UNIVERSITY OF CALIFORNIA SAN DIEGO

Multimodal Data Integration Applied to Cancer Evolution and Signaling

A Dissertation submitted in partial satisfaction of the requirements

for the degree Doctor of Philosophy

in

Bioinformatics

and Systems Biology with a Specialization in Biomedical Informatics

by

Adam Officer

Committee in charge:

Professor Pablo Tamayo, Chair

Professor J. Silvio Gutkind, Co-Chair

Professor Hannah Carter

Professor Alon Goren

Professor Matthew Hangauer

2024

The Dissertation of Adam Officer is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2024

Table of Contents

List of Figures

List of Tables

Acknowledgements

A deep thank you to all of my thesis advisors over the last few years: Dr. Olivier Harismendy, Dr. Pablo Tamayo and Dr. J. Silvio Gutkind. Without your guidance, mentorship, expertise and enthusiasm over the last few years none of this would have been possible. I would also like to thank all of the members of my committee: Dr. Alon Goren, Dr. Hannah Carter and Dr. Matthew Hangauer, for their input at every stage of this work. Additionally I would like to thank the BISB community for their support, both tangible and intangible, as I have navigated through my PhD.

Thank you to all of the members of the Gutkind, Tamayo and former Harismendy lab, especially Dana, Tim and all of Bay I for all of your helpful advice and friendship. A sincere thank you to all of my friends who have listened to my research and supported me through the last few years.

And, finally, a special thank you to my family. To my parents, Carol and Keith, to my brother Mark and to my loving and supportive wife Diana.

Chapter 2, in full, is a reprint of the material as it appears in biorXiv. A. Officer, A. M. Dempsey, L. M. Murrow, Z. Gartner, P. Tamayo, C. Yau, and O. Harismendy, Differential analysis of stromal-epithelial interactions between in situ and invasive breast cancer using gene expression profiling, bioRxiv, 2022. Additional figures and text were taken from A. Officer, E. Armingol, O. Harismendy, and N. E. Lewis, Deciphering cell-cell interactions and communication from gene expression," Nat. Rev. Genet., 2021. The dissertation author was the primary researcher and author of these papers.

Chapter 3 is currently being prepared for publication. A. Officer, R.D. Cervantes-Villagrana, S. Lubrano, N. Arang, P. Tamayo, J.S. Gutkind. The GNAQ-p53 Signaling

Axis in Uveal Melanoma. The dissertation author was the primary researcher and author of this paper.

Chapter 4 is currently in the review process at Cancer Stem Cell. F. Faraji, S. I. Ramirez, L. M. Clubb, K. Sato, P. Y. A. Quiroz, W. M. G. Galloway, Z. Mikulski, T. S. Hoang, K. Medetgul-Ernar, P. Marangoni, K. B. Jones, A. Officer, A. A. Molinolo, K. Kim, K. Sakaguchi, J. A. Califano, Q. Smith, O. D. Klein, P. Tamayo, and J. S. Gutkind, Direct reprogramming of oral epithelial progenitor cells to cancer stem cells at single cell resolution in vivo (2024). The dissertation author was a contributing researcher to this paper.

Vita

2010 - 2014 Bachelor of Science in Chemistry and Biochemistry, University of Washington, Seattle

2014 - 2018 Research Associate, Proteomics Platform, Broad Institute of MIT and Harvard

2018 - 2024 PhD in Bioinformatics and Systems Biology with a Specialization in Biomedical Informatics

Publications

A. Officer, A. M. Dempsey, L. M. Murrow, Z. Gartner, P. Tamayo, C. Yau, and O. Harismendy, “Differential analysis of stromal-epithelial interactions between in situ and invasive breast cancer using gene expression profiling,” bioRxiv, 2022. doi: 10.1101/2022.07.30.502150.

A. Officer, E. Armingol, O. Harismendy, and N. E. Lewis, “Deciphering cell-cell interactions and communication from gene expression,” Nat. Rev. Genet., vol. 22, no. 2, pp. 71–88, Feb. 2021.

F. Faraji, S. I. Ramirez, L. M. Clubb, K. Sato, P. Y. A. Quiroz, W. M. G. Galloway, Z. Mikulski, T. S. Hoang, K. Medetgul-Ernar, P. Marangoni, K. B. Jones, **A. Officer**, A. A. Molinolo, K. Kim, K. Sakaguchi, J. A. Califano, Q. Smith, O. D. Klein, P. Tamayo, and J. S. Gutkind, “Direct reprogramming of oral epithelial progenitor cells to cancer stem cells at single cell resolution in vivo,” en, bioRxiv, Jul. 2023.

A. Semaan, V. Bernard, J. Wong, Y. Makino, D. B. Swartzlander, K. I. Rajapakshe, J. J. Lee, **A. Officer**, C. M. Schmidt, H. H. Wu, C. L. Scaife, K. E. Affolter, D. Nachmanson, M. A. Firpo, M. Yip-Schneider, A. M. Lowy, O. Harismendy, S. Sen, A. Maitra, Y. A. Jakubek, and P. A. Guerrero, “Integrated molecular characterization of intraductal papillary mucinous neoplasms: An NCI cancer moonshot precancer atlas pilot project,” en, Cancer Res. Commun., vol. 3, no. 10, pp. 2062–2073, Oct. 2023.

A. Glencer, K. Ramalingam, N. Schindler, H. Mori, P. Ghule, K. Lee, D. Nachmanson, **A. Officer**, O. Harismendy, J. Stein, G. Stein, M. Evans, D. Weaver, C. Yau, G. L. Hirst, M. J. Campbell, L. J. Esserman, and A. D. Borowsky, “Tumor microenvironmental determinants of high-risk DCIS progression,” bioRxiv, 2024. doi: 10.1101/2023.12.01.569676. eprint: https://www.biorxiv.org/content/early/2024/03/19/2023.12.01.569676.full.pdf.

V. Burghi, J. S. Paradis, **A. Officer**, S. Adame Garcia, X. Wu, E. S. Matthees, B. Barsi-Rhyne, D. J. Ramms, L. Clubb, M. Acosta, P. Tamayo, M. Bouvier, A. Inoue, M. von Zastrow, C.

Hoffmann, and J. S. G. Gutkind, “Gαs is dispensable for β-arrestin coupling but dictates GRK selectivity and is predominant for gene expression regulation by β2-adrenergic receptor,” en, J. Biol. Chem., no. 105293, p. 105 293, Sep. 2023.

D. Nachmanson, **A. Officer**, H. Mori, J. Gordon, M. F. Evans, J. Steward, H. Yao, T. O’Keefe, F. Hasteh, G. S. Stein, K. Jepsen, D. L. Weaver, G. L. Hirst, B. L. Sprague, L. J. Esserman, A. D. Borowsky, J. L. Stein, and O. Harismendy, “The breast pre-cancer atlas illustrates the molecular and micro-environmental diversity of ductal carcinoma in situ,” NPJ Breast Cancer, vol. 8, no. 1, p. 6, Jan. 2022.

D. Nachmanson, J. Steward, H. Yao, **A. Officer**, E. Jeong, T. J. O’Keefe, F. Hasteh, K. Jepsen, G. L. Hirst, L. J. Esserman, A. D. Borowsky, and O. Harismendy, “Mutational profiling of micro-dissected pre-malignant lesions from archived specimens,” BMC Med. Genomics, vol. 13, no. 1, p. 173, Nov. 2020.

A. Papadas, G. Deb, A. Cicala, **A. Officer**, C. Hope, A. Pagenkopf, E. Flietner, Z. T. Morrow, P. Emmerich, J. Wiesner, G. Arauz, V. Bansal, K. Esbona, C. M. Capitini, K. A. Matkowskyj, D. A. Deming, K. Politi, S. I. Abrams, O. Harismendy, and F. Asimakopoulos, “Stromal remodeling regulates dendritic cell abundance and activity in the tumor microenvironment,” Cell Rep., vol. 40, no. 7, p. 111 201, Aug. 2022.

Q. Sun, Y. Wang, **A. Officer**, B. Pecknold, G. Lee, O. Harismendy, and J. S. Desgrosellier, “Stem-like breast cancer cells in the activated state resist genetic stress via TGFBI-ZEB1,” NPJ Breast Cancer, vol. 8, no. 1, p. 5, Jan. 2022.

D. O. Dele-Oni, K. E. Christianson, S. B. Egri, A. S. Vaca Jacome, K. C. DeRuff, J. Mullahoo, V. Sharma, D. Davison, T. Ko, M. Bula, J. Blanchard, J. Z. Young, L. Litichevskiy, X. Lu, D. Lam, J. K. Asiedu, C. Toder, **A. Officer**, R. Peckner, M. J. MacCoss, L.-H. Tsai, S. A. Carr, M.

Papanastasiou, and J. D. Jaffe, “Proteomic profiling dataset of chemical perturbations in multiple biological backgrounds,” Sci. Data, vol. 8, no. 1, p. 226, Aug. 2021.

A. S. Vaca Jacome, R. Peckner, N. Shulman, K. Krug, K. C. DeRuff, **A. Officer**, K. E. Christianson, B. MacLean, M. J. MacCoss, S. A. Carr, and J. D. Jaffe, “Avant-garde: An automated data-driven DIA data curation tool,” Nat. Methods, vol. 17, no. 12, pp. 1237–1244, Dec. 2020.

S. Gopal, Q. Lu, J. J. Man, W. Baur, S. P. Rao, L. Litichevskiy, M. Papanastasiou, A. L. Creech, K. C. DeRuff, J. Mullahoo, **A. Officer**, S. B. Egri, D. Davison, J. D. Jaffe, and I. Z. Jaffe, “A phosphoproteomic signature in endothelial cells predicts vascular toxicity of tyrosine kinase inhibitors used in CML,” Blood Adv., vol. 2, no. 14, pp. 1680–1684, Jul. 2018.

L. Litichevskiy, R. Peckner, J. G. Abelin, J. K. Asiedu, A. L. Creech, J. F. Davis, D. Davison, C. M. Dunning, J. D. Egertson, S. Egri, J. Gould, T. Ko, S. A. Johnson, D. L. Lahr, D. Lam, Z. Liu, N. J. Lyons, X. Lu, B. X. MacLean, A. E. Mungenast, **A. Officer**, T. E. Natoli, M. Papanastasiou, J. Patel, V. Sharma, C. Toder, A. A. Tubelli, J. Z. Young, S. A. Carr, T. R. Golub, A. Subramanian, M. J. MacCoss, L.-H. Tsai, and J. D. Jaffe, “A library of phosphoproteomic and chromatin signatures for characterizing cellular responses to drug perturbations,” Cell Syst., vol. 6, no. 4, 424–443.e7, Apr. 2018.

S. M. Horwitz, R. Koch, P. Porcu, Y. Oki, A. Moskowitz, M. Perez, P. Myskowski, **A. Officer**, J. D. Jaffe, S. N. Morrow, K. Allen, M. Douglas, H. Stern, J. Sweeney, P. Kelly, V. Kelly, J. C. Aster, D. Weaver, F. M. Foss, and D. M. Weinstock, “Activity of the PI3K-δ,γ inhibitor duvelisib in a phase 1 trial and preclinical models of t-cell lymphoma,” Blood, vol. 131, no. 8, pp. 888–898, Feb. 2018.

J. G. Abelin, J. Patel, X. Lu, C. M. Feeney, L. Fagbami, A. L. Creech, R. Hu, D. Lam, D. Davison, L. Pino, J. W. Qiao, E. Kuhn, **A. Officer**, J. Li, S. Abbatiello, A. Subramanian, R.

Sidman, E. Snyder, S. A. Carr, and J. D. Jaffe, “Reduced-representation phosphosignatures measured by quantitative targeted MS capture cellular states and enable large-scale comparison of drug-induced phenotypes,” Mol. Cell. Proteomics, vol. 15, no. 5, pp. 1622–1641, May 2016

Abstract of the Dissertation

Multimodal Data Integration Applied to Cancer Evolution and Signaling

by

Adam Officer

Doctor of Philosophy in Bioinformatics

and Systems Biology with a Specialization in Biomedical Informatics

University of California San Diego, 2024

Professor Pablo Tamayo, Chair

Professor J. Silvio Gutkind, Co-Chair

Tumors evolve from normal cells due to aberrant activation or repression of cell signaling. Our understanding of how these signaling pathways affect the phenotype of the malignant cells themselves, as well as the surrounding microenvironment, is lacking. As the era of precision medicine approaches, a more holistic understanding of cancer intrinsic signaling and its effect on the epigenome of tumor cells, as well as how these tumor cells interact with surrounding normal cells, is key to unlocking novel therapeutic targets and biomarkers. Through

three examples of cancer evolution and signaling I will show the value of integrative data modeling to better understand the biology of this disease.

First, in the context of breast cancer progression, I developed a microenvironment modeling approach to identify several novel intercellular signaling pathways that are altered in the transition from in situ to invasive disease. Second, in uveal melanoma, a cancer caused by aberrant GNAQ signaling, I generated and integrated transcriptional and protein level datasets together to nominate p53, and other pathways, as novel downstream targets of GNAQ. And third, in the context of head and neck cancer progression, I characterized the role of YAP activation in malignant transformation and cell signaling through EGFR and mTOR.

Chapter 1: Introduction

1.1: Cancers evolve from normal cells and precancerous lesions

The transition from normal cells, through premalignancies, to invasive carcinomas is an intricate process. This evolution may take years or even decades before lesions can be observed and treated[1]. Precancer is considered a stage between normal and cancer where there is a growth of abnormal cells that are confined “in situ” to one tissue structure. The transition from precancer to cancer occurs when these abnormal cells invade into the surrounding tissue. Individuals with precancer have a higher risk of developing invasive carcinomas later on, but there is emerging evidence that this is not always the case[2]. Studying precancer as an intermediate stage in the trajectory of cancer evolution offers insights into the evolutionary pressures and mechanisms of normal cell reprogramming that occur during tumorigenesis[3]. These pressures may remain constant over the evolution of these abnormal cells, or may ebb and flow as the surrounding tissue responds to their presence[4]. All cancer evolution shares some similar factors, all cancers are reprogrammed to divide more frequently, evade surveillance by the immune system, co-opt resources from the body for their growing nutrient needs and more. These hallmarks of cancer include a “hijacking” of normal growth and signaling pathways that lead to malignant transformation[5].

1.2. Cancer evolution associated with aberrant cell signaling

Hallmarks of cancer include factors intrinsic and extrinsic to the tumor cells themselves. All tumor cells need to have a fitness advantage over normal cells, most often this occurs via reprogramming of intrinsic signaling pathways. Across a wide range of cancers the most commonly altered signaling pathway is the DNA repair pathway. Somatic mutations in TP53, another gene involved in DNA damage response, are the most common, with 50% of all tumors carrying mutated p53[6]. Approximately 73% of all head and neck cancers carry mutations in TP53[7], but this is not the only p53 suppression mechanism in the oral cavity. The majority of

tumors without TP53 mutations were HPV-positive and the E6 and E7 proteins from HPV16/18 are believed to be p53 inhibitors[8]. An additional 30% of head and neck cancers have mutations in FAT1, a receptor involved in activating YAP downstream through the Hippo pathway[9]. A better understanding of the interplay between p53 activity and Hippo/YAP activation as it relates to tumorigenesis in head and neck cancer is needed[10].

One of the clearest cases of altered intrinsic cell signaling is in uveal melanoma. Uveal melanoma is the most common carcinoma of the eye with approximately 3000 cases per year in the United States alone[11]. Nearly 95% of all uveal melanoma cases present with mutations in the GNAQ or GNA11 gene[12,13]. These genes encode G-coupled proteins which bind to G-couple protein receptors (GPCR's) to transmit extracellular signals to intracellular signaling[14]. Notably, the GNAQ/11 mutations observed in uveal melanoma occur in the GTP binding pocket of these proteins which inhibits their ability to hydrolyse GTP meaning these proteins are stuck in their active conformation[15]. This aberrant activation of the GNAQ pathway has been used to identify novel therapeutic targets of downstream signaling members[16]. Prior work has identified MEK and FAK as downstream of GNAQ signaling, both of which are currently targets of ongoing clinical trials in this disease[17–20]. Further interrogation of the GNAQ signaling pathway will lead to a better understanding of downstream signaling members and shed light on novel therapeutic targets and mechanisms of tumorigenesis[21].

1.3. Holistic modeling of biological data

Although intrinsic signaling pathway alteration is frequently studied there are many other hallmarks of cancer that extend beyond gene/protein expression in the malignant tissue. Epigenetic dysregulation through chromatin remodeling or DNA methylation is a hallmark of cancer[22]. Characterizing and understanding the epigenetic mechanisms through which malignant cells evolve and survive will improve our ability to detect and treat cancer[23,24]. Epigenetic reprogramming can either be the cause of aberrant cell signaling or be caused by

aberrant cell signaling as many signaling pathways involve epigenetic and transcription factor regulation[25]. Cells control gene activity through epigenetic, transcriptional and protein-level events, all of which work systematically to generate a malignant, or non-malignant, phenotype. Modern experimental techniques such as ATAC-seq[26], DNase-seq[27], ChIP-seq[28], CUT&Tag[29] and CUT&RUN[30], have been developed to measure the epigenetic landscape. Integration of existing transcriptional data with epigenetic data gives a more holistic view of the systematic reprogramming of normal cells during tumor development[31].

This holistic modeling approach of cancer as a system extends beyond the cancer cell itself. With the advent of microenvironment-derived cancer treatment, such as immunotherapy, a better understanding of the tumor niche will inform better treatment options. Excreted factors from malignant cells have been shown to reprogram surrounding normal fibroblasts into "cancer-associated fibroblasts[32]." This phenotype is also observed with immune cells and several other stromal cell types in the tumor microenvironment. Cancer-dependent reprogramming of surrounding normal cell types has been implicated in cancer progression and invasion and is dubbed "reactive stroma[33]." Reactive stroma has been observed in prostate cancer, where it is most well studied, and is associated with a higher risk of invasion from the premalignant to invasive stage[34,35].

In the context of breast cancer there is a well-defined precancer, dubbed DCIS for ductal carcinoma in situ, that shares many similar features with prostate precancer. Notably, breast precancer and breast cancer have similar gene expression profiles with few gene expression changes in the malignant cells associated with the pro-invasive phenotype. These small changes in gene expression suggest that there may be malignant cell extrinsic factors at play, notably a change in the surrounding stroma that allows for an invasive phenotype. Indeed this is what is observed, in DCIS the basal myoepithelial layer of the mammary duct is still intact, while

in IDC the integrity of this layer is compromised allowing malignant cells to invade into the surrounding tissue.

1.4. Figures

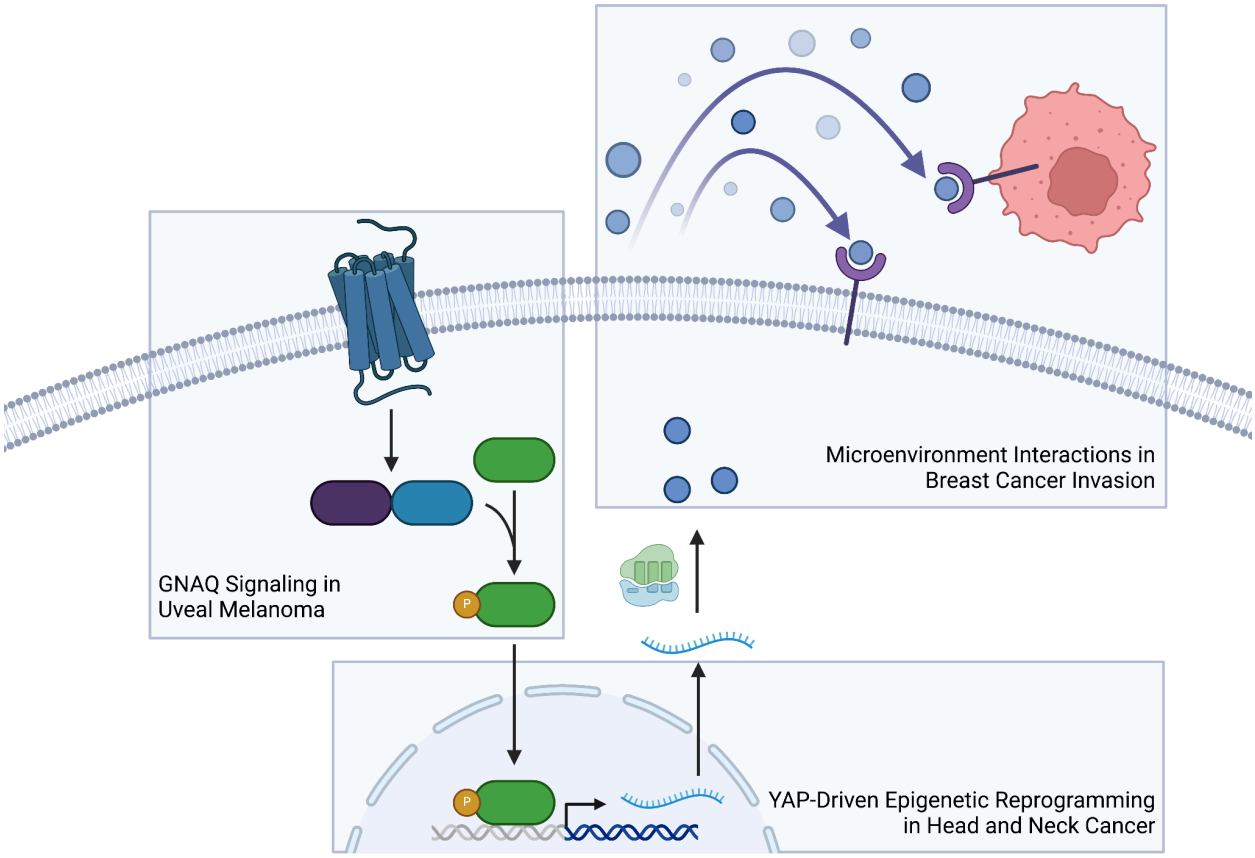

Figure 1.1: Graphical abstract of the dissertation

Chapter 2: Differential Analysis of Stromal-Epithelial Interactions between In Situ and Invasive Breast Cancer using Gene Expression Profiling

2.1. Introduction

Ductal carcinoma in situ (DCIS) is considered a non-obligate precursor to invasive ductal carcinoma (IDC), but the mechanisms driving invasion are not fully understood. Observational studies suggest that a large fraction of DCIS lesions may never progress to IDC[36,37]. Multiple genomic and transcriptomic studies have investigated the molecular differences between DCIS and IDC and highlighted key differences and similarities between these two diseases. Prior work has shown that DCIS synchronous with IDC lesions harbor similar copy number alterations (CNAs) and mutations.[38] Gene expression profiling of DCIS and IDC has identified similar molecular subtypes in both stages suggesting similarities between the premalignant and malignant stage[39]. Additionally, there are small gene expression changes between DCIS and IDC with only a handful of significantly differential genes identified across several studies.[40,41]

Among the gene expression differences identified, many involve signaling pathways participating in intercellular communication.[42,43] Within the epithelial compartment, extracellular matrix and cell adhesion processes are significantly changing between DCIS and IDC, suggesting that the surrounding intercellular space is being restructured. Several studies of microdissected histological regions suggest that more gene expression changes take place in the surrounding stroma as compared to malignant epithelial cells[40]. Follow-up studies have repeatedly identified changes in signaling pathways involved in oncogenesis and invasion, including extracellular matrix, cell adhesion and intercellular communication.[40,44] These results have somewhat been inconsistent, although not necessarily conflicting. Such gene expression micro-environmental studies have remained small and hard to validate due to the small size,

systemic archival of DCIS tissue, and the labor required for microdissection, hampering the generalization and further exploration of the findings.

With recent advances in methodologies and analytical tools, it is now possible to measure microenvironment signaling, specifically through cell-cell interactions (CCI), using gene expression from single-cell RNA-seq as well as bulk gene expression.[45,46] In general, CCI methods quantify coordinated expression between known ligand and receptor pairs and compare these values to a background model in order to identify significantly active CCI. However, the vast majority of these methods are not able to compare CCI between two different groups, in this case DCIS and IDC, to identify differential CCIs that may be mediating phenotypic changes accompanying breast cancer progression.

Here we present the results of a meta-analysis that compares stromal-epithelial CCI between DCIS and IDC using publicly available gene expression studies of micro-dissected stromal and epithelial regions from 163 patients. We characterize differences in gene expression and cell-cell interactions occurring at the interface between stromal and epithelial compartment, highlighting the contribution of microenvironment factors to breast cancer invasion, validating previous findings, prioritizing ligand-receptor pairs in their micro-environment context and assessing their contribution to overall survival.

2.2. Results

A large gene expression dataset was assembled from 6 independent gene expression studies and merged after correcting for batch effects. A total of 293 distinct gene expression profiles from epithelial and stroma regions were identified, including 46 from normal biopsies, 123 from DCIS, 108 from IDC and 16 from other breast malignancies representing a total of 163 distinct patients (Table 2.1). Using principal component analysis, samples corresponding to stromal and epithelial regions are well separated while study of origin cannot be distinguished (Figure 2.1A,B), therefore indicating the aggregation of the data was effective and did not

introduce spurious variation. Furthermore, hierarchical clustering suggests that within both stromal and epithelial compartments the normal regions are distinct from non-normal regions with small differences between DCIS and IDC observed (Figure 2.1C,D).

Gene expression signatures were used to classify epithelial regions by intrinsic subtypes (PAM50 - Table 2.2). Her2 and ER status was determined by expression of ERBB2 and ESR1, respectively (see Methods). We observed the expected enrichment of ER samples in Luminal subtypes, Her2 enrichment in Her2-like subtype and high prevalence of ER-/Her2- in Basal-like subtypes, indicating the validity of the expression based subtyping. Luminal A was more frequently observed in DCIS lesions (21/30, $p=1.7*10^{-3}$, Fisher exact test) and Luminal B was more frequently observed in IDC lesions (38/59, $p=5.6*10^{-5}$, Fisher exact test). Furthermore, there was a trend towards the enrichment of the Her2-like subtype in DCIS samples (17/92, p=0.045, Fisher exact test). Hence, while the original studies were not designed to faithfully represent a breast oncology clinic, the composition of the aggregated cohort was broadly consistent with previous studies[39] and likely captured the histological and molecular diversity of breast DCIS and IDC.

2.2.1. Changes in Stromal and Epithelial Cellular Composition

To explore the role of the tumor microenvironment in breast cancer invasion, we first used the bulk expression profiles to estimate and compare the abundance of multiple cell lineages in both the epithelial and stromal compartments with CIBERSORTx. Computational deconvolution of cell types by CIBERSORTx works optimally when provided with high-quality cell lineage reference gene expression profiles, ideally derived from large clusters of single-cell datasets. With no large-scale DCIS single-cell gene expression data publicly available and the risk of skewing the results for IDC biology, we chose instead to use the profiles of 86,136 cells observed in normal breast tissue specimens obtained from 28 breast reduction mammoplasty donors. This single-cell dataset consists of a breadth of both stromal and epithelial cell types, is

of high data quality and has been fully characterized[47]. The cell type clusters were aggregated into 6 broad cell lineages - luminal, myoepithelial, macrophages, lymphocytes,endothelial and fibroblasts - which were used to estimate cell abundance in normal, DCIS, and IDC stromal and epithelial regions.

Hierarchical clustering of the epithelial samples based on cell lineage abundance estimates identified four different clusters labeled according to their dominant lineage: Myoepi-Enriched, Infiltrated, Vascularized and Luminal-Enriched (Figure 2.2A). The Infiltrated and Vascularized clusters were enriched for IDC samples (16/78, p=0.01 and 17/78 p=0.018, respectively, Fisher Exact Test, Figure 2.2B). In contrast, the Myoepi-Enriched cluster was enriched in DCIS samples (28/92 p=0.009, Figure 2.2B). These observations suggest that epithelial IDC samples have proportionally more lymphocytes, macrophages and endothelial cells and fewer myoepithelial cells. A supervised univariate approach corroborated these results (Figure 2.2C) and further showed that fibroblasts are 67% more abundant in IDC compared to DCIS (FDR<0.1).

Similarly, hierarchical clustering of the stromal samples based on cell lineage abundance estimates subset of the cell abundance matrix identified four different clusters: Immune Hot, Fibroblast-Enriched, Vascularized and Epithelial-Enriched (Figure 2.2D). None of the clusters showed a significant enrichment between DCIS and IDC stromal regions, to the exception of a trend toward higher fraction of IDC in Fibroblast-Enriched stroma (9/30 p=0.059 Fisher Exact test, Figure 2.2E). Univariate analysis identified a median of 34% more luminal cells and 27% and 23% fewer lymphocytes and vascular and endothelial cells in the stroma adjacent to IDC regions (Figure 2.2F). This may be due to contamination from imperfect microdissection, as previously reported[40], or imperfect cell lineage annotation by the CIBERSORTx algorithm[48]. Importantly, the important differences in cellular composition of DCIS and IDC identified can be

a major confounding factor when trying to use bulk expression profiling to investigate molecular changes associated with invasion.

2.2.2 Stromal and epithelial functional differences between in situ and invasive lesions

In order to explore DCIS-IDC molecular differences beyond differences in cellular composition, we corrected the bulk gene expression matrix by relying on the linear properties of the CIBERSORTx method. Specifically, we derived a cell abundance corrected gene expression matrix from a normalized cell abundance matrix and the cell specific expression tensor. The resulting corrected gene expression matrix was then used to investigate functional differences using gene set enrichment analysis (GSEA).

We first compared epithelial regions. The large size of the assembled cohort allowed us to split the analysis according to ER and HER2 status. We compared the enrichment for 803 gene sets (Cancer Hallmark and Reactome from MSigDB[49]) across four subtypes of epithelial regions: ER+ (40 IDC samples, 41 DCIS samples), ER- (51 IDC, 38 DCIS), Her2+ (7 IDC, 17 DCIS), Her2- (71 IDC, 75 DCIS), recognizing the overlaps between ER and Her2 classifications. There were 215 gene sets whose expression was significantly altered (FDR<0.01) in at least one of the comparisons and these gene sets grouped according to 4 enrichment patterns (Figure 2.3A) : 1) Enriched in IDC included gene sets related to extracellular matrix organization and degradation which likely reflect how the epithelial cells actively participate in the degradation of the matrix in invasive disease; 2) Depleted in IDC included gene sets such as regulation of Tp53 activity, expression of ion and vitamin transporters or G protein signaling likely reflecting the stronger reliance of invasive disease on the loss of G1 checkpoint and changes in cellular metabolism and environment sensing; 3) Enriched in IDC in ER- and Her2- context - the largest group - included gene sets related to proliferation and cell cycle control, indicating strong proliferative differences from DCIS to IDC in triple negative cancer which represent the majority of this category; and finally 4) Depleted in ER+ IDC, were genesets

associated with Nucleotide Excision Repair, targets of Myc transcription factors and mitochondrial protein import suggesting that in the context of ER signaling, these processes were dispensable, or perhaps detrimental, to IDC growth.

In the stromal region a similar analysis was conducted albeit limited to 2 subtypes: ER+ (17 IDC, 14 DCIS) and ER- (13 IDC, 17 DCIS) with sufficient number of samples (Figure 2.3B). There were 167 gene sets whose expression was significantly altered in one or both comparisons (FDR<0.01). Notably, after correcting for cell type abundance differences, the differences observed reflect phenotypic changes in the stroma, or cell abundance differences not accounted for (e.g. B vs T lymphocytes). The significant gene sets included increased TCR, BCR and Interleukin signaling in the ER- context, likely reflecting an immunoactive state, in addition to the known higher infiltration. In the ER+ context, semaphorin and complement signaling were enriched in IDC. Finally, a number of pathways were depleted in IDC irrespective of the ER status. These processes are involved in the remodeling of cell-cell communication and adhesion - notably with Cancer Associated Fibroblasts - and mediated by signaling through Ephrin, Rho GTPase, C-type lectin or Notch1, all previously associated with breast cancer invasion. These pathways may therefore be dysregulated in fibroblasts in reaction to the invasion, and/or in the residual epithelial cells remaining at the invasive front.

The results generally validate previous gene expression studies. The size of the assembled cohort, combined with separation of epithelial and stromal regions allows a more refined perspective, especially distinguishing the relative importance of processes in different histological subtypes. In turn, the correction for cell-type abundance enabled the specific assessment of phenotypic signaling differences in cells in the micro-environment, pinpointing to the importance of interactions with fibroblasts in invasion.

2.2.3 Differences in Stromal-Epithelial Interactions between DCIS and IDC

Gene set enrichment analysis highlighted important differences in the functional states of the DCIS and IDC epithelial and stromal regions, but did not capture possible synergies or antagonism between them and how interaction between the malignant and non-malignant compartments can drive or repress invasion. We hypothesize that several of the changes observed are triggered by ligand-receptor interactions occurring between cellular compartments or between a cell and its environment. To investigate this, we selected 139 pairs of ligands and receptors from CellPhoneDB[50] and used their gene expression level to estimate the strength of their stromal-epithelial interaction in both IDC and DCIS contexts. Each known ligand-receptor pair was tested in both directions: epithelial ligand and stromal receptor (ES) and the reciprocal epithelial receptor and stromal ligand (SE), for a total of 278 interactions measured. Each interaction was estimated separately in DCIS and IDC context and significant differences were assessed using an empirical permutation strategy of the DCIS and IDC labels to compute the false discovery rate.

Using this framework a total of 10 interactions, involving 8 ligands and 8 receptors significantly different (FDR<0.1) between DCIS and IDC (Figure 2.4A). Six interactions were stronger in DCIS and all involved ephrins, themselves both ligands and receptors depending on the context[51]. Ephrins participate in cell localization, guidance and compartmentalization of tissues and have previously been implicated in breast cancer progression and invasiveness[52]. Of the 4 interactions estimated to be stronger in IDC, the WNT2-FZD2 interaction was the strongest in both the ES and SE directions (normalized difference in interaction score of -0.88 and -0.63, respectively, Figure 2.4A). This SE finding is consistent with the role of Fzd2 and Wnt signaling in mediating EMT, invasion and metastasis in cancer[53,54]. Furthermore, 2 out of 4 interactions stronger in IDC involved immune-receptors: CD86-CD28 (ES), is a co-stimulatory interaction required for TCR activation by T cells[55], while CD226-PVR (ES), is an interaction

known to sensitize NK and T-cell mediated lysis of tumor cells[56]. Given that the expression values were corrected to account for variation in cellular composition, these interaction changes may reflect a change of state of the cell in a particular histological region, or can be the consequences of sub-lineage differences not accounted for (T vs B lymphocytes).

In order to understand at which point in breast carcinogenesis these interactions shift we used the adjacent normal stromal and epithelial regions as an outgroup. By comparing the strength of the interactions between normal and DCIS or normal and IDC we can determine how early in oncogenesis such interactions are being remodeled. Of the 278 interactions tested, there were 98 and 102 interactions that are identified as significantly changing between normal-DCIS and normal-IDC, respectively. This is an order of magnitude more interactions than those changing in the DCIS-IDC comparison (Figure 2.4B). This suggests that the CCI remodeling occurs early in oncogenesis, prior to invasion. This was the case for all of the ephrin-mediated interactions that showed a monotonic decrease from normal to DCIS to IDC suggesting that these interactions are progressively changing between both the normal to DCIS transition as well as the DCIS to IDC transition (Figure 2.4C). In contrast, the CD226-PVR (ES) interaction was identified as significantly stronger in IDC compared to either DCIS or normal, timing this interaction to change with the DCIS to IDC transition. A similar observation was made for CD86-CD28 (ES), indicating that the activation of immune-related CCI is a later event in the disease progression.

2.2.4 Association with Breast Cancer Survival

The changes of CCI between DCIS and IDC, while identified from the microdissection of local disease, may also mediate metastatic spread and have important consequences on survival. We analyzed The Cancer Genome Atlas (TCGA) to determine, within an IDC cohort, if these interactions had prognostic significance. In absence of microdissection, we approximated

ES and SE CCI scores using the bulk expression profile. To account for known prognostic factors such as ER positivity, tumor stage, advanced age, immune infiltration, we used Cox proportional hazards (CoxPH) regression to determine the prognostic value of candidate CCI score (Figure 2.5A). Of the 10 ligand-receptor interactions that were significantly different between DCIS and IDC, we identified one interaction, EPHB4-EFNB1, that was significantly associated with survival (HR=1.47, p=0.04, Figure 2.5B) and the 3 year survival of patients patient with high EPHB4-EFNB1 expression product was worse (76% vs 88% - Figure 2.5C). Notably the interaction between the two is required as neither the expression of EPHB4 nor EFNB1 alone significantly affected survival (p=0.18 and 0.34, respectively, Figure 2.5D,E,F).

2.3. Discussion

As gene expression data is widely available for a range of diseases of the breast, we can answer more nuanced questions using comparative microenvironment modeling approaches. In particular, questions related to the molecular and functional landscape surrounding early cancer lesions and their role in cancer progression are becoming more pressing.[3,57] We performed a meta-analysis of published gene expression studies that specifically separated the stromal and epithelial compartments in DCIS and IDC contexts. With a sufficient number of samples, we investigated functional differences using state of the art methods accounting for differences in intrinsic subtypes and cellular composition and gaining insights on the role of stromal-epithelial interactions.

Prior work has suggested modest gene expression differences between DCIS and IDC[40], and while our observations help revisit that postulate, hierarchical clustering confirmed that in comparison to normal breast, DCIS and IDC are highly similar. The first analytical improvement we brought was to correct for cellular composition using computational deconvolution. Indeed, while the histological regions studied were microdissected, it does not mean they represent a

pure population of cells. The cell-type relative abundance across samples and regions followed expected trends in cellular composition. For example in stromal regions, it is well established that TNBC and Her2 positive subtypes have higher levels of immune cells.[39,58] In epithelial regions, infiltrating lymphocytes and macrophages, endothelial cells from increased vascularization, or fibroblasts at the invasive front can all be present.[58,59] Similarly, DCIS lesions tended to have higher prevalence of myoepithelial cells.[60] The computational deconvolution has limitations, especially in comparison to emerging single-cell or spatial profiling approaches. In particular, at the resolution considered, we are not able to distinguish between B- and T-lymphocytes, macrophage polarity, or fibroblast subtypes, all possibly contributing to invasion and immune escape. On the other hand, such high resolution assays have not yet been conducted on large cohorts and, for single-cell RNA-seq, are critically dependent on the availability of fresh specimens, which are impossible to get for most pure DCIS cases.

Akin to gene expression comparison, gene set enrichment analysis have been also conducted between DCIS and IDC gene expression datasets, including by the very studies that were sourced to assemble our meta-analysis. Correcting for cellular abundance, even at limited resolution, helped identify intrinsic differences in pathway activities, in contrast to differences due to variable cellular composition which may have driven the expression differences observed in previous studies. Notably Lee et al[41] and Knudsen et al[61] previously identified progression-associated expression patterns of genes expressed in the vasculature. Our results do not identify these same trends, likely due to our cell abundance normalization procedure. Knudsen et al additionally identified genes from a myoepithelial signature as significantly higher in invasive disease. Instead, we observed lower levels of myoepithelial gene expression in IDC as a result of cell abundance correction. In agreement with prior studies, processes linked to proliferation and ECM were dysregulated in IDC. Indeed, these processes are likely dominant in

the epithelial compartment, which is the least affected by the cell abundance correction we implemented.

Thanks to the aggregation of multiple datasets and the resulting large number of samples, we were able to distinguish dysregulated processes by subtypes, revealing compelling contrasts in both epithelial and stromal functional states. The most prominent finding in the epithelium was that the majority of subtype-dependent dysregulation of cell cycle and DNA repair occurred in the ER- and Her2- context, and by extension is likely driven by the triple negative subtype. Estrogen receptor status has been associated with DNA repair capacity,[62] but it is interesting to observe that this dependency may be limited to IDC and that proliferation and DNA repair may become increasingly dysregulated as the disease progresses. As a corollary, it may suggest that ER- and/or Her2- DCIS may not be as proliferative and genetically unstable as their invasive counterparts. Interestingly, telomere maintenance was also specifically induced in IDC in the ER- context, consistent with the evidence that altered telomere maintenance associated with invasion and stemness.[63,64] Our results suggest that the dysregulation of this process is more prominent in IDC than DCIS.

In the stroma, we identify increased Notch or Rho GTP-ase signaling in DCIS, irrespective of subtypes. Notch signaling has been implicated in both breast cancer progression and inhibition[65] as well as in tumor-stroma crosstalk.[66] Notch signaling in stromal cells (fibroblasts or immune cells) contributes to carcinogenesis and drug resistance.[67] Such pleiotropy of Notch signaling as well as the lack of information on which Notch ligands and receptors are at play renders the interpretation of our results difficult, but it may suggest that down regulation of certain stromal Notch signals are associated with invasive disease, perhaps supporting immune escape. Subtype-specific dysregulated processes in the stroma were dominated by immune signaling, stronger in IDC in ER- context, but as mentioned before may

be due to residual, or uncorrected differences in cell abundance, which would predominantly affect ER- disease, where immune-infiltration is higher.

The evaluation of ligand-receptor interactions between stroma and epithelium is the most novel aspect of our report as it required the availability of expression datasets from microdissected regions in a large enough number of samples to support a robust statistical analysis. Such discovery efforts were further aided by the restriction to well-curated pairs of ligands and receptors. While much more comprehensive databases of ligand and receptor exist,[45] they are not well validated (e.g. computationally derived) or overwhelmed with LR pairs irrelevant to breast cancer biology (e.g. neuronal communications). In contrast, we chose to restrict the analysis to experimentally proven LR pairs, for which one member was expressed in the studied dataset. While this approach is likely less sensitive, evaluating fewer pairs and relying on known biology, it facilitated the implementation of an empirical framework to test for the significance of the interactions via permutation and increased our confidence in the differential interaction observed. Notably, previous approaches to measure cell-cell interactions did not necessarily propose solutions for differential interaction testing, hence making it hard to compare interaction scores between samples and conditions. Importantly, conducting cell abundance deconvolution prior to cell-cell interaction scoring can help normalize heterogeneous samples and hence measure genuine changes in cell-cell interaction instead of those driven by a shift in cell composition. Similar approaches have been used in bulk expression studies,[68] but, here again, the availability of single-cell RNA-seq could alleviate the need for such computational tricks and directly measure LR interaction from pure populations of cells as conducted by multiple methods recently published.[69–71] As mentioned above, however, single-cell gene expression data is not available for pure DCIS specimens, and approaches like ours are needed to overcome the resulting lack of data. Emerging spatial profiling methods may soon replace or complement this approach by eliminating the imprecise and poorly scalable

micro-dissection step and also by measuring the physical proximity and transcriptional state of cells in the stroma and epithelium. Spatial profiling could provide an ideal method to validate and extend our initial observation of the remodeling of cell-cell interaction.

The WNT2-FZD2 interaction was the most notably changing stromal-epithelial interaction, stronger in IDC in both SE and ES directions, suggesting reciprocal signaling of these molecules. Increased expression of Wnt family members has been shown to induce breast cancer in mouse models[72] but this gene family is also involved in normal breast development and lactation. Prior gene expression studies have identified dysregulation of the Wnt pathway in breast cancer compared to healthy tissue[73] and an upregulation of Wnt signaling in early carcinogenesis. Furthermore, the level of WNT2-FZD2 interaction is not different between DCIS and adjacent normal epithelium, suggesting the change occurs later in breast cancer progression or specifically affects patients diagnosed with IDC. Of the nine Wnt-Fzd family interactions investigated, this is the only pair that is significantly higher in IDC indicating that this specific exocrine Wnt signal may be relevant to the IDC phenotype.

Two immune-mediated interactions that were significantly stronger in IDC compared to DCIS-- CD28-CD86 (ES) and CD226-PVR (ES) - highlight immunological differences between DCIS and IDC. Both of these interactions are stimulatory interactions involving T-cells and/or NK cells and increase the susceptibility of cancer cells to death, hence consistent with their possible role in more advanced invasive disease. The CD28-CD86 (ES) directions and cell-type specificity suggest that T-cells in the epithelium are being stimulated by antigen presenting cells in the stroma, hence reflective of an ongoing active immune response[74]. Similarly, CD226-PVR (ES) stronger interaction in IDC is also mediated by immune cells from the epithelial compartment. Acknowledging that immune-cell abundance differences cannot be completely corrected computationally, these findings likely reflect the overall higher immunoreactive microenvironment environment observed in IDC. In general the IDC environment is more

immunosuppressive and with T cells that are less primed for activation[58]. Consistently, the stronger CD86-CD28 interaction observed in IDC could mediate FOXP3+ regulatory T-cell homeostasis[75] contributing to this immunosuppressive environment. Of note, the somewhat redundant CD80-CD28 interaction, which affects the same T cell homeostatic pathway, is not identified as significantly differential across normal, DCIS or IDC.

The most consistent result to come out of the microenvironment modeling approach is the 6 DCIS-enriched ephrin interactions. These are of particular interest to breast cancer as different ephrin signaling members have been implicated both as promoters or suppressors of tumor invasion.[52,76] Ephrins themselves can be ligands or receptors based on their function and have several pleiotropic binding partners. These ephrin interactions represent 4 of the 49 unique ephrin interactions in CellPhoneDB. By using the normal samples as an outgroup we identified that 3 out of 4 the significantly changing ephrin interactions have significant monotonically decreasing trends from normal to DCIS to IDC suggesting a progressive erosion of ephrin interactions as the disease progresses. Remarkably, EPHB4-EFNB1 interaction was associated with poor survival in TCGA, indicating that this interaction may have prognostic value at multiple stages of the disease. The true prognostic value after a DCIS diagnosis would however need to be properly validated, ideally through independent stromal and epithelial gene or protein expression measurements. Given the slow and rare progression, such cohorts would have to be established prospectively across multiple institutions. Finally, given some of the nuanced phenotypes mediated by ephrins, it may be difficult to get precise mechanistic insights or determine of this erosion is truly causal to the disease progression in patients - or rather the results from high growth and proliferation fueled by other drivers and further experiments and validation will be required.[51,52,77]

Despite its innovative aspects, it is important to highlight some of the most important limitations of an analysis like the one presented here. First and foremost is the reliance on gene

expression data. Most CCI interaction analysis rely on the expression of ligand and receptor genes, but the CCI are mediated at the protein level and the two do not always correlate. While changes in gene expression may be sufficient to capture long-term changes that affect cell identity and cell states, more transient changes may be mediated only at the protein level, or by changes in subcellular distribution or post-translational modifications, including glycosylation, which are ignored by our approach. Cell biology experiments will be required to faithfully capture the CCI changes. Furthermore, the study included cross-sectional samples from DCIS or IDC patients, and therefore may not truly capture changes that may occur longitudinally in individuals. This limitation is systematic in the vast majority of studies of DCIS, as longitudinal sampling cannot be performed. But since not all DCIS patients progress to IDC, there is increasing evidence that host specific factors contribute to DCIS progression or containment [78,79] and cross sectional studies cannot capture their contributions.

Despite the above limitations, the stromal-epithelial modeling approach considerably enriched the context of previous findings, characterizing better the source of expression differences between DCIS and IDC from a functional and micro-environmental viewpoint. Additional studies using spatial profiling and protein based cell-cell interaction observation will be needed to confirm the findings and determine their prognostic value or therapeutic utility to prevent breast cancer progression.

Chapter 2, in full, is a reprint of the material as it appears in biorXiv. A. Officer, A. M. Dempsey, L. M. Murrow, Z. Gartner, P. Tamayo, C. Yau, and O. Harismendy, Differential analysis of stromal-epithelial interactions between in situ and invasive breast cancer using gene expression profiling, bioRxiv, 2022. Additional figures and text were taken from A. Officer, E. Armingol, O. Harismendy, and N. E. Lewis, Deciphering cell-cell interactions and communication from gene expression," Nat. Rev. Genet., 2021. The dissertation author was the primary researcher and author of these papers.

2.4. Methods

2.4.1. Preparation of the gene expression datasets

Gene expression datasets were obtained from each of the 6 studies available on NCBI Gene Expression Omnibus (GEO) database.[40,41,44,61,80,81] The studies were selected with the following criteria: (1) gene expression profiling by Affymetrix microarray (2) laser-capture microdissection or manual macro-dissection of breast tissue specimens into stromal and epithelial compartments and (3) included samples of invasive ductal carcinoma (IDC) or ductal carcinoma in situ (DCIS). Where available, CEL files were downloaded directly from GEO, background corrected and normalized using the Robust Multi-Array Average (RMA)[82] and multiple probes for the same gene were collapsed to the mean value. Where CEL files were unavailable the processed RMA values on GEO were used. The RMA expression levels from each of the separate datasets were combined and batch effects were corrected using ComBat from the sva package[83] accounting for differences in tissue compartment and disease state in the design matrix ("mod" argument). The resulting batch-corrected RMA matrix was used as-is for all subsequent analyses except for the principal component analysis (PCA) and hierarchical clustering where z-scoring was performed prior to analysis. Cosine distance with complete linkage was used as the distance metric for hierarchical clustering.

2.4.2. Molecular subtype classification

The "molecular.subtyping" function in the genefu package (version 2.22.1) was used to determine the PAM50 intrinsic subtypes for the epithelial regions of DCIS and IDC biopsies from normalized expression values. Her2 status and ER status were assigned based on the expression of the ERBB2 ($z>1$ is positive) and ESR1 genes ($z>0$ is positive), respectively.

2.4.3. Accounting for cellular composition

Cell abundance measurement: CIBERSORTx[48] was used to estimate cell abundance from the normalized expression values. The reference dataset was derived from single-cell

RNA-seq data from 28 normal breast reduction mammoplasty donors for a total of 86,136 cells assigned to 10 cell type clusters. Histologically related clusters were aggregated into 6 major cell lineages: fibroblasts, myoepithelial cells, luminal cells, lymphocytes, macrophages and vascular and endothelial cells. The CIBERSORTx HiRes Docker (release date Sep 2019)[84] was used in S-mode to quantify cell abundance from the batch-corrected expression matrix, G_{IxJ}, with genes I and samples J, using the cell lineages C, identified in the single-cell RNA-seq data. This method estimates a matrix of cell lineage specific expression, A^{j}_{IxC}, and a vector of cell lineage abundance, B^{j}_{C}, according to the following equation for every sample j in J: $G_{Ixj} = A^{j}_{IxC}$ B^{j}_{C}. CIBERSORTx uses the marker genes for each cell lineage in the single-cell RNA-seq data to constrain this equation and identify bulk gene expression variation that can be attributed to each of the cell lineages provided..

Normalization of the gene expression matrix: The gene expression matrix was normalized to account for differences in cell abundance between samples. First, a median cell abundance vector for each tissue compartment is computed by taking the median of all cell abundance vectors within stromal and epithelial samples separately giving B^{str}_{C} and B^{epi}_{C}. We expect the stromal and epithelial compartments to have different cell composition so we normalize them separately. Then the G_{IxJ} matrix is normalized for cell abundance differences within each tissue compartment giving G^{N}_{IxJ} according to the following procedure: if a sample, j, is from the stromal compartment then $G^{N}_{Ixj} = A^{j}_{IxC}$ B^{str}_{C}, otherwise: $G^{N}_{Ixj} = A^{j}_{IxC}$ B^{epi}_{C} where $A^{j}_{I,C}$ is the corresponding cell lineage specific expression matrix for sample j. Finally, G^{N}_{IxJ} is log-transformed to return to RMA space.

2.4.4. Gene Set Enrichment Analysis

MSigDB Hallmarks (H) and REACTOME (C2 REACTOME) collections were used in all gene set centric analyses. Gene set enrichment analysis (GSEA) was performed using cell-abundance normalized expression values from $G^{N}_{i,j}$. The GSEA implementation in gseapy[85]

was used to identify gene sets significantly changing between DCIS and IDC for each subtype and histological region considered. Gene sets with a minimum of 15 genes that were detected in the dataset were retained for permutation analysis. The significance was determined using a permutation test on the phenotype labels for a total of 10,000 permutations. Gene sets with an FDR of less than 0.01 were considered significant.

2.4.5. Differential Cell-Cell Interaction Measurement

The CellPhoneDB ligand-receptor (LR) database (N=279 pairs) was filtered for detected interactions between ligands and receptors in the cohort, resulting in 139 ligand-receptor pairs. Two separate ligand-receptor interaction models were considered: one for stromal-epithelial (SE) and another for epithelial-stromal (ES). For each LR pair in the ES model, the interaction score was computed from the outer sum of the ligand RMA expression in the epithelium and the receptor RMA expression in the stroma (and vice-versa for the SE model). The median score among DCIS-DCIS sample pairs and IDC-IDC sample pairs were compared. The significance of the observed DCIS-IDC difference was determined in comparison to a null distribution of interaction scores differences obtained from 10,000 permutations randomly shuffling within epithelial and stroma regions. Two-sided p-values were computed from this null distribution and corrected for multiple hypothesis testing using the Benjamini-Hochberg procedure. Interactions with an empirical FDR of less than 0.1 were considered significant. The same analytical workflow was used to compare IDC to normal and DCIS to normal.

2.4.6. Survival Analysis in TCGA Breast Cancer Cohort

The PanCan TCGA gene expression data RNA-seq V2.0 was downloaded from the PanCan GDC portal. Clinical annotations from the same website were used. Univariate Cox proportional hazards (CoxPH) models were used to select important clinical features for the final multivariate model. Analyses were performed in Python 3.9.12 using the lifelines package version 0.27.0. Hormone receptor results were only used if the sample was labeled “Positive” or

"Negative". T, N and M pathological staging values were simplified to remove sub-stages (e.g. T1a becomes T1), and age was thresholded at 60 years old in line with previous studies.[86] The multivariate CoxPH model included only clinical variables that were significant in univariate analyses. For each ligand-receptor pair tested the interaction score was derived from the product of RSEM values and binarized into "Low" and "High" groups according to the median interaction score.

2.5. Tables

Table 2.1: Distribution of disease states of samples used in meta-analysis

	Normal	DCIS	IDC
Epithelial	29	92	78
Stromal	17	31	30

Table 2.2: Dataset HER2 and ER status by disease state

	DCIS	IDC
ER Positive	41	40
ER Negative	38	51
HER2 Positive	17	7
HER2 Negative	75	71

2.6. Figures

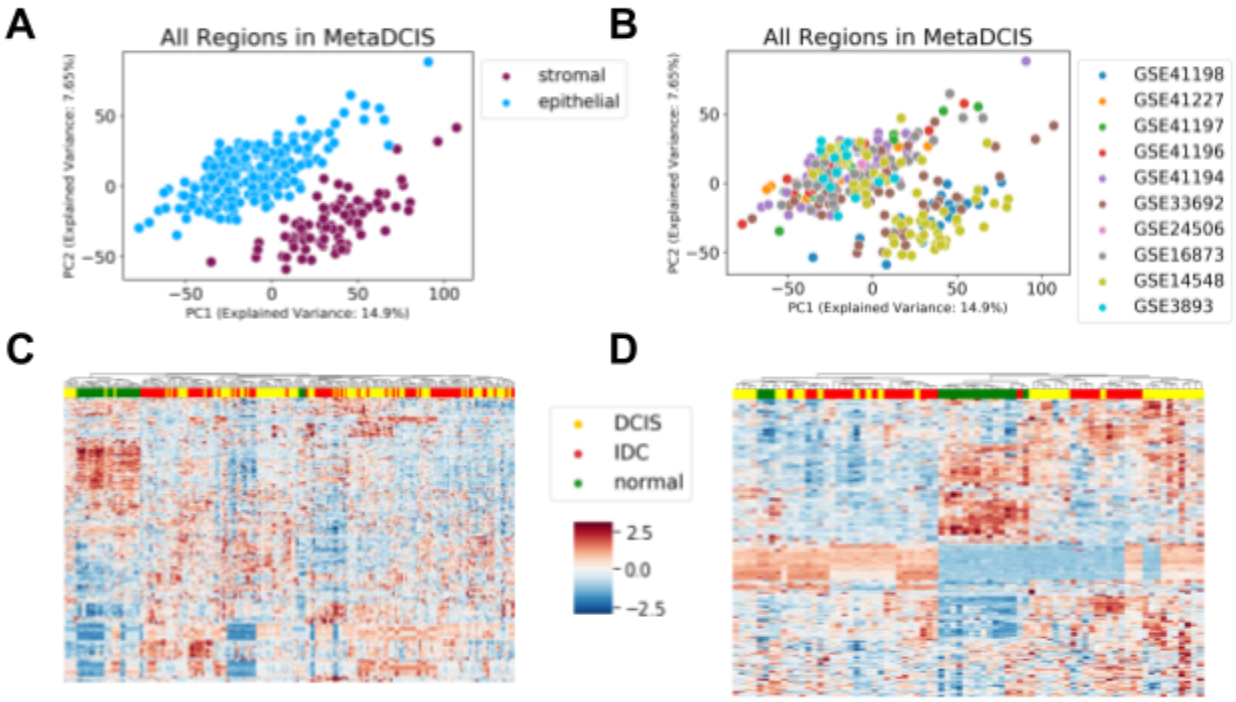

Figure 2.1: Unsupervised classification of gene expression profiles. (A-B) Principal Component Analysis (PCA) of all gene expression datasets included in the study after batch-effect correction and colored by tissue compartment (A) or source dataset (B). (C-D) Hierarchical clustering of the epithelial (C) or stromal (D) expression profiles using the expression of the 500 most variable genes.

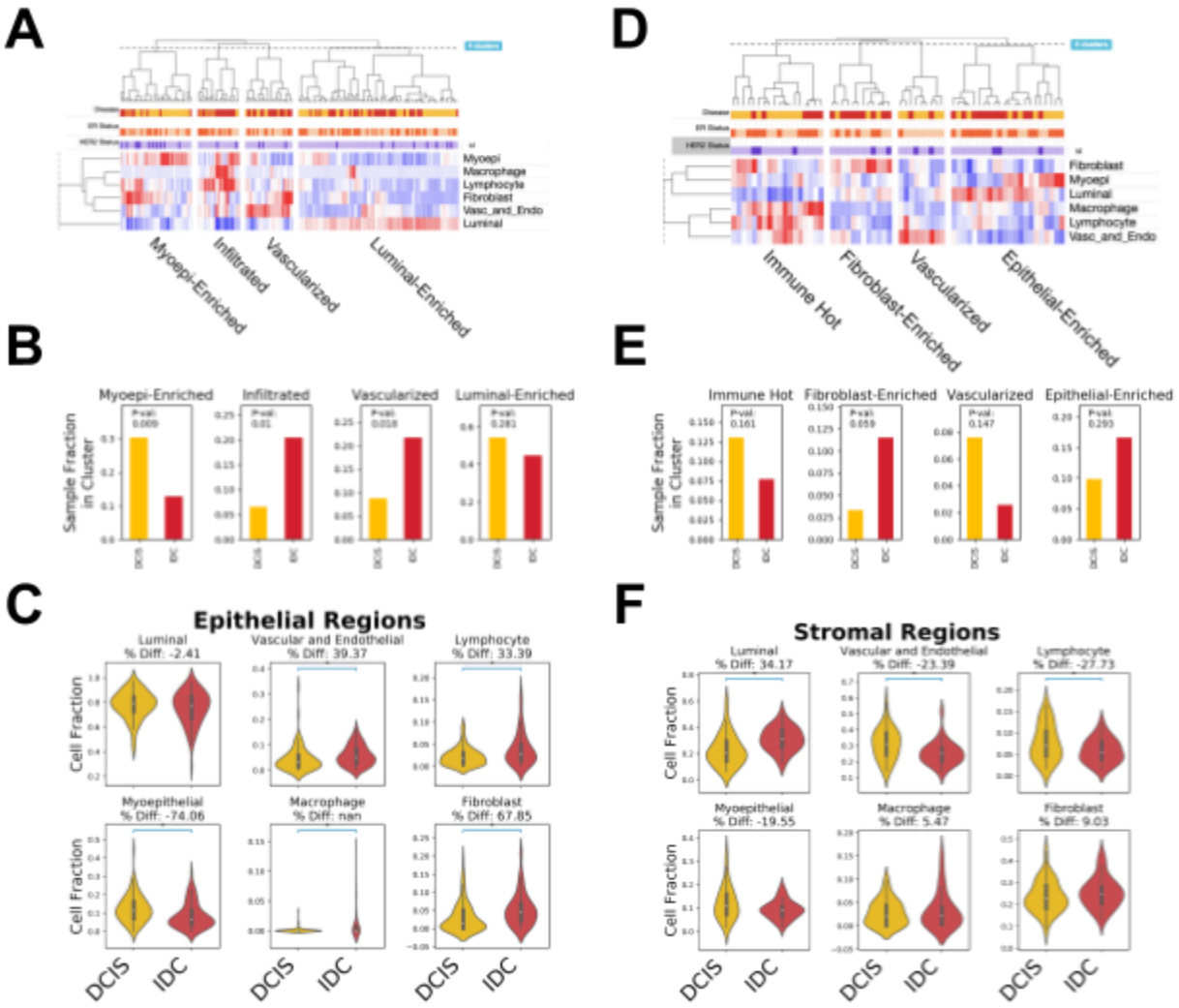

Figure 2.2: DCIS-IDC expression differences associated with cellular composition. (A,D). Hierarchical clustering of epithelial (A) or stromal (D) regions according to the prevalence of 6 cell lineages as measured using CIBERSORTx. Clusters are labeled according to the dominant pattern in cellular composition. (B,E) Prevalence of DCIS (yellow) and IDC (red) region in each cluster for epithelial (B) or stromal (E) regions. P-values were obtained using the Fisher Exact test. (C,F) Univariate comparison of fraction of each cell lineage between DCIS and IDC in epithelial (C) and stromal (F) regions.

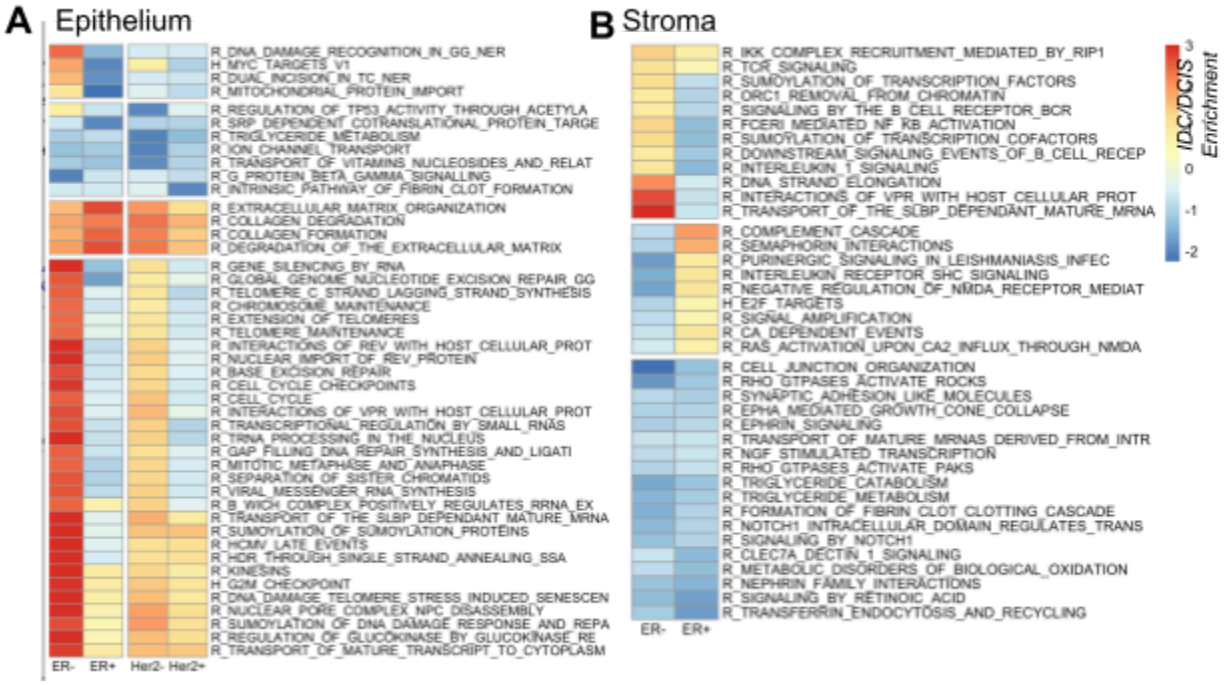

Figure 2.3: Context specific functional differences between DCIS and IDC. (A-B) Heatmap of GSEA normalized enrichment scores (positive/red: enriched in IDC, negative/blue: enriched in DCIS) from the differential gene expression between DCIS and IDC epithelial (A) or stromal (B) samples from different subtypes (columns, ER and Her2 overlap). Gene sets significant (FDR<0.05) in at least one context and with highest absolute enrichment scores are displayed (epithelium: 2.4, stromal: 1.8) . Gene set names are truncated and/or abbreviated (R: Reactome, H: Hallmarks) for visualization clarity.

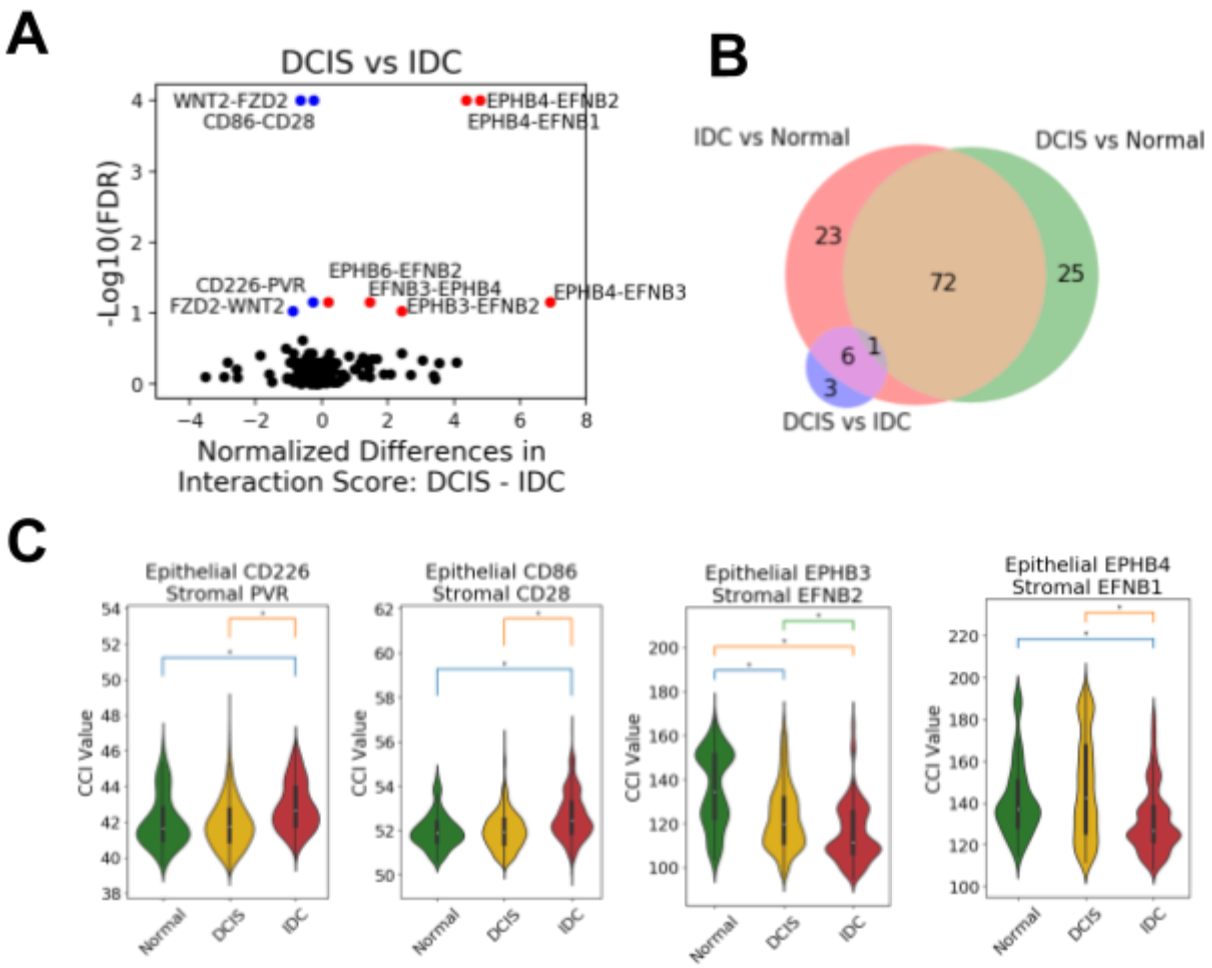

Figure 2.4: Differences in stromal-epithelial cell-cell interactions. (A) Differences in DCIS-IDC interaction scores (x-axis) for all tested LR pairs (points) in both SE and ES directions. Significance was estimated using False Discovery Rate (y-axis) and significant LR pairs stronger in DCIS (red) or IDC (blue) are labeled). (B) Venn-diagram displaying the number and overlap of significantly different interactions observed in DCIS-IDC, DCIS-normal and IDC-normal comparisons. (C) Distribution of the stromal-epithelial LR Interaction scores (y-axis) measured in normal (green), DCIS (orange) and IDC (red) context for 4 selected LR pairs significant in the DCIS-IDC comparison. (*) indicate FDR<0.1.

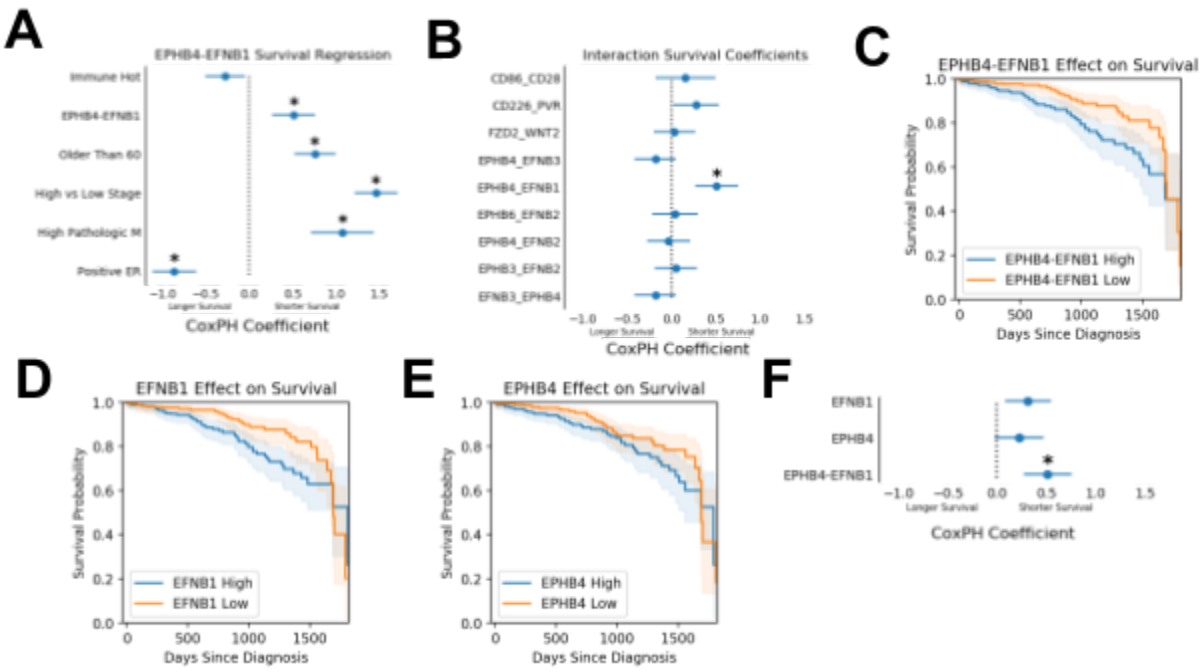

Figure 2.5: Prognostic significance of LR interactions in IDC: (A) Results of a multivariate Cox proportional hazards regression analysis (Hazard ratio: x-axis) including EPHB4-EFNB1 co-expression in the TCGA breast cancer cohort. Clinical covariates were selected using univariate analysis. Error bars represent the 95% confidence interval (*p < 0.05). (B) Overall survival hazard ratio (x-axis) attributed to the LR pairs significant in DCIS-IDC comparison and included in 9 independent CoxPH IDC survival models. (*p=0.04). (C-E) Kaplan-Meier plots showing the difference in survival probability of IDC patients with high and low EPHB4-EFNB1 co-expression (C), EFNB1 (D) or EPHB4 (E) single gene expression (CoxPH p=0.04, p=0.18 and p=0.36, respectively, CoxPH). (F) CoxPH coefficients for separate and combined interaction models for the EPHB4-EFNB1 interaction.

.

Chapter 3: Oncogenic GNAQ Signaling Suppresses p53 in Uveal Melanoma

3.1. Introduction

G protein–coupled receptors (GPCR) are the largest family of cell surface proteins with over 800 members[87], and their dysregulation contributes to some of the most prevalent human diseases[88–90]. GPCRs represent the largest family of targets for approved drugs.

A recent analysis of human cancer genomes revealed that nearly 30% of human cancers occur with mutations in G proteins or GPCRs[16]. In particular, uveal melanoma is enriched for GNAQ or GNA11 mutations with approximately 93% of uveal melanoma lesions harbor activating mutations in these genes. GNAQ and GNA11 genes encode the alpha subunits Gaq and Ga11 of the heterotrimeric G protein complex. Uveal melanoma is diagnosed in about 3,000 adults in the United States every year, and is the most common primary cancer of the eye in adults and the second most common melanoma subtype after skin cutaneous melanoma[11]. These activating mutations in GNAQ/11 cause aberrant signaling into downstream pathways. Prior work in our lab has characterized several canonical and non-canonical proteins involved in this signal transduction including YAP, PI3K/AKT and PKC[17,18,21,91].

Although the majority of uveal melanoma patients will be treated by radiation or surgery, approximately 50% of the patients will metastasize, primarily to the liver, within 5 to 10 years after diagnosis[92]. Inactivating mutations or copy loss of the BAP1 gene, which is located on chromosome 3p21, are strongly associated with metastasis in patients with uveal melanoma[93], supporting that BAP1 functions as a metastasis suppressor. Most patients with metastatic uveal melanoma (mUM) are refractory to current chemotherapies and immune checkpoint blockade. Ultimately, the majority of advanced disease patients succumb within a year due to the suboptimal efficacy of these treatments, often combined with severe toxicities, underlying the

high unmet medical need for new therapeutic strategies. Most of the recent clinical research efforts in uveal melanoma have focused on inhibiting the Gaq classical signaling pathway, which signals downstream into MAPK (Figure 3.1). Targeting of this pathway with MEK inhibitors such as selumetinib and trametinib are currently being evaluated in clinical trials for efficacy towards metastatic uveal melanoma.

Prior studies of the GNAQ signaling pathway have focused on specific hypotheses and targeted screens to identify additional downstream signaling members. One prior experiment performed an unbiased screen using RNA-seq to identify response to GNAQ inhibition, but the time point chosen, 3 days of drug treatment, showed strong epigenetic reprogramming and a repression of the cell cycle[94]. To identify earlier signaling events we propose to use the GNAQ inhibitor tool compound, YM-254890[95], to treat uveal melanoma cells acutely and measure early transcriptional and protein response using RNA-seq and RPPA.

3.2. Results

3.2.1 GNAQ Inhibition in Uveal Melanoma Cell Lines Induces Transcriptional Response through MAPK, PKA and DNA Damage Response

To identify transcriptional targets of the GNAQ signaling pathway we treated OMM1.3, MP46 and MP38 cells with 1 uM of YM-254890 or vehicle control for 3 hours and generated RNA-seq data from these samples in quadruplicate. PCA analysis of the protein-coding genes in this dataset show clear cell line separation suggesting cell of origin is the main source of variation in this dataset, not GNAQ pathway inhibition. To account for this cell line heterogeneity we used DESeq2 to identify differentially expressed genes within each cell line independently for a total of 3 independent DESeq2 models. For OMM1.3 we found 3701 differential genes, MP46 2571 genes and MP38 2618 differential genes. 729 genes were differentially expressed (adjusted p-value<0.1) in all three cell lines. Among the downregulated consensus genes were MAPK-associated genes such as DUSP6, FOS and JUN suggesting MAPK signaling and

downstream AP-1 activity is inhibited by GNAQ inhibition. Additionally we see GPCR negative regulators SPRY2, RGS16 and RGS1 downregulated by drug treatment showing that general GPCR signaling is altered. CCN1 and CCN2 levels were also decreased by YM-254890 showing that YAP, which is downstream of the non-canonical pathway, activity is also repressed. Several immune-related genes were downregulated including REL, a member of the NFkB transcription factor family, as well as the chemokines CXCL1 and CXCL2. Among the genes upregulated by GNAQ inhibition included several members of the DNA damage response pathway including POLD3, GEN1, RMI2 and SPIDR suggesting that this pathway is repressed by GNAQ signaling.

We also performed GSEA against the REACTOME collection to identify signaling pathways affected by GNAQ inhibition. Unsurprisingly, the majority of downregulated signaling pathways were MAPK or AP-1 associated further confirming the role of MAPK in GNAQ signaling. Several gene sets for NFkB signaling were also downregulated in this comparison, consistent with the gene-level analyses presented earlier. Upregulated pathways included several chromatin-associated gene sets and processes associated with DNA damage response showing that systematic alteration of DNA damage response is associated with GNAQ inhibition.

To further characterize the transcriptional response to GNAQ inhibition we applied the VIPER method to infer regulon transcription factor activity. This method shows the transcriptional activity of several AP-1 members is downregulated including ATF2, ATF4, and JUNB as well as TEAD1, the DNA binding domain of the YAP/TEAD complex. REL activity is also inhibited suggesting NFkB signaling is also detected in this complementary analytical method. DNA damage response regulon TP53 is activated by GNAQ inhibition. Novel to this computational method CREB1 activity is observed to be repressed by GNAQ inhibition.

3.2.2. Protein-Level Measurements Show Altered MAPK, PKA Signaling and DNA Damage Response

To validate the transcriptional response next RPPA was performed using the same drug treatment: 1 uM YM-254890 for 3 hours and vehicle control, in 92.1, OMM1.3 and MP46 cell lines. Differential protein expression was analyzed using a multiple linear regression to control for cell line effects and identify shared changes across all cell lines. A total of 265 proteins were upregulated by GNAQ inhibition and 234 were downregulated. The top 15 up- or down-regulated proteins are shown in Figure 3.2. Among the most differentially regulated proteins include pFAK and pERK confirming the known downstream members of the GNAQ signaling pathway. Additionally several proteins and phosphosites involved in DNA damage response were upregulated including pATM, gamma-H2AX and p53 suggesting damage increases upon GNAQ inhibition, consistent with what was observed in the RNA-seq dataset. Conversely pMDM2 and MDM2 were downregulated confirming the regulatory role of this protein in p53 activity. Furthermore pCREB1 is downregulated by GNAQ inhibition confirming the results found in the RNA-seq transcription factor activity analysis.

3.2.3. GNAQ Associated with p53 and MDM2 in Public Datasets

To identify potential gene-gene interactions between GNAQ and p53 The Cancer Genome Atlas (TCGA) is analyzed for mutational co-occurrence between these two genes. Notably not all GNAQ or GNA11 mutations are considered, only those that are considered “hotspot” activating mutations in the GTP binding pocket[96]. As can be seen in Figure 3.4 there is significant mutual exclusivity between p53 mutations and GNAQ/11 hotspot mutations: of 83 tumors with GNAQ/11 mutations only 3 also contain a p53 mutation compared to the 10670 GNAQ/11 WT tumors of which 4338 are p53 mutant ($p<0.001$, Fisher exact test).

Using this RNA-seq dataset I define a GNAQ signature as the top 100 most downregulated genes in response to GNAQ inhibition. This GNAQ signature response is then

measured in TCGA and the DepMap dataset. In TCGA there is a clear correlation between the GNAQ signature and the Hallmark p53 signature suggesting p53 activity is correlated with GNAQ activity (Figure 3.4, $p<0.001$, linear regression). The DepMap dataset can be split into two subgroups based on TP53 mutational status. As shown in Figure 3.3 there is a negative correlation between the GNAQ signature and the gene essentiality of TP53 (Figure 3.4, $p<0.001$, linear regression) suggesting cell lines with high levels of GNAQ signaling rely less on intact p53 protein to survive. Notably this relationship is nonexistent in CCLE lines where TP53 is mutated ($p=0.06$). TP53 mutations are believed to alter the transcriptional targets of p53 by inhibiting the ability for p53 to bind to DNA on its own and instead causing p53 to bind different transcription factors to induce a different phenotype. This difference in correlation values (Figure 3.4, $p<0.001$, linear regression) in the DepMap is consistent with a model in which GNAQ signaling alters p53 activity, but in the presence of TP53 mutations that link is broken. Furthermore, looking at the relationship between GNAQ signaling and MDM2 essentiality an opposite effect is observed: that cells with high GNAQ signaling are more reliant on MDM2 to divide (Figure 3.4, $p<0.001$, linear regression) because MDM2 is repressing p53 protein levels.

Additionally when comparing uveal melanoma cell lines to all other cell lines in the DepMap data portal the UVM cell lines are less reliant on TP53 for growth and more reliant on MDM2 for growth (Figure 3.4). In fact, these two genes are among the top 5 most differentially essential genes in the entire comparison, meaning their effect on the survival of uveal melanoma cell lines is unique.

3.2.4. UVM Cell Lines Have Repressed p53

To experimentally validate if uveal melanoma cell lines, which have highly active GNAQ, do indeed have repressed p53 through MDM2 we treated 92.1 and OMM1.3 cells with nutlin 3a, an inhibitor of the interaction between MDM2 and p53, which is known to activate p53 signaling. These cells were treated for 24 hours with vehicle control or an increasing dose of nutlin 3a.

In both OMM1.3 and 92.1 the nutlin 3a treatment increased p53 protein levels in a dose-dependent manner as well as increased the level of p21, a known downstream target of p53. Additionally, as MDM2 expression is induced by p53 in a positive feedback loop, total MDM2 and phospho-MDM2 were both increased in the same manner. Furthermore we can observe that these cells actively induce cell death by measuring cleaved-PARP which can be observed at nutlin 3a doses of ~1 uM and above.

3.2.5. p53 Activation Induces Cell Death Synergistically with Canonical Pathway Inhibitors

To determine if nutlin 3a treatment, and subsequent p53 activation, is phenotypically relevant to the growth of uveal melanoma cell lines we performed drug synergism assays with nutlin 3a and other GNAQ pathway inhibitors. Based on previous work we identified three targeted inhibitors in uveal melanoma: MEK, PKC and FAK, and chose the drugs trametinib, sotrastaurin and VS-4718, respectively, to inhibit these proteins.

Drug synergism assays were performed in OMM1.3 and synergism was determined using the Loewe score. All three drugs were synergistic with p53 activation, though the synergism scores with trametinib and sotrastaurin were significantly higher than the synergism scores for VS-4718. This result suggests that p53 activation through the non-canonical pathway affects the growth of uveal melanoma cells more than the canonical pathway.

Drug synergism can be induced through senescence or growth inhibition, to confirm that these cells are actively inducing cell death; the level of caspase activity was measured using Caspase-Glo. Cells treated with nutlin 3a and MEK inhibitors, VS-6766 and trametinib, caused a strong increase in caspase activity that is higher than the additive effect from each drug alone. This result suggests that these cells are actively dying in response to p53 activation combined with canonical pathway inhibition.

Chapter 3 is currently being prepared for publication. A. Officer, R.D. Cervantes-Villagrana, S. Lubrano, N. Arang, P. Tamayo, J.S. Gutkind. The GNAQ-p53 Signaling Axis in Uveal Melanoma. The dissertation author was the primary researcher and author of this paper.

3.3. Methods

Cell Culture

92.1 and OMM1.3 cells were cultured in RPMI (catalog no.: D6429; Sigma) supplemented with 10% fetal bovine serum (FBS) (catalog no.: F2442; Sigma) and 1% penicillin/streptomycin (catalog no.: A5955; Sigma). MP46 and MP38 were cultured in the same RPMI media supplemented with 20% FBS.

Sample Preparation

To determine the transcriptional response to GNAQ inhibition all the cell lines were plated in 6-well plates. The following day, the cells were serum starved overnight and the cells were incubated with 1 μM YM-254890 or vehicle for 3 hours. After drug treatment the cells were washed with PBS, total RNA was isolated with an RNeasy Mini Kit (catalog no.: 74104; Qiagen) including an on-column DNase I digestion and quantified using a Nanodrop ND-1000 (Thermo Scientific). Library preparation and paired-end 150 bp (catalog no.: PE150; Illumina) RNA-Seq was performed by Novogene Corporation.

RNA-seq alignment and analysis

The quantification of transcripts was calculated using Salmon[97] (version 1.7.0). Differential gene expression analysis including quality control, model fitting, and hypothesis testing was conducted using DESeq2[98]. Prerank gene set enrichment analysis was performed on the DESeq2-estimated LFC values as previously described using the gseapy[85] prerank function (version 1.0.0) in Python (version 3.10.5). The MSigDB[49] REACTOME collection was

compared, and significance was assessed using 10000 permutations, and q-values less than 0.25 were considered significant.

TCGA and CCLE Analysis

TCGA gene expression data, RNA-seq V2.0, was downloaded from the GDC portal. CCLE gene expression data was downloaded from the DepMap data portal[99]. The GNAQ signature was defined as the top 100 downregulated genes upon GNAQ inhibition. Signatures were measured by single-sample GSEA in the TCGA and CCLE datasets using the ssgsea function in the gseapy package. Gene set levels were z-scored and correlation measured using a linear regression. For the p53 mutation analysis an interaction term was used to quantify the difference between the two regression lines. P-values less than 0.05 were considered significant.

RPPA Data Generation and Analysis

Lysis in recommended buffer, quant using Biorad Bradford, normalized concentration, sent to MD Anderson using their standard workflow[100].

Nutlin dose-response in OMM1.3 and 92.1

OMM1.3 and 92.1 cell lines were grown in RPMI containing 10% FBS supplementation. Prio rot treatment cells were starved overnight in serum-free media and treated with nutlin 3a for 24 hours.

Drug Synergism with the Loewe Method

The Loewe additivity synergy model[101] was used to determine possible synergistic effects of selected kinase inhibitor combinations. Briefly, cells were seeded at a density of 3×10^3 to 5×10^3 cells/well in 96-well white plates (CulturePlate; PerkinElmer Inc.). Cells were treated with either single inhibitors or combinations thereof using eight different dilutions of each inhibitor and in technical triplicate. Cell viability was measured, after 72-hour treatment, with the

CellTiter Glo Luminescent Cell Viability Assay on a Spark microplate reader (Tecan). The Synergy Score (SS) was calculated using the SynergyFinder+ web application[102].

Caspase-Glo Assay

Cells were seeded at a density of 10000 cells/well in 96-well white plates. After 24 h, drug treatment or vehicle was added and cells were assayed as indicated. Apoptosis was measured using the Promega CaspaseGlo3/7 Assay System (G8090) as per manufacturer's instructions.

3.4. Discussion

Mutations in the GNAQ or 11 oncogenes are the main driver for uveal melanoma. Because of the limited mutational landscape in UVM the aberrant GNAQ signaling, and Gq-coupled GPCR's, is likely the key driver of tumor initiation and progression. A deeper understanding of the signaling mechanisms of GNAQ/11 will reveal novel pathway-based targeted therapies for UVM.

The work presented herein suggests that using an unbiased approach to dissect the GNAQ signaling network can nominate novel downstream targets, chiefly MDM2/p53. On the transcriptional and protein level we show p53, PKA and NFkB signaling are all altered by GNAQ inhibition in UVM cells. On a gene, pathway, transcription factor and protein level these results are consistent and warrant additional investigation. Further interrogation into the p53 axis shows that two UVM cell lines have repressed p53 through MDM2-dependent ubiquitination. Phenotypic analysis of the GNAQ-p53 relationship suggests that cell growth is arrested by p53 activation and, using drug synergism experiments, this effect is stronger downstream of the non-canonical pathway.

Taken together, these results suggest that MDM2 phosphorylation is driven by GNAQ activation, likely through signaling downstream of FAK. FAK has been previously shown to

signal downstream into PI3K/mTOR, which itself is known to activate MDM2 through direct phosphorylation. MDM2-dependent p53 degradation is a mechanism through which uveal melanoma tumors can functionally inhibit p53 activation without having to resort to mutational inactivation. Uveal melanoma tumors have low mutational burden, but even so we can shown that activation of p53 is enough to induce apoptosis in uveal melanoma cells, suggesting that there is an evolutionary pressure against p53 mutations in uveal melanoma as p53 mutations are known to inhibit the regulatory activity of MDM2.

3.5. Figures

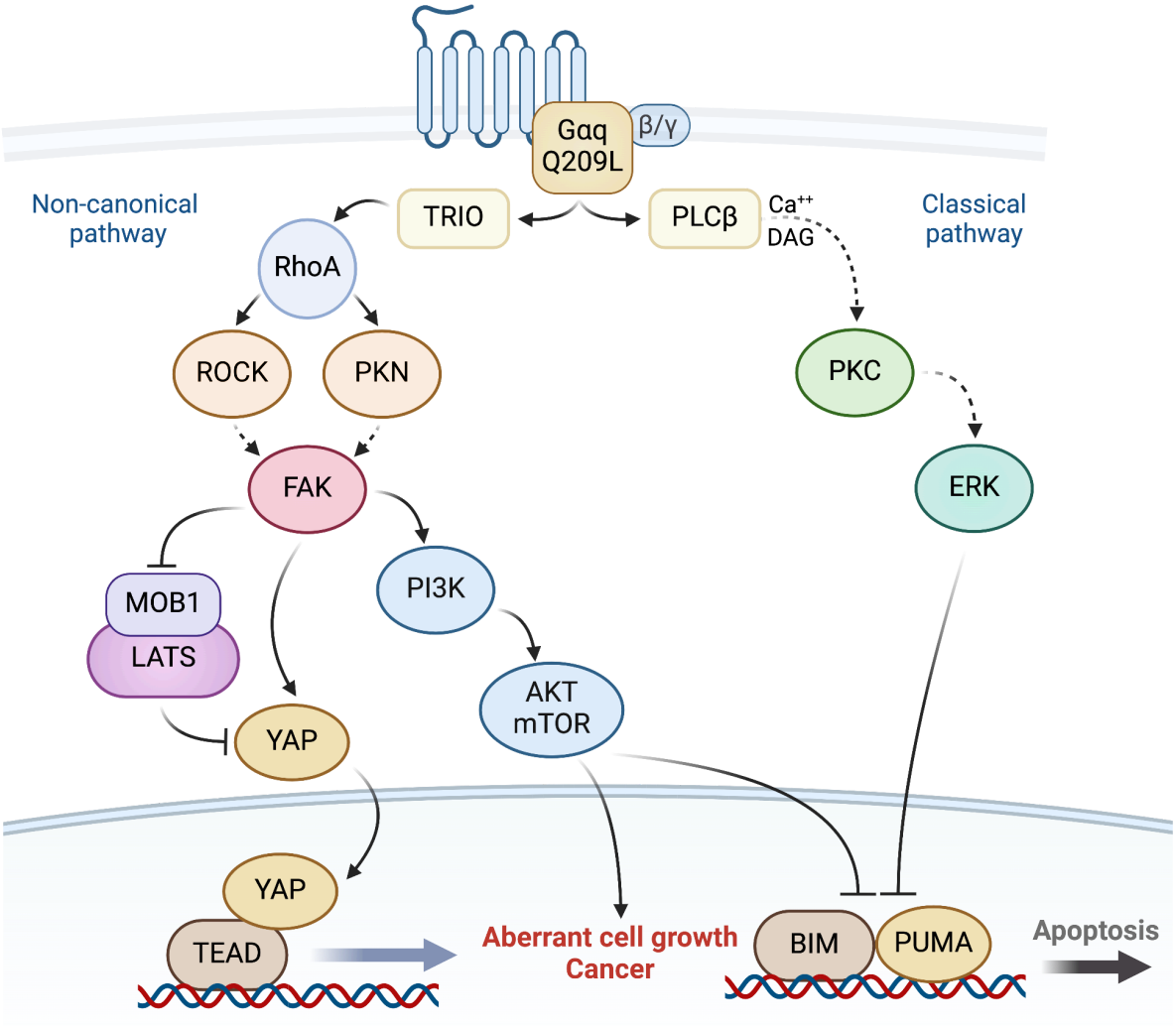

Figure 3.1: Schematic of the GNAQ Signaling Pathway

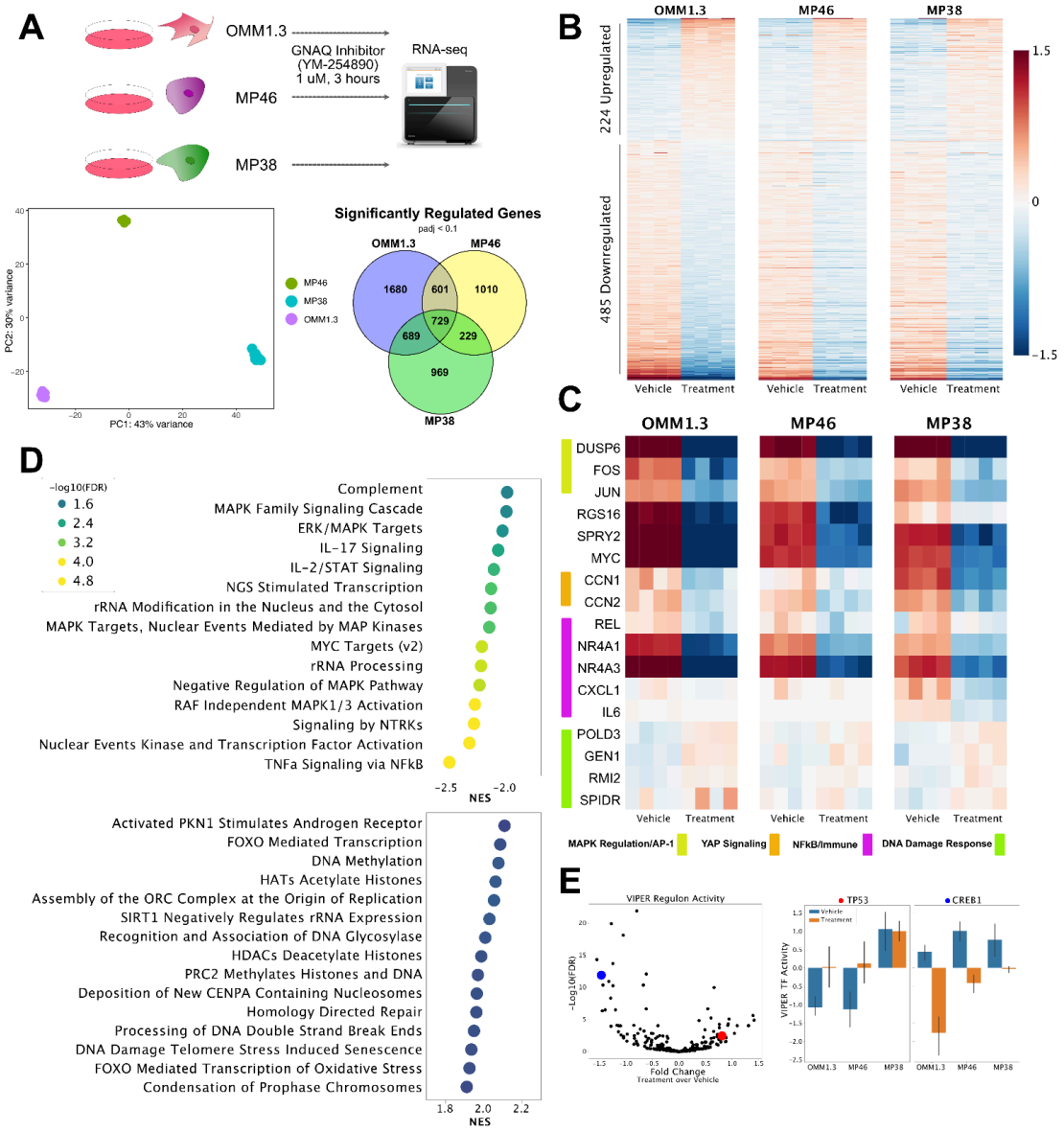

Figure 3.2. RNA-seq analysis of transcriptional response to GNAQ inhibition. A) RNA-seq experimental design, clustering analysis and differential expression overlap. (B) Heatmap of median normalized gene expression. (C) Gene expression of selected genes representative of modulated pathways. (D) Gene Set Enrichment Analysis (GSEA) results showing MAPK downregulated (left) and DNA damage response upregulated (right) upon GNAQ inhibition. (E)

VIPER analysis for transcriptional regulon activity shows increased p53 and inhibited CREB1 activity upon GNAQ inhibition.

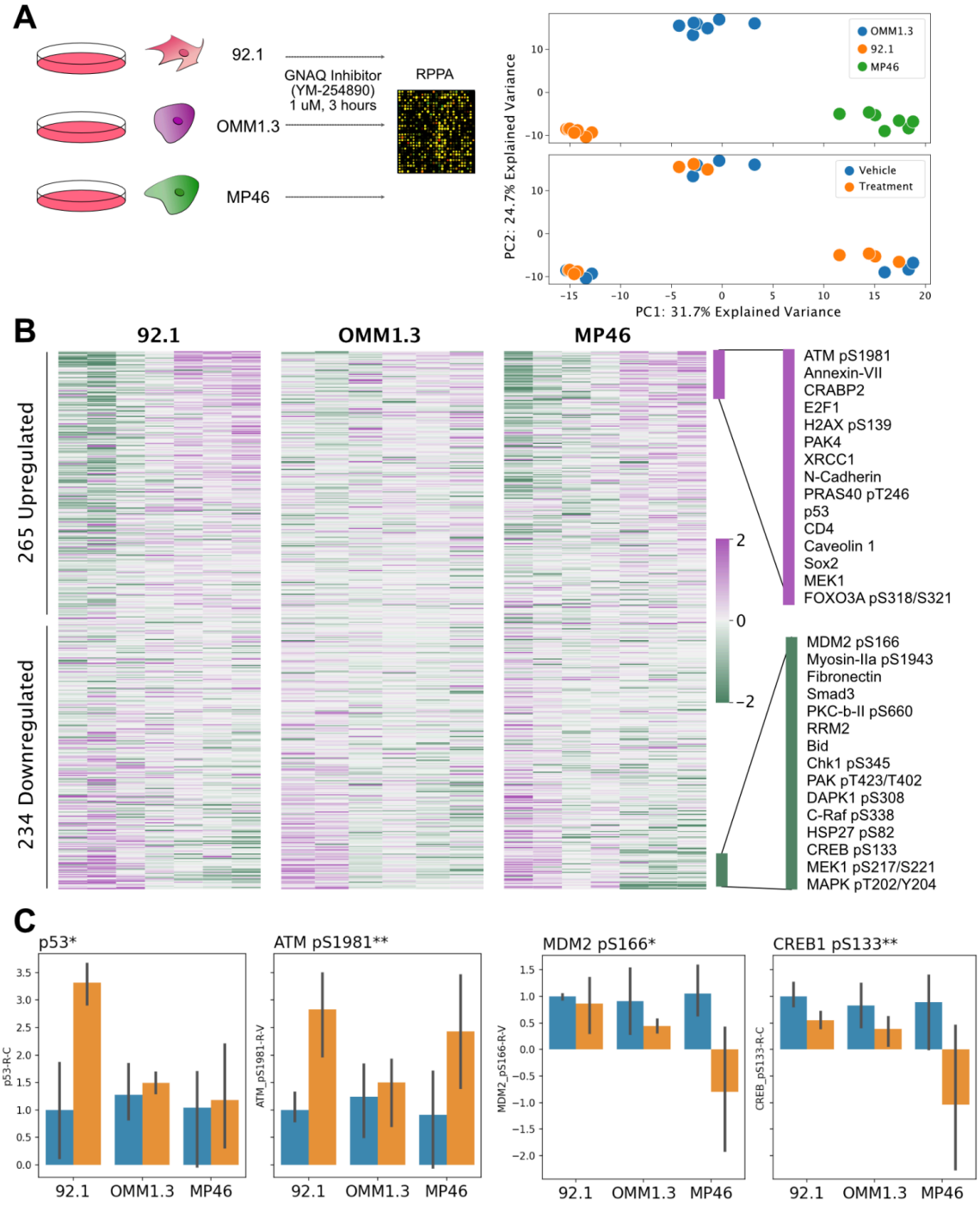

Figure 3.3. RPPA analysis of GNAQ inhibition. (A) PCA analysis showing cell line differences largest factor. (B) Heatmap of differential proteins showing top 10 up/downregulated. (C) Selected protein level barplot of DNA damage response and PKA associated proteins.

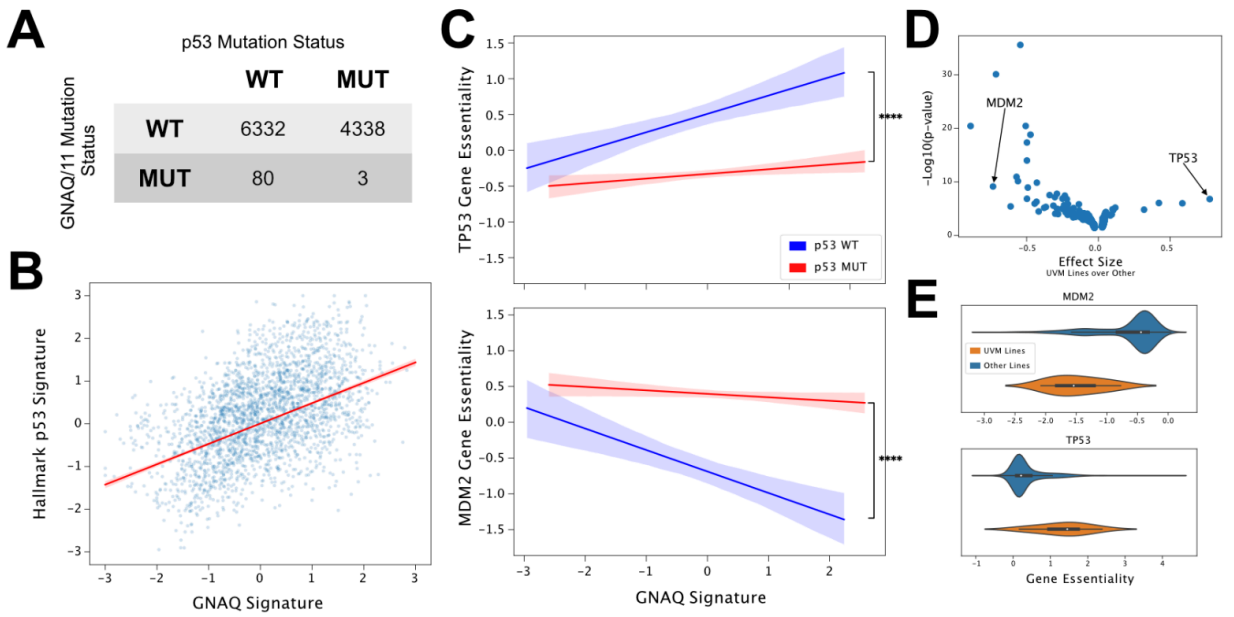

Figure 3.4. GNAQ signature associated with p53 Signaling. (A) Mutations in p53 and GNAQ/11 mutually exclusive in TCGA ($p<0.0001$, Fisher Exact test). (B) GNAQ signature correlated with p53 signature in TCGA dataset ($p<0.0001$, linear regression). (C) GNAQ signature correlated with MDM2 and TP53 gene essentiality, but only in p53 WT cell lines. p-values computed using an interaction test for differences of slopes. (D) MDM2 and TP53 gene essentiality unique to uveal melanoma (UVM) cell lines, they are 2 of the top 3 most differentially essential genes in the DepMap dataset. (E) Violinplots showing essentiality differences between UVM and other cell lines.

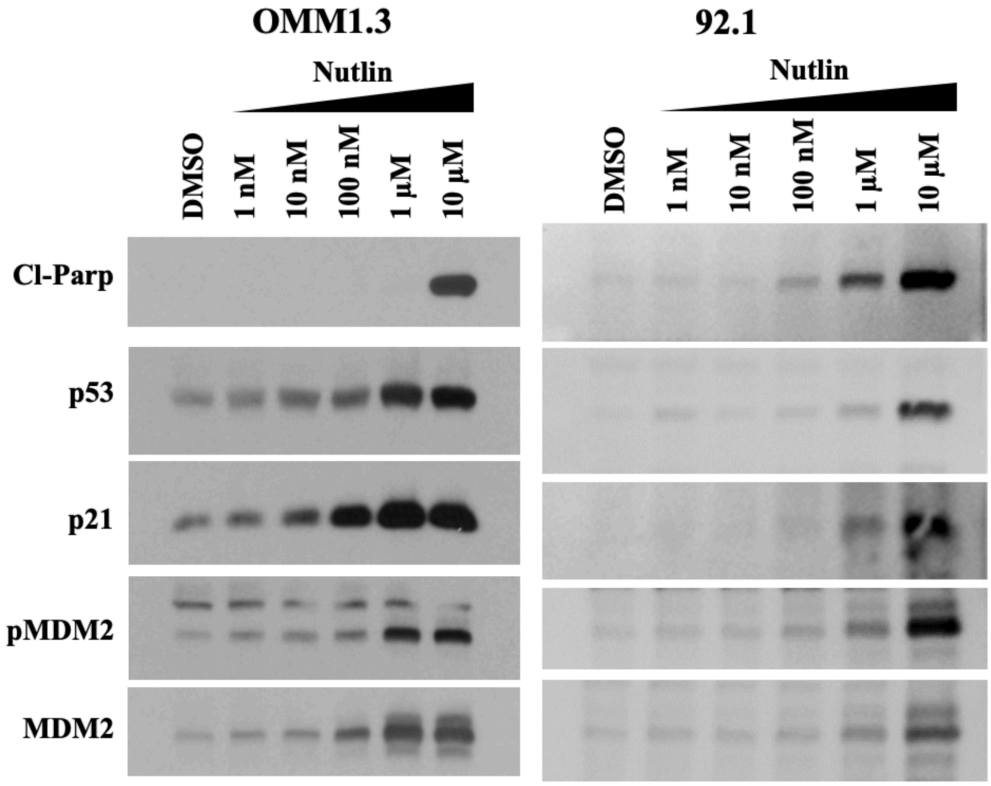

Figure 3.5. Uveal melanoma cell lines repress p53. Western blot showing at the 24 hour timepoint pMDM2, p21 and p53 levels increase in response to nutlin (MDM2 inhibitor) in OMM1.3 and 92.1 cell lines.

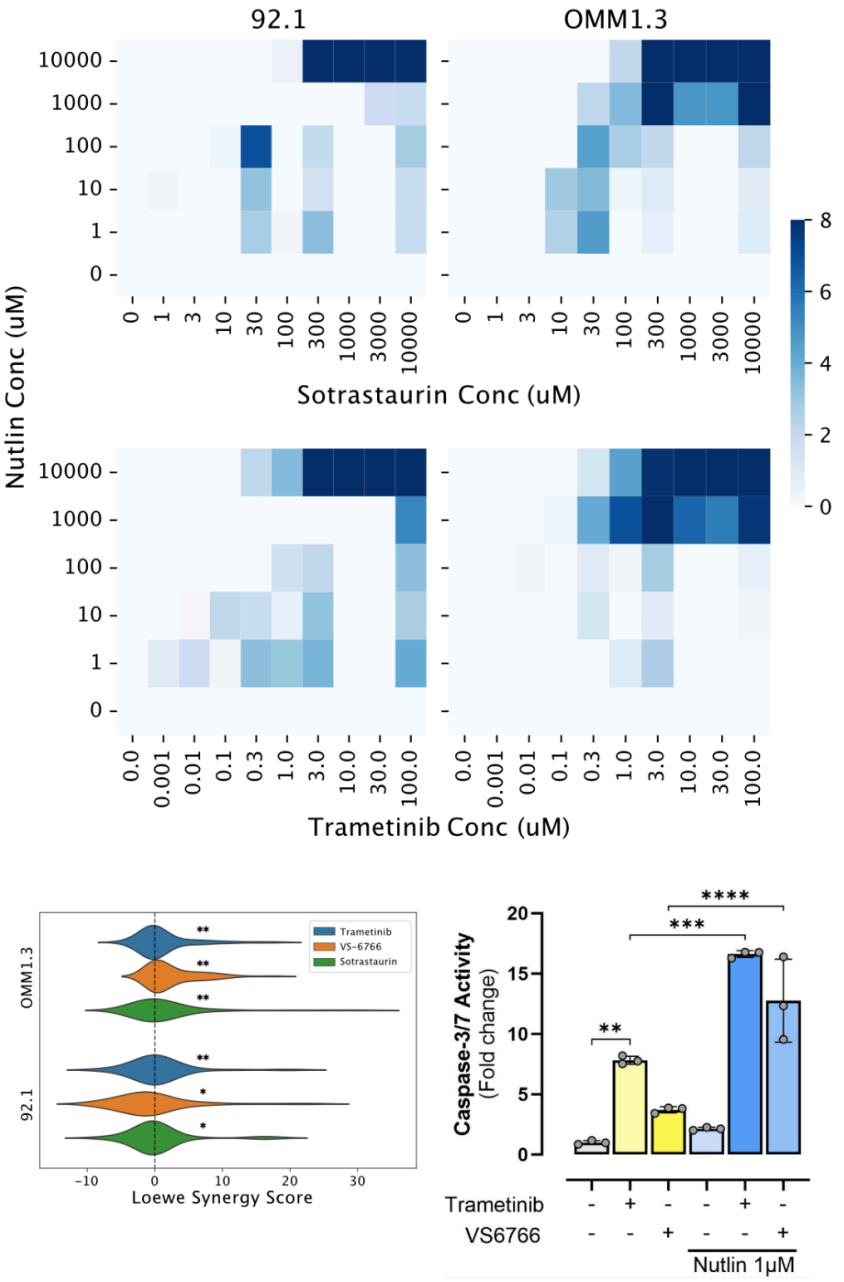

Figure 3.6. Inhibition of canonical pathway and activation of p53 signaling sufficient to inhibit UVM cell growth and induce apoptosis. (A) MEK inhibitor and MDM2 inhibitor synergistic in 92.1 and OMM1.3 UVM cell lines. (B) MDM2 synergism distributions for classical

(MEK, PKC) inhibitors are higher than for non-canonical (FAK) inhibitors. (C) MDM2 inhibition with MEK inhibition significantly increases caspase activity in 92.1.

Chapter 4: YAP-Driven Epigenetic Reprogramming and mTOR Signaling in a Mouse Model of Head and Neck Cancer

4.1 Introduction

Stem cells in the oral mucosa reside in the basal layer of the stratified squamous epithelium, and consist of a single pool of self-renewing oral epithelial progenitor cells (OEPCs)[103]. This single progenitor cell population renders the oral epithelium an ideal system to elucidate mechanisms underlying malignant reprogramming,[104] unlike other tissue systems such as the skin or gut, in which elegant studies have revealed multiple unique stem cell pools that play distinct roles in cancer initiation. Head and neck squamous cell carcinoma (HNSC) represents the most common malignancy arising from the upper aerodigestive epithelia[105]. Extensive molecular characterization of HNSC has revealed that while each individual tumor exhibits alterations in a large number of genes, these alterations converge to impact a finite set of oncogenic molecular pathways[7]. HNSC is characterized by near universal loss-of-function of the TP53 tumor suppressor by genomic alteration or human papillomavirus (HPV)-mediated inhibition. Notably, alterations in FAT1, observed in nearly 30% of HNSC[7], disrupt Hippo pathway signaling and result in unrestrained activation of the transcriptional co-activator YAP[106]. We now combine knowledge of the landscape of oncogenic pathway alterations in HNSC with genetically engineered animal models and multiomics to unveil the underpinnings of cancer initiation in vivo.

4.2. Results

4.2.1. Mouse Model Developed to Study the Role of HPV and YAP in Tumorigenesis

Tumor initiation represents the crucial first step in tumorigenesis during which normal progenitor cells undergo cell fate transition to cancer. To investigate this process, we developed genetically engineered murine systems focusing on prevalent and co-occurring genomic

alterations in HNSC (Figure 4.1). While genomic alterations involving FAT1 are observed in ~30% of HNSC patients, this represents one mechanism of YAP activation in HNSC. YAP activation may also occur through a variety of mechanisms, including amplification that is mutually exclusive with FAT1 genomic alterations, and is observed in even higher percentages of HNSC patients[107–109]. In addition, HNSC is characterized by widespread amplification of the YAP paralog TAZ (WWTR1). We thus investigated the status of YAP activation in human malignancies by evaluating immunohistochemical (IHC) staining for YAP activation in human HNSC tissue microarrays using nuclear localization as a surrogate for YAP activation[110]. Consistent with its physiologic role in stem cell maintenance and self-renewal, nuclear YAP was detected primarily in basal cells in normal oral epithelial tissue (Figure 4.1). Compared to normal tissue, HNSC had a significantly greater fraction of YAP activated cells, which were distributed throughout the tumors (Figure 4.1).

To investigate tumorigenesis in the context of a minimum complement of pathway alterations, we employed transgenic expression of the HPV16 E6-E7 oncogene, which concomitantly inhibits the TP53 and CDKN2A tumor suppressors[111], and the constitutively active $YAP1^{S127A}$ allele[9] . Keratin 14 (KRT14) is expressed in the basal layer of oral epithelia, which contains OEPCs that may represent the cell of origin for HNSC[112]. Utilizing a tamoxifen-inducible Cre-recombinase (CreERT) driven by the Krt14 promoter, genomic alterations were targeted to KRT14+ OEPCs. We bred mice bearing E6-E7 (“E”), $YAP1^{S127A}$ (“Y”), or both transgenes (“EY”). Littermates lacking these transgenes but possessing Krt14-CreERT, LSL-rtTA regulatory transgenes, and the H2B-GFP reporter were used as normal controls (“N”). Local administration of tamoxifen activated CreERT-mediated recombination of a floxed STOP cassette (LSL) and enabled transcription of the reverse tetracycline-controlled transactivator (rtTA) in KRT14+ OEPCs. Administration of doxycycline chow then induced expression of the tetracycline response element-regulated $HPV16^{E6\text{-}E7}$, $YAP1^{S127A}$, and H2B-GFP transgenes (Figure 4.1).

4.2.2. Cell Lines Used to Generate ATAC-seq and CUT&Tag Data

We utilized N and EY primary cell cultures for a multiomics approach to perform transcriptomic profiling (RNA-seq), map YAP genomic localization (YAP CUT&Tag), assess activity of regulatory elements (H3K27ac CUT&Tag), and explore chromatin accessibility (ATAC-seq). We first evaluated whether N and EY cells in primary culture were transcriptomically comparable to N and EY whole tissue epithelia samples by comparing differentially expressed genes, and observed high agreement (~90%) among EY DEGs across cultured cells and epithelia (Figure 4.2). Further, examining transcripts detected exclusively in whole tissue or primary cultured cells revealed that tissue-specific transcripts included stromal and immune cell transcripts, transcripts that are not expected to be observed in FACS-purified cultured cells (Figure 4.2). To confirm that this is indeed due to a loss of immune cells and not a loss of chemokine expression the expression of specific genes were analyzed. Immune cell markers Ptprc (Cd45), Csf3r, Cxcr2 and Ccr1 were not expressed in the primary culture but were expressed in the whole tissue samples, consistent with a loss of the immune cell population. Notably the cognate ligands Cxcl1, Cxcl2 and Csf3, which are usually expressed in epithelial malignant cells, were detected and shared similar expression patterns between EY and N.

4.2.3. Epigenetic Reprogramming Core to Tumorigenesis

After confirming the primary culture is a reasonable system in which to study this tumorigenesis mechanism we then investigated the YAP CUT&Tag data. Including N and EY samples there were 38,002 YAP binding sites (‘peaks’) identified (Figure 4.3). Consistent with published YAP ChIPseq data, annotation of YAP CUT&Tag genome-wide peak distribution showed that ~43% of YAP peaks occurred in intergenic regions, and approximately half of YAP peaks occurred at 10-100kb from transcription start sites (TSS, Figure 4.3)[113]. We classified YAP peaks into those significantly gained, lost, or unchanged in EY compared to N cells according to

their detection and occupancy observed in the two genotypes (Figure 4.3). A total of 11,169 (29%) gained, 3109 (8%) lost, and 23,724 unchanged peaks were observed in EY compared with N (Figure 4.3).

We next evaluated the chromatin accessibility in this model. 86809 ATAC-seq peaks were detected using MACS2 and DESeq2 used to call differentially accessible peaks, of which 5435 (6.3%) were gained in EY and 2960 (3.4%) were lost in EY (Figure 4.3, see Methods). Motif enrichment shows several TEAD and AP-1 motifs enriched in the peaks gained in the EY (Figure 4.3) suggesting that YAP binding is associated with increased chromatin accessibility.

Indeed, YAP CUT&Tag peaks tended to overlap with ATAC-seq peaks; 28,986 of the 38,002 YAP CUT&Tag peaks (76%) overlapped with an ATAC-seq peak. Analysis of ATAC-seq at gained YAP peaks in EY showed that EY gained YAP peaks more often overlapped with ATAC sites gained in EY compared with N (8% of EY gained versus 3% of unchanged verus <1% of EY lost YAP peaks), indicating gained YAP sites at newly opened chromatin regions in EY (Figure 4.3). Examining transcriptional regulatory activity at YAP binding sites, we observed a subtle increase in H3K27ac CUT&Tag signal intensity at gained YAP binding sites in EY (Figure 4.3). Together with the ATAC-seq results, these findings suggest that EY expression leads to increased chromatin accessibility and activation of a subset of YAP-regulated genes without exerting strong global genome wide effects on chromatin accessibility or transcriptional activity.

Examination of transcription factor motifs enriched in YAP gained peaks in EY identified consensus motifs for TEAD, AP-1, Sp2, KLF, p63, and NRF2 as transcription factors that may cooperate with YAP in transcriptional regulation (Figure 4.3). Indeed, TEAD family DNA binding proteins are required for YAP-mediated gene expression[114], and YAP/TEAD complex with AP-1 factors, KLF4, and p63 to regulate transcriptional programs[113,115].

4.2.4. YAP Binding Associated with mTOR Activation

To gain further insight into specific genes and pathways regulated by YAP, we integrated RNAseq and YAP CUT&Tag data using Binding and Expression Target Analysis (BETA), and found that YAP predominantly acts as a transcriptional activator (Figure 4.4; p = 4.1x10-14). Intriguingly, the top 300 YAP-activated genes showed marked enrichment for three MSigDB Hallmark Pathways: TNFa signalling via NFkB, mTORC1 signalling, and inflammatory response (Figure 4.4). Using an independent approach we performed multiomic Venn analysis to identify YAP-bound genes that showed transcriptional upregulation and newly gained chromatin accessibility in EY (Figure 4.4). The 849 genes identified by this approach showed enrichment for similar pathways as for EY tissue by RNAseq or EY cells by BETA, with mTORC1 and PI3K/AKT/mTOR Hallmark pathways again representing highly enriched pathways (Figure 4.4).

4.2.5. EGFR Ligand Expression Induced by YAP Binding

Upregulation of YAP transcriptional signatures by GSEA was observed in YAP-expressing tongue epithelia from Y and EY mice (Figure 4.5). Y and EY mice also showed enrichment for gene sets for mTOR pathway activation, a commonly activated signaling mechanism in HNSC (Figure 4.5). Consistently, IHC of epithelia from EY and Y mice showed a pronounced increase in phospho-S6 (pS6) ribosomal protein levels, a downstream marker of mTOR activity (Figure 4.5).

Using siRNA-mediated activation of YAP, via LATS1/2 knockdown, in the HN12 human cell line shows activation of EGFR signaling and pS6 suggesting that YAP transcriptional programs may be activating EGFR signaling (Figure 4.5). Additionally, increased expression of several EGFR ligands including Axl, Nrg1 and Epgn all contained local YAP peaks and had increased expression in the Y and EY epithelia suggesting YAP is directly responsible for the regulation of these genes (Figure 4.5).

We then performed siRNA-mediated knockdown of YAP and/or its paralog TAZ (WWTR1) in HN12 to test if YAP indeed regulates mTOR activity. Consistent with reports that YAP and TAZ serve mutually compensatory functions, combined knockdown of YAP and TAZ was required to diminish expression of the YAP target CYR61 and mTOR activity, based on pS6 abundance (Figure 4.5). Combined knockdown also resulted in diminished pEGFR but not total EGFR, potentially related to decreased expression of YAP-regulated EGFR ligands, including Epgn, Nrg1, and Nrg4 (Figure 4.5), suggesting that YAP initiates multiple transcriptional mechanisms resulting in mTOR activation (Figure 4.5). These findings were in accordance with the two multiomics analyses that also pointed to YAP activation driving mTOR activity.

4.3. Methods

Husbandry and genotyping

We bred mice expressing E6-E7 (Krt14-CreERT/LSL-rtTA/tetON_H2B-GFP/tetON_E6-E7, “E”), YAP1S127A (Krt14-CreERT/LSL-rtTA/tetON_H2B-GFP/tetON_YAP1S127A, “Y”), or both transgenes (Krt14-CreERT/LSL-rtTA/tetON_H2B-GFP/tetON_E6-E7/tetON_YAP1S127A, “EY”). Littermates that bore neither tetON_E6-E7 nor tetON_YAP1S127A effector transgenes but possessed the Krt14-CreERT and LSL-rtTA regulatory transgenes were used as the normal condition (Krt14-CreERT/LSL-rtTA/tetON_H2B-GFP, “N”). Intralingual injection of tamoxifen was performed to achieve reliable transgene induction. Mice were started on a doxycycline-containing diet on the first day of tamoxifen treatment. This treatment regimen resulted in consistent CreERT-mediated excision of the floxed STOP cassette and expression of effector and reporter transgenes in KRT14+ basal cells.

Krt14-CreERT+/+/LSL-rtTA+/+/H2B-GFP+/+/E6-E7+/− mice were crossed to Krt14-CreERT+/+/LSL-rtTA+/+/H2B-GFP+/+/YAP1S127A+/− resulting in Mendelian proportions of N, E, Y, and EY littermates. At 3–4 weeks of age, a tail fragment was obtained for initial

screening genotype confirmation. Directly prior to transgene induction for experiments, mice were assigned to age and sex-balanced groups, and an ear fragment was obtained for confirmatory genotyping. Genomic DNA was isolated by incubating tissue in 25mM NaOH and 0.2 mM EDTA at 100°C for 1 hour, followed by neutralization with an equal volume of 40 mM Tris-HCl (pH 5.5). Multiplex polymerase chain reaction (PCR)-based genotyping was performed using REDTaq® polymerase per manufacturer recommendations (Millipore Sigma). Oligonucleotides were multiplexed as follows: (1) LSL-rtTA and E6-E7 and Il2 (positive control), (2) Yap1S127A and Trp53 (positive control), (3) Krt14-CreERT and Il2 (positive control). All PCR products were subjected to electrophoresis on 2% agarose gel in Tris acetate EDTA buffer.

Transgene induction

Mice were anesthetized with isoflurane and 100μL of tamoxifen solution (20 mg/mL in miglyol) was administered into the tongue under stereomicroscopic visualization. One dose of tamoxifen was administered every other day for a total of 3 doses.

Epithelia isolation

After in situ infiltration with 500uL collagenase+dispase solution (1mg/mL, 2.5 mg/mL) (Millipore Sigma), the tongues of euthanized mice were dissected free and incubated for 30 minutes at 37°C. The tongue epithelium was then dissected free from the underlying muscle under stereomicroscopic visualization.

Generation of epithelial cell suspensions

Isolated epithelia were minced in 0.25% trypsin-EDTA (Thermo) and subjected to mechanical dissociation in the gentleMACS dissociator C tubes (Miltenyi #130-095-937) for 12 minutes at 37°C, followed by inactivation of trypsin and filtration.

Primary epithelial cell culture

Mouse tongue epithelial cells were isolated from mice following transgene induction as described above. Cells were grown on collagen coated plates in complete DermaCult keratinocyte basal expansion medium (STEMCELL Technologies). Medium contained the manufacturer's provided supplements, plus 5 ng/mL mouse EGF (Gibco), 50 pM cholera toxin (Sigma), 1x antibiotic/antimycotic solution [Gibco], and 2 uM doxycycline hyclate (Sigma, to maintain transgene activation) at 37°C with 5% CO2.

CUT&Tag Data Processing and Analysis

Raw reads were aligned using Bowtie2 (version 2.2.5[116]) to build version mm10 of the mouse genome. Peaks were called independently in each replicate against the corresponding IgG control using SEACR (version 1.3[117]) in relaxed mode. Peaks with RPKM < 10 were filtered out. Consensus peaks were merged for each genotype, EY or N, by combining all filtered peaks using bedtools merge (version 2.27.1[118]). Tornado plots were generated using deeptools (version 3.3.5[119]). Differential acetylation was called using DESeq2 (version 1.42.0[98,120]) and apeglm (version 1.24.0) in R (version 4.3.2). Peaks with adjusted p-values less than 0.05 were considered significant. Motif enrichment was performed using the findMotifsGenome.pl script in the HOMER package (version 4.11[121]). Peaks were annotated using the annotatePeaks.pl script in the HOMER package. Peaks were annotated if they lie within the gene body or closer than 10 kb to the annotated TSS.

ATAC-seq Data Processing and Analysis

Raw reads were aligned using BWA (version 0.7.17[122]) to build version mm10 of the mouse genome. Peaks were called using MACS2 (version 2.2.7.1[123]) in narrow peak mode with a False Discovery Rate threshold of less than 0.01. Consensus peaks were merged for all samples by combining all called peaks using bedtools merge (version 2.27.1). Reads were recounted in consensus peaks using bedtools coverage (version 2.27.1). DESeq2 (version 1.42.0) and apeglm (version 1.24.0) in R (version 4.3.2) were used to call differential chromatin

accessibility, peaks with adjusted p-value of less than 0.05 were considered significant. Motifs and peak annotation was performed as with the CUT&Tag data using HOMER.

RNA-seq of primary tongue epithelial cells and derived cell lines

Paired-end reads were aligned using STAR (version 2.7.9) using default settings. STAR index was created using the GRCm39 (mm10) primary genome FASTA and annotation files. The resulting BAM files were sorted by name using samtools (version 1.7) then gene counts were quantified using HTSeq-count (version 0.13.5). Pairwise differential expression was calculated using DESeq2 (version 1.42.0). Genes with >10 counts were considered detected. Comparative differential expression between the tumor and cell line was performed using an interaction in the DESeq2 design. Genes with $|LFC|>1$ and $p.adj<0.05$ were considered significant in all analyses.

Gene set overlap analysis

Significant overrepresentation of gene sets was determined using a hypergeometric overlap test as implemented in the scipy package (version 1.10.0) in Python (version 3.10.5). One-sided p-values and the expected number of genes were computed using the hypergeometric distribution.

4.4. Discussion

The overwhelming majority of work investigating cancer-driving mechanisms has relied on established tumors, which requires retrospective inference of tumor initiating events and limits distinction between processes governing tumor initiation from progression. In HNSC, despite this mutational and cellular heterogeneity, the diverse genomic alterations observed in HNSC converge to activate a limited number of oncogenic signaling pathways. Through the use of a spatiotemporally controlled murine system targeting genomic alterations to a single pool of epithelial progenitor cells, we now show that unrestrained YAP activation in the context of TP53 inhibition by HPV E6-E7 oncogenes induces carcinoma with rapid kinetics and nearly complete

penetrance, enabling in-depth investigation of tumor initiation. The ability to combine this genetically-defined system with multi-omic analysis provided a unique opportunity to investigate the conversion of epithelial progenitor cells into CSC as it occurs in vivo.

One finding was that tumor initiation could be achieved in the absence of genomic alterations in the PI3K/AKT/mTOR signaling axis, given the extensive evidence implicating widespread mTOR activation in HNSC[124–127]. However, we observed robust mTOR activation in pre-invasive lesions and throughout SCC in YAP-expressing epithelia, suggesting that YAP can activate mTOR signaling. Indeed, mTOR program enrichment represented a distinguishing feature of the EY-induced CSC state. Mechanistically, we found that YAP drives the transcriptional upregulation of Axl and multiple EGFR ligands, which may explain EY-mediated mTOR activation. The fact that >70% of HNSCs do not harbor genomic alterations in the PI3K-mTOR pathway but exhibit a widespread activation of YAP[9,10] and mTOR is aligned with a potential signaling cross-talk in which YAP can activate mTOR. Specifically, our findings support a model in which YAP:TEAD-mediated transcription activates an AXL- and EGFR-initiated signaling cascade resulting in the activation of mTOR in epithelial cells, thus representing an actionable target to prevent tumor initiation.

In summary, we demonstrate that a genetically-defined, traceable system simultaneously activating oncogenic pathways and disabling tumor suppressive mechanisms in normal oral epithelial progenitor cells induces the emergence of a distinct cancer initiating stem-like cell state. Through multimodal analysis of nascent cancer cells at the single cell level in vivo, we define tumor-autonomous transcriptional programs during invasive carcinoma formation. This conceptual framework of cancer initiation has the potential to open multiple novel avenues for early intervention, including precision targeting of tumor cell-autonomous cancer initiating signaling pathways.

Chapter 4 is currently in the review process at Cancer Stem Cell. F. Faraji, S. I. Ramirez, L. M. Clubb, K. Sato, P. Y. A. Quiroz, W. M. G. Galloway, Z. Mikulski, T. S. Hoang, K. Medetgul-Ernar, P. Marangoni, K. B. Jones, A. Officer, A. A. Molinolo, K. Kim, K. Sakaguchi, J. A. Califano, Q. Smith, O. D. Klein, P. Tamayo, and J. S. Gutkind, Direct reprogramming of oral epithelial progenitor cells to cancer stem cells at single cell resolution in vivo (2024). The dissertation author was a contributing researcher to this paper.

4.5. Figures

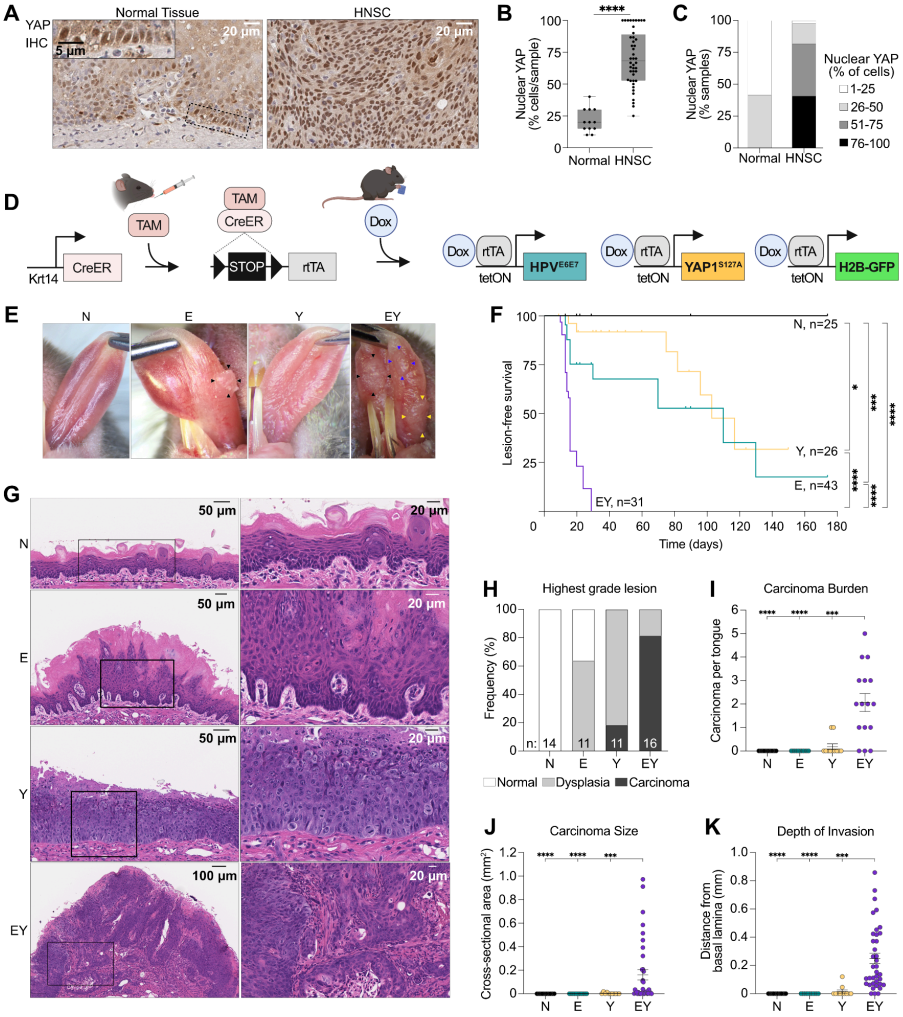

Figure 4.1. Characterization of the HPV/YAP mouse model. (A) YAP staining of normal and cancerous lesions show increased signal intensity in cancer. (B) and (C) Quantification of nuclear YAP shows significantly higher levels in HNSC. (D) Schematic of HPV- and YAP-driven mouse model expressed in Krt14 positive basal epithelial cells. (E) Representative images of cancer burden in mouse tongues. (F) Lesion-free survival across all 4 genotypes showing strong synergistic effects on tumorgenesis. (G) IHC staining of lesions at 14 days showing normal, in situ and invasive lesions. (H/I/J/K) EY genotype mice show increased degree of carcinoma, lesion size, number of lesions and more invasive lesions.

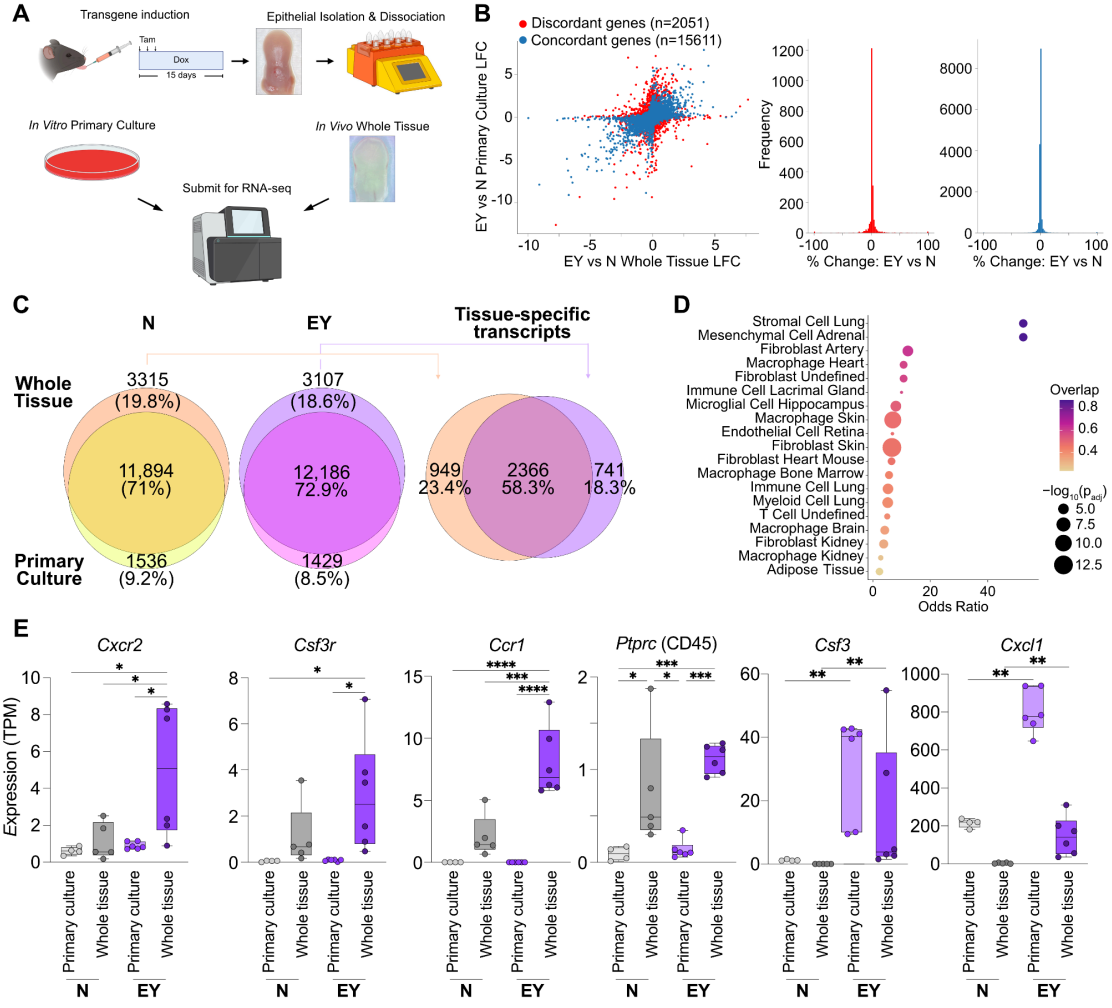

Figure 4.2. Tumor-derived primary culture is a reasonable proxy for whole tissue. (A) Data generation schematic from whole tissue and tumor-derived primary culture lines. (B) Differential expression (DESeq2, see methods) agreement between whole tissue and primary culture shows the majority of genes agree between modalities, disagreeing genes differ by single-digit percentage points. (C) Consensus detected gene lists between EY and N genotype tissue and primary culture. (D) Gene set overlap for lineage-related gene sets showing undetected genes in the primary culture associated with stromal, immune and fibroblast cell types. (E) Expression levels of selected genes showing loss of immune lineage genes while retained expression of tumor-derived factors.

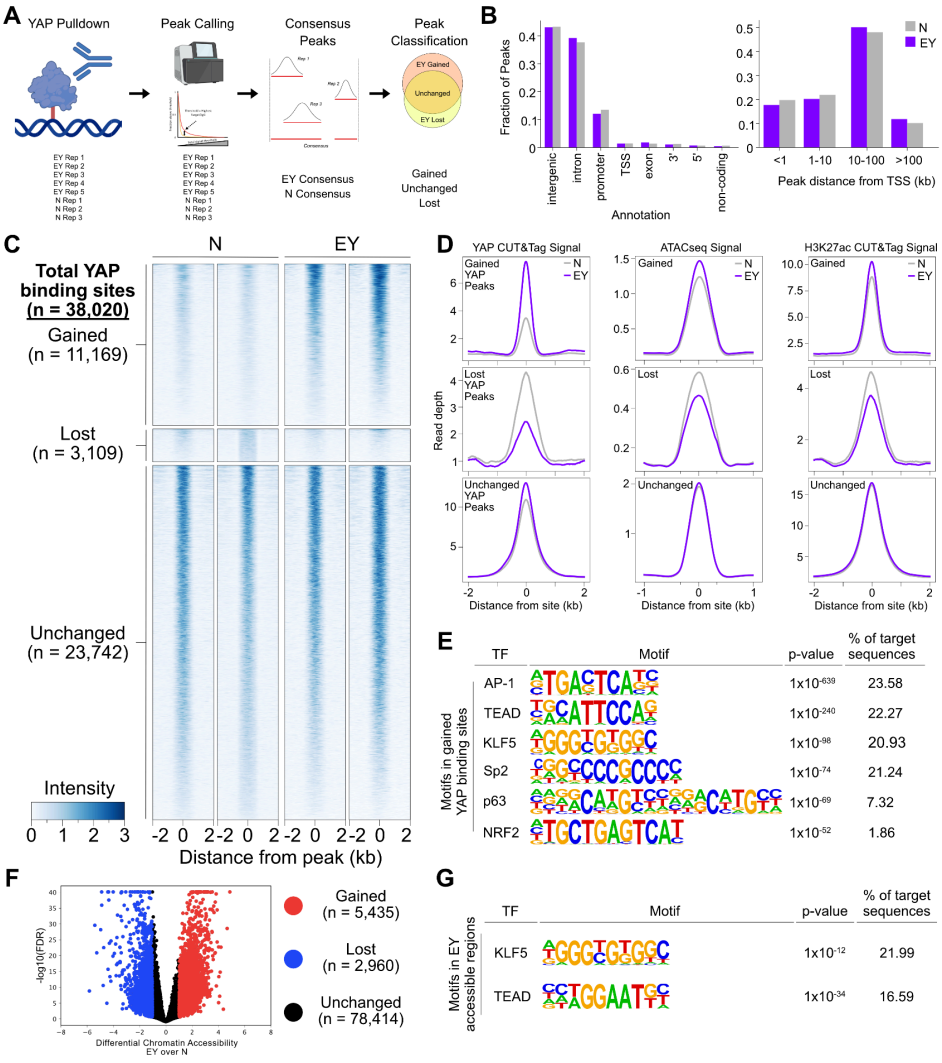

Figure 4.3. Epigenetic data generation and characterization of YAP binding. (A) Workflow for YAP peak calling and categorization using SEACR and DESeq2. (B) YAP peaks tend to lie in intergenic regions far from genes. (C) Tornado plot showing YAP peak occupancy between different types of YAP peaks. (D) EY gained YAP peaks tend to occur in regions with higher levels of EY H3K27ac and ATAC-seq read density. (E) HOMER motif enrichment analysis in EY gained peaks shows clear signal for AP-1, TEAD, KLF family, p63 and other motifs. (F) ATAC-seq differential peak quantification using MACS2 and DESeq2 calls several thousand differentially accessible regions. (G) Motifs enriched in EY gained peaks associated with TEAD and KLF family sequence logos.

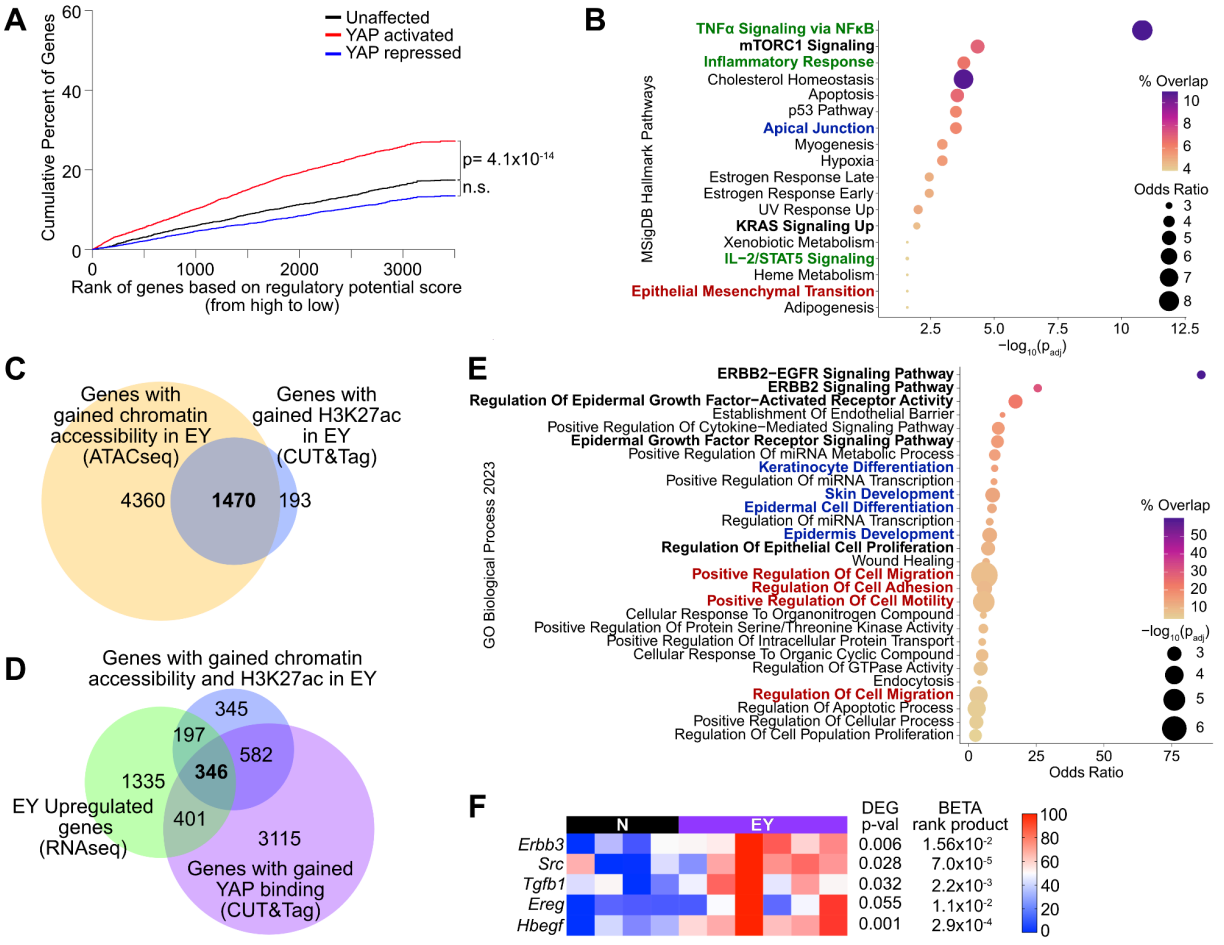

Figure 4.4. Multiomics integration identifies mTOR signaling as a target of YAP binding. (A) Binding and Expression Target Analysis (BETA) analysis of YAP CUT&Tag data shows YAP activates gene expression at local genes. (B) Overlap analysis against MSigDB Hallmark collection for YAP-controlled genes shows inflammatory and mTOR signaling controlled by YAP binding. (C) Overlap between genes in opening chromatin regions (ATAC-seq, EY gain) and genes with more active chromatin (H3K27ac, EY gain). (D) Further gene filtering by differential RNA-seq and EY gained YAP binding. (E) Gene set overlap test of the 346 genes from (D) against GO Biological Processes showing overrepresentation of genes involved in HER2/EGFR signaling, differentiation pathways and cell migration. (F) Primary culture gene expression values from selected relevant genes for HER2/EGFR signaling controlled by YAP binding.

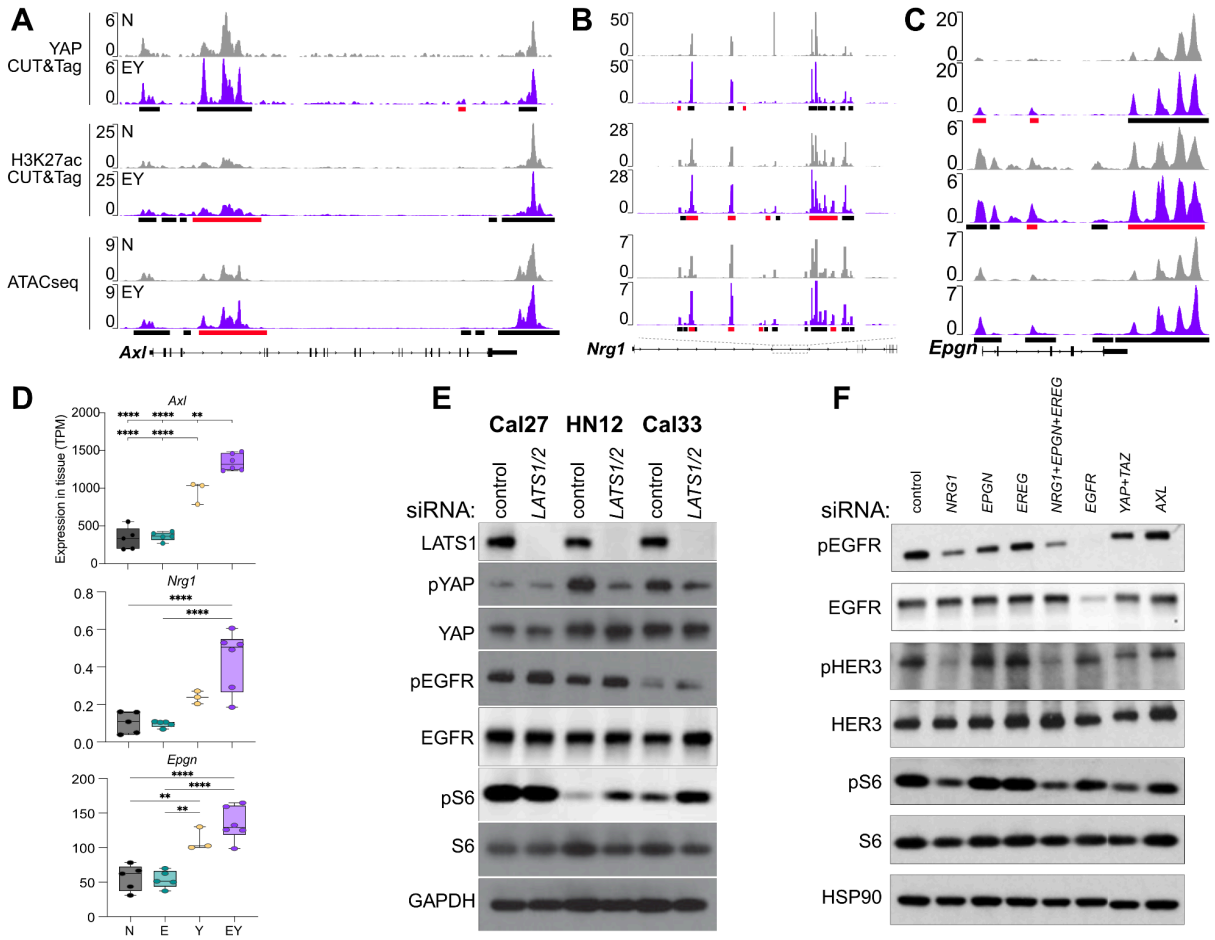

Figure 4.5. EGFR ligands controlled by YAP induce mTOR activation. (A/B/C) IGV screenshots showing EY gained chromatin accessibility (ATAC-seq), EY gained active chromatin regions (H3K27ac CUT&Tag) and YAP binding (YAP CUT&Tag). (D) Same genes showing increased expression in whole tissue RNA-seq dataset. (E) siRNA-mediated knockdown of LATS1/2, and subsequent YAP activation, across several human HNSC cell lines showing higher levels of mTOR signaling (pS6) and EGFR activation (pEGFR). (F) Knockdown of potential YAP-controlled EGFR ligands in HN12 shows NRG1, EPGN and EREG all affect EGFR and mTOR signaling.

Chapter 5: Conclusions

Through these three examples I have shown that using a holistic approach to modeling cancer evolution and signaling provides a more well-rounded picture of the state of cancer cells. Using novel statistical methods that integrate multiple datasets together I have demonstrated the value of biologically inspired modeling in three distinct areas of research:

First, in the case of microenvironment signaling associated with breast cancer invasion. Data curation from literature and integration with single-cell RNA-seq datasets allows for a novel cell-type aware gene expression comparison between epithelial and stromal regions of DCIS and IDC. This approach shows that there are small but important differences in the local microenvironment composition of DCIS and IDC and that when accounting for these differences I can nominate novel invasion-associated signaling pathways. Additionally I developed a microenvironment modeling approach to measure gene-gene interactions across different tissue compartments that shows changing Wnt, ephrin and T cell activation associated with an

invasive phenotype. Ephrin signaling was validated in an independent dataset using survival modeling suggesting that cell homing signals are lost in invasive ductal carcinomas.

Second, this time in the context of uveal melanoma where GNAQ signaling is the primary driver of tumorigenesis I generated RNA-seq and RPPA data and analyzed them simultaneously to identify transcriptional and protein-level changes associated with GNAQ signaling inhibition. This unbiased experiment, the first of its kind on the GNAQ pathway, nominated several novel downstream signaling pathways including PKA, NFkB and p53 signaling. Focusing on p53 signaling I showed through an independent computational analysis that the GNAQ signature is correlated with p53 signaling, p53 gene essentiality and MDM2 gene essentiality in a TP53 mutation-dependent fashion. Confirming this computational work I then show that uveal melanoma cell lines have repressed p53 signaling through MDM2 and that induction of p53 causes apoptosis. Using p53 as a potential therapeutic target I then show using in vitro drug synergism assays that using nutlin, a p53 activator, in combination with canonical pathway inhibition is synergistic and induces cell death in these cell lines.

And finally, in the last chapter of my dissertation I worked to characterize the YAP-driven epigenetic and gene expression changes associated with tumorigenesis in a mouse model of oral cancer. I generated ATAC-seq, CUT&Tag and RNA-seq data from both primary tissue and tumor-derived cell lines. This data is integrated together to show that YAP-associated motifs, namely TEAD and AP-1, are associated with increased chromatin accessibility in these YAP-driven tumors. Additionally YAP tends to bind in areas of open chromatin that have active transcription, as measured by H3K27 acetylation. Through an integrative modeling approach the ATAC-seq, CUT&Tag and RNA-seq data are combined to show that YAP is driving gene expression programs associated with mTOR signaling. Interrogation of specific EGFR ligands shows that YAP binds to the local genomic region and induces expression of Nrg1, Epgn, Axl

and more. siRNA-mediated knockdown of these ligands shows that Nrg1 is the ligand with the strongest effect on EGFR, and downstream mTOR signaling.

References

1. Wang, J., Li, B., Luo, M., Huang, J., Zhang, K., Zheng, S., Zhang, S. & Zhou, J. Progression from ductal carcinoma in situ to invasive breast cancer: molecular features and clinical significance. *Signal Transduction and Targeted Therapy* **9**, 1–28 (2024).
2. Chen, H., Bai, F., Wang, M., Zhang, M., Zhang, P. & Wu, K. The prognostic significance of co-existence ductal carcinoma in situ in invasive ductal breast cancer: a large population-based study and a matched case-control analysis. *Ann Transl Med* **7**, 484 (2019).
3. Srivastava, S., Ghosh, S., Kagan, J. & Mazurchuk, R. The PreCancer Atlas (PCA). *Trends Cancer Res.* **4**, 513–514 (2018).
4. Curtius, K., Wright, N. A. & Graham, T. A. Evolution of Premalignant Disease. *Cold Spring Harb. Perspect. Med.* **7**, (2017).
5. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
6. Olivier, M., Hollstein, M. & Hainaut, P. TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harb. Perspect. Biol.* **2**, a001008 (2010).
7. Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517**, 576–582 (2015).
8. Zhou, G., Liu, Z. & Myers, J. N. TP53 Mutations in Head and Neck Squamous Cell Carcinoma and Their Impact on Disease Progression and Treatment Response. *J. Cell. Biochem.* **117**, 2682–2692 (2016).
9. Martin, D., Degese, M. S., Vitale-Cross, L., Iglesias-Bartolome, R., Valera, J. L. C., Wang, Z., Feng, X., Yeerna, H., Vadmal, V., Moroishi, T., Thorne, R. F., Zaida, M., Siegele, B.,

Cheong, S. C., Molinolo, A. A., Samuels, Y., Tamayo, P., Guan, K. L., Lippman, S. M., Lyons, J. G. & Gutkind, J. S. Assembly and activation of the Hippo signalome by FAT1 tumor suppressor. *Nat. Commun.* **9**, 2372 (2018).
10. Faraji, F., Ramirez, S. I., Anguiano Quiroz, P. Y., Mendez-Molina, A. N. & Gutkind, J. S. Genomic Hippo Pathway Alterations and Persistent YAP/TAZ Activation: New Hallmarks in Head and Neck Cancer. *Cells* **11**, (2022).
11. Singh, A. D., Turell, M. E. & Topham, A. K. Uveal melanoma: trends in incidence, treatment, and survival. *Ophthalmology* **118**, 1881–1885 (2011).
12. Van Raamsdonk, C. D., Bezrookove, V., Green, G., Bauer, J., Gaugler, L., O'Brien, J. M., Simpson, E. M., Barsh, G. S. & Bastian, B. C. Frequent somatic mutations of GNAQ in uveal melanoma and blue naevi. *Nature* **457**, 599–602 (2009).
13. Robertson, A. G., Shih, J., Yau, C., Gibb, E. A., Oba, J., Mungall, K. L., Hess, J. M., Uzunangelov, V., Walter, V., Danilova, L., Lichtenberg, T. M., Kucherlapati, M., Kimes, P. K., Tang, M., Penson, A., Babur, O., Akbani, R., Bristow, C. A., Hoadley, K. A., Iype, L., Chang, M. T., TCGA Research Network, Cherniack, A. D., Benz, C., Mills, G. B., Verhaak, R. G. W., Griewank, K. G., Felau, I., Zenklusen, J. C., Gershenwald, J. E., Schoenfield, L., Lazar, A. J., Abdel-Rahman, M. H., Roman-Roman, S., Stern, M.-H., Cebulla, C. M., Williams, M. D., Jager, M. J., Coupland, S. E., Esmaeli, B., Kandoth, C. & Woodman, S. E. Integrative Analysis Identifies Four Molecular and Clinical Subsets in Uveal Melanoma. *Cancer Cell* **32**, 204–220.e15 (2017).
14. Arang, N. & Gutkind, J. S. G Protein-Coupled receptors and heterotrimeric G proteins as cancer drivers. *FEBS Lett.* **594**, 4201–4232 (2020).
15. Larribère, L. & Utikal, J. Update on GNA Alterations in Cancer: Implications for Uveal Melanoma Treatment. *Cancers* **12**, (2020).
16. Wu, V., Yeerna, H., Nohata, N., Chiou, J., Harismendy, O., Raimondi, F., Inoue, A., Russell, R. B., Tamayo, P. & Gutkind, J. S. Illuminating the Onco-GPCRome: Novel G protein-coupled receptor-driven oncocrine networks and targets for cancer immunotherapy. *J. Biol. Chem.* **294**, 11062–11086 (2019).
17. Feng, X., Arang, N., Rigiracciolo, D. C., Lee, J. S., Yeerna, H., Wang, Z., Lubrano, S., Kishore, A., Pachter, J. A., König, G. M., Maggiolini, M., Kostenis, E., Schlaepfer, D. D., Tamayo, P., Chen, Q., Ruppin, E. & Gutkind, J. S. A Platform of Synthetic Lethal Gene Interaction Networks Reveals that the GNAQ Uveal Melanoma Oncogene Controls the Hippo Pathway through FAK. *Cancer Cell* **35**, 457–472.e5 (2019).
18. Paradis, J. S., Acosta, M., Saddawi-Konefka, R., Kishore, A., Gomes, F., Arang, N., Tiago, M., Coma, S., Lubrano, S., Wu, X., Ford, K., Day, C.-P., Merlino, G., Mali, P., Pachter, J. A., Sato, T., Aplin, A. E. & Gutkind, J. S. Synthetic Lethal Screens Reveal Cotargeting FAK and MEK as a Multimodal Precision Therapy for GNAQ-Driven Uveal Melanoma. *Clin. Cancer Res.* **27**, 3190–3200 (2021).
19. Zuidervaart, W., van Nieuwpoort, F., Stark, M., Dijkman, R., Packer, L., Borgstein, A.-M., Pavey, S., van der Velden, P., Out, C., Jager, M. J., Hayward, N. K. & Gruis, N. A. Activation of the MAPK pathway is a common event in uveal melanomas although it rarely occurs through mutation of BRAF or RAS. *Br. J. Cancer* **92**, 2032–2038 (2005).
20. Edmunds, S. C., Cree, I. A., Dí Nícolantonío, F., Hungerford, J. L., Hurren, J. S. & Kelsell, D. P. Absence of BRAF gene mutations in uveal melanomas in contrast to cutaneous melanomas. *Br. J. Cancer* **88**, 1403–1405 (2003).
21. Arang, N., Lubrano, S., Rigiracciolo, D. C., Nachmanson, D., Lippman, S. M., Mali, P., Harismendy, O. & Gutkind, J. S. Whole-genome CRISPR screening identifies PI3K/AKT as a downstream component of the oncogenic GNAQ-focal adhesion kinase signaling circuitry. *J. Biol. Chem.* **299**, 102866 (2023).
22. Sharma, S., Kelly, T. K. & Jones, P. A. Epigenetics in cancer. *Carcinogenesis* **31**, 27–36 (2010).

23. Verma, M. & Srivastava, S. Epigenetics in cancer: implications for early detection and prevention. *Lancet Oncol.* **3**, 755–763 (2002).
24. Yu, X., Zhao, H., Wang, R., Chen, Y., Ouyang, X., Li, W., Sun, Y. & Peng, A. Cancer epigenetics: from laboratory studies and clinical trials to precision medicine. *Cell Death Discovery* **10**, 1–12 (2024).
25. Totaro, A., Panciera, T. & Piccolo, S. YAP/TAZ upstream signals and downstream responses. *Nat. Cell Biol.* **20**, 888–899 (2018).
26. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1–21.29.9 (2015).
27. Song, L. & Crawford, G. E. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb. Protoc.* **2010**, db.prot5384 (2010).
28. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497–1502 (2007).
29. Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., Ahmad, K. & Henikoff, S. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 (2019).
30. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, (2017).
31. Cazaly, E., Saad, J., Wang, W., Heckman, C., Ollikainen, M. & Tang, J. Making Sense of the Epigenome Using Data Integration Approaches. *Front. Pharmacol.* **10**, 126 (2019).
32. Ping, Q., Yan, R., Cheng, X., Wang, W., Zhong, Y., Hou, Z., Shi, Y., Wang, C. & Li, R. Cancer-associated fibroblasts: overview, progress, challenges, and directions. *Cancer Gene Ther.* **28**, 984–999 (2021).
33. Tuxhorn, J. A., Ayala, G. E., Smith, M. J., Smith, V. C., Dang, T. D. & Rowley, D. R. Reactive stroma in human prostate cancer: induction of myofibroblast phenotype and extracellular matrix remodeling. *Clin. Cancer Res.* **8**, 2912–2923 (2002).
34. Andersen, M. K., Rise, K., Giskeødegård, G. F., Richardsen, E., Bertilsson, H., Størkersen, Ø., Bathen, T. F., Rye, M. & Tessem, M.-B. Integrative metabolic and transcriptomic profiling of prostate cancer tissue containing reactive stroma. *Sci. Rep.* **8**, 1–11 (2018).
35. Rowley, D. R. Reactive Stroma and Evolution of Tumors: Integration of Transforming Growth Factor-β, Connective Tissue Growth Factor, and Fibroblast Growth Factor-2 Activities. in *Transforming Growth Factor-β in Cancer Therapy, Volume II: Cancer Treatment and Therapy* (ed. Jakowlew, S. B.) 475–505 (Humana Press, Totowa, NJ, 2008).
36. Page, D. L., Dupont, W. D., Rogers, L. W., Jensen, R. A. & Schuyler, P. A. Continued local recurrence of carcinoma 15-25 years after a diagnosis of low grade ductal carcinoma in situ of the breast treated only by biopsy. *Cancer* **76**, 1197–1200 (1995).
37. Collins, L. C., Tamimi, R. M., Baer, H. J., Connolly, J. L., Colditz, G. A. & Schnitt, S. J. Outcome of patients with ductal carcinoma in situ untreated after diagnostic biopsy: results from the Nurses' Health Study. *Cancer* **103**, 1778–1784 (2005).
38. Kim, S. Y., Jung, S.-H., Kim, M. S., Baek, I.-P., Lee, S. H., Kim, T.-M., Chung, Y.-J. & Lee, S. H. Genomic differences between pure ductal carcinoma in situ and synchronous ductal carcinoma in situ with invasive breast cancer. *Oncotarget* **6**, 7597–7607 (2015).
39. Lesurf, R., Aure, M. R. & Sørlie, T. Molecular Features of Subtype-Specific Progression from Ductal Carcinoma In Situ to Invasive Breast Resource Molecular Features of Subtype-Specific Progression from Ductal Carcinoma In Situ to Invasive Breast Cancer. 1166–1179 (2016).
40. Ma, X.-J., Dahiya, S., Richardson, E., Erlander, M. & Sgroi, D. C. Gene expression profiling of the tumor microenvironment during breast cancer progression. *Breast Cancer Res.* **11**, R7 (2009).

41. Lee, S., Stewart, S., Nagtegaal, I., Luo, J., Wu, Y., Colditz, G., Medina, D. & Allred, D. C. Differentially Expressed Genes Regulating the Progression of Ductal Carcinoma In Situ to Invasive Breast Cancer. **72**, 4574–4587 (2012).
42. Allinen, M., Beroukhim, R., Cai, L., Brennan, C., Lahti-Domenici, J., Huang, H., Porter, D., Hu, M., Chin, L., Richardson, A., Schnitt, S., Sellers, W. R. & Polyak, K. Molecular characterization of the tumor microenvironment in breast cancer. *Cancer Cell* **6**, 17–32 (2004).
43. Porter, D. A., Krop, I. E., Nasser, S., Sgroi, D., Kaelin, C. M., Marks, J. R., Riggins, G. & Polyak, K. A SAGE (serial analysis of gene expression) view of breast tumor progression. *Cancer Res.* **61**, 5697–5702 (2001).
44. Emery, L. A., Tripathi, A., King, C., Kavanah, M., Mendez, J., Stone, M. D., Morenas, A. D., Sebastiani, P. & Rosenberg, C. L. Early Dysregulation of Cell Adhesion and Extracellular Matrix Pathways in Breast Cancer Progression. *Am. J. Pathol.* **175**, 1292–1302 (2009).
45. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* (2020) doi:10.1038/s41596-020-0292-x.
46. Armingol, E., Officer, A., Harismendy, O. & Lewis, N. E. Deciphering cell–cell interactions and communication from gene expression. *Nat. Rev. Genet.* **22**, 71–88 (2020).
47. Murrow, L. M., Weber, R. J., Caruso, J. A., McGinnis, C. S., Phong, K., Gascard, P., Rabadam, G., Borowsky, A. D., Desai, T. A., Thomson, M., Tlsty, T. & Gartner, Z. J. Mapping hormone-regulated cell-cell interaction networks in the human breast at single-cell resolution. *Cell Syst* (2022) doi:10.1016/j.cels.2022.06.005.
48. Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., Khodadoust, M. S., Esfahani, M. S., Luca, B. A., Steiner, D., Diehn, M. & Alizadeh, A. A. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* (2019) doi:10.1038/s41587-019-0114-2.
49. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P. & Tamayo, P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417–425 (2015).
50. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB v2.0: Inferring cell-cell communication from combined expression of multi-subunit receptor-ligand complexes. *bioRxiv* 680926 (2019) doi:10.1101/680926.
51. Kania, A. & Klein, R. Mechanisms of ephrin–Eph signalling in development, physiology and disease. *Nat. Rev. Mol. Cell Biol.* **17**, 240–256 (2016).
52. Kaenel, P., Mosimann, M. & Andres, A.-C. The multifaceted roles of Eph/ephrin signaling in breast cancer. *Cell Adh. Migr.* **6**, 138–147 (2012).
53. Fu, Y., Zheng, Q., Mao, Y., Jiang, X., Chen, X., Liu, P., Lv, B., Huang, T., Yang, J., Cheng, Y., Dai, X., Dai, C., Wang, X., Yin, Y., Song, T., Jin, W., Zou, C., Chen, T., Fu, L. & Chen, Z. WNT2-Mediated FZD2 Stabilization Regulates Esophageal Cancer Metastasis via STAT3 Signaling. *Front. Oncol.* **10**, 1168 (2020).
54. Gujral, T. S., Chan, M., Peshkin, L., Sorger, P. K., Kirschner, M. W. & MacBeath, G. A noncanonical Frizzled2 pathway regulates epithelial-mesenchymal transition and metastasis. *Cell* **159**, 844–856 (2014).
55. Shi, Z., Rifa'i, M., Lee, Y. H., Shiku, H., Isobe, K.-I. & Suzuki, H. Importance of CD80/CD86-CD28 interactions in the recognition of target cells by CD8+CD122+ regulatory T cells. *Immunology* **124**, 121–128 (2008).
56. Pende, D., Bottino, C., Castriconi, R., Cantoni, C., Marcenaro, S., Rivera, P., Spaggiari, G. M., Dondero, A., Carnemolla, B., Reymond, N., Mingari, M. C., Lopez, M., Moretta, L. & Moretta, A. PVR (CD155) and Nectin-2 (CD112) as ligands of the human DNAM-1 (CD226) activating receptor: involvement in tumor cell lysis. *Mol. Immunol.* **42**, 463–469 (2005).
57. Nachmanson, D., Officer, A., Mori, H., Gordon, J., Evans, M. F., Steward, J., Yao, H.,

O'Keefe, T., Hasteh, F., Stein, G. S., Jepsen, K., Weaver, D. L., Hirst, G. L., Sprague, B. L., Esserman, L. J., Borowsky, A. D., Stein, J. L. & Harismendy, O. The breast pre-cancer atlas illustrates the molecular and micro-environmental diversity of ductal carcinoma in situ. *npj Breast Cancer* **8**, 1–13 (2022).
58. Gil Del Alcazar, C. R., Huh, S. J., Ekram, M. B., Trinh, A., Liu, L. L., Beca, F., Zi, X., Kwak, M., Bergholtz, H., Su, Y., Ding, L., Russnes, H. G., Richardson, A. L., Babski, K., Min Hui Kim, E., McDonnell, C. H., 3rd, Wagner, J., Rowberry, R., Freeman, G. J., Dillon, D., Sorlie, T., Coussens, L. M., Garber, J. E., Fan, R., Bobolis, K., Allred, D. C., Jeong, J., Park, S. Y., Michor, F. & Polyak, K. Immune Escape in Breast Cancer During In Situ to Invasive Carcinoma Transition. *Cancer Discov.* **7**, 1098–1115 (2017).
59. Kristensen, V. N., Vaske, C. J., Ursini-Siegel, J., Van Loo, P., Nordgard, S. H., Sachidanandam, R., Sørlie, T., Wärnberg, F., Haakensen, V. D., Helland, Å., Naume, B., Perou, C. M., Haussler, D., Troyanskaya, O. G. & Børresen-Dale, A.-L. Integrated molecular profiles of invasive breast tumors and ductal carcinoma in situ (DCIS) reveal differential vascular and interleukin signaling. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 2802–2807 (2012).
60. Lo, P.-K., Zhang, Y., Yao, Y., Wolfson, B., Yu, J., Han, S.-Y., Duru, N. & Zhou, Q. Tumor-associated myoepithelial cells promote the invasive progression of ductal carcinoma in situ through activation of TGFβ signaling. *J. Biol. Chem.* **292**, 11466–11484 (2017).
61. Knudsen, E. S., Ertel, A., Davicioni, E. & Witkiewicz, A. K. Progression of ductal carcinoma in situ to invasive breast cancer is associated with gene expression programs of EMT and myoepithelia. 1009–1024 (2012).
62. Matta, J., Morales, L., Ortiz, C., Adams, D., Vargas, W., Casbas, P., Dutil, J., Echenique, M. & Suárez, E. Estrogen Receptor Expression Is Associated with DNA Repair Capacity in Breast Cancer. *PLoS One* **11**, e0152422 (2016).
63. Wang, Z., Zhang, Z., Guo, Y., Shui, H., Liu, G., Jin, T. & Wang, H. Shorter Telomere Length Is Associated with Increased Breast Cancer Risk in a Chinese Han Population: A Case-Control Analysis. *J. Breast Cancer* **21**, 391–398 (2018).
64. Herbert, B. S., Wright, W. E. & Shay, J. W. Telomerase and breast cancer. *Breast Cancer Res.* **3**, 146–149 (2001).
65. Meurette, O. & Mehlen, P. Notch Signaling in the Tumor Microenvironment. *Cancer Cell* **34**, 536–548 (2018).
66. Liubomirski, Y., Lerrer, S., Meshel, T., Morein, D., Rubinstein-Achiasaf, L., Sprinzak, D., Wiemann, S., Körner, C., Ehrlich, M. & Ben-Baruch, A. Notch-Mediated Tumor-Stroma-Inflammation Networks Promote Invasive Properties and CXCL8 Expression in Triple-Negative Breast Cancer. *Front. Immunol.* **10**, 804 (2019).
67. Nandi, A. & Chakrabarti, R. The many facets of Notch signaling in breast cancer: toward overcoming therapeutic resistance. *Genes Dev.* **34**, 1422–1438 (2020).
68. Wang, K., Patkar, S., Lee, J. S., Michael Gertz, E., Robinson, W., Schischlik, F., Crawford, D. R., Schäffer, A. A. & Ruppin, E. Deconvolving clinically relevant cellular immune crosstalk from bulk gene expression using CODEFACS and LIRICS. Preprint at https://doi.org/10.1101/2021.01.20.427515.
69. Armingol, E., Baghdassarian, H. M., Martino, C., Perez-Lopez, A., Knight, R. & Lewis, N. E. Context-aware deconvolution of cell-cell communication with Tensor-cell2cell. *bioRxiv* 2021.09.20.461129 (2021) doi:10.1101/2021.09.20.461129.
70. Li, D., Velazquez, J. J., Ding, J., Hislop, J., Ebrahimkhani, M. R. & Bar-Joseph, Z. Inferring cell-cell interactions from pseudotime ordering of scRNA-Seq data. *bioRxiv* 2021.07.28.454054 (2021) doi:10.1101/2021.07.28.454054.
71. Noël, F., Massenet-Regad, L., Carmi-Levy, I., Cappuccio, A., Grandclaudon, M., Trichot, C., Kieffer, Y., Mechta-Grigoriou, F. & Soumelis, V. ICELLNET: a transcriptome-based framework to dissect intercellular communication. *bioRxiv* 2020.03.05.976878 (2020) doi:10.1101/2020.03.05.976878.

72. Zhan, T., Rindtorff, N. & Boutros, M. Wnt signaling in cancer. *Oncogene* **36**, 1461–1473 (2017).
73. Koval, A. & Katanaev, V. L. Dramatic dysbalancing of the Wnt pathway in breast cancers. *Sci. Rep.* **8**, 7329 (2018).
74. Sousa, S., Brion, R., Lintunen, M., Kronqvist, P., Sandholm, J., Mönkkönen, J., Kellokumpu-Lehtinen, P.-L., Lauttia, S., Tynninen, O., Joensuu, H., Heymann, D. & Määttä, J. A. Human breast cancer cells educate macrophages toward the M2 activation status. *Breast Cancer Res.* **17**, 101 (2015).
75. Halliday, N., Williams, C., Kennedy, A., Waters, E., Pesenacker, A. M., Soskic, B., Hinze, C., Hou, T. Z., Rowshanravan, B., Janman, D., Walker, L. S. K. & Sansom, D. M. CD86 is a selective CD28 ligand supporting FoxP3+ regulatory T cell homeostasis in the presence of high levels of CTLA-4. *Front. Immunol.* **11**, 600000 (2020).
76. Vaught, D., Brantley-Sieders, D. M. & Chen, J. Eph receptors in breast cancer: roles in tumor promotion and tumor suppression. *Breast Cancer Res.* **10**, 217 (2008).
77. Park, I. & Lee, H.-S. EphB/ephrinB signaling in cell adhesion and migration. *Mol. Cells* **38**, 14–19 (2015).
78. Lips, E. H., Kumar, T., Megalios, A., Visser, L. L., Sheinman, M., Fortunato, A., Shah, V., Hoogstraat, M., Sei, E., Mallo, D., Roman-Escorza, M., Ahmed, A. A., Xu, M., van den Belt-Dusebout, A. W., Brugman, W., Casasent, A. K., Clements, K., Davies, H. R., Fu, L., Grigoriadis, A., Hardman, T. M., King, L. M., Krete, M., Kristel, P., de Maaker, M., Maley, C. C., Marks, J. R., Menegaz, B. A., Mulder, L., Nieboer, F., Nowinski, S., Pinder, S., Quist, J., Salinas-Souza, C., Schaapveld, M., Schmidt, M. K., Shaaban, A. M., Shami, R., Sridharan, M., Zhang, J., Stobart, H., Collyar, D., Nik-Zainal, S., Wessels, L. F. A., Hwang, E. S., Navin, N. E., Futreal, P. A., Grand Challenge PRECISION consortium, Thompson, A. M., Wesseling, J. & Sawyer, E. J. Genomic analysis defines clonal relationships of ductal carcinoma in situ and recurrent invasive breast cancer. *Nat. Genet.* **54**, 850–860 (2022).
79. Nachmanson, D., Pagadala, M., Steward, J., Cheung, C., Bruce, L. K., Lee, N. Q., O'Keefe, T. J., Lin, G. Y., Hasteh, F., Morris, G. P., Carter, H. & Harismendy, O. Accurate genome-wide germline profiling from decade-old archival tissue DNA reveals the contribution of common variants to precancer disease outcome. *bioRxiv* (2022) doi:10.1101/2022.03.31.22273116.
80. Schuetz, C. S., Bonin, M., Clare, S. E., Nieselt, K., Sotlar, K., Walter, M., Fehm, T., Solomayer, E., Riess, O., Wallwiener, D., Kurek, R. & Neubauer, H. J. Progression-Specific Genes Identified by Expression Profiling of Matched Ductal Carcinomas In situ and Invasive Breast Tumors , Combining Laser Capture Microdissection and Oligonucleotide Microarray Analysis. 5278–5287 (2006).
81. Hannafon, B. N., Sebastiani, P., de las Morenas, A., Lu, J. & Rosenberg, C. L. Expression of microRNA and their gene targets are dysregulated in preinvasive breast cancer. *Breast Cancer Res.* **13**, R24 (2011).
82. Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U. & Speed, T. P. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).
83. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
84. Steen, C. B., Liu, C. L., Alizadeh, A. A. & Newman, A. M. Profiling Cell Type Abundance and Expression in Bulk Tissues with CIBERSORTx. *Methods Mol. Biol.* **2117**, 135–157 (2020).
85. Fang, Z., Liu, X. & Peltz, G. GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* **39**, (2023).
86. Zhang, D., Yang, S., Li, Y., Yao, J., Ruan, J., Zheng, Y., Deng, Y., Li, N., Wei, B., Wu, Y.,

Zhai, Z., Lyu, J. & Dai, Z. Prediction of Overall Survival Among Female Patients With Breast Cancer Using a Prognostic Signature Based on 8 DNA Repair-Related Genes. *JAMA Netw Open* **3**, e2014622 (2020).
87. Fredriksson, R., Lagerström, M. C., Lundin, L.-G. & Schiöth, H. B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* **63**, 1256–1272 (2003).
88. Schöneberg, T., Schulz, A., Biebermann, H., Hermsdorf, T., Römpler, H. & Sangkuhl, K. Mutant G-protein-coupled receptors as a cause of human diseases. *Pharmacol. Ther.* **104**, 173–206 (2004).
89. Rosenbaum, D. M., Rasmussen, S. G. F. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459**, 356–363 (2009).
90. Simon, M. I., Strathmann, M. P. & Gautam, N. Diversity of G proteins in signal transduction. *Science* **252**, 802–808 (1991).
91. Feng, X., Degese, M. S., Iglesias-Bartolome, R., Vaque, J. P., Molinolo, A. A., Rodrigues, M., Zaidi, M. R., Ksander, B. R., Merlino, G., Sodhi, A., Chen, Q. & Gutkind, J. S. Hippo-independent activation of YAP by the GNAQ uveal melanoma oncogene through a trio-regulated rho GTPase signaling circuitry. *Cancer Cell* **25**, 831–845 (2014).
92. Carvajal, R. D., Schwartz, G. K., Tezel, T., Marr, B., Francis, J. H. & Nathan, P. D. Metastatic disease from uveal melanoma: treatment options and future prospects. *Br. J. Ophthalmol.* **101**, 38–44 (2017).
93. Shain, A. H., Bagger, M. M., Yu, R., Chang, D., Liu, S., Vemula, S., Weier, J. F., Wadt, K., Heegaard, S., Bastian, B. C. & Kiilgaard, J. F. The genetic evolution of metastatic uveal melanoma. *Nat. Genet.* **51**, 1123–1130 (2019).
94. Onken, M. D., Makepeace, C. M., Kaltenbronn, K. M., Choi, J., Hernandez-Aya, L., Weilbaecher, K. N., Piggott, K. D., Kumar Rao, P., Yuede, C. M., Dixon, A. J., Osei-Owusu, P., Cooper, J. A. & Blumer, K. J. Targeting primary and metastatic uveal melanoma with a G protein inhibitor. *J. Biol. Chem.* **296**, (2021).
95. Takasaki, J., Saito, T., Taniguchi, M., Kawasaki, T., Moritani, Y., Hayashi, K. & Kobori, M. A novel Galphaq/11-selective inhibitor. *J. Biol. Chem.* **279**, 47438–47445 (2004).
96. de Lange, M. J., Razzaq, L., Versluis, M., Verlinde, S., Dogrusöz, M., Böhringer, S., Marinkovic, M., Luyten, G. P. M., de Keizer, R. J. W., de Gruijl, F. R., Jager, M. J. & van der Velden, P. A. Distribution of GNAQ and GNA11 Mutation Signatures in Uveal Melanoma Points to a Light Dependent Mutation Mechanism. *PLoS One* **10**, e0138002 (2015).
97. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
98. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
99. Vanden Heuvel, J. P., Maddox, E., Maalouf, S. W. & Reproducibility Project: Cancer Biology. Replication Study: Systematic identification of genomic markers of drug sensitivity in cancer cells. *Elife* **7**, (2018).
100. Siwak, D. R., Li, J., Akbani, R., Liang, H. & Lu, Y. Analytical Platforms 3: Processing Samples via the RPPA Pipeline to Generate Large-Scale Data for Clinical Studies. *Adv. Exp. Med. Biol.* **1188**, 113–147 (2019).
101. Loewe, S. The problem of synergism and antagonism of combined drugs. *Arzneimittelforschung* **3**, 285–290 (1953).
102. Zheng, S., Wang, W., Aldahdooh, J., Malyutina, A., Shadbahr, T., Tanoli, Z., Pessia, A. & Tang, J. SynergyFinder Plus: Toward Better Interpretation and Annotation of Drug Combination Screening Datasets. *Genomics Proteomics Bioinformatics* **20**, 587–596 (2022).
103. Jones, K. B., Furukawa, S., Marangoni, P., Ma, H., Pinkard, H., D'Urso, R., Zilionis, R., Klein, A. M. & Klein, O. D. Quantitative Clonal Analysis and Single-Cell Transcriptomics

Reveal Division Kinetics, Hierarchy, and Fate of Oral Epithelial Progenitor Cells. *Cell Stem Cell* **24**, 183–192.e8 (2019).
104. Chen, D., Wu, M., Li, Y., Chang, I., Yuan, Q., Ekimyan-Salvo, M., Deng, P., Yu, B., Yu, Y., Dong, J., Szymanski, J. M., Ramadoss, S., Li, J. & Wang, C.-Y. Targeting BMI1+ Cancer Stem Cells Overcomes Chemoresistance and Inhibits Metastases in Squamous Cell Carcinoma. *Cell Stem Cell* **20**, 621–634.e6 (2017).
105. Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Piñeros, M., Znaor, A. & Bray, F. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int. J. Cancer* **144**, 1941–1953 (2019).
106. Lian, I., Kim, J., Okazawa, H., Zhao, J., Zhao, B., Yu, J., Chinnaiyan, A., Israel, M. A., Goldstein, L. S. B., Abujarour, R., Ding, S. & Guan, K.-L. The role of YAP transcription coactivator in regulating stem cell self-renewal and differentiation. *Genes Dev.* **24**, 1106–1118 (2010).
107. Bartkova, J., Rezaei, N., Liontos, M., Karakaidos, P., Kletsas, D., Issaeva, N., Vassiliou, L.-V. F., Kolettas, E., Niforou, K., Zoumpourlis, V. C., Takaoka, M., Nakagawa, H., Tort, F., Fugger, K., Johansson, F., Sehested, M., Andersen, C. L., Dyrskjot, L., Ørntoft, T., Lukas, J., Kittas, C., Helleday, T., Halazonetis, T. D., Bartek, J. & Gorgoulis, V. G. Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. *Nature* **444**, 633–637 (2006).
108. Blanpain, C. & Fuchs, E. Epidermal homeostasis: a balancing act of stem cells in the skin. *Nat. Rev. Mol. Cell Biol.* **10**, 207–217 (2009).
109. Schwitalla, S., Fingerle, A. A., Cammareri, P., Nebelsiek, T., Göktuna, S. I., Ziegler, P. K., Canli, O., Heijmans, J., Huels, D. J., Moreaux, G., Rupec, R. A., Gerhard, M., Schmid, R., Barker, N., Clevers, H., Lang, R., Neumann, J., Kirchner, T., Taketo, M. M., van den Brink, G. R., Sansom, O. J., Arkan, M. C. & Greten, F. R. Intestinal tumorigenesis initiated by dedifferentiation and acquisition of stem-cell-like properties. *Cell* **152**, 25–38 (2013).
110. Puram, S. V., Tirosh, I., Parikh, A. S., Patel, A. P., Yizhak, K., Gillespie, S., Rodman, C., Luo, C. L., Mroz, E. A., Emerick, K. S., Deschler, D. G., Varvares, M. A., Mylvaganam, R., Rozenblatt-Rosen, O., Rocco, J. W., Faquin, W. C., Lin, D. T., Regev, A. & Bernstein, B. E. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **171**, 1611–1624.e24 (2017).
111. Callejas-Valera, J. L., Iglesias-Bartolome, R., Amornphimoltham, P., Palacios-Garcia, J., Martin, D., Califano, J. A., Molinolo, A. A. & Gutkind, J. S. mTOR inhibition prevents rapid-onset of carcinogen-induced malignancies in a novel inducible HPV-16 E6/E7 mouse model. *Carcinogenesis* **37**, 1014–1025 (2016).
112. Tang, X.-H., Scognamiglio, T. & Gudas, L. J. Basal stem cells contribute to squamous cell carcinomas in the oral cavity. *Carcinogenesis* **34**, 1158–1164 (2013).
113. Zanconato, F., Forcato, M., Battilana, G., Azzolin, L., Quaranta, E., Bodega, B., Rosato, A., Bicciato, S., Cordenonsi, M. & Piccolo, S. Genome-wide association between YAP/TAZ/TEAD and AP-1 at enhancers drives oncogenic growth. *Nat. Cell Biol.* **17**, 1218–1227 (2015).
114. Zhao, B., Ye, X., Yu, J., Li, L., Li, W., Li, S., Yu, J., Lin, J. D., Wang, C.-Y., Chinnaiyan, A. M., Lai, Z.-C. & Guan, K.-L. TEAD mediates YAP-dependent gene induction and growth control. *Genes Dev.* **22**, 1962–1971 (2008).
115. Yuan, Y., Park, J., Feng, A., Awasthi, P., Wang, Z., Chen, Q. & Iglesias-Bartolome, R. YAP1/TAZ-TEAD transcriptional networks maintain skin homeostasis by regulating cell proliferation and limiting KLF4 activity. *Nat. Commun.* **11**, 1472 (2020).
116. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
117. Meers, M. P., Tenenbaum, D. & Henikoff, S. Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics Chromatin* **12**, 42 (2019).

118. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
119. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–91 (2014).
120. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35**, 2084–2092 (2018).
121. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. & Glass, C. K. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
122. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
123. Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W. & Liu, X. S. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
124. Amornphimoltham, P., Patel, V., Sodhi, A., Nikitakis, N. G., Sauk, J. J., Sausville, E. A., Molinolo, A. A. & Gutkind, J. S. Mammalian target of rapamycin, a molecular target in squamous cell carcinomas of the head and neck. *Cancer Res.* **65**, 9953–9961 (2005).
125. Molinolo, A. A., Hewitt, S. M., Amornphimoltham, P., Keelawat, S., Rangdaeng, S., Meneses García, A., Raimondi, A. R., Jufe, R., Itoiz, M., Gao, Y., Saranath, D., Kaleebi, G. S., Yoo, G. H., Leak, L., Myers, E. M., Shintani, S., Wong, D., Massey, H. D., Yeudall, W. A., Lonardo, F., Ensley, J. & Gutkind, J. S. Dissecting the Akt/mammalian target of rapamycin signaling network: emerging results from the head and neck cancer tissue array initiative. *Clin. Cancer Res.* **13**, 4964–4973 (2007).
126. Molinolo, A. A., Marsh, C., El Dinali, M., Gangane, N., Jennison, K., Hewitt, S., Patel, V., Seiwert, T. Y. & Gutkind, J. S. mTOR as a molecular target in HPV-associated oral and cervical squamous carcinomas. *Clin. Cancer Res.* **18**, 2558–2568 (2012).
127. Iglesias-Bartolome, R., Martin, D. & Gutkind, J. S. Exploiting the head and neck cancer oncogenome: widespread PI3K-mTOR pathway alterations and novel molecular targets. *Cancer Discov.* **3**, 722–725 (2013).