinorelbine Chemotherapy,” *Clinical Cancer Research*, vol. 14, no. 23, pp. 7871–7877, Dec. 2008, doi: 10.1158/1078-0432.CCR-08-0593.
- [41] B. Gordon and V. K. Gadi, “The Role of the Tumor Microenvironment in Developing Successful Therapeutic and Secondary Prophylactic Breast Cancer Vaccines,” *Vaccines (Basel)*, vol. 8, no. 3, p. 529, Sep. 2020, doi: 10.3390/vaccines8030529.
- [42] J. Su *et al.*, “Cell–cell communication: new insights and clinical implications,” *Signal Transduct Target Ther*, vol. 9, no. 1, p. 196, Aug. 2024, doi: 10.1038/s41392-024-01888-z.
- [43] F. M. Watt and H. Fujiwara, “Cell-Extracellular Matrix Interactions in Normal and Diseased Skin,” *Cold Spring Harb Perspect Biol*, vol. 3, no. 4, pp. a005124–a005124, Apr. 2011, doi: 10.1101/cshperspect.a005124.
- [44] M. W. Renshaw, L. S. Price, and M. A. Schwartz, “Focal Adhesion Kinase Mediates the Integrin Signaling Requirement for Growth Factor Activation of Map Kinase,” *J Cell Biol*, vol. 147, no. 3, pp. 611–618, Nov. 1999, doi: 10.1083/jcb.147.3.611.
- [45] J. L. Andrews, A. C. Kim, and J. R. Hens, “The role and function of cadherins in the mammary gland,” *Breast Cancer Res*, vol. 14, no. 1, p. 203, Feb. 2012, doi: 10.1186/bcr3065.
- [46] D.-A. D. Nguyen and M. C. Neville, “Tight Junction Regulation in the Mammary Gland,” *J Mammary Gland Biol Neoplasia*, vol. 3, no. 3, pp. 233–246, 1998, doi: 10.1023/A:1018707309361.
- [47] M. A. Garcia, W. J. Nelson, and N. Chavez, “Cell-Cell Junctions Organize Structural and Signaling Networks,” *Cold Spring Harb Perspect Biol*, vol. 10, no. 4, Apr. 2018, doi: 10.1101/cshperspect.a029181.
- [48] G. Maheshwari, H. S. Wiley, and D. A. Lauffenburger, “Autocrine epidermal growth factor signaling stimulates directionally persistent mammary epithelial cell migration,” *J Cell Biol*, vol. 155, no. 7, pp. 1123–1128, Dec. 2001, doi: 10.1083/jcb.200109060.
- [49] J. M. Rosen and K. Roarty, “Paracrine signaling in mammary gland development: what can we learn about intratumoral heterogeneity?,” *Breast Cancer Research*, vol. 16, no. 1, p. 202, Jan. 2014, doi: 10.1186/bcr3610.
- [50] M. A. Lemmon and J. Schlessinger, “Cell Signaling by Receptor Tyrosine Kinases,” *Cell*, vol. 141, no. 7, pp. 1117–1134, Jun. 2010, doi: 10.1016/j.cell.2010.06.011.
- [51] H. Lu *et al.*, “A breast cancer stem cell niche supported by juxtacrine signalling from monocytes and macrophages,” *Nat Cell Biol*, vol. 16, no. 11, pp. 1105–1117, Nov. 2014, doi: 10.1038/ncb3041.
- [52] A. Nair, P. Chauhan, B. Saha, and K. F. Kubatzky, “Conceptual Evolution of Cell Signaling,” *Int J Mol Sci*, vol. 20, no. 13, p. 3292, Jul. 2019, doi: 10.3390/ijms20133292.
- [53] C. J. Watson, C. H. Oliver, and W. T. Khaled, “Cytokine signalling in mammary gland development,” *J Reprod Immunol*, vol. 88, no. 2, pp. 124–129, Mar. 2011, doi: 10.1016/j.jri.2010.11.006.

- [54] P. C. HEINRICH, I. BEHRMANN, G. MÜLLER-NEWEN, F. SCHAPER, and L. GRAEVE, "Interleukin-6-type cytokine signalling through the gp130/Jak/STAT pathway," *Biochemical Journal*, vol. 334, no. 2, pp. 297–314, Sep. 1998, doi: 10.1042/bj3340297.
- [55] C. E. Hughes and R. J. B. Nibbs, "A guide to chemokines and their receptors," *FEBS J*, vol. 285, no. 16, pp. 2944–2971, Aug. 2018, doi: 10.1111/febs.14466.
- [56] D. F. Legler and M. Thelen, "New insights in chemokine signaling.," *F1000Res*, vol. 7, p. 95, 2018, doi: 10.12688/f1000research.13130.1.
- [57] F. Borellini and T. Oka, "Growth control and differentiation in mammary epithelial cells.," *Environ Health Perspect*, vol. 80, pp. 85–99, Mar. 1989, doi: 10.1289/ehp.898085.
- [58] R. A. Weinberg, *The Biology of Cancer*. W.W. Norton & Company, 2013. doi: 10.1201/9780429258794.
- [59] L. M. Arendt and C. Kuperwasser, "Form and Function: how Estrogen and Progesterone Regulate the Mammary Epithelial Hierarchy," *J Mammary Gland Biol Neoplasia*, vol. 20, no. 1–2, pp. 9–25, Jun. 2015, doi: 10.1007/s10911-015-9337-0.
- [60] M. J. Higgins and J. Baselga, "Targeted therapies for breast cancer," *Journal of Clinical Investigation*, vol. 121, no. 10, pp. 3797–3803, Oct. 2011, doi: 10.1172/JCI57152.
- [61] T. W. Hodo, M. T. P. de Aquino, A. Shimamoto, and A. Shanker, "Critical Neurotransmitters in the Neuroimmune Network," *Front Immunol*, vol. 11, Aug. 2020, doi: 10.3389/fimmu.2020.01869.
- [62] R. K. Singh, R. Tandon, S. G. Dastidar, and A. Ray, "A review on leukotrienes and their receptors with reference to asthma," *Journal of Asthma*, vol. 50, no. 9, pp. 922–931, Nov. 2013, doi: 10.3109/02770903.2013.823447.
- [63] P. G. Tiffen, N. Omidvar, N. Marquez-Almuina, D. Croston, C. J. Watson, and R. W. E. Clarkson, "A Dual Role for Oncostatin M Signaling in the Differentiation and Death of Mammary Epithelial Cells in Vivo," *Molecular Endocrinology*, vol. 22, no. 12, pp. 2677–2688, Dec. 2008, doi: 10.1210/me.2008-0097.
- [64] H. Moses and M. H. Barcellos-Hoff, "TGF- Biology in Mammary Development and Breast Cancer," *Cold Spring Harb Perspect Biol*, vol. 3, no. 1, pp. a003277–a003277, Jan. 2011, doi: 10.1101/cshperspect.a003277.
- [65] K. Plaut, "Role of Epidermal Growth Factor and Transforming Growth Factors in Mammary Development and Lactation," *J Dairy Sci*, vol. 76, no. 6, pp. 1526–1538, Jun. 1993, doi: 10.3168/jds.S0022-0302(93)77485-8.
- [66] A. Morato, E. Martignani, S. Miretti, M. Baratta, and P. Accornero, "External and internal EGFR-activating signals drive mammary epithelial cells proliferation and viability," *Mol Cell Endocrinol*, vol. 520, p. 111081, Jan. 2021, doi: 10.1016/j.mce.2020.111081.
- [67] R. S. Herbst, "Review of epidermal growth factor receptor biology," *Int J Radiat Oncol Biol Phys*, vol. 59, no. 2 SUPPL., pp. S21–S26, 2004, doi: 10.1016/j.ijrobp.2003.11.041.

- [68] J. O. Humtsoe and R. H. Kramer, "Differential epidermal growth factor receptor signaling regulates anchorage-independent growth by modulation of the PI3K/AKT pathway," *Oncogene*, vol. 29, no. 8, pp. 1214–1226, Feb. 2010, doi: 10.1038/onc.2009.419.
- [69] H.-H. Lee, Y.-N. Wang, and M.-C. Hung, "Non-canonical signaling mode of the epidermal growth factor receptor family," *Am J Cancer Res*, vol. 5, no. 10, pp. 2944–58, 2015.
- [70] A. Weiss and L. Attisano, "The TGFbeta superfamily signaling pathway," Jan. 2013. doi: 10.1002/wdev.86.
- [71] J. Massagué, "TGFβ signalling in context," *Nat Rev Mol Cell Biol*, vol. 13, no. 10, pp. 616–630, Oct. 2012, doi: 10.1038/nrm3434.
- [72] J. Xu, S. Lamouille, and R. Derynck, "TGF-β-induced epithelial to mesenchymal transition," *Cell Res*, vol. 19, no. 2, pp. 156–172, Feb. 2009, doi: 10.1038/cr.2009.5.
- [73] Y. E. Zhang, "Non-Smad pathways in TGF-β signaling," *Cell Res*, vol. 19, no. 1, pp. 128–139, Jan. 2009, doi: 10.1038/cr.2008.328.
- [74] G. Dey *et al.*, "Signaling network of Oncostatin M pathway," *J Cell Commun Signal*, vol. 7, no. 2, pp. 103–108, Jun. 2013, doi: 10.1007/s12079-012-0186-y.
- [75] M. Tanaka and A. Miyahima, "Onconstatin M, a multifunctional cytokine," in *Reviews of Physiology, Biochemistry and Pharmacology*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 39–52. doi: 10.1007/s10254-003-0013-1.
- [76] K. A. Schafer, "The Cell Cycle: A Review," *Vet Pathol*, vol. 35, no. 6, pp. 461–478, Nov. 1998, doi: 10.1177/030098589803500601.
- [77] H. K. Matthews, C. Bertoli, and R. A. M. de Bruin, "Cell cycle control in cancer," *Nat Rev Mol Cell Biol*, vol. 23, no. 1, pp. 74–88, Jan. 2022, doi: 10.1038/s41580-021-00404-3.
- [78] K. J. Barnum and M. J. O'Connell, "Cell Cycle Regulation by Checkpoints," 2014, pp. 29–40. doi: 10.1007/978-1-4939-0888-2\_2.
- [79] S. A. Ezhevsky, H. Nagahara, A. M. Vocero-Akbani, D. R. Gius, M. C. Wei, and S. F. Dowdy, "Hypo-phosphorylation of the retinoblastoma protein (pRb) by cyclin D:Cdk4/6 complexes results in active pRb," *Proceedings of the National Academy of Sciences*, vol. 94, no. 20, pp. 10699–10704, Sep. 1997, doi: 10.1073/pnas.94.20.10699.
- [80] K. Engeland, "Cell cycle regulation: p53-p21-RB signaling," *Cell Death Differ*, vol. 29, no. 5, pp. 946–960, May 2022, doi: 10.1038/s41418-022-00988-z.
- [81] D. Cannata *et al.*, "Elevated Circulating IGF-I Promotes Mammary Gland Development and Proliferation," *Endocrinology*, vol. 151, no. 12, pp. 5751–5761, Dec. 2010, doi: 10.1210/en.2010-0792.
- [82] K. Oda, Y. Matsuoka, A. Funahashi, and H. Kitano, "A comprehensive pathway map of epidermal growth factor receptor signaling," *Mol Syst Biol*, vol. 1, 2005, doi: 10.1038/msb4100014.

- [83] J. A. I. Johnson *et al.*, “Digitize your Biology! Modeling multicellular systems through interpretable cell behavior,” Sep. 17, 2023. doi: 10.1101/2023.09.17.557982.
- [84] S. Hauguel-Demouzon, P. Csermely, G. Zoppini, and C. R. Kahn, “Quantitative dissociation between EGF effects on *c - myc* and *c - fos* gene expression, DNA synthesis, and epidermal growth factor receptor tyrosine kinase activity,” *J Cell Physiol*, vol. 150, no. 1, pp. 180–187, Jan. 1992, doi: 10.1002/jcp.1041500124.
- [85] T. Mosmann, “Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays,” *J Immunol Methods*, vol. 65, no. 1–2, pp. 55–63, Dec. 1983, doi: 10.1016/0022-1759(83)90303-4.
- [86] G. A. Romar, T. S. Kupper, and S. J. Divito, “Research Techniques Made Simple: Techniques to Assess Cell Proliferation,” *Journal of Investigative Dermatology*, vol. 136, no. 1, pp. e1–e7, Jan. 2016, doi: 10.1016/j.jid.2015.11.020.
- [87] M. Maška *et al.*, “The Cell Tracking Challenge: 10 years of objective benchmarking,” *Nat Methods*, vol. 20, no. 7, pp. 1010–1020, Jul. 2023, doi: 10.1038/s41592-023-01879-y.
- [88] T. R. Geiger and D. S. Peeper, “Metastasis mechanisms,” *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, vol. 1796, no. 2, pp. 293–308, Dec. 2009, doi: 10.1016/j.bbcan.2009.07.006.
- [89] P. S. Kunwar, D. E. Siekhaus, and R. Lehmann, “In Vivo Migration: A Germ Cell Perspective,” *Annu Rev Cell Dev Biol*, vol. 22, no. 1, pp. 237–265, Nov. 2006, doi: 10.1146/annurev.cellbio.22.010305.103337.
- [90] T. Lämmermann and M. Sixt, “Mechanical modes of ‘amoeboid’ cell migration,” *Curr Opin Cell Biol*, vol. 21, no. 5, pp. 636–644, Oct. 2009, doi: 10.1016/j.ceb.2009.05.003.
- [91] M. Innocenti, “New insights into the formation and the function of lamellipodia and ruffles in mesenchymal cell migration.,” *Cell Adh Migr*, vol. 12, no. 5, pp. 401–416, 2018, doi: 10.1080/19336918.2018.1448352.
- [92] V. Gifford and Y. Itoh, “MT1-MMP-dependent cell migration: proteolytic and non-proteolytic mechanisms,” *Biochem Soc Trans*, vol. 47, no. 3, pp. 811–826, Jun. 2019, doi: 10.1042/BST20180363.
- [93] A. J. Ewald, A. Brenot, M. Duong, B. S. Chan, and Z. Werb, “Collective Epithelial Migration and Cell Rearrangements Drive Mammary Branching Morphogenesis,” *Dev Cell*, vol. 14, no. 4, pp. 570–581, Apr. 2008, doi: 10.1016/j.devcel.2008.03.003.
- [94] K. J. Cheung, E. Gabrielson, Z. Werb, and A. J. Ewald, “Collective Invasion in Breast Cancer Requires a Conserved Basal Epithelial Program,” *Cell*, vol. 155, no. 7, pp. 1639–1651, Dec. 2013, doi: 10.1016/j.cell.2013.11.029.
- [95] N. V. Krakhmal, M. V. Zavyalova, E. V. Denisov, S. V. Vtorushin, and V. M. Perelmuter, “Cancer Invasion: Patterns and Mechanisms.,” *Acta Naturae*, vol. 7, no. 2, pp. 17–28, 2015.

- [96] A. Yamamoto, A. E. Doak, and K. J. Cheung, "Orchestration of Collective Migration and Metastasis by Tumor Cell Clusters," *Annual Review of Pathology: Mechanisms of Disease*, vol. 18, no. 1, pp. 231–256, Jan. 2023, doi: 10.1146/annurev-pathmechdis-031521-023557.
- [97] B. M. Szczerba *et al.*, "Neutrophils escort circulating tumour cells to enable cell cycle progression," *Nature*, vol. 566, no. 7745, pp. 553–557, Feb. 2019, doi: 10.1038/s41586-019-0915-y.
- [98] P. Hurtado, I. Martínez-Pena, and R. Piñeiro, "Dangerous Liaisons: Circulating Tumor Cells (CTCs) and Cancer-Associated Fibroblasts (CAFs)," *Cancers (Basel)*, vol. 12, no. 10, p. 2861, Oct. 2020, doi: 10.3390/cancers12102861.
- [99] Q. Liu, Q. Liao, and Y. Zhao, "Myeloid-derived suppressor cells (MDSC) facilitate distant metastasis of malignancies by shielding circulating tumor cells (CTC) from immune surveillance," *Med Hypotheses*, vol. 87, pp. 34–39, Feb. 2016, doi: 10.1016/j.mehy.2015.12.007.
- [100] K. J. Cheung and A. J. Ewald, "A collective route to metastasis: Seeding by tumor cell clusters," *Science (1979)*, vol. 352, no. 6282, pp. 167–169, Apr. 2016, doi: 10.1126/science.aaf6546.
- [101] N. Tapon, "Rho, Rac and Cdc42 GTPases regulate the organization of the actin cytoskeleton," *Curr Opin Cell Biol*, vol. 9, no. 1, pp. 86–92, 1997, doi: 10.1016/S0955-0674(97)80156-1.
- [102] N. Tapon, "Rho, Rac and Cdc42 GTPases regulate the organization of the actin cytoskeleton," *Curr Opin Cell Biol*, vol. 9, no. 1, pp. 86–92, 1997, doi: 10.1016/S0955-0674(97)80156-1.
- [103] K. Burridge and K. Wennerberg, "Rho and Rac Take Center Stage," *Cell*, vol. 116, no. 2, pp. 167–179, Jan. 2004, doi: 10.1016/S0092-8674(04)00003-0.
- [104] O. M. Yu and J. H. Brown, "G Protein-Coupled Receptor and RhoA-Stimulated Transcriptional Responses: Links to Inflammation, Differentiation, and Cell Proliferation.," *Mol Pharmacol*, vol. 88, no. 1, pp. 171–80, Jul. 2015, doi: 10.1124/mol.115.097857.
- [105] J. Z. Kechagia, J. Ivaska, and P. Roca-Cusachs, "Integrins as biomechanical sensors of the microenvironment," *Nat Rev Mol Cell Biol*, vol. 20, no. 8, pp. 457–473, Aug. 2019, doi: 10.1038/s41580-019-0134-2.
- [106] Q. Cao, R. T. Mertens, K. N. Sivanathan, X. Cai, and P. Xiao, "Macrophage orchestration of epithelial and stromal cell homeostasis in the intestine," *J Leukoc Biol*, vol. 112, no. 2, pp. 313–331, Jul. 2022, doi: 10.1002/JLB.3RU0322-176R.
- [107] F. Guo, Y. Wang, J. Liu, S. C. Mok, F. Xue, and W. Zhang, "CXCL12/CXCR4: a symbiotic bridge linking cancer cells and their stromal neighbors in oncogenic communication networks," *Oncogene*, vol. 35, no. 7, pp. 816–826, Feb. 2016, doi: 10.1038/onc.2015.139.
- [108] H. Ha, B. Debnath, and N. Neamati, "Role of the CXCL8-CXCR1/2 Axis in Cancer and Inflammatory Diseases.," *Theranostics*, vol. 7, no. 6, pp. 1543–1588, 2017, doi: 10.7150/thno.15625.
- [109] K. Campbell, "Contribution of epithelial-mesenchymal transitions to organogenesis and cancer metastasis," *Curr Opin Cell Biol*, vol. 55, pp. 30–35, Dec. 2018, doi: 10.1016/j.ceb.2018.06.008.

- [110] T. A. Leathers and C. D. Rogers, "Time to go: neural crest cell epithelial-to-mesenchymal transition," *Development*, vol. 149, no. 15, Aug. 2022, doi: 10.1242/dev.200712.
- [111] Y. Nakaya and G. Sheng, "Epithelial to mesenchymal transition during gastrulation: An embryological view," *Dev Growth Differ*, vol. 50, no. 9, pp. 755–766, Dec. 2008, doi: 10.1111/j.1440-169X.2008.01070.x.
- [112] S. Lamouille, J. Xu, and R. Derynck, "Molecular mechanisms of epithelial–mesenchymal transition," *Nat Rev Mol Cell Biol*, vol. 15, no. 3, pp. 178–196, Mar. 2014, doi: 10.1038/nrm3758.
- [113] C.-Y. Loh *et al.*, "The E-Cadherin and N-Cadherin Switch in Epithelial-to-Mesenchymal Transition: Signaling, Therapeutic Implications, and Challenges," *Cells*, vol. 8, no. 10, p. 1118, Sep. 2019, doi: 10.3390/cells8101118.
- [114] V. Kumar *et al.*, "The Role of Notch, Hedgehog, and Wnt Signaling Pathways in the Resistance of Tumors to Anticancer Therapies.," *Front Cell Dev Biol*, vol. 9, p. 650772, 2021, doi: 10.3389/fcell.2021.650772.
- [115] Z. Wang, Y. Li, D. Kong, and F. H. Sarkar, "The role of Notch signaling pathway in epithelial-mesenchymal transition (EMT) during development and tumor aggressiveness.," *Curr Drug Targets*, vol. 11, no. 6, pp. 745–51, Jun. 2010, doi: 10.2174/138945010791170860.
- [116] A. Moustakas and P. Heldin, "TGF $\beta$  and matrix-regulated epithelial to mesenchymal transition," *Biochimica et Biophysica Acta (BBA) - General Subjects*, vol. 1840, no. 8, pp. 2621–2634, Aug. 2014, doi: 10.1016/j.bbagen.2014.02.004.
- [117] J. Yang *et al.*, "Guidelines and definitions for research on epithelial–mesenchymal transition," *Nat Rev Mol Cell Biol*, vol. 21, no. 6, pp. 341–352, Jun. 2020, doi: 10.1038/s41580-020-0237-9.
- [118] K. Saxena, M. K. Jolly, and K. Balamurugan, "Hypoxia, partial EMT and collective migration: Emerging culprits in metastasis," *Transl Oncol*, vol. 13, no. 11, p. 100845, Nov. 2020, doi: 10.1016/j.tranon.2020.100845.
- [119] W. Zhang *et al.*, "HIF-1 $\alpha$  Promotes Epithelial-Mesenchymal Transition and Metastasis through Direct Regulation of ZEB1 in Colorectal Cancer," *PLoS One*, vol. 10, no. 6, p. e0129603, Jun. 2015, doi: 10.1371/journal.pone.0129603.
- [120] M. Singh *et al.*, "Targeting the Temporal Dynamics of Hypoxia-Induced Tumor-Secreted Factors Halts Tumor Migration," *Cancer Res*, vol. 79, no. 11, pp. 2962–2977, Jun. 2019, doi: 10.1158/0008-5472.CAN-18-3151.
- [121] Z. Cao, T. Livas, and N. Kyprianou, "Anoikis and EMT: Lethal 'Liaisons' during Cancer Progression," *Crit Rev Oncog*, vol. 21, no. 3–4, pp. 155–168, 2016, doi: 10.1615/CritRevOncog.2016016955.
- [122] O. H. Ocaña *et al.*, "Metastatic Colonization Requires the Repression of the Epithelial-Mesenchymal Transition Inducer Prrx1," *Cancer Cell*, vol. 22, no. 6, pp. 709–724, Dec. 2012, doi: 10.1016/j.ccr.2012.10.012.

- [123] Q. Zhao, B. Li, Q. Gao, Y. Luo, and L. Ming, "Prognostic value of epithelial-mesenchymal transition circulating tumor cells in female breast cancer: A meta-analysis.," *Front Oncol*, vol. 12, p. 1024783, 2022, doi: 10.3389/fonc.2022.1024783.
- [124] A.-J. Twigger *et al.*, "Transcriptional changes in the mammary gland during lactation revealed by single cell sequencing of cells from human milk," *Nat Commun*, vol. 13, no. 1, p. 562, Jan. 2022, doi: 10.1038/s41467-021-27895-0.
- [125] P. F. Slepicka, A. V. H. Somasundara, and C. O. Dos Santos, "The molecular basis of mammary gland development and epithelial differentiation.," *Semin Cell Dev Biol*, vol. 114, pp. 93–112, Jun. 2021, doi: 10.1016/j.semcdb.2020.09.014.
- [126] W. W. Pang and P. E. Hartmann, "Initiation of Human Lactation: Secretory Differentiation and Secretory Activation," *J Mammary Gland Biol Neoplasia*, vol. 12, no. 4, pp. 211–221, Nov. 2007, doi: 10.1007/s10911-007-9054-4.
- [127] H. Moses and M. H. Barcellos-Hoff, "TGF- Biology in Mammary Development and Breast Cancer," *Cold Spring Harb Perspect Biol*, vol. 3, no. 1, pp. a003277–a003277, Jan. 2011, doi: 10.1101/cshperspect.a003277.
- [128] C. Sirinian *et al.*, "Cellular Senescence in Normal Mammary Gland and Breast Cancer. Implications for Cancer Therapy.," *Genes (Basel)*, vol. 13, no. 6, Jun. 2022, doi: 10.3390/genes13060994.
- [129] X. Yang, H. Wang, and B. Jiao, "Mammary gland stem cells and their application in breast cancer.," *Oncotarget*, vol. 8, no. 6, pp. 10675–10691, Feb. 2017, doi: 10.18632/oncotarget.12893.
- [130] Q. H. Nguyen *et al.*, "Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity," *Nat Commun*, vol. 9, no. 1, p. 2028, May 2018, doi: 10.1038/s41467-018-04334-1.
- [131] C.-M. Clausson, I. Grundberg, I. Weibrecht, M. Nilsson, and O. Söderberg, "Methods for analysis of the cancer microenvironment and their potential for disease prediction, monitoring and personalized treatments.," *EPMA J*, vol. 3, no. 1, p. 7, Mar. 2012, doi: 10.1007/s13167-012-0140-3.
- [132] T. M. M. S. R. W. W. D. P. Jr. , R. B. C. M. M. J. R. R. J. P. R. F. J. and S. C. B. Herbert D. Soule, "Isolation and Characterization of a Spontaneously Immortalized Human Breast Epithelial Cell Line, MCF-101," 1990. [Online]. Available: <http://aacrjournals.org/cancerres/article-pdf/50/18/6075/2441394/cr0500186075.pdf>
- [133] J. Debnath, S. K. Muthuswamy, and J. S. Brugge, "Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures," *Methods*, vol. 30, no. 3, pp. 256–268, Jul. 2003, doi: 10.1016/S1046-2023(03)00032-X.
- [134] Y. Qu *et al.*, "Evaluation of MCF10A as a Reliable Model for Normal Human Mammary Epithelial Cells," *PLoS One*, vol. 10, no. 7, p. e0131285, Jul. 2015, doi: 10.1371/journal.pone.0131285.
- [135] K. M. Imbalzano, I. Tatarkova, A. N. Imbalzano, and J. A. Nickerson, "Increasingly transformed MCF-10A cells have a progressively tumor-like phenotype in three-dimensional basement membrane culture," *Cancer Cell Int*, vol. 9, no. 1, p. 7, 2009, doi: 10.1186/1475-2867-9-7.

- [136] H. Katsuno-Kambe, J. L. Teo, R. J. Ju, J. Hudson, S. J. Stehbens, and A. S. Yap, "Collagen polarization promotes epithelial elongation by stimulating locoregional cell proliferation.," *Elife*, vol. 10, Oct. 2021, doi: 10.7554/eLife.67915.
- [137] R. S. Stowers *et al.*, "Extracellular Matrix Stiffening Induces a Malignant Phenotypic Transition in Breast Epithelial Cells.," *Cell Mol Bioeng*, vol. 10, no. 1, pp. 114–123, Feb. 2017, doi: 10.1007/s12195-016-0468-1.
- [138] S. C. Mohan, T.-Y. Lee, A. E. Giuliano, and X. Cui, "Current Status of Breast Organoid Models," *Front Bioeng Biotechnol*, vol. 9, Nov. 2021, doi: 10.3389/fbioe.2021.745943.
- [139] L. Yuan, "A dynamic culture system that models the intricacy of the mammary gland," *Nat Rev Mol Cell Biol*, vol. 25, no. 4, pp. 248–248, Apr. 2024, doi: 10.1038/s41580-024-00704-4.
- [140] N. Zhao *et al.*, "Morphological screening of mesenchymal mammary tumor organoids to identify drugs that reverse epithelial-mesenchymal transition," *Nat Commun*, vol. 12, no. 1, p. 4262, Jul. 2021, doi: 10.1038/s41467-021-24545-3.
- [141] K. M. McAndrews, K. K. Mahadevan, and R. Kalluri, "Mouse Models to Evaluate the Functional Role of the Tumor Microenvironment in Cancer Progression and Therapy Responses.," *Cold Spring Harb Perspect Med*, vol. 14, no. 7, Jul. 2024, doi: 10.1101/cshperspect.a041411.
- [142] F. Löönd *et al.*, "Distinct contributions of partial and full EMT to breast cancer malignancy," *Dev Cell*, vol. 56, no. 23, pp. 3203–3221.e11, Dec. 2021, doi: 10.1016/j.devcel.2021.11.006.
- [143] K. L. Betterman *et al.*, "Remodeling of the Lymphatic Vasculature during Mouse Mammary Gland Morphogenesis Is Mediated via Epithelial-Derived Lymphangiogenic Stimuli," *Am J Pathol*, vol. 181, no. 6, pp. 2225–2238, Dec. 2012, doi: 10.1016/j.ajpath.2012.08.035.
- [144] B. Lloyd-Lewis *et al.*, "Imaging the mammary gland and mammary tumours in 3D: optical tissue clearing and immunofluorescence methods," *Breast Cancer Research*, vol. 18, no. 1, p. 127, Dec. 2016, doi: 10.1186/s13058-016-0754-9.
- [145] A. D. Elliott, "Confocal Microscopy: Principles and Modern Practices.," *Curr Protoc Cytom*, vol. 92, no. 1, p. e68, Mar. 2020, doi: 10.1002/cpcy.68.
- [146] G. A. Smolen *et al.*, "A genome-wide RNAi screen identifies multiple RSK-dependent regulators of cell migration.," *Genes Dev*, vol. 24, no. 23, pp. 2654–65, Dec. 2010, doi: 10.1101/gad.1989110.
- [147] S. M. Gross *et al.*, "A multi-omic analysis of MCF10A cells provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses," *Commun Biol*, vol. 5, no. 1, Dec. 2022, doi: 10.1038/s42003-022-03975-9.
- [148] M. A. Schwartz, "Super-Resolution Microscopy: A New Dimension in Focal Adhesions," *Current Biology*, vol. 21, no. 3, pp. R115–R116, Feb. 2011, doi: 10.1016/j.cub.2010.12.025.
- [149] M. D. Johnson and S. C. Mueller, "Three dimensional multiphoton imaging of fresh and whole mount developing mouse mammary glands.," *BMC Cancer*, vol. 13, p. 373, Aug. 2013, doi: 10.1186/1471-2407-13-373.

- [150] J. Eng, G. Thibault, S.-W. Luoh, J. W. Gray, Y. H. Chang, and K. Chin, "Cyclic Multiplexed-Immunofluorescence (cmIF), a Highly Multiplexed Method for Single-Cell Analysis," 2020, pp. 521–562. doi: 10.1007/978-1-4939-9773-2\_24.
- [151] K. E. G. Magnusson, J. Jalden, P. M. Gilbert, and H. M. Blau, "Global Linking of Cell Tracks Using the Viterbi Algorithm," *IEEE Trans Med Imaging*, vol. 34, no. 4, pp. 911–929, Apr. 2015, doi: 10.1109/TMI.2014.2370951.
- [152] M. Pachitariu and C. Stringer, "Cellpose 2.0: how to train your own model," *Nat Methods*, vol. 19, no. 12, pp. 1634–1641, Dec. 2022, doi: 10.1038/s41592-022-01663-4.
- [153] J. J. Provost, "Principles and techniques of biochemistry and molecular biology (6th Ed.): Wilson, Keith, and Walker, John," *Biochemistry and Molecular Biology Education*, vol. 33, no. 5, pp. 379–380, Sep. 2005, doi: 10.1002/bmb.2005.49403305379.
- [154] M. Morikawa, D. Koinuma, K. Miyazono, and C.-H. Heldin, "Genome-wide mechanisms of Smad binding," *Oncogene*, vol. 32, no. 13, pp. 1609–1615, Mar. 2013, doi: 10.1038/onc.2012.191.
- [155] H. Wang *et al.*, "Development of small molecule inhibitors targeting TGF- $\beta$  ligand and receptor: Structures, mechanism, preclinical studies and clinical usage," *Eur J Med Chem*, vol. 191, p. 112154, Apr. 2020, doi: 10.1016/j.ejmech.2020.112154.
- [156] J. W. Harper and E. J. Bennett, "Proteome complexity and the forces that drive proteome imbalance," *Nature*, vol. 537, no. 7620, pp. 328–338, Sep. 2016, doi: 10.1038/nature19947.
- [157] V. Blay, B. Tolani, S. P. Ho, and M. R. Arkin, "High-Throughput Screening: today's biochemical and cell-based approaches," *Drug Discov Today*, vol. 25, no. 10, pp. 1807–1821, Oct. 2020, doi: 10.1016/j.drudis.2020.07.024.
- [158] D. C. Koboldt, K. M. Steinberg, D. E. Larson, R. K. Wilson, and E. R. Mardis, "The next-generation sequencing revolution and its impact on genomics," *Cell*, vol. 155, no. 1, pp. 27–38, Sep. 2013, doi: 10.1016/j.cell.2013.09.006.
- [159] A. Kulkarni, A. G. Anderson, D. P. Merullo, and G. Konopka, "Beyond bulk: a review of single cell transcriptomics methodologies and applications," *Curr Opin Biotechnol*, vol. 58, pp. 129–136, Aug. 2019, doi: 10.1016/j.copbio.2019.03.001.
- [160] S. Z. Wu *et al.*, "A single-cell and spatially resolved atlas of human breast cancers," *Nat Genet*, vol. 53, no. 9, pp. 1334–1347, Sep. 2021, doi: 10.1038/s41588-021-00911-1.
- [161] Cancer Genome Atlas Research Network *et al.*, "The Cancer Genome Atlas Pan-Cancer analysis project," *Nat Genet*, vol. 45, no. 10, pp. 1113–20, Oct. 2013, doi: 10.1038/ng.2764.
- [162] T. S. Andrews, V. Y. Kiselev, D. McCarthy, and M. Hemberg, "Tutorial: guidelines for the computational analysis of single-cell RNA sequencing data," *Nat Protoc*, vol. 16, no. 1, pp. 1–9, Jan. 2021, doi: 10.1038/s41596-020-00409-w.
- [163] L. Van Der Maaten and G. Hinton, "Visualizing Data using t-SNE," 2008.
- [164] V. A. Traag, L. Waltman, and N. J. van Eck, "From Louvain to Leiden: guaranteeing well-connected communities," *Sci Rep*, vol. 9, no. 1, p. 5233, Mar. 2019, doi: 10.1038/s41598-019-41695-z.

- [165] Y. Hao *et al.*, “Dictionary learning for integrative, multimodal and scalable single-cell analysis,” *Nat Biotechnol*, vol. 42, no. 2, pp. 293–304, Feb. 2024, doi: 10.1038/s41587-023-01767-y.
- [166] D. Lähnemann *et al.*, “Eleven grand challenges in single-cell data science,” *Genome Biol*, vol. 21, no. 1, p. 31, Feb. 2020, doi: 10.1186/s13059-020-1926-6.
- [167] F. C. Grandi, H. Modi, L. Kampman, and M. R. Corces, “Chromatin accessibility profiling by ATAC-seq,” *Nat Protoc*, vol. 17, no. 6, pp. 1518–1552, Jun. 2022, doi: 10.1038/s41596-022-00692-9.
- [168] D. A. Cusanovich *et al.*, “A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility,” *Cell*, vol. 174, no. 5, pp. 1309–1324.e18, Aug. 2018, doi: 10.1016/j.cell.2018.06.052.
- [169] S. R. Shuken, “An Introduction to Mass Spectrometry-Based Proteomics,” *J Proteome Res*, vol. 22, no. 7, pp. 2151–2171, Jul. 2023, doi: 10.1021/acs.jproteome.2c00838.
- [170] G. Jiang *et al.*, “Single-cell transcriptomics reveal the heterogeneity and dynamic of cancer stem-like cells during breast tumor progression,” *Cell Death Dis*, vol. 12, no. 11, p. 979, Oct. 2021, doi: 10.1038/s41419-021-04261-y.
- [171] A. Spivey, “Systems Biology: The Big Picture,” *Environ Health Perspect*, vol. 112, no. 16, Nov. 2004, doi: 10.1289/ehp.112-a938.
- [172] P. Kohl, E. J. Crampin, T. A. Quinn, and D. Noble, “Systems Biology: An Approach,” *Clin Pharmacol Ther*, vol. 88, no. 1, pp. 25–33, Jul. 2010, doi: 10.1038/clpt.2010.92.
- [173] M. Zitnik *et al.*, “Current and future directions in network biology,” *Bioinformatics Advances*, vol. 4, no. 1, Jan. 2024, doi: 10.1093/bioadv/vbae099.
- [174] N. Safari-Alighiarloo, M. Taghizadeh, M. Rezaei-Tavirani, B. Goliaei, and A. A. Peyvandi, “Protein-protein interaction networks (PPI) and complex diseases,” *Gastroenterol Hepatol Bed Bench*, vol. 7, no. 1, pp. 17–31, 2014.
- [175] W. M. Yashar *et al.*, “Predicting transcription factor activity using prior biological information,” *iScience*, vol. 27, no. 3, p. 109124, Mar. 2024, doi: 10.1016/j.isci.2024.109124.
- [176] Ö. Babur *et al.*, “Causal interactions from proteomic profiles: Molecular data meet pathway knowledge,” *Patterns*, vol. 2, no. 6, p. 100257, Jun. 2021, doi: 10.1016/j.patter.2021.100257.
- [177] J. Yan, S. L. Risacher, L. Shen, and A. J. Saykin, “Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data,” *Brief Bioinform*, Jun. 2017, doi: 10.1093/bib/bbx066.
- [178] M. Bansal *et al.*, “A community computational challenge to predict the activity of pairs of compounds,” *Nat Biotechnol*, vol. 32, no. 12, pp. 1213–1222, Dec. 2014, doi: 10.1038/nbt.3052.
- [179] M. C. Berenbaum, “What is synergy?,” *Pharmacol Rev*, vol. 41, no. 2, pp. 93–141, Jun. 1989.
- [180] Q. Liu, X. Yin, L. R. Languino, and D. C. Altieri, “Evaluation of Drug Combination Effect Using a Bliss Independence Dose–Response Surface Model,” *Stat Biopharm Res*, vol. 10, no. 2, pp. 112–122, Apr. 2018, doi: 10.1080/19466315.2018.1437071.

- [181] S. LOEWE, "The problem of synergism and antagonism of combined drugs.," *Arzneimittelforschung*, vol. 3, no. 6, pp. 285–90, Jun. 1953.
- [182] J. EL Diaz *et al.*, "The transcriptomic response of cells to a drug combination is more than the sum of the responses to the monotherapies," *Elife*, vol. 9, Sep. 2020, doi: 10.7554/eLife.52707.
- [183] P. S. Ward and C. B. Thompson, "Signaling in control of cell growth and metabolism," *Cold Spring Harb Perspect Biol*, vol. 4, no. 7, pp. a006783–a006783, Jul. 2012, doi: 10.1101/cshperspect.a006783.
- [184] R. J. Duronio and Y. Xiong, "Signaling pathways that control cell proliferation," *Cold Spring Harb Perspect Biol*, vol. 5, no. 3, pp. a008904–a008904, Mar. 2013, doi: 10.1101/cshperspect.a008904.
- [185] P. Devreotes and A. R. Horwitz, "Signaling networks that regulate cell migration," *Cold Spring Harb Perspect Biol*, vol. 7, no. 8, pp. a005959–a005959, Aug. 2015, doi: 10.1101/cshperspect.a005959.
- [186] N. Perrimon, C. Pitsouli, and B.-Z. Shilo, "Signaling mechanisms controlling cell fate and embryonic patterning," *Cold Spring Harb Perspect Biol*, vol. 4, no. 8, pp. a005975–a005975, Aug. 2012, doi: 10.1101/cshperspect.a005975.
- [187] "PNAS Plus Significance Statements," *Proceedings of the National Academy of Sciences*, vol. 109, no. 52, pp. 21190–21190, 2012, doi: 10.1073/pnas.ss10952.
- [188] J. C. Costello *et al.*, "A community effort to assess and improve drug sensitivity prediction algorithms," *Nat Biotechnol*, vol. 32, no. 12, pp. 1202–1212, Dec. 2014, doi: 10.1038/nbt.2877.
- [189] L. Litichevskiy *et al.*, "A Library of Phosphoproteomic and Chromatin Signatures for Characterizing Cellular Responses to Drug Perturbations," *Cell Syst*, vol. 6, no. 4, pp. 424–443.e7, Apr. 2018, doi: 10.1016/j.cels.2018.03.012.
- [190] R. M. Neve *et al.*, "A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes," *Cancer Cell*, vol. 10, no. 6, pp. 515–527, Dec. 2006, doi: 10.1016/j.ccr.2006.10.008.
- [191] A. Tsherniak *et al.*, "Defining a Cancer Dependency Map," *Cell*, vol. 170, no. 3, pp. 564–576.e16, Jul. 2017, doi: 10.1016/j.cell.2017.06.010.
- [192] S. S. Watson *et al.*, "Microenvironment-Mediated Mechanisms of Resistance to HER2 Inhibitors Differ between HER2+ Breast Cancer Subtypes," *Cell Syst*, vol. 6, no. 3, pp. 329–342.e6, Mar. 2018, doi: 10.1016/j.cels.2018.02.001.
- [193] T. R. Wilson *et al.*, "Widespread potential for growth-factor-driven resistance to anticancer kinase inhibitors," *Nature*, vol. 487, no. 7408, pp. 505–509, Jul. 2012, doi: 10.1038/nature11249.
- [194] D. K. Morrison, "MAP kinase pathways," *Cold Spring Harb Perspect Biol*, vol. 4, no. 11, p. a011254, Nov. 2012, doi: 10.1101/cshperspect.a011254.
- [195] R. Nusse, "Wnt signaling," *Cold Spring Harb Perspect Biol*, vol. 4, no. 5, p. a011163, May 2012, doi: 10.1101/cshperspect.a011163.

- [196] C. J. David and J. Massagué, "Contextual determinants of TGF $\beta$  action in development, immunity and cancer," *Nat Rev Mol Cell Biol*, vol. 19, no. 7, pp. 419–435, Jul. 2018, doi: 10.1038/s41580-018-0007-0.
- [197] H. Liang *et al.*, "Abstract 397: Systematic functional annotation of somatic mutations in cancer," *Cancer Res*, vol. 78, no. 13\_Supplement, pp. 397–397, 2018, doi: 10.1158/1538-7445.am2018-397.
- [198] C. Bock *et al.*, "The Organoid Cell Atlas," *Nat Biotechnol*, vol. 39, no. 1, pp. 13–17, Jan. 2021, doi: 10.1038/s41587-020-00762-x.
- [199] J. Drost and H. Clevers, "Organoids in cancer research," *Nat Rev Cancer*, vol. 18, no. 7, pp. 407–418, 2018, doi: 10.1038/s41568-018-0007-6.
- [200] L. F. Sullivan, "Rewiring the Drosophila Brain With Genetic Manipulations in Neural Lineages," *Front Mol Neurosci*, vol. 12, p. 82, Apr. 2019, doi: 10.3389/fnmol.2019.00082.
- [201] H. E. Kinser and Z. Pincus, "High-throughput screening in the *C. elegans* nervous system," *Mol Cell Neurosci*, vol. 80, pp. 192–197, Apr. 2017, doi: 10.1016/j.mcn.2016.06.001.
- [202] J. Srinivasan *et al.*, "A modular library of small molecule signals regulates social behaviors in *Caenorhabditis elegans*," *PLoS Biol*, vol. 10, no. 1, pp. e1001237–e1001237, Jan. 2012, doi: 10.1371/journal.pbio.1001237.
- [203] K. Kersten, K. E. de Visser, M. H. van Miltenburg, and J. Jonkers, "Genetically engineered mouse models in oncology research and cancer medicine," *EMBO Mol Med*, vol. 9, no. 2, pp. 137–153, Feb. 2017, doi: 10.15252/emmm.201606857.
- [204] N. Yosef *et al.*, "Dynamic regulatory network controlling TH17 cell differentiation," *Nature*, vol. 496, no. 7446, pp. 461–468, Apr. 2013, doi: 10.1038/nature11981.
- [205] V. Jojic *et al.*, "Identification of transcriptional regulators in the mouse immune system," *Nat Immunol*, vol. 14, no. 6, pp. 633–643, Jun. 2013, doi: 10.1038/ni.2587.
- [206] W. Saelens, R. Cannoodt, and Y. Saeys, "A comprehensive evaluation of module detection methods for gene expression data," *Nat Commun*, vol. 9, no. 1, p. 1090, Mar. 2018, doi: 10.1038/s41467-018-03424-4.
- [207] L. Alsina *et al.*, "A narrow repertoire of transcriptional modules responsive to pyogenic bacteria is impaired in patients carrying loss-of-function mutations in MYD88 or IRAK4," *Nat Immunol*, vol. 15, no. 12, pp. 1134–1142, Dec. 2014, doi: 10.1038/ni.3028.
- [208] E. P. Consortium, "An integrated encyclopedia of DNA elements in the human genome," *Nature*, vol. 489, no. 7414, pp. 57–74, Sep. 2012, doi: 10.1038/nature11247.
- [209] C. G. A. R. Network *et al.*, "The Cancer Genome Atlas Pan-Cancer analysis project," *Nat Genet*, vol. 45, no. 10, pp. 1113–1120, Oct. 2013, doi: 10.1038/ng.2764.
- [210] Gte. Consortium *et al.*, "Genetic effects on gene expression across human tissues," *Nature*, vol. 550, no. 7675, pp. 204–213, Oct. 2017, doi: 10.1038/nature24277.

- [211] J. M. Smith and R. M. Conroy, "The NIH Common Fund Human Biomolecular Atlas Program (HuBMAP): Building a Framework for Mapping the Human Body," *The FASEB Journal*, vol. 32, no. S1, 2018, doi: 10.1096/fasebj.2018.32.1\_supplement.818.2.
- [212] M. Melani, K. J. Simpson, J. S. Brugge, and D. Montell, "Regulation of Cell Adhesion and Collective Cell Migration by Hindsight and Its Human Homolog RREB1," *Current Biology*, vol. 18, no. 7, pp. 532–537, 2008, doi: 10.1016/j.cub.2008.03.024.
- [213] S. E. Seton-Rogers *et al.*, "Cooperation of the ErbB2 receptor and transforming growth factor beta in induction of migration and invasion in mammary epithelial cells," *Proc Natl Acad Sci U S A*, vol. 101, no. 5, pp. 1257–1262, Feb. 2004, doi: 10.1073/pnas.0308090100.
- [214] J. Debnath, S. K. Muthuswamy, and J. S. Brugge, "Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures," *Methods*, vol. 30, no. 3, pp. 256–268, 2003, doi: 10.1016/s1046-2023(03)00032-x.
- [215] J. Debnath, K. R. Mills, N. L. Collins, M. J. Reginato, S. K. Muthuswamy, and J. S. Brugge, "The Role of Apoptosis in Creating and Maintaining Luminal Space within Normal and Oncogene-Expressing Mammary Acini," *Cell*, vol. 111, no. 1, pp. 29–40, 2002, doi: 10.1016/s0092-8674(02)01001-2.
- [216] S. Sampattavanich, B. Steiert, B. A. Kramer, B. M. Gyori, J. G. Albeck, and P. K. Sorger, "Encoding Growth Factor Identity in the Temporal Dynamics of FOXO3 under the Combinatorial Control of ERK and AKT Kinases," *Cell Syst*, vol. 6, no. 6, pp. 664–678.e9, Jun. 2018, doi: 10.1016/j.cels.2018.05.004.
- [217] M. Caldera *et al.*, "Mapping the perturbome network of cellular perturbations," *Nat Commun*, vol. 10, no. 1, p. 5140, Nov. 2019, doi: 10.1038/s41467-019-13058-9.
- [218] M. Niepel *et al.*, "A Multi-center Study on the Reproducibility of Drug-Response Assays in Mammalian Cell Lines," *Cell Syst*, vol. 9, no. 1, pp. 35–48.e5, Jul. 2019, doi: 10.1016/j.cels.2019.06.005.
- [219] R. Smith *et al.*, "Using Microarrays to Interrogate Microenvironmental Impact on Cellular Phenotypes in Cancer," *J Vis Exp*, no. 147, p. 10.3791/58957, May 2019, doi: 10.3791/58957.
- [220] C. Allan *et al.*, "OMERO: flexible, model-driven data management for experimental biology," *Nat Methods*, vol. 9, no. 3, pp. 245–253, Feb. 2012, doi: 10.1038/nmeth.1896.
- [221] C. McQuin *et al.*, "CellProfiler 3.0: Next-generation image processing for biology," *PLoS Biol*, vol. 16, no. 7, pp. e2005970–e2005970, Jul. 2018, doi: 10.1371/journal.pbio.2005970.
- [222] R. Molania, J. A. Gagnon-Bartsch, A. Dobrovic, and T. P. Speed, "A new normalization for Nanostring nCounter gene expression data," *Nucleic Acids Res*, vol. 47, no. 12, pp. 6073–6083, Jul. 2019, doi: 10.1093/nar/gkz433.
- [223] K. E. G. Magnusson, J. Jalden, P. M. Gilbert, and H. M. Blau, "Global Linking of Cell Tracks Using the Viterbi Algorithm," *IEEE Trans Med Imaging*, vol. 34, no. 4, pp. 911–929, Apr. 2015, doi: 10.1109/TMI.2014.2370951.

- [224] J. Schindelin *et al.*, “Fiji: an open-source platform for biological-image analysis,” *Nat Methods*, vol. 9, no. 7, pp. 676–682, Jul. 2012, doi: 10.1038/nmeth.2019.
- [225] S. Berg *et al.*, “ilastik: interactive machine learning for (bio)image analysis,” *Nat Methods*, vol. 16, no. 12, pp. 1226–1232, 2019, doi: 10.1038/s41592-019-0582-9.
- [226] D. Ebdon, *Statistics in Geography*, 2nd ed. Hoboken: Blackwell Publishing, 1985.
- [227] R. Tibes *et al.*, “Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells,” *Mol Cancer Ther*, vol. 5, no. 10, pp. 2512–2521, Oct. 2006, doi: 10.1158/1535-7163.MCT-06-0334.
- [228] M. I. Love, W. Huber, and S. Anders, “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2,” *Genome Biol*, vol. 15, no. 12, p. 550, Dec. 2014, doi: 10.1186/s13059-014-0550-8.
- [229] A. B. Keenan *et al.*, “ChEA3: transcription factor enrichment analysis by orthogonal omics integration,” *Nucleic Acids Res*, vol. 47, no. W1, pp. W212–W224, Jul. 2019, doi: 10.1093/nar/gkz446.
- [230] S. Hänzelmann, R. Castelo, and J. Guinney, “GSVA: gene set variation analysis for microarray and RNA-seq data,” *BMC Bioinformatics*, vol. 14, p. 7, Jan. 2013, doi: 10.1186/1471-2105-14-7.
- [231] M. R. Corces *et al.*, “An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues,” *Nat Methods*, vol. 14, no. 10, pp. 959–962, Oct. 2017, doi: 10.1038/nmeth.4396.
- [232] C. S. Ross-Innes *et al.*, “Differential oestrogen receptor binding is associated with clinical outcome in breast cancer,” *Nature*, vol. 481, no. 7381, pp. 389–393, Jan. 2012, doi: 10.1038/nature10730.
- [233] J. Ou *et al.*, “ATACseqQC: a Bioconductor package for post-alignment quality assessment of ATAC-seq data,” *BMC Genomics*, vol. 19, no. 1, p. 169, Mar. 2018, doi: 10.1186/s12864-018-4559-3.
- [234] J. D. Buenrostro, B. Wu, H. Y. Chang, and W. J. Greenleaf, “ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide,” *Curr Protoc Mol Biol*, vol. 109, pp. 21.29.1-21.29.9, Jan. 2015, doi: 10.1002/0471142727.mb2129s109.
- [235] A. N. Schep, B. Wu, J. D. Buenrostro, and W. J. Greenleaf, “chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data,” *Nat Methods*, vol. 14, no. 10, pp. 975–978, Oct. 2017, doi: 10.1038/nmeth.4401.
- [236] A. Khan *et al.*, “JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework,” *Nucleic Acids Res*, vol. 46, no. D1, pp. D260–D266, Jan. 2018, doi: 10.1093/nar/gkx1126.
- [237] A. L. Creech *et al.*, “Building the Connectivity Map of epigenetics: chromatin profiling by quantitative targeted mass spectrometry,” *Methods*, vol. 72, pp. 57–64, Jan. 2015, doi: 10.1016/j.ymeth.2014.10.033.

- [238] Z. Wang, A. Lachmann, A. B. Keenan, and A. Ma'ayan, "L1000FWD: fireworks visualization of drug-induced transcriptomic signatures," *Bioinformatics*, vol. 34, no. 12, pp. 2150–2152, Jun. 2018, doi: 10.1093/bioinformatics/bty060.
- [239] J.-R. Lin, M. Fallahi-Sichani, and P. K. Sorger, "Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method," *Nat Commun*, vol. 6, p. 8390, Sep. 2015, doi: 10.1038/ncomms9390.
- [240] T. Peng *et al.*, "A BaSiC tool for background and shading correction of optical microscopy images," *Nat Commun*, vol. 8, p. 14836, Jun. 2017, doi: 10.1038/ncomms14836.
- [241] D. M. Moskowitz and W. J. Greenleaf, "Nonparametric analysis of contributions to variance in genomics and epigenomics data," May 04, 2018. doi: 10.1101/314112.
- [242] G. J. Hunt, M. A. Dane, J. E. Korkola, L. M. Heiser, and J. A. Gagnon-Bartsch, "Automatic Transformation and Integration to Improve Visualization and Discovery of Latent Effects in Imaging Data," *J Comput Graph Stat*, vol. 29, no. 4, pp. 929–941, 2020, doi: 10.1080/10618600.2020.1741379.
- [243] M. Pierre-Jean, J.-F. Deleuze, E. Le Floch, and F. Mauger, "Clustering and variable selection evaluation of 13 unsupervised methods for multi-omics data integration," *Brief Bioinform*, vol. 21, no. 6, pp. 2011–2030, 2019, doi: 10.1093/bib/bbz138.
- [244] C. Meng, D. Helm, M. Frejno, and B. Kuster, "moCluster: Identifying Joint Patterns Across Multiple Omics Data Sets," *J Proteome Res*, vol. 15, no. 3, pp. 755–765, 2015, doi: 10.1021/acs.jproteome.5b00824.
- [245] P. Tatlow and S. R. Piccolo, "A cloud-based workflow to quantify transcript-expression levels in public cancer compendia," *Sci Rep*, vol. 6, no. 1, p. 39259, Dec. 2016, doi: 10.1038/srep39259.
- [246] D. M. Abd El-Rehim *et al.*, "Expression of luminal and basal cytokeratins in human breast carcinoma," *J Pathol*, vol. 203, no. 2, pp. 661–671, 2004, doi: 10.1002/path.1559.
- [247] A. Subramanian *et al.*, "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles," *Proc Natl Acad Sci U S A*, vol. 102, no. 43, pp. 15545–15550, Oct. 2005, doi: 10.1073/pnas.0506580102.
- [248] V. Stathias *et al.*, "Drug and disease signature integration identifies synergistic combinations in glioblastoma," *Nat Commun*, vol. 9, no. 1, p. 5315, Dec. 2018, doi: 10.1038/s41467-018-07659-z.
- [249] L. B. Ivashkiv, "IFN $\gamma$ : signalling, epigenetics and roles in immunity, metabolism, disease and cancer immunotherapy," *Nat Rev Immunol*, vol. 18, no. 9, pp. 545–558, Sep. 2018, doi: 10.1038/s41577-018-0029-z.
- [250] F. Belinky *et al.*, "PathCards: multi-source consolidation of human biological pathways," *Database (Oxford)*, vol. 2015, p. bav006, Feb. 2015, doi: 10.1093/database/bav006.
- [251] S. Carrasco Pro *et al.*, "Global landscape of mouse and human cytokine transcriptional regulation," *Nucleic Acids Res*, vol. 46, no. 18, pp. 9321–9337, Oct. 2018, doi: 10.1093/nar/gky787.

- [252] S. Mok *et al.*, “Inhibition of CSF-1 receptor improves the antitumor efficacy of adoptive cell transfer immunotherapy,” *Cancer Res*, vol. 74, no. 1, pp. 153–161, Jan. 2014, doi: 10.1158/0008-5472.CAN-13-1816.
- [253] Y. Zhu *et al.*, “CSF1/CSF1R blockade reprograms tumor-infiltrating macrophages and improves response to T-cell checkpoint immunotherapy in pancreatic cancer models,” *Cancer Res*, vol. 74, no. 18, pp. 5057–5069, Sep. 2014, doi: 10.1158/0008-5472.CAN-13-3723.
- [254] M. Zhao *et al.*, “Development of a recombinant human IL-15-sIL-15R $\alpha$ /Fc superagonist with improved half-life and its antitumor activity alone or in combination with PD-1 blockade in mouse model,” *Biomedicine & Pharmacotherapy*, vol. 112, p. 108677, 2019, doi: 10.1016/j.biopha.2019.108677.
- [255] P. Berraondo, I. Etxeberria, M. Ponz-Sarvisé, and I. Melero, “Revisiting Interleukin-12 as a Cancer Immunotherapy Agent,” *Clinical Cancer Research*, vol. 24, no. 12, pp. 2716–2718, 2018, doi: 10.1158/1078-0432.ccr-18-0381.
- [256] J. A. Flores-Toro *et al.*, “CCR2 inhibition reduces tumor myeloid cells and unmasks a checkpoint inhibitor effect to slow progression of resistant murine gliomas,” *Proc Natl Acad Sci U S A*, vol. 117, no. 2, pp. 1129–1138, Jan. 2020, doi: 10.1073/pnas.1910856117.
- [257] R. Singh, K. Nannuru, B. Sharma, and M. Varney, “Role of chemokine receptor CXCR2 expression in mammary tumor growth, angiogenesis and metastasis,” *J Carcinog*, vol. 10, no. 1, p. 40, 2011, doi: 10.4103/1477-3163.92308.
- [258] R. R. Starks, A. Biswas, A. Jain, and G. Tuteja, “Combined analysis of dissimilar promoter accessibility and gene expression profiles identifies tissue-specific genes and actively repressed networks,” *Epigenetics Chromatin*, vol. 12, no. 1, p. 16, Feb. 2019, doi: 10.1186/s13072-019-0260-2.
- [259] R. Tibshirani, G. Walther, and T. Hastie, “Estimating the Number of Clusters in a Data Set Via the Gap Statistic,” *J R Stat Soc Series B Stat Methodol*, vol. 63, no. 2, pp. 411–423, Jul. 2001, doi: 10.1111/1467-9868.00293.
- [260] F. M. R. Witjas, B. M. van den Berg, C. W. van den Berg, M. A. Engelse, and T. J. Rabelink, “Concise Review: The Endothelial Cell Extracellular Matrix Regulates Tissue Homeostasis and Repair,” *Stem Cells Transl Med*, vol. 8, no. 4, pp. 375–382, Apr. 2019, doi: 10.1002/sctm.18-0155.
- [261] J. E. Wagenseil and R. P. Mecham, “Vascular extracellular matrix and arterial mechanics,” *Physiol Rev*, vol. 89, no. 3, pp. 957–989, Jul. 2009, doi: 10.1152/physrev.00041.2008.
- [262] C. A. Squier and M. J. Kremer, “Biology of Oral Mucosa and Esophagus,” *JNCI Monographs*, vol. 2001, no. 29, pp. 7–15, 2001, doi: 10.1093/oxfordjournals.jncimonographs.a003443.
- [263] B. C. Heng *et al.*, “Role of YAP/TAZ in Cell Lineage Fate Determination and Related Signaling Pathways,” *Front Cell Dev Biol*, vol. 8, p. 735, Jul. 2020, doi: 10.3389/fcell.2020.00735.
- [264] D. R. Cohen and T. Curran, “fra-1: a serum-inducible, cellular immediate-early gene that encodes a fos-related antigen,” *Mol Cell Biol*, vol. 8, no. 5, pp. 2063–2069, May 1988, doi: 10.1128/mcb.8.5.2063-2069.1988.

- [265] D. R. Cohen, P. C. Ferreira, R. Gentz, B. R. Franza, and T. Curran, "The product of a fos-related gene, fra-1, binds cooperatively to the AP-1 site with Jun: transcription factor AP-1 is comprised of multiple protein complexes.," *Genes & Development*, vol. 3, no. 2, pp. 173–184, 1989, doi: 10.1101/gad.3.2.173.
- [266] T. E. Gillies, M. Pargett, M. Minguet, A. E. Davies, and J. G. Albeck, "Linear Integration of ERK Activity Predominates over Persistence Detection in Fra-1 Regulation," *Cell Syst*, vol. 5, no. 6, pp. 549–563.e5, Dec. 2017, doi: 10.1016/j.cels.2017.10.019.
- [267] J. Rutter, C. H. Michnoff, S. M. Harper, K. H. Gardner, and S. L. McKnight, "PAS kinase: an evolutionarily conserved PAS domain-regulated serine/threonine kinase," *Proc Natl Acad Sci U S A*, vol. 98, no. 16, pp. 8991–8996, Jul. 2001, doi: 10.1073/pnas.161284798.
- [268] W. A. Wilson, A. V Skurat, B. Probst, A. de Paoli-Roach, P. J. Roach, and J. Rutter, "Control of mammalian glycogen synthase by PAS kinase," *Proc Natl Acad Sci U S A*, vol. 102, no. 46, pp. 16596–16601, Nov. 2005, doi: 10.1073/pnas.0508481102.
- [269] M. Hafner *et al.*, "Multiomics Profiling Establishes the Polypharmacology of FDA-Approved CDK4/6 Inhibitors and the Potential for Differential Clinical Activity," *Cell Chem Biol*, vol. 26, no. 8, pp. 1067–1080.e8, Aug. 2019, doi: 10.1016/j.chembiol.2019.05.005.
- [270] S. A. Jones and B. J. Jenkins, "Recent insights into targeting the IL-6 cytokine family in inflammatory diseases and cancer," *Nat Rev Immunol*, vol. 18, no. 12, pp. 773–789, 2018, doi: 10.1038/s41577-018-0066-7.
- [271] M. R. Ng, A. Besser, G. Danuser, and J. S. Brugge, "Substrate stiffness regulates cadherin-dependent collective migration through myosin-II contractility," *J Cell Biol*, vol. 199, no. 3, pp. 545–563, Oct. 2012, doi: 10.1083/jcb.201207148.
- [272] H. Poplimont *et al.*, "Neutrophil Swarming in Damaged Tissue Is Orchestrated by Connexins and Cooperative Calcium Alarm Signals," *Curr Biol*, vol. 30, no. 14, pp. 2761–2776.e7, Jul. 2020, doi: 10.1016/j.cub.2020.05.030.
- [273] R. Huang *et al.*, "The NCATS BioPlanet - An Integrated Platform for Exploring the Universe of Cellular Signaling Pathways for Toxicology, Systems Biology, and Chemical Genomics," *Front Pharmacol*, vol. 10, p. 445, Apr. 2019, doi: 10.3389/fphar.2019.00445.
- [274] U. Amara *et al.*, "Molecular intercommunication between the complement and coagulation systems," *J Immunol*, vol. 185, no. 9, pp. 5628–5636, Nov. 2010, doi: 10.4049/jimmunol.0903678.
- [275] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2–3, pp. 191–203, 1984, doi: 10.1016/0098-3004(84)90020-7.
- [276] P. Friedl and D. Gilmour, "Collective cell migration in morphogenesis, regeneration and cancer," *Nat Rev Mol Cell Biol*, vol. 10, no. 7, pp. 445–457, Jul. 2009, doi: 10.1038/nrm2720.
- [277] R. Mayor and S. Etienne-Manneville, "The front and rear of collective cell migration," *Nat Rev Mol Cell Biol*, vol. 17, no. 2, pp. 97–109, Feb. 2016, doi: 10.1038/nrm.2015.14.

- [278] L. C. Abbott and F. Nigussie, "Adult neurogenesis in the mammalian dentate gyrus," *Anat Histol Embryol*, vol. 49, no. 1, pp. 3–16, Jan. 2020, doi: 10.1111/ahe.12496.
- [279] J. M. Reinke and H. Sorg, "Wound Repair and Regeneration," *European Surgical Research*, vol. 49, no. 1, pp. 35–43, 2012, doi: 10.1159/000339613.
- [280] F. Peglion, F. Llense, and S. Etienne-Manneville, "Adherens junction treadmill during collective migration," *Nat Cell Biol*, vol. 16, no. 7, pp. 639–651, Jul. 2014, doi: 10.1038/ncb2985.
- [281] N. Yamaguchi, T. Mizutani, K. Kawabata, and H. Haga, "Leader cells regulate collective cell migration via Rac activation in the downstream signaling of integrin  $\beta$ 1 and PI3K," *Sci Rep*, vol. 5, no. 1, p. 7656, Jan. 2015, doi: 10.1038/srep07656.
- [282] M. M. Zegers and P. Friedl, "Rho GTPases in collective cell migration," *Small GTPases*, vol. 5, no. 3, p. e983869, Jul. 2014, doi: 10.4161/sgtp.28997.
- [283] N. M. Neumann, D. M. Kim, R. J. Huebner, and A. J. Ewald, "Collective cell migration is spatiotemporally regulated during mammary epithelial bifurcation," *J Cell Sci*, vol. 136, no. 1, Jan. 2023, doi: 10.1242/jcs.259275.
- [284] G. K. Chan, J. A. McGrath, and M. Parsons, "Spatial activation of ezrin by epidermal growth factor receptor and focal adhesion kinase co-ordinates epithelial cell migration," *Open Biol*, vol. 11, no. 8, Aug. 2021, doi: 10.1098/rsob.210166.
- [285] A. Nawshad, D. LaGamba, A. Polad, and E. D. Hay, "Transforming Growth Factor- $\beta$  Signaling during Epithelial-Mesenchymal Transformation: Implications for Embryogenesis and Tumor Metastasis," *Cells Tissues Organs*, vol. 179, no. 1–2, pp. 11–23, 2005, doi: 10.1159/000084505.
- [286] K. J. Cheung *et al.*, "Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters," *Proceedings of the National Academy of Sciences*, vol. 113, no. 7, Feb. 2016, doi: 10.1073/pnas.1508541113.
- [287] N. M. Aiello *et al.*, "EMT Subtype Influences Epithelial Plasticity and Mode of Cell Migration," *Dev Cell*, vol. 45, no. 6, pp. 681–695.e4, Jun. 2018, doi: 10.1016/j.devcel.2018.05.027.
- [288] K. Campbell *et al.*, "Collective cell migration and metastases induced by an epithelial-to-mesenchymal transition in *Drosophila* intestinal tumors," *Nat Commun*, vol. 10, no. 1, p. 2311, May 2019, doi: 10.1038/s41467-019-10269-y.
- [289] K. Saito, A. Mitsui, I. W. Sumardika, Y. Yokoyama, M. Sakaguchi, and E. Kondo, "PLOD2-driven IL-6/STAT3 signaling promotes the invasion and metastasis of oral squamous cell carcinoma via activation of integrin  $\beta$ 1," *Int J Oncol*, vol. 58, no. 6, Jun. 2021, doi: 10.3892/ijo.2021.5209.
- [290] A. Righi, I. Sarotto, L. Casorzo, S. Cavalchini, E. Frangipane, and M. Risio, "Tumour budding is associated with hypoxia at the advancing front of colorectal cancer," *Histopathology*, vol. 66, no. 7, pp. 982–990, Jun. 2015, doi: 10.1111/his.12602.
- [291] J.-A. Montero, B. Kilian, J. Chan, P. E. Bayliss, and C.-P. Heisenberg, "Phosphoinositide 3-Kinase Is Required for Process Outgrowth and Cell Polarization of Gastrulating Mesendodermal Cells," *Current Biology*, vol. 13, no. 15, pp. 1279–1289, Aug. 2003, doi: 10.1016/S0960-9822(03)00505-0.

- [292] G. Valentin, P. Haas, and D. Gilmour, "The Chemokine SDF1a Coordinates Tissue Migration through the Spatially Restricted Activation of Cxcr7 and Cxcr4b," *Current Biology*, vol. 17, no. 12, pp. 1026–1031, Jun. 2007, doi: 10.1016/j.cub.2007.05.020.
- [293] D. C. Ardila, V. Aggarwal, M. Singh, A. Chattopadhyay, S. Chaparala, and S. Sant, "Identifying Molecular Signatures of Distinct Modes of Collective Migration in Response to the Microenvironment Using Three-Dimensional Breast Cancer Models," *Cancers (Basel)*, vol. 13, no. 6, p. 1429, Mar. 2021, doi: 10.3390/cancers13061429.
- [294] C. D. Richards, "The Enigmatic Cytokine Oncostatin M and Roles in Disease," *ISRN Inflamm*, vol. 2013, pp. 1–23, Dec. 2013, doi: 10.1155/2013/512103.
- [295] A. Masjedi *et al.*, "Oncostatin M: A mysterious cytokine in cancers," *Int Immunopharmacol*, vol. 90, p. 107158, Jan. 2021, doi: 10.1016/j.intimp.2020.107158.
- [296] Y. Zhong, J. Li, Y. Zhang, W. Qiu, and Y. Luo, "The polymorphisms of oncostatin M receptor gene associated with increased risk of lung cancer," 2018. [Online]. Available: [www.ijcem.com/](http://www.ijcem.com/)
- [297] A. Geethadevi *et al.*, "Oncostatin M Receptor–Targeted Antibodies Suppress STAT3 Signaling and Inhibit Ovarian Cancer Growth," *Cancer Res*, vol. 81, no. 20, pp. 5336–5352, Oct. 2021, doi: 10.1158/0008-5472.CAN-21-0483.
- [298] A. M. Araujo *et al.*, "Stromal oncostatin M cytokine promotes breast cancer progression by reprogramming the tumor microenvironment," *Journal of Clinical Investigation*, vol. 132, no. 7, Apr. 2022, doi: 10.1172/JCI148667.
- [299] B. Rah *et al.*, "JAK/STAT Signaling: Molecular Targets, Therapeutic Opportunities, and Limitations of Targeted Inhibitions in Solid Malignancies.," *Front Pharmacol*, vol. 13, p. 821344, 2022, doi: 10.3389/fphar.2022.821344.
- [300] R Core Team, "R: A language and environment for statistical computing. R Foundation for Statistical Computing," 2022, *Vienna, Austria*.
- [301] H. Qian, M. P. Sheetz, and E. L. Elson, "Single particle tracking. Analysis of diffusion and flow in two-dimensional systems," *Biophys J*, vol. 60, no. 4, pp. 910–921, Oct. 1991, doi: 10.1016/S0006-3495(91)82125-7.
- [302] C. W. Dunnett, "A Multiple Comparison Procedure for Comparing Several Treatments with a Control," *J Am Stat Assoc*, vol. 50, no. 272, pp. 1096–1121, Dec. 1955, doi: 10.1080/01621459.1955.10501294.
- [303] P. Shannon *et al.*, "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks," *Genome Res*, vol. 13, no. 11, pp. 2498–2504, Nov. 2003, doi: 10.1101/gr.1239303.
- [304] I. H. Goenawan, K. Bryan, and D. J. Lynn, "DyNet: visualization and analysis of dynamic molecular interaction networks," *Bioinformatics*, vol. 32, no. 17, pp. 2713–2715, Sep. 2016, doi: 10.1093/bioinformatics/btw187.

- [305] Student, "THE PROBABLE ERROR OF A MEAN," *Biometrika*, vol. 6, no. 1, pp. 1–25, Mar. 1908, doi: 10.1093/biomet/6.1.1.
- [306] C. S. McGinnis *et al.*, "MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices," *Nat Methods*, vol. 16, no. 7, pp. 619–626, Jul. 2019, doi: 10.1038/s41592-019-0433-8.
- [307] Y. Hao *et al.*, "Integrated analysis of multimodal single-cell data," *Cell*, vol. 184, no. 13, pp. 3573–3587.e29, Jun. 2021, doi: 10.1016/j.cell.2021.04.048.
- [308] G. Yu, L.-G. Wang, Y. Han, and Q.-Y. He, "clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters," *OMICS*, vol. 16, no. 5, pp. 284–287, May 2012, doi: 10.1089/omi.2011.0118.
- [309] J.-R. Lin, M. Fallahi-Sichani, J.-Y. Chen, and P. K. Sorger, "Cyclic Immunofluorescence (CyclIF), A Highly Multiplexed Method for Single-cell Imaging," *Curr Protoc Chem Biol*, vol. 8, no. 4, pp. 251–264, Dec. 2016, doi: 10.1002/cpch.14.
- [310] A. S. Neish, L. M. Khachigian, A. Park, V. R. Baichwal, and T. Collins, "Sp1 Is a Component of the Cytokine-inducible Enhancer in the Promoter of Vascular Cell Adhesion Molecule-1," *Journal of Biological Chemistry*, vol. 270, no. 48, pp. 28903–28909, Dec. 1995, doi: 10.1074/jbc.270.48.28903.
- [311] A. Szabowski *et al.*, "c-Jun and JunB Antagonistically Control Cytokine-Regulated Mesenchymal–Epidermal Interaction in Skin," *Cell*, vol. 103, no. 5, pp. 745–755, Nov. 2000, doi: 10.1016/S0092-8674(00)00178-1.
- [312] M. Stellato *et al.*, "The AP-1 transcription factor Fosl-2 drives cardiac fibrosis and arrhythmias under immunofibrotic conditions," *Commun Biol*, vol. 6, no. 1, p. 161, Feb. 2023, doi: 10.1038/s42003-023-04534-6.
- [313] X. Li, Y. Xiang, F. Li, C. Yin, B. Li, and X. Ke, "WNT/ $\beta$ -Catenin Signaling Pathway Regulating T Cell Inflammation in the Tumor Microenvironment," *Front Immunol*, vol. 10, Sep. 2019, doi: 10.3389/fimmu.2019.02293.
- [314] J. S. Beeler *et al.*, "p73 regulates epidermal wound healing and induced keratinocyte programming," *PLoS One*, vol. 14, no. 6, p. e0218458, Jun. 2019, doi: 10.1371/journal.pone.0218458.
- [315] T. Ozaki and A. Nakagawara, "p73, a sophisticated p53 family member in the cancer world," *Cancer Sci*, vol. 96, no. 11, pp. 729–737, Nov. 2005, doi: 10.1111/j.1349-7006.2005.00116.x.
- [316] R. Sowamber *et al.*, "CCAAT/enhancer binding protein delta (C/EBP $\delta$ ) demonstrates a dichotomous role in tumour initiation and promotion of epithelial carcinoma.," *EBioMedicine*, vol. 44, pp. 261–274, Jun. 2019, doi: 10.1016/j.ebiom.2019.05.002.
- [317] G. W. Robinson, P. F. Johnson, L. Hennighausen, and E. Sterneck, "The C/EBP $\beta$  transcription factor regulates epithelial cell proliferation and differentiation in the mammary gland," *Genes Dev*, vol. 12, no. 12, pp. 1907–1916, Jun. 1998, doi: 10.1101/gad.12.12.1907.

- [318] R. Shrivastava, V. Singh, M. Asif, M. P. S. Negi, and S. Bhadauria, "Oncostatin M upregulates HIF-1 $\alpha$  in breast tumor associated macrophages independent of intracellular oxygen concentration," *Life Sci*, vol. 194, pp. 59–66, Feb. 2018, doi: 10.1016/j.lfs.2017.12.017.
- [319] N. Battello *et al.*, "The role of HIF-1 in oncostatin M-dependent metabolic reprogramming of hepatic cells," *Cancer Metab*, vol. 4, no. 1, p. 3, Dec. 2016, doi: 10.1186/s40170-016-0141-0.
- [320] J. P. Zhang *et al.*, "Down-regulation of Sp1 suppresses cell proliferation, clonogenicity and the expressions of stem cell markers in nasopharyngeal carcinoma," *J Transl Med*, vol. 12, no. 1, Aug. 2014, doi: 10.1186/s12967-014-0222-1.
- [321] M. Piechaczyk and R. Farràs, "Regulation and function of JunB in cell proliferation," *Biochem Soc Trans*, vol. 36, no. 5, pp. 864–867, Oct. 2008, doi: 10.1042/BST0360864.
- [322] Y. Teng, J. L. Ross, and J. K. Cowell, "The involvement of JAK-STAT3 in cell motility, invasion, and metastasis.," *JAKSTAT*, vol. 3, no. 1, p. e28086, Jan. 2014, doi: 10.4161/jkst.28086.
- [323] J. D. Gordan, J. A. Bertout, C.-J. Hu, J. A. Diehl, and M. C. Simon, "HIF-2 $\alpha$  Promotes Hypoxic Cell Proliferation by Enhancing c-Myc Transcriptional Activity," *Cancer Cell*, vol. 11, no. 4, pp. 335–347, Apr. 2007, doi: 10.1016/j.ccr.2007.02.006.
- [324] R. Kemler, "From cadherins to catenins: cytoplasmic protein interactions and regulation of cell adhesion," *Trends in Genetics*, vol. 9, no. 9, pp. 317–321, Sep. 1993, doi: 10.1016/0168-9525(93)90250-L.
- [325] A. Weidemann *et al.*, "HIF-1 $\alpha$  activation results in actin cytoskeleton reorganization and modulation of Rac-1 signaling in endothelial cells," *Cell Communication and Signaling*, vol. 11, no. 1, p. 80, Dec. 2013, doi: 10.1186/1478-811X-11-80.
- [326] D. M. Gilkes, S. Bajpai, P. Chaturvedi, D. Wirtz, and G. L. Semenza, "Hypoxia-inducible Factor 1 (HIF-1) Promotes Extracellular Matrix Remodeling under Hypoxic Conditions by Inducing P4HA1, P4HA2, and PLOD2 Expression in Fibroblasts," *Journal of Biological Chemistry*, vol. 288, no. 15, pp. 10819–10829, Apr. 2013, doi: 10.1074/jbc.M112.442939.
- [327] A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov, and P. Tamayo, "The Molecular Signatures Database Hallmark Gene Set Collection," *Cell Syst*, vol. 1, no. 6, pp. 417–425, Dec. 2015, doi: 10.1016/j.cels.2015.12.004.
- [328] Q. Wang *et al.*, "HIF-1 $\alpha$  up-regulates NDRG1 expression through binding to NDRG1 promoter, leading to proliferation of lung cancer A549 cells," *Mol Biol Rep*, vol. 40, no. 5, pp. 3723–3729, May 2013, doi: 10.1007/s11033-012-2448-4.
- [329] T. van den Beucken *et al.*, "Hypoxia-induced Expression of Carbonic Anhydrase 9 Is Dependent on the Unfolded Protein Response," *Journal of Biological Chemistry*, vol. 284, no. 36, pp. 24204–24212, Sep. 2009, doi: 10.1074/jbc.M109.006510.
- [330] M. Ashburner *et al.*, "Gene Ontology: tool for the unification of biology," *Nat Genet*, vol. 25, no. 1, pp. 25–29, May 2000, doi: 10.1038/75556.

- [331] S. J. Kierans and C. T. Taylor, "Regulation of glycolysis by the hypoxia-inducible factor (HIF): implications for cellular physiology," *J Physiol*, vol. 599, no. 1, pp. 23–37, Jan. 2021, doi: 10.1113/JP280572.
- [332] E. L. Campbell, "Hypoxia-recruited angiogenic neutrophils," *Blood*, vol. 126, no. 17, pp. 1972–1973, Oct. 2015, doi: 10.1182/blood-2015-09-666578.
- [333] J. C. Morote-Garcia, D. Napiwotzky, D. Köhler, and P. Rosenberger, "Endothelial Semaphorin 7A promotes neutrophil migration during hypoxia," *Proceedings of the National Academy of Sciences*, vol. 109, no. 35, pp. 14146–14151, Aug. 2012, doi: 10.1073/pnas.1202165109.
- [334] S. P. Colgan, A. L. Dzus, and C. A. Parkos, "Epithelial exposure to hypoxia modulates neutrophil transepithelial migration.," *J Exp Med*, vol. 184, no. 3, pp. 1003–1015, Sep. 1996, doi: 10.1084/jem.184.3.1003.
- [335] J. F. Foley, "Complement, Attraction, and Collective Migration," *Sci Signal*, vol. 4, no. 204, Dec. 2011, doi: 10.1126/scisignal.4204ec352.
- [336] C. Carmona-Fontaine *et al.*, "Complement Fragment C3a Controls Mutual Cell Attraction during Collective Cell Migration," *Dev Cell*, vol. 21, no. 6, pp. 1026–1037, Dec. 2011, doi: 10.1016/j.devcel.2011.10.012.
- [337] S. Vandendriessche, S. Cambier, P. Proost, and P. E. Marques, "Complement Receptors and Their Role in Leukocyte Recruitment and Phagocytosis," *Front Cell Dev Biol*, vol. 9, Feb. 2021, doi: 10.3389/fcell.2021.624025.
- [338] D. C. Mastellos *et al.*, "Compstatin: a C3-targeted complement inhibitor reaching its prime for bedside intervention," *Eur J Clin Invest*, vol. 45, no. 4, pp. 423–440, Apr. 2015, doi: 10.1111/eci.12419.
- [339] J. C. Jun, A. Rathore, H. Younas, D. Gilkes, and V. Y. Polotsky, "Hypoxia-Inducible Factors and Cancer.," *Curr Sleep Med Rep*, vol. 3, no. 1, pp. 1–10, Mar. 2017, doi: 10.1007/s40675-017-0062-7.
- [340] E. Tátrai *et al.*, "Cell type-dependent HIF1  $\alpha$ -mediated effects of hypoxia on proliferation, migration and metastatic potential of human tumor cells.," *Oncotarget*, vol. 8, no. 27, pp. 44498–44510, Jul. 2017, doi: 10.18632/oncotarget.17806.
- [341] P. Wang *et al.*, "Hypoxia inducible factor-1 $\alpha$  regulates autophagy via the p27-E2F1 signaling pathway.," *Mol Med Rep*, vol. 16, no. 2, pp. 2107–2112, Aug. 2017, doi: 10.3892/mmr.2017.6794.
- [342] C.-C. Chou, H.-C. Chuang, S. B. Salunke, S. K. Kulp, and C.-S. Chen, "A novel HIF-1 $\alpha$ -integrin-linked kinase regulatory loop that facilitates hypoxia-induced HIF-1 $\alpha$  expression and epithelial-mesenchymal transition in cancer cells.," *Oncotarget*, vol. 6, no. 10, pp. 8271–85, Apr. 2015, doi: 10.18632/oncotarget.3186.
- [343] T. Sakamoto *et al.*, "Hypoxia-inducible factor 1 regulation through cross talk between mTOR and MT1-MMP.," *Mol Cell Biol*, vol. 34, no. 1, pp. 30–42, Jan. 2014, doi: 10.1128/MCB.01169-13.

- [344] K. M. Hardy, B. W. Booth, M. J. C. Hendrix, D. S. Salomon, and L. Strizzi, "ErbB/EGF Signaling and EMT in Mammary Development and Breast Cancer," *J Mammary Gland Biol Neoplasia*, vol. 15, no. 2, pp. 191–199, Jun. 2010, doi: 10.1007/s10911-010-9172-2.
- [345] J. Donovan and J. Slingerland, "Transforming growth factor- $\beta$  and breast cancer: Cell cycle arrest by transforming growth factor- $\beta$  and its disruption in cancer," *Breast Cancer Research*, vol. 2, no. 2, p. 116, Apr. 2000, doi: 10.1186/bcr43.
- [346] D. E. Johnson, R. A. O'Keefe, and J. R. Grandis, "Targeting the IL-6/JAK/STAT3 signalling axis in cancer," *Nat Rev Clin Oncol*, vol. 15, no. 4, pp. 234–248, Apr. 2018, doi: 10.1038/nrclinonc.2018.8.
- [347] N. Geary, "Understanding synergy," *American Journal of Physiology-Endocrinology and Metabolism*, vol. 304, no. 3, pp. E237–E253, Feb. 2013, doi: 10.1152/ajpendo.00308.2012.
- [348] M. Niepel *et al.*, "Common and cell-type specific responses to anti-cancer drugs revealed by high throughput transcript profiling," *Nat Commun*, vol. 8, no. 1, p. 1186, Oct. 2017, doi: 10.1038/s41467-017-01383-w.
- [349] M. Eslami *et al.*, "Prediction of whole-cell transcriptional response with machine learning," *Bioinformatics*, vol. 38, no. 2, pp. 404–409, Jan. 2022, doi: 10.1093/bioinformatics/btab676.
- [350] M. Lotfollahi *et al.*, "Predicting cellular responses to complex perturbations in high-throughput screens," *Mol Syst Biol*, vol. 19, no. 6, Jun. 2023, doi: 10.15252/msb.202211517.
- [351] C. M. McQueen *et al.*, "PER2 regulation of mammary gland development," *Development*, vol. 145, no. 6, Mar. 2018, doi: 10.1242/dev.157966.
- [352] M. Melani, K. J. Simpson, J. S. Brugge, and D. Montell, "Regulation of Cell Adhesion and Collective Cell Migration by Hindsight and Its Human Homolog RREB1," *Current Biology*, vol. 18, no. 7, pp. 532–537, Apr. 2008, doi: 10.1016/j.cub.2008.03.024.
- [353] S. E. Seton-Rogers *et al.*, "Cooperation of the ErbB2 receptor and transforming growth factor  $\beta$  in induction of migration and invasion in mammary epithelial cells," *Proceedings of the National Academy of Sciences*, vol. 101, no. 5, pp. 1257–1262, Feb. 2004, doi: 10.1073/pnas.0308090100.
- [354] P. Pierozan and O. Karlsson, "PFOS induces proliferation, cell-cycle progression, and malignant phenotype in human breast epithelial cells.," *Arch Toxicol*, vol. 92, no. 2, pp. 705–716, Feb. 2018, doi: 10.1007/s00204-017-2077-8.
- [355] K. G. Joreskog and H. O. A. Wold, *Systems under indirect observation. Causality-structure-prediction*, vol. 1–2. Amsterdam: North-Holland, 1982.
- [356] A. Signorell and et mult. al., "DescTools: Tools for descriptive statistics," 2017, R package version 0.99.23.
- [357] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple," 1995.
- [358] Z. Gu, R. Eils, and M. Schlesner, "Complex heatmaps reveal patterns and correlations in multidimensional genomic data," *Bioinformatics*, vol. 32, no. 18, pp. 2847–2849, Sep. 2016, doi: 10.1093/bioinformatics/btw313.

- [359] M. Ashburner *et al.*, “Gene ontology: tool for the unification of biology. The Gene Ontology Consortium,” *Nat Genet*, vol. 25, no. 1, pp. 25–29, May 2000, doi: 10.1038/75556.
- [360] L. Kolberg, U. Raudvere, I. Kuzmin, P. Adler, J. Vilo, and H. Peterson, “g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update),” *Nucleic Acids Res*, vol. 51, no. W1, pp. W207–W212, Jul. 2023, doi: 10.1093/nar/gkad347.
- [361] T. Stuart *et al.*, “Comprehensive Integration of Single-Cell Data,” *Cell*, vol. 177, no. 7, pp. 1888–1902.e21, Jun. 2019, doi: 10.1016/j.cell.2019.05.031.
- [362] B.-H. Mevik and R. Wehrens, “The **pls** Package: Principal Component and Partial Least Squares Regression in R,” *J Stat Softw*, vol. 18, no. 2, 2007, doi: 10.18637/jss.v018.i02.
- [363] S. Wold, M. Sjöström, and L. Eriksson, “PLS-regression: a basic tool of chemometrics,” *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, Oct. 2001, doi: 10.1016/S0169-7439(01)00155-1.
- [364] S. A. Aleksander *et al.*, “The Gene Ontology knowledgebase in 2023,” *Genetics*, vol. 224, no. 1, May 2023, doi: 10.1093/genetics/iyad031.
- [365] M. Ashburner *et al.*, “Gene Ontology: tool for the unification of biology,” *Nat Genet*, vol. 25, no. 1, pp. 25–29, May 2000, doi: 10.1038/75556.
- [366] L. M. Heiser *et al.*, “Subtype and pathway specific responses to anticancer compounds in breast cancer,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 8, pp. 2724–2729, Feb. 2012, doi: 10.1073/pnas.1018854108.
- [367] N. U. Nair *et al.*, “Migration rather than proliferation transcriptomic signatures are strongly associated with breast cancer patient survival,” *Sci Rep*, vol. 9, no. 1, p. 10989, Jul. 2019, doi: 10.1038/s41598-019-47440-w.
- [368] B. DepMap, “Current DepMap Release data, including CRISPR Screens, PRISM Drug Screens, Copy Number, Mutation, Expression, and Fusions,” 2024. doi: <https://doi.org/10.25452/figshare.plus.25880521.v1>.
- [369] A. Tsherniak *et al.*, “Defining a Cancer Dependency Map,” *Cell*, vol. 170, no. 3, pp. 564–576.e16, Jul. 2017, doi: 10.1016/j.cell.2017.06.010.
- [370] G. Blair *et al.*, “estimatr: Fast Estimators for Design-Based Inference,” 2024, R package version 1.0.2.
- [371] R. V. Lenth, “emmeans: Estimated Margin2l Means, aka Least-Squares Means,” 2024, R package version 1.10.3-090006.
- [372] R. Akbani *et al.*, “A pan-cancer proteomic perspective on The Cancer Genome Atlas,” *Nat Commun*, vol. 5, no. 1, p. 3887, May 2014, doi: 10.1038/ncomms4887.
- [373] K. J. Livak and T. D. Schmittgen, “Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2– $\Delta\Delta$ CT Method,” *Methods*, vol. 25, no. 4, pp. 402–408, Dec. 2001, doi: 10.1006/meth.2001.1262.

- [374] G. Tarcic *et al.*, “EGR1 and the ERK-ERF axis drive mammary cell migration in response to EGF,” *The FASEB Journal*, vol. 26, no. 4, pp. 1582–1592, Apr. 2012, doi: 10.1096/fj.11-194654.
- [375] E.-S. Kim, M.-S. Kim, and A. Moon, “TGF- $\beta$ -induced upregulation of MMP-2 and MMP-9 depends on p38 MAPK, but not ERK signaling in MCF10A human breast epithelial cells,” *Int J Oncol*, Nov. 2004, doi: 10.3892/ijo.25.5.1375.
- [376] E.-S. Kim, M.-S. Kim, and A. Moon, “Transforming growth factor (TGF)- $\beta$  in conjunction with H-ras activation promotes malignant progression of MCF10A breast epithelial cells,” *Cytokine*, vol. 29, no. 2, pp. 84–91, Jan. 2005, doi: 10.1016/j.cyto.2004.10.001.
- [377] A. Sundqvist *et al.*, “TGF $\beta$  and EGF signaling orchestrates the AP-1- and p63 transcriptional regulation of breast cancer invasiveness,” *Oncogene*, vol. 39, no. 22, pp. 4436–4449, May 2020, doi: 10.1038/s41388-020-1299-z.
- [378] S. Lloyd, “Least squares quantization in PCM,” *IEEE Trans Inf Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982, doi: 10.1109/TIT.1982.1056489.
- [379] E. M. Sanford, B. L. Emert, A. Coté, and A. Raj, “Gene regulation gravitates toward either addition or multiplication when combining the effects of two signals,” *Elife*, vol. 9, Dec. 2020, doi: 10.7554/eLife.59388.
- [380] H.-H. Dong *et al.*, “The epithelial-mesenchymal transition promotes transdifferentiation of subcutaneously implanted hepatic oval cells into mesenchymal tumor tissue.,” *Stem Cells Dev*, vol. 18, no. 9, pp. 1293–8, Nov. 2009, doi: 10.1089/scd.2008.0321.
- [381] T. Hastie, R. Tibshirani, and J. Friedman, “Model Assessment and Selection,” 2009, pp. 219–259. doi: 10.1007/978-0-387-84858-7\_7.
- [382] J. H. Chung and F. Bunz, “Cdk2 Is Required for p53-Independent G2/M Checkpoint Control,” *PLoS Genet*, vol. 6, no. 2, p. e1000863, Feb. 2010, doi: 10.1371/journal.pgen.1000863.
- [383] J. Liu, C. L. Smith, D. DeRyckere, K. DeAngelis, G. S. Martin, and J. M. Berger, “Structure and Function of Cdc6/Cdc18,” *Mol Cell*, vol. 6, no. 3, pp. 637–648, Sep. 2000, doi: 10.1016/S1097-2765(00)00062-9.
- [384] R. Hayashi, Y. Goto, R. Ikeda, K. K. Yokoyama, and K. Yoshida, “CDCA4 is an E2F transcription factor family-induced nuclear factor that regulates E2F-dependent transcriptional activation and cell proliferation.,” *J Biol Chem*, vol. 281, no. 47, pp. 35633–48, Nov. 2006, doi: 10.1074/jbc.M603800200.
- [385] B. Giotti *et al.*, “Assembly of a parts list of the human mitotic cell cycle machinery,” *J Mol Cell Biol*, vol. 11, no. 8, pp. 703–718, Aug. 2019, doi: 10.1093/jmcb/mjy063.
- [386] C. Berasain and M. A. Avila, “Amphiregulin,” *Semin Cell Dev Biol*, vol. 28, pp. 31–41, Apr. 2014, doi: 10.1016/j.semcdb.2014.01.005.
- [387] M. Li *et al.*, “p16 promotes proliferation in cervical carcinoma cells through CDK6-HuR-IL1A axis,” *J Cancer*, vol. 11, no. 6, pp. 1457–1467, 2020, doi: 10.7150/jca.35479.

- [388] A. Chakraborty and S. Guha, "Granulocyte Colony-Stimulating Factor/Granulocyte Colony-Stimulating Factor Receptor Biological Axis Promotes Survival and Growth of Bladder Cancer Cells," *Urology*, vol. 69, no. 6, pp. 1210–1215, Jun. 2007, doi: 10.1016/j.urology.2007.02.035.
- [389] S. G. Mathiassen, D. De Zio, and F. Cecconi, "Autophagy and the Cell Cycle: A Complex Landscape," *Front Oncol*, vol. 7, Mar. 2017, doi: 10.3389/fonc.2017.00051.
- [390] C. Klec *et al.*, "ALYREF, a novel factor involved in breast carcinogenesis, acts through transcriptional and post-transcriptional mechanisms selectively regulating the short NEAT1 isoform," *Cellular and Molecular Life Sciences*, vol. 79, no. 7, p. 391, Jul. 2022, doi: 10.1007/s00018-022-04402-2.
- [391] S. Yu *et al.*, "Mining the prognostic significance of the GINS2 gene in human breast cancer using bioinformatics analysis," *Oncol Lett*, vol. 20, no. 2, pp. 1300–1310, May 2020, doi: 10.3892/ol.2020.11651.
- [392] H. Sun *et al.*, "Cyclic AMP-Responsive Element Binding Protein- and Nuclear Factor- $\kappa$ B-Regulated CXC Chemokine Gene Expression in Lung Carcinogenesis," *Cancer Prevention Research*, vol. 1, no. 5, pp. 316–328, Oct. 2008, doi: 10.1158/1940-6207.CAPR-07-0002.
- [393] M. Tanaka and A. Miyahima, "Onconstatin M, a multifunctional cytokine," in *Reviews of Physiology, Biochemistry and Pharmacology*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 39–52. doi: 10.1007/s10254-003-0013-1.
- [394] D. Bressan, G. Battistoni, and G. J. Hannon, "The dawn of spatial omics," *Science (1979)*, vol. 381, no. 6657, Aug. 2023, doi: 10.1126/science.abq4964.
- [395] B. Güvenç Paltun, S. Kaski, and H. Mamitsuka, "Machine learning approaches for drug combination therapies," *Brief Bioinform*, vol. 22, no. 6, Nov. 2021, doi: 10.1093/bib/bbab293.
- [396] A. P. Lind and P. C. Anderson, "Predicting drug activity against cancer cells by random forest models based on minimal genomic information and chemical properties," *PLoS One*, vol. 14, no. 7, p. e0219774, Jul. 2019, doi: 10.1371/journal.pone.0219774.
- [397] A. Torkamannia, Y. Omid, and R. Ferdousi, "SYNDEEP: a deep learning approach for the prediction of cancer drugs synergy," *Sci Rep*, vol. 13, no. 1, p. 6184, Apr. 2023, doi: 10.1038/s41598-023-33271-3.
- [398] M. Ali and T. Aittokallio, "Machine learning and feature selection for drug response prediction in precision oncology applications," *Biophys Rev*, vol. 11, no. 1, pp. 31–39, Feb. 2019, doi: 10.1007/s12551-018-0446-z.
- [399] A. Banerjee, V. Pirrone, B. Wigdahl, and M. R. Nonnemacher, "Transcriptional regulation of the chemokine co-receptor CCR5 by the cAMP/PKA/CREB pathway," *Biomedicine & Pharmacotherapy*, vol. 65, no. 4, pp. 293–297, Jul. 2011, doi: 10.1016/j.biopha.2011.03.009.
- [400] A. Y. Wen *et al.*, "Increased Abscess Formation and Defective Chemokine Regulation in CREB Transgenic Mice," *PLoS One*, vol. 8, no. 2, p. e55866, Feb. 2013, doi: 10.1371/journal.pone.0055866.

- [401] D. Y. Grinman *et al.*, "Role of 3'-5'-cyclic adenosine monophosphate on the epidermal growth factor dependent survival in mammary epithelial cells," *Mol Cell Endocrinol*, vol. 419, pp. 259–267, Jan. 2016, doi: 10.1016/j.mce.2015.10.026.
- [402] L. R. Bohrer and K. L. Schwertfeger, "Macrophages Promote Fibroblast Growth Factor Receptor-Driven Tumor Cell Migration and Invasion in a Cxcr2-Dependent Manner," *Molecular Cancer Research*, vol. 10, no. 10, pp. 1294–1305, Oct. 2012, doi: 10.1158/1541-7786.MCR-12-0275.
- [403] L. J. Miller, S. H. Kurtzman, Y. Wang, K. H. Anderson, R. R. Lindquist, and D. L. Kreutzer, "Expression of interleukin-8 receptors on tumor cells and vascular endothelial cells in human breast cancer tissue.," *Anticancer Res*, vol. 18, no. 1A, pp. 77–81, 1998.
- [404] M. R. Zanutelli, J. Zhang, and C. A. Reinhart-King, "Mechanoresponsive metabolism in cancer cell migration and metastasis.," *Cell Metab*, vol. 33, no. 7, pp. 1307–1321, Jul. 2021, doi: 10.1016/j.cmet.2021.04.002.
- [405] C. D. Gregory, C. A. Ford, and J. J. L. P. Voss, "Microenvironmental Effects of Cell Death in Malignant Disease," 2016, pp. 51–88. doi: 10.1007/978-3-319-39406-0\_3.
- [406] K. M. Bussard and G. H. Smith, "The Mammary Gland Microenvironment Directs Progenitor Cell Fate *In Vivo*," *Int J Cell Biol*, vol. 2011, pp. 1–11, 2011, doi: 10.1155/2011/451676.
- [407] M. Nava, P. Dutta, N. R. Zemke, R. Farias-Eisner, J. V. Vadgama, and Y. Wu, "Transcriptomic and ChIP-sequence interrogation of EGFR signaling in HER2+ breast cancer cells reveals a dynamic chromatin landscape and S100 genes as targets," *BMC Med Genomics*, vol. 12, no. 1, p. 32, Dec. 2019, doi: 10.1186/s12920-019-0477-8.
- [408] M. Richter, O. Piwocka, M. Musielak, I. Piotrowski, W. M. Suchorska, and T. Trzeciak, "From Donor to the Lab: A Fascinating Journey of Primary Cell Lines," *Front Cell Dev Biol*, vol. 9, Jul. 2021, doi: 10.3389/fcell.2021.711381.
- [409] J. M. Rosenbluth *et al.*, "Organoid cultures from normal and cancer-prone human breast tissues preserve complex epithelial lineages," *Nat Commun*, vol. 11, no. 1, p. 1711, Apr. 2020, doi: 10.1038/s41467-020-15548-7.
- [410] X. Trepac, Z. Chen, and K. Jacobson, "Cell Migration," in *Comprehensive Physiology*, Wiley, 2012, pp. 2369–2392. doi: 10.1002/cphy.c110012.
- [411] B. A. Camley, "Collective gradient sensing and chemotaxis: modeling and recent developments," *Journal of Physics: Condensed Matter*, vol. 30, no. 22, p. 223001, Jun. 2018, doi: 10.1088/1361-648X/aabd9f.
- [412] S. Martinotti and E. Ranzato, "Scratch Wound Healing Assay," 2019, pp. 225–229. doi: 10.1007/7651\_2019\_259.
- [413] M. Ravi, V. Paramesh, S. R. Kaviya, E. Anuradha, and F. D. P. Solomon, "3D Cell Culture Systems: Advantages and Applications," *J Cell Physiol*, vol. 230, no. 1, pp. 16–26, Jan. 2015, doi: 10.1002/jcp.24683.

- [414] C. T. Guy, R. D. Cardiff, and W. J. Muller, "Induction of Mammary Tumors by Expression of Polyomavirus Middle T Oncogene: A Transgenic Mouse Model for Metastatic Disease," *Mol Cell Biol*, vol. 12, no. 3, pp. 954–961, Mar. 1992, doi: 10.1128/mcb.12.3.954-961.1992.
- [415] B. Györfy, "Survival analysis across the entire transcriptome identifies biomarkers with the highest prognostic power in breast cancer," *Comput Struct Biotechnol J*, vol. 19, pp. 4101–4109, 2021, doi: 10.1016/j.csbj.2021.07.014.
- [416] L. A. Matisse *et al.*, "Lack of transforming growth factor- $\beta$  signaling promotes collective cancer cell invasion through tumor-stromal crosstalk," *Breast Cancer Research*, vol. 14, no. 4, p. R98, Jul. 2012, doi: 10.1186/bcr3217.
- [417] E. Deguchi *et al.*, "Low-affinity ligands of the epidermal growth factor receptor are long-range signal transmitters in collective cell migration of epithelial cells," *Cell Rep*, vol. 43, no. 11, p. 114986, Nov. 2024, doi: 10.1016/j.celrep.2024.114986.
- [418] N. Hino *et al.*, "A feedback loop between lamellipodial extension and HGF-ERK signaling specifies leader cells during collective cell migration," *Dev Cell*, vol. 57, no. 19, pp. 2290–2304.e7, Oct. 2022, doi: 10.1016/j.devcel.2022.09.003.
- [419] U. M. Neelathi, D. Dalle Nogare, and A. B. Chitnis, "Cxcl12a induces *snail1b* expression to initiate collective migration and sequential Fgf-dependent neuromast formation in the zebrafish posterior lateral line primordium," *Development*, vol. 145, no. 14, Jul. 2018, doi: 10.1242/dev.162453.
- [420] C. Engel-Pizcueta and C. Pujades, "Interplay Between Notch and YAP/TAZ Pathways in the Regulation of Cell Fate During Embryo Development," *Front Cell Dev Biol*, vol. 9, Aug. 2021, doi: 10.3389/fcell.2021.711531.
- [421] V. Miskolci, L. C. Klemm, and A. Huttenlocher, "Cell Migration Guided by Cell–Cell Contacts in Innate Immunity," *Trends Cell Biol*, vol. 31, no. 2, pp. 86–94, Feb. 2021, doi: 10.1016/j.tcb.2020.11.002.
- [422] G.-S. Shang, L. Liu, and Y.-W. Qin, "IL-6 and TNF- $\alpha$  promote metastasis of lung cancer by inducing epithelial-mesenchymal transition.," *Oncol Lett*, vol. 13, no. 6, pp. 4657–4660, Jun. 2017, doi: 10.3892/ol.2017.6048.
- [423] R. Bayat Mokhtari *et al.*, "Combination therapy in combating cancer.," *Oncotarget*, vol. 8, no. 23, pp. 38022–38043, Jun. 2017, doi: 10.18632/oncotarget.16723.
- [424] G. Vert and J. Chory, "Crosstalk in cellular signaling: background noise or the real thing?," *Dev Cell*, vol. 21, no. 6, pp. 985–991, Dec. 2011, doi: 10.1016/j.devcel.2011.11.006.
- [425] M. P. Delghandi, M. Johannessen, and U. Moens, "The cAMP signalling pathway activates CREB through PKA, p38 and MSK1 in NIH 3T3 cells," *Cell Signal*, vol. 17, no. 11, pp. 1343–1351, Nov. 2005, doi: 10.1016/j.cellsig.2005.02.003.