ADVANCED SCREENING AND INTEGRATIVE ANALYSIS FOR BIOTECHNOLOGY AND DISEASE

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Miranda Burnette

______________________________
Jeremiah J. Zartman, Director

Graduate Program in Chemical and Biomolecular Engineering
Notre Dame, Indiana
June 2016
ADVANCED SCREENING AND INTEGRATIVE ANALYSIS FOR BIOTECHNOLOGY AND DISEASE

Abstract

by

Miranda Burnette

In the era of big data, there is an ever-increasing breadth of both open-source data and advanced quantitative tools available for studying biology; from RNA-sequencing to high-throughput screening (HTS), advanced technologies have changed the landscape of biological research and medicine. However, there are still many challenges associated with fully harnessing the available resources for particular applications; specifically, comparatively little has been done to integrate data and disparate tools for systems-level solutions. The work herein describes multiple case studies where integrative approaches were applied to synthesize and build on the wealth of resources available to advance biotechnology, preclinical medicine, and more fundamentally in developmental biology. This thesis describes several milestone applications that advance the state-of-the-art in each field: the generation of the first chemically defined cell culture medium enabling long term culture of a Drosophila cell lines (Chapter 2); building on these results led to unexpected discoveries in the field of developmental biology with the discovery of novel spatiotemporal patterning of
polyamines during organogenesis (Chapter 3); the optimization and application of an *in vivo* tumor screening approach to efficiently identify novel functional targets during secondary metastatic outgrowth in tumors (Chapter 4); and through a close, multidisciplinary collaboration, the characterization of novel compounds through both cytotoxicity and mechanism of action studies, along with the preliminary creation of an ultra-high-content screening platform for structure-activity analysis of compound libraries (Chapter 5). For each application, several avenues are available to further progress the technology, which is discussed in the concluding chapter (Chapter 6).
For everyone who helped me get here
| Chapter 3: Spatiotemporal Patterning of Polyamines in Drosophila Development | 62 |
| Chapter 4: Rapid Transcriptomics with High-Throughput Functional Validation Elucidates Novel Therapeutic Targets for Breast Cancer Brain Metastasis | 77 |
| Chapter 5: Integrative Evaluation and Characterization of Novel Spirooxindoles for Treating Breast Cancer Brain Metastasis | 106 |
| Chapter 6: Conclusions and Discussion | 145 |

### Chapter 3: Spatiotemporal Patterning of Polyamines in Drosophila Development
- **3.1 Introduction** ......................................................... 62
- **3.2 Results** ........................................................................ 63
- **3.3 Discussion** ............................................................... 73
- **3.4 Methods** ..................................................................... 75
- **3.5 Acknowledgments** ..................................................... 76

### Chapter 4: Rapid Transcriptomics with High-Throughput Functional Validation Elucidates Novel Therapeutic Targets for Breast Cancer Brain Metastasis
- **4.1 Introduction** ............................................................ 77
- **4.2 Results** ....................................................................... 84
- **4.3 Methods** .................................................................. 100
  - 4.3.1 Identification of *Drosophila* screening reagents .......... 100
  - 4.3.2 Tumor screen ........................................................... 100
  - 4.3.3 PCR .......................................................................... 101
- **4.4 Discussion** ............................................................... 102
- **4.5 Acknowledgments** ..................................................... 105

### Chapter 5: Integrative Evaluation and Characterization of Novel Spirooxindoles for Treating Breast Cancer Brain Metastasis
- **5.1 Introduction** ............................................................ 106
- **5.2 Results** .................................................................. 110
  - 5.2.1 Assessing half-maximal effective concentration (EC\textsubscript{50}) .......... 110
  - 5.2.2 Detecting mechanism of action by network dysregulation (DeMAND) .................. 114
  - 5.2.3 Analyzing structure-activity relationships by uHCS .......... 127
- **5.3 Methods** ................................................................ 136
  - 5.3.1 Cell culture ............................................................... 136
  - 5.3.2 Randomized compound stock plates .......................... 136
  - 5.3.3 Determination of EC\textsubscript{50} .................................. 136
  - 5.3.4 Prediction of mechanism of action (DeMAND) ................ 137
  - 5.3.5 Protein assay ........................................................... 139
  - 5.3.6 Phalloidin staining .................................................... 140
  - 5.3.7 Classification of MDA-231-Br and BT-474-Br2 cell lines .......... 140
  - 5.3.8 Ultra-high-content screen ........................................... 141
- **5.4 Discussion** ............................................................ 143
- **5.5 Acknowledgments** ................................................ 144

### Chapter 6: Conclusions and Discussion
- **6.1 Summary of findings** ............................................. 145
- **6.2 New questions and future directions** ...................... 148
FIGURES

Figure 1.1: RNA interference. (A) RNAi pathway. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Immunology Ref. 64, copyright (2003). (B) SiRNAs, shRNAs, shmiRNAs, and long dsRNAs pathways. Reprinted by permission from Cold Spring Harb Perspect Biol 2010;2:a003640, copyright to Cold Spring Harbor Laboratory Press (2010).

Figure 1.2: Drosophila melanogaster as a model organism. (A) Attractive features of Drosophila melanogaster as a model organism. (B-D) Genetic tools to control transgene expression in Drosophila. (B) The yeast transcription activator protein GAL4 is expressed under the control of an endogenous promoter. GAL4 is expressed and binds to the GAL4 specific upstream activating site (UAS) ahead of a target transgene, activating its transcription specifically in the tissue in which GAL4 is expressed. More recently developed alternative binary expression systems LexA/opA and Q operate analogously to GAL4/UAS. (C) GAL4- and Q-mediated expression can be further tuned through the repressors GAL80 or QS, respectively. (D) Further control can be exerted through use of site-specific recombinases in a technique called Flp-out, where site specific recombinases remove a stop sequence that prevent transgene expression. Thus expression is limited to where GAL4 and Flp are present but GAL80 is not (green area in D'). Panels B-D Reprinted by permission from Macmillan Publishers Ltd: Nature Methods Ref. 64, copyright (2016).

Figure 2.1: (A) Advantages and limitations of in vivo and in vitro experimentation. The precise control offered by in vitro culture is abrogated by the required medium supplementation with undefined extracts. (B) Comparison of mammalian versus Drosophila cell culture. (C) Comparison of Cl.8 growth kinetics in complete serum containing media, ZO media unsupplemented, and “ZO Fortified.” ZO Fortified supports initial attachment and proliferation of Cl.8 cells whereas ZO unsupplemented does not (Figure A.1). Single and double asterisks denote p < 0.05 and p < 0.01, respectively, for two-tailed t-test for unmatched pairs. (D) Screening pipeline. Small molecule libraries are screened on two cell lines for their ability to promote proliferation in a minimal, serum-free medium. Putative hit compounds are then compared between cell lines to minimize false positives, and analysis on putative targets is conducted. The screen can be iterated upon to
identify compound synergies by incorporating hit compounds into the background media and rescreening. (E) Target scoring pipeline. Compound-protein interaction scores ($s_i$) are identified from a database and linked to compound z-scores. Non-Drosophila protein targets are “translated” to their Drosophila orthologs retaining their percent identity match score, $q_i$. Drosophila protein target p-values are then calculated by summing for each interaction with a compound its compound-protein interaction score, percent identity match scores, and squared compound z-scores (which follows a chi-squared distribution). This analysis yields a list of significantly targeted proteins for a process of interest. Protein target lists can be also converted to their encoding genes and gene ontology enrichment can be performed. More details can be found in the materials and methods.

Figure 2.2: (A-B) Z-scores for all screened compounds on (A) Cl.8 cells and (B) S2 cells. (C-D) Representative images of hit compounds on Cl.8 (C) and S2 (D) cells. (E-F) Venn diagrams for unique positive (E) and negative (F) hit compounds.

Figure 2.3: Growth kinetics of Clone 8, Schneider 2, and Kc167 cells in ZO unsupplemented, ZO Fortified, ZB Media, and their respective complete, serum-containing media. Error bars represent standard deviations. (A-C) Changes in averaged fluorescent intensity measurements indicate changes in DNA content from viable cells. Error bars represent standard deviations. (D-F) Fold changes in intensity from day 0 are taken to represent fold changes in cell number. Each of the four cell lines proliferate significantly better in ZB Media (with spermidine supplementation) than ZO Fortified by the fifth day. Kc167 cells in particular proliferate at comparable rates to the complete, serum-containing medium by day 5. Error bars represent standard deviations for which propagation of error is accounted. Single and double asterisks denote $p < 0.05$ and $p < 0.01$, respectively, for two-tailed t-test for unmatched pairs. Corresponding p-values can be found in the electronic supplementary information. (G-I) Cells from day five of the same experiment. The apparent low confluency of Kc167 cells (I) is due to a high number of suspension cells that cannot be seen in one plane of view.

Figure 2.4: (A-B) Proliferation of polyamine-depleted Cl.8 cells increases in a dose-dependent manner with spermidine supplementation. Error bars represent standard deviations with for which propagation of error is accounted. Single and double asterisks denote $p < 0.05$ and $p < 0.01$, respectively, for two-tailed t-test for unmatched pairs. (C) Spermidine supplementation (ZB Media) results in increased EdU incorporation in both Cl.8 and S2 cells. Error bars represent standard deviations. Single and double asterisks denote $p < 0.05$ and $p < 0.01$, respectively, for two-tailed t-test for unmatched pairs. (D-E) Representative
images of EdU (magenta) and DAPI (blue) stained Cl.8 (D) and S2 (E) cells in each culture medium tested. ................................................................. 49

Figure 2.5: (A-B) Spermidine supplementation (ZB Media) results in increases in ERK phosphorylation (dpERK) for Cl.8 cells (p < 0.012), but not for S2 cells (p > 0.75). (B-C) Although ERK expression does not change significantly with spermidine supplementation for either Cl.8 (p > 0.55) or S2 (p < 0.45) cells, there is a trend of increasing ERK expression across ZO Media (unsupplemented) improvements. Error bars represent standard deviations across three replicates. Single and double asterisks denote p < 0.05 and p < 0.01, respectively, for two-tailed t-test for unmatched pairs. Corresponding p-values can be found in the electronic supplementary information. ................................................................. 52

Figure 2.6: REVIGO scatter plots of enriched ontology terms for targets of (A) positive scoring compounds for Cl.8 cells, (B) positive scoring compound for S2 cells, (C) negative scoring compounds for Cl.8 cells, and (D) negative scoring compounds for S2 cells. Enriched ontology annotations are grouped into terms by semantic similarity and plotted on semantic axes, where similar terms are clustered closely on the plot. Circles representing terms are colour-coded according to p-value and sized according to the number of ontology annotations per term. .......................... 55

Figure 3.1: (A) Polyamine metabolic pathway, adapted. Metabolites in white boxes, enzymes in ovals—human gene top, Drosophila ortholog bottom (bold). Enzymes are color-coded to indicate observed regulatory clusters. (B) Principle component analysis (PCA) of PA genes across development reveals four clusters that follow the regulatory logic of the metabolic pathway (black arrow): upstream (red) and downstream (blue) biosynthesis, interconversion (blue), and inhibition (yellow). .................................................................................. 65

Figure 3.2: Normalized gene cluster dynamics and measured PA accumulation profiles (left and right axes, respectively) across stages of Drosophila embryonic (top) and larval (bottom) development (gaps represent data gaps from the literature). Average cluster dynamics align to predict basic dynamics of PA accumulation through development. ......................................................................... 68

Figure 3.3: (A) The embryonic trachea is a branched tubular organ that develops in the latter half of embryogenesis. (B) Tube maturation depends on the formation of a chitinous matrix (shown in green) that expands the lumen diameter. This matrix is generated when btl-expressing tracheal cells (orange) experience a “secretion pulse”, inducing luminal deposition of chitin-binding proteins that expand the tube diameter. A subsequent endocytosis pulse rapidly clears the luminal matrix. Figure adapted. (C-D) Analysis of PA accumulation in developing embryos revealed PA deposition within the lumen of viii
expanding trachea. Polyamines were accumulated with the same dynamics of 2A12 (magenta), a luminal marker for GASP, a chitin-binding protein. (scale bars 50 µm). (E-F) Differentiation of the larval eye-antennal imaginal disc is driven by the morphogenetic furrow (orange), which initiates at the posterior part of the disc and proceeds to the anterior side. The progression of the morphogenetic furrow is dependent on well characterized morphogens including Decapentaplegic (Dpp) and Hedgehog (Hh) (blue) as well as and Wingless (Wg) (red), Notch (purple), and Epidermal Growth Factor Receptor (EGFR) (green). Figure adapted . (G-I) PAs (red) are accumulated along the morphogenetic furrow in eye discs expressing E-cadherin::GFP (green, apical cell boundaries), with differentiated ommatidia present to the posterior (g-h, day 4; i-j, day six after egg laying). The PA pattern is initially more widespread (G-H), and becomes progressively more confined to the furrow (I-J). Scale bars 20 µm (G,I), 5 µm (H,J).

Figure 4.1: Schematic of transcriptome deep-sequencing experiment identifying BrainMets signature. Single (7 dpi) or actively proliferating (40 dpi) RFP labeled MDA-231 tumor cells were isolated from the brain and subjected to RNA-seq. Actively proliferating metastases were classified as either small, medium, or large by visual inspection. Differential expression analysis revealed 537 genes that were significantly different between d7 and d40 small, medium, or large metastases. These 537 genes were taken to be the “BrainMets signature.” Figure graciously provided by Dr. Siyuan Zhang.

Figure 4.2: *Drosophila* tumor screening model. Schematic of a *Drosophila* larva and relevant tissues (top). The model we utilized expresses UAS-transgenes under control of a FLP-out GAL4 transgene that stably expresses GAL4 after FLP-induced recombination. FLP-induced recombination, and thus transgene expression, is restricted to the eye discs using an eyeless promoter-driven FLP transgene. Spatial extent of transgene expression is visualized using UAS-GFP to mark GAL4-expressing cells (green, bottom left). Additional expression of oncogenic Ras results in neoplastic overgrowth of the tissue but no metastatic outgrowth. Expression of shRNA against the epithelial polarity marker discs large (dlg) results in elimination of those cells by neighboring cells without FLP-induced transgene expression. Concurrent expression of Ras and dlg results in neoplastic overgrowth of the eye discs and metastatic invasion into the neighboring brain and ventral nerve chord (bottom right). Figure adapted from Halder and Mills .

Figure 4.3: Pipeline for identification of screening reagents. Briefly, the 537 differentially regulated genes between small, medium, and large overt metastases were reduced to 126 genes using Fisher’s exact test to combine
significance. Conversion of human genes to *Drosophila* genes by homology resulted in 241 *Drosophila* genes representing 81 of the 126 signature genes. Finally, we acquired 368 RNAi reagents to functionally test those 126 human genes in our *Drosophila* tumor screen. ................................................................. 90

Figure 4.4: SSMD plot from tumor screen. Flies from the GFP expressing tumor line were crossed to flies bearing RNAi constructs under control of UAS, enabling spatially restricted expression of shRNA in the tumor region. Tumors were given six days to develop in larva before larva were harvested and imaged. Integrated intensity of GFP was calculated for at least 15 larvae per cross, and SSMD scores were calculated with respect to the negative control (yw, no additional RNAi). shRNA against the tumor suppressor PTEN was used as a positive control. Hits are characterized as moderate or strong negatives as indicated. .......................... 93

Figure 4.5: (A-D) Integrated GFP intensity for SERPINF1, RAB11B, MRPL37, and SNRPD2 RNAi lines and controls. Error bars represent standard deviation. (E) Representative images of larvae for controls (yw, negative; shPTEN, positive) and shSERPINF1, shMRPL37, shSNRPD2, and shRAB11b. ............................................. 95

Figure 4.6: (A) qRT-PCR validation of tumor mediating gene mRNA levels in primary mammary gland tumors versus intracranially-injected brain metastases. Error bars represent standard deviation; asterisk denotes p<0.05 (t-test). (B) Rab11b and SERPINF1 are up-regulated in brain metastases. ............................................. 98

Figure 5.1: General molecular framework of spirooxindoles (A) and novel bis-spirooxindoles (B) with highlighted points of structural diversity (Table 5.1). .. 108

Figure 5.2: EC50 curves for top performing compounds. The top two most efficacious compounds against MDA-231 cells—13b (A) and 12b (C)—demonstrated comparable efficacy to the tyrosine kinase inhibitor, lapatanib. MDA-231-Br cells are, in general, more resistant to our compounds than their parental line, but 13b (B) and 12i (D) were more effective than temozolamide.............................. 113

Figure 5.3: Experimental basis for DeMAND118. (A) DeMAND determines targeted genes by network analysis. In this example, while G0 expression is not affected by treatment, the regulon components G2, G5, and G7 are down-regulated, while G3 and G4 are up-regulated. Although differential expression analysis would not identify G0 as a potential target, given the degree of dysregulation of the G0 regulon, DeMAND identifies G0 as a target. (B) Experimental pipeline. (C) Cytoscape visualization of targets of 13b implicated from DeMAND with FDR q-value < 0.001. ..................................................................................................... 116
Figure 5.4: DeMAND\textsuperscript{118}. (A) Normalized expression of top 100 DeMAND implicated genes at 6, 12, and 24 hours post-treatment. (B) Rank in DeMAND output of top 30 DeMAND implicated genes using combined versus individual timepoint expression. ................................................................. 118

Figure 5.5: DeMAND implicates ribosomes as target of 13b. (A) Top 15 most significant negatively-(yellow) and positively-(blue) regulated gene sets from GSEA. (B) Overlap between genes in the top 15 most negatively-enriched gene sets. (C) Relative rank of predicted effector genes across custom classes of common MoAs. Genes predicted by DeMAND to most significantly contribute to 13b MoA cluster strongly in the ribosome regulation class................................................. 121

Figure 5.6: 13b inhibits protein translation. (A) Total protein levels following 15 hours of treatment with DMSO (negative), dox, taxol, or 13b. Compound 13b significantly decreases total protein levels whereas dox (DNA intercalation) and taxol (microtubule stabilization) do not. (B) Phalloidin staining to visualize acting cytoskeleton following 15 hours of treatment with DMSO (negative), dox, taxol, or 13b. No treatments caused dramatic morphological changes indicative of apoptosis, indicating that 15 hours is early enough to observe MoA and not cell-death-induced changes. Taken together, this strongly supports the predicted MoA that 13b functions by inhibiting protein translation. ........................................ 125

Figure 5.7: HCS approach. (A) Representative images of MDA-231-Br and BT474-Br2 and segmented nuclei. Fluorescent images represent maximum z-projections. (B) PCA on summary statistics of the nuclear segmentation revealed a clear separation between MDA-231-Br and BT474-Br2 cell lines along the first two principle components (representing 80.1% of the data variability). Scale bars represent 100 µm. (C) General approach: drug-induced cell characteristics will be quantified using a range of features from image analysis of segmented cell nuclear areas. Briefly, drugs are screened in dose-response on MDA-231-Br cells, which are then fixed, stained, and imaged. A custom image processing pipeline is employed to segment cells and extract statistical features describing the images. Machine learning is used to reduce data dimensionality to visualize dose-response trajectories in n-dimensional space (first three components shown here). These response modalities can then be correlated based on their corresponding drug’s structural or synthetic properties, enabling us to generate a “map” of generalized cell responses that describe specific drug features....... 129

Figure 5.8: (A) Cell segmentation. Representative images and segmentation for DMSO control cells versus dox, taxol, and 13b treated cells (10 µM). Scale bar represents 50 µm. (B) Differential dose-response trajectories of induced cells’ morphological responses in principle component space. (C) Drug clustering
based on induced dose-response trajectories, as measured by complete linkage of the coefficient of determination\(^1\), which quantifies how well a trajectory for a given compound predicts cell responses induced by another compound. Drug labels are colored by the drug’s EC\(_{50}\), green to red being low to high. ............ 133

Figure A.1: Comparison of Cl.8 cell proliferation in complete serum containing media, ZO media unsupplemented, and “ZO Fortified.” Cells were seeded at 50,000 cells/well and imaged after three days. ZO Fortified supports initial attachment and proliferation of Cl.8 cells whereas ZO unsupplemented does not. ............ 151

Figure A.2: Cl.8 and S2 cells during long term culture in ZO Fortified versus ZB Media; images at the point of discernable difference in cell growth between ZO Fortified and ZB Media. S2 cells passaged into ZO Fortified were unable to reach passage two (confluency) whereas those supplemented with spermidine (ZB Media) were able to reach confluency and undergo two population doublings before growth rates stalled (passage four). Cl.8 cells passaged into ZO Fortified became confluent and underwent one population doubling (passage three) before growth stopped. Cl.8 cells cultured in ZB Media are able to proliferate long-term (currently passage 35). Cl.8 cells after adaptation (ten passages in ZB Media) can be frozen and thawed successfully. ........................................................................................................ 152

Figure A.3: Cl8 growth kinetics are dependent on seeding density. Spermidine causes significant increase in growth for all seeding densities tested. Corresponding p-values across multiple days (data shown in Figure A.4) can be found in the electronic supplementary information. ............................................................. 153

Figure A.4: Cl.8 growth kinetics are dependent on seeding density. Even at 125,000 cell/mL seeding density, ZB Media promotes cell growth (d5/d0 intensity fold change of 1.12 versus 1.01 for ZO Fortified). However, growth at this low of seeding densities is low enough that cells will likely deplete culture medium before reaching confluency. Thus seeding density must be considered when using ZB Medium................................................................. 154

Figure A.5: Supplementing ZB Media with spent media ZB media from adapted or unadapted Cl.8 cells does not induce proliferation of Cl.8 or S2 cells. Adapted (passage 33 in ZB Media) and unadapted Cl.8 cells were seeded in ZB Media at 1x10\(^6\) cell/mL and allowed to proliferate for five days. Spent media was then harvested and Cl.8 and S2 cells were seeded at 50,000 cell/well in 96-well plates with either 0, 25, 50, 75, or 100% spent media from adapted or unadapted cells. Cells were allowed to proliferate for eight days and then assayed with CyQUANT.155

Figure A.6: (A) Hierarchical clustering of known PA pathway genes over developmental time reveals that genes at similar levels of metabolic flow show similar
expression profiles (cluster colors consistent with coloring in Figure 3.1, Figure 3.2). Clusters can be interpreted to represent different roles in the metabolic pathway: pink is upstream biosynthesis, green is downstream biosynthesis, blue is interconversion, and yellow is inhibition. (B) Violin plots depicting the time-course expression of the average mean-centered expression of each of the four gene clusters across developmental stages. Each violin plot shows the frequency distribution of the mean expression level of all genes per cluster. .......................... 156

Figure A.7: Orthogonal z-projection reveals that PAs are located apical to E-cadherin (five days after egg laying, same disc as Figure 3.3 G). Scale bar 5 µm................. 157
TABLES

Table 5.1 Structures of novel compounds................................................................. 109
Table 5.2 Compound EC_{50} values ................................................................. 111
Table A.1 Top 50 negatively enriched terms from GSEA........................................ 158
Table A.2 Top 50 positively enriched terms from GSEA........................................ 162
Table A.3 Top 100 DeMAND genes from time-points combined............................. 165
ACKNOWLEDGMENTS

I would like to thank my advisor, Jeremiah Zartman, without whom this work would not have been possible. I also am also eternally grateful to Erin Howe for her hard work, friendship, and support through the final years of this work. Thanks for being my scientific big sis and bestie. To my other labmates, Cody Narciso, Qinfeng Wu, and Pavel Brodskiy: thank you for your valuable feedback and help in the lab.

I was extremely lucky to be involved in a number of collaborative projects, giving me insight and exposure to a variety of disciplines: thank you Siyuan Zhang and his lab members (in particular Patricia Skallos and Xeujuan Tan), Brandon Ashfeld and his lab members (in particular Kevin Rodriguez and Jennifer Meloche), and Jun Li. I would not be the scientist I am today without having had the opportunity to work with you and learn from you.

I never would have made it this far without the support of my parents, Wayne and Suzanne, who instilled in me the ambition, work ethic, and resilience required to make it through a PhD program. Even more than that, the love and support you’ve shown throughout my studies and life has provided me a constant sense a security that helped get me through my roughest days. I love you both so much and am so thankful to have you as my parents.
Ditto to my sister. April, thank you for being the perfectionist that you are and for always being willing to harness your powers to proofread things for me. If it weren’t for you there would be a lot more typos to follow.

To Thomas. Thank you for being my partner in life and my best friend. Thank you for being patient with me when I came home stressed out and complaining about experiments that wouldn’t work. Thank you for coming with me to the lab at 10PM or a Saturday afternoon when I needed to check on something really quickly, and for not being more irritated when it inevitably took longer than I originally led you to believe. Thank you for giving me a hug, making me a drink, and cheering me up on the bad days. I’m not sure I would have made it without you. I love you so much.

And finally to Merrick. Thank you for being the greatest.
1.1 Overview

Advanced technologies such as RNA-sequencing and high-throughput screening (HTS) have increasingly spawned “big data” approaches that have changed the landscape of biological research and medicine\(^1\text{--}^4\). However, there are still many challenges associated with fully harnessing these resources. In particular, comparatively little has been done to integrate data and disparate tools for systems-level solutions. This dissertation presents an overview of integrating advanced screening techniques (reviewed in sections 1.2 and 1.3) with bioinformatics analysis of gene expression data (section 1.4), while harnessing the power of the data available in protein-protein and protein-drug interaction databases (section 1.5). This chapter details the central background to these topics and introduces their application within the context of our work (section 1.6).

1.2 Cell-based screening

Because human experimentation is usually impractical and expensive (not to mention unethical), the scientific community relies on models to study specific aspects
of biology and physiology. Experimentally, the most commonly used models are in vitro cell-based models and in vivo organism-based models. Cell-based models are attractive for many reasons: there are multiple human cell lines available for studying diverse aspects of physiology, and in vitro culture offers a degree of control, resolution, and reproducibility that is unachievable in an organism-based model. Cell-based models are particularly well suited for screening, where multiple gene or drug effects can be probed in a single experiment by arraying cells in multi-well plates. Cell-based screening is the central foundation for the identification of lead compounds within the pharmaceutical industry; however, it is increasingly being used outside of this industry for discovery based science, most commonly through RNA interference (RNAi) based genetic screening and chemical-genetics-based drug screening$^{5-8}$.

In RNAi cell-based screens, target genes are post-transcriptionally silenced via RNAi and the biological effect of that silencing is assayed. RNAi is a form of post-transcriptional gene silencing that functions through the binding of small (or short) interfering RNAs (siRNAs) to specific messenger RNAs (mRNAs). First described in the roundworm Caenorhabditis elegans (C. elegans), RNAi was found to only require a few molecules of double stranded (dsRNA) per cell to silence expression, not only in those cells but throughout the worm and even in the next generation$^9$. RNAi-induced gene silencing was later found to mimic a well conserved endogenous gene suppression mechanism, and relies on two steps (Figure 1.1)$^{10}$. First, dsRNA is cleaved into siRNA that are 21-25 nucleotide long guide sequences targeting the mRNA to be silenced$^{11-14}$.  

2
Cleavage is dependent on Dicer enzymes that are required for successful RNAi\textsuperscript{13–16}. In the second step, siRNAs are incorporated into the RNA-induced silencing complex (RISC), which recruits the target mRNA and silences expression through translational repression or target cleavage (RNAse activity)\textsuperscript{11,16–18,13}. 
Figure 1.1: RNA interference. (A) RNAi pathway\textsuperscript{19}. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Immunology Ref. 64, copyright (2003). (B) SiRNAs, shRNAs, shmiRNAs, and long dsRNAs pathways\textsuperscript{20}. Reprinted by permission from Cold Spring Harb Perspect Biol 2010;2:a003640, copyright to Cold Spring Harbor Laboratory Press (2010).
Transient RNAi can be induced in human cells through the introduction of synthetic siRNA through transfection, or through direct introduction of dsRNA in *Drosophila* or *C. elegans*; introduction of dsRNA directly to mammals induces an interferon response that results in nonspecific changes in transcription and translation\(^{10,21,22}\). Continuous induction of RNAi in cells can be achieved by transcription of short hairpin RNA (shRNA) can be used to mimic pri-microRNA (pri-miRNA), which is processed by the RNAse III enzyme, Drosha, before being cleaved into siRNA by Dicer\(^{23–25}\). These techniques are frequently used for cell-based RNAi screens, enabling genome-wide loss-of-function screening on cell lines from mice, rats, humans, and *Drosophila*\(^{26–29,20}\). RNAi screening is now used extensively for HTS on human and *Drosophila* cells\(^{30–36}\).

Analogous to RNAi screens, chemical genetics has emerged as a complimentary method that utilizes small molecules with known targets to assay for biological activity\(^{37–40}\). Because screened compounds all have known or putative targets, compound “hits” provide insights into the biological pathways involved in the process of interest\(^{37,39–42}\). One particularly promising form of chemical genetics is inverse drug screening, where a large panel of known bioactive compounds are screened for their effect on a given biological phenotype to implicate compound targets in a phenomena of interest\(^{37,39–42}\). This approach is inverse to screens performed in the pharmaceutical industry that seek to link new drugs with unknown targets to their biological activity. Some advantages of the small molecule approach versus RNAi are speed, reversibility, wide applicability across species, and efficiency (one compound can probe multiple
putative targets\textsuperscript{40}. Further, because the approach is based on the use of established drugs with known targets, targets implicated to impact specific biological processes have \textit{de novo} therapeutic options.

Whether chemical or genetic-based, all screens require an assay readout; along these lines, cell-based screens can be thought of as high-throughput or high-content. While to be truly high-throughput, a screen must test a minimum of thousands of conditions, the vast majority of screens conducted today can be thought of as high-throughput in spirit, yielding a single quantitative, automated assay readout\textsuperscript{5–8}. In contrast, high-content screening (HCS) relies on automated microscopy and image analysis for the quantitation of multiparametric assay readouts describing induced changes at the cellular level\textsuperscript{43–45}. An advantage of this macroscopic approach is the ability of HCS to capture and quantify complex phenotypes that can more closely represent a disease state than a single readout from HTS\textsuperscript{46}. Because of this increased complexity, HCS is generally used for smaller, more focused screens or secondary characterization of induced phenotypes \textit{after} HTS\textsuperscript{5}. However, although HCS is receiving increasingly more attention as an alternative or complimentary method to HTS, the majority (60-80\% by an estimate in 2015) of papers presenting high-content screens only actually used one to two image readouts for screen analysis\textsuperscript{46}. Thus a large amount of the data available during a HCS remains unleveraged. Recently, the term ultra-high-content screen (uHCS) was coined to describe use of more than 500 cell-morphological parameters to describe cell response paradigms to drug perturbations, including
morphometric features such as eccentricity, area, roughness, and circularity\textsuperscript{47}. Extraction of this many features requires calculation of multiple cell and often neighborhood level statistics, requiring robust segmentation algorithms and classification of distinct subpopulations of cells\textsuperscript{47–49}. Classification of cells into subpopulations can be a powerful tool in itself for the assessment of heterogeneity in induced cellular responses\textsuperscript{47–49}. While it might seem that a higher number of features is always better, correlated or non-informative features can introduce noise. Therefore, machine-learning is often used in uHCS screens to reduce the dimensionality to only the most predictive features. A recent paper found that only 10-15\% of the original 300+ features quantitated were actually necessary to characterize drug mechanism of action\textsuperscript{50}. Selection of features to use in downstream statistical hit selection is generally done using machine learning algorithms, and often relies on a reduction of data dimensionality. Principle component analysis (PCA) is very frequently to reduce data dimensionality. PCA uses coordinate transformation to convert a set of observations to principle component space, in which each principle component (PC) axis is a linear combination of the original variables. In PC space, the first PC axis (PC1) accounts for the highest amount of variability in the data, followed by the second highest (PC2) and so on.

There are multiple statistical approaches for quantifying the effect of treatments and assessing quality in a HTS or HCS. Two of the most common metrics are Z-factor and strictly standardized mean difference (SSMD) score\textsuperscript{51}. In the context of determining the
quality of a screen, both of these metrics measure the difference between effects of the positive and negative controls on the assay readout, and have guidelines to classify quality (excellent, good, inferior, poor) based on the score. While Z-factor is historically most frequently used, SSMD is being utilized more due to its ability to classify effect size. This enables SSMD to be used for screens using strong or moderate positive control, as opposed to the very strong control that is required to achieve good quality in a screen based on Z-factor. In addition to screen quality, Z-factor and SSMD are used to measure effects of screened conditions and for subsequent hit selection based on those scores\textsuperscript{51}.

1.3 Screening in model organisms

While attractive on numerous levels, cell-based models can often be too simplistic, and \textit{in vitro} results often are not validated \textit{in vivo}. Thus model organisms can provide complimentary data with increased physiological relevance. Model organisms have been used to gain much insight into fundamental biological processes such as development and disease\textsuperscript{52–57}. There are a wide variety of model organisms available, each with their own degree of evolutionary conservation with humans. One way of estimating this conservation is through molecular clocks, which estimate when two organisms diverged\textsuperscript{58}. The bacterial model \textit{E. coli} is more than 2.5 billion years removed from humans, followed by \textit{Saccharomyces cerevisae} (baker’s yeast) and \textit{C. elegans} (roundworm), between one and two billion years; and \textit{Drosophila melanogaster} (fruit-
fly), *Danio rerio* (zebrafish), and *Mus musculus* (mouse), one billion years or less. Aside from evolutionary conservation, other considerations in choosing a model organism are the experimental tools available; ease, speed, and cost of care; the extent of genetic redundancy; and relevancy to the biological question of choice. Each model organism has its own particular strengths and weaknesses; for example, baker’s yeast has been widely used to study the cell cycle due to its high protein conservation and rapid growth, but lacks applicability for developmental studies. Thus there is a tradeoff between simplicity, cost, and *in vivo* relevancy such that each biological question has an ideal model system.

Model organisms are also being increasingly used in the context of screening, due to increased physiological relevancy. For genetic loss of function screens, RNAi can be conducted *in vivo*, although for most model organisms the methods are less well-developed. While *in vivo* RNAi is achievable in murine and other models, it is low throughput and expensive. Currently, genome-wide RNAi screening is possible *in vivo* in *C. elegans* and *Drosophila* only. However, as discussed in Cell-based screens, RNAi in *C. elegans* is non-cell autonomous, meaning that RNAi is systemic and cannot be spatially confined. In contrast, RNAi in *Drosophila* is cell-autonomous when produced from transgene expression, enabling spatial control over RNAi-mediated gene silencing. In *Drosophila*, RNAi can be continuously induced though transcription of long hairpin mRNAs that mimic dsRNA, transcription of an inverted repeat or complementary strands through opposing promoters, or viral vectors expressing dsRNAs
against a specific gene\textsuperscript{13,14,63–65}. Spatial control of RNAi is achievable through the advanced genetic tool-kit available for use in \textit{Drosophila melanogaster}\textsuperscript{66,67}.
Figure 1.2: *Drosophila melanogaster* as a model organism. (A) Attractive features of *Drosophila melanogaster* as a model organism. (B-D) Genetic tools to control transgene expression in *Drosophila*. (B) The yeast transcription activator protein GAL4 is expressed under the control of an endogenous promoter. GAL4 is expressed and binds to the GAL4 specific upstream activating site (UAS) ahead of a target transgene, activating its transcription specifically in the tissue in which GAL4 is expressed. More recently developed alternative binary expression systems LexA/opA and Q operate analogously to GAL4/UAS. (C) GAL4- and Q-mediated expression can be further tuned through the repressors GAL80 or QS, respectively. (D) Further control can be exerted through use of site-specific recombinases in a technique called Flp-out, where site specific recombinases remove a stop sequence that prevent transgene expression. Thus expression is limited to where GAL4 and Flp are present but GAL80 is not (green area in D'). Panels B-D Reprinted by permission from Macmillan Publishers Ltd: Nature Methods Ref. 64, copyright (2016).
In particular, the well-developed GAL4/UAS binary expression system is the hallmark genetic tool in *Drosophila*; it enables spatiotemporally controlled manipulations of gene expression (e.g. fluorescent labeling, constitutively active transgenes for overexpression studies, RNAi for gene knockdown studies)\(^{66,71}\). A fly can be engineered to express the yeast transcription activator protein, GAL4, under the control of an endogenous promoter (generally a promoter for a gene expressed in specific tissues in the organism) as well as the GAL4-specific upstream activating site (UAS) ahead of a target transgene; in this case, GAL4 is expressed and binds to the UAS site, activating transcription of the target transgene specifically in the tissue in which GAL4 is expressed (Figure 1.2)\(^71\). Thousands of GAL4 and UAS-RNAi lines have been generated and are publically available\(^{72,73}\). Transgene expression through GAL4 can be controlled further through GAL80 or site-specific recombinases in a technique called Flp-out (Figure 1.2)\(^{68-70}\). GAL80 is a GAL4 repressor protein which also exists as a temperature sensitive mutant that is active and represses GAL4 at 18\(^\circ\)C but not at 29\(^\circ\)C\(^{68}\). Flp-out relies on the yeast recombinase flipase (Flp) and its target sequence (*FRT*), and a transcriptional “stop signal” surrounded by *FRT* sites such that expression of Flp removes the stop signal, resulting in transcription of the gene of interest\(^{69,70}\). The LexA-*lexAop* and QF-*QAS* systems are analogous orthogonal binary expression systems that enable sophisticated intersectional strategies for the generation of genetic mosaics\(^{74-76}\).
Along with the genetic tools available to *Drosophila*, there are a large number of reagents available to the community, including UAS-RNAi stocks representing a large number of genes of interest, and a concerted effort is currently underway to generate stocks for all 14,000+ annotated protein-coding genes in the *Drosophila* genome\(^72,73\).

In the context of drug screens, *Drosophila* has a digestive system and many analogous tissues to humans, including a brain protected by a blood brain barrier, making it an ideal model for preclinical drug testing\(^57,77–79\). Because of these similarities, *Drosophila* is an ideal system for modeling a wide variety of diseases. More than 80% of human disease-related genes having a *Drosophila* homolog, and a large number of screens have been performed in *Drosophila*, most notably from the Perrimon lab (genetic screens) and the Cagan lab (drug screens)\(^20,57,80\).

1.4 Gene expression profiling

While both cell-based and *in vivo* screening have become central tools for studying biology, the availability of high-dimensional -omics and interaction data has provided an opportunity to further advance these technologies. Thanks to new and rapidly evolving technologies, genomics, transcriptomics, proteomics, and metabolomics can be used to characterize the cell state with an unprecedented level of detail.

Gene expression profiling in particular has become a major tool in biology and medicine\(^81,82\). A vast amount of transcriptomics data is publicly available describing
expression of a variety of disease states, treatment, and developmental stages through various repositories, including the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO), the Cancer Gene Expression Database (CGED), the Gene Expression Atlas, and the cBioportal for Cancer Genomics, among others\textsuperscript{83–86}. The most commonly available data are complimentary DNA (cDNA) microarrays and massively parallel RNA-sequencing (RNA-seq), which is increasingly replacing cDNA microarrays for gene expression studies due to its higher specificity, lower background signal, and absolute rather than relative expression levels\textsuperscript{87–89}.

Gene-expression microarrays consist of arrays of microscopic DNA probes that are used to hybridize experimental samples’ cDNA, which is usually fluorescently labeled\textsuperscript{87–89}. The microarray is then scanned and each probe yields a fluorescent intensity measurement based on the amount of cDNA present in the sample for each probe. In contrast, RNA-sequencing allows for both discovery and quantification since sequences are directly read, thus not requiring a sequence probe, enabling later mapping for transcript identification\textsuperscript{90–92}. Briefly, sample RNA is isolated and enriched against ribosomal RNA (rRNA) before being converted into a library of cDNA fragments tagged with adaptors on one or both ends. These fragments are then sequenced on a high-throughput sequencing platform to obtain reads, generally around 30-300 base pairs (bp) long, from one or both ends (single-end sequencing or pair end-sequencing). Sequencing reads are subsequently aligned to the genome or transcript reference\textsuperscript{91–93}. There are many open-source analysis pipelines for pre-processing and aligning reads
prior to assembling transcripts, determining read counts, and conducting expression analysis\textsuperscript{92,94–96}. The accuracy of RNA-seq has been reported to be greater than 80% for splicing events and differentially expressed exons compared to reverse transcription polymerase chain reaction (RT-PCR) or qPCR\textsuperscript{94,97}.

1.5 Protein interaction databases

The Search Tool for Recurring Instances of Neighboring Genes (STRING) is a database cataloguing known and predicted protein-protein interactions, including both physical and functional associations coming from computational prediction, extrapolation between organisms, and aggregation from other protein-protein interaction databases. STRING catalogues almost ten million proteins interactions in over 2,000 organisms\textsuperscript{98–101}. Comparably, the Search Tool for Interactions of Chemicals (STITCH) curates known and predicted chemical-protein interactions\textsuperscript{102–104}. Similar to STRING, STITCH integrates data from multiple databases, cataloguing the interactions between more than 300,000 compounds and 2.6 million proteins in 1,133 different organisms\textsuperscript{104}. Because STITCH is rooted in text-mining, interaction scores do not provide any information about whether a given compound acts as an inhibitor or activator of its target protein.
1.6 Integrated approaches

Each chapter to follow will detail our approach to integrate aspects of these technologies to address needs in the biotechnology (Chapters 2-3) and healthcare (Chapters 4-5) sectors.

In Chapter 2, we describe our integrated approach for the development of a chemically defined medium (CDM) for *Drosophila* cells. Traditionally, the unsurpassed genetic and molecular tools available for *in vivo* studies in *Drosophila* has relegated the development of *in vitro* tools to a secondary role. However, for increased throughput and ease, *Drosophila* cell and organ culture is becoming more widespread, especially in RNAi and compound screens. Unfortunately, the historical lack of emphasis on *in vitro* *Drosophila* cell culture has resulted in a stark underdevelopment of tools compared to mammalian cell culture, limiting its utility as a model system. In particular, no chemically defined medium (CDM) was available for the culture of *Drosophila* cell lines; all media either included undefined extracts (e.g., yeast extract) or required supplementation with undefined and highly variable serum such as fetal bovine serum (FBS) or fly extract (FEX). These undefined supplements limit the control and reproducibility of cell culture experiments, and due to their complex nature hinder proteomic analyses. To characterize the minimal requirements for long-term maintenance of *Drosophila* cell lines, we developed an inverse screening strategy to identify small molecules and synergies stimulating proliferation in a CDM. In this chemical-genetics approach, a compound-protein interaction database is used to
systematically score genetic targets on a screen-wide scale to extract further information about cell growth. In the pilot screen, we focused on two well-characterized cell lines. Validated factors were investigated for their ability to maintain cell growth over multiple passages in the CDM. The polyamine spermidine proved to be the critical component that enables the CDM to support the long-term maintenance of multiple cell lines. Spermidine supplementation was found to upregulate DNA synthesis and increases mitogen activated protein kinase (MAPK) signaling. The CDM also supports the long-term growth of Kc167 cells. Our target scoring approach validated the importance of polyamines, with enrichment for multiple polyamine ontologies found for both cell lines.

Building on findings generated during our design of the CDM, we present a side story on the role of polyamines in developmental biology\textsuperscript{112} (Chapter 3:). Polyamines (PAs) are ubiquitous organic molecules that influence many cellular processes, including gene regulation, signal transduction, cell growth, and cell proliferation. Previous research has demonstrated that the metabolic regulation of PAs plays important roles in multiple diseases, including cancer and Alzheimer’s. The current paradigm is that polyamines and other metabolites primarily play roles in cell homeostasis and when dysregulated contribute to disease. This idea is at odds, however, with studies in multiple model organisms demonstrating dynamic levels of polyamines throughout development and their requirement in embryonic development. We integrate previously published whole genome expression and PA metabolic data across
Drosophila organismal development to show developmental-stage-specific regulation of the PA metabolic pathway. Combining this transcriptomic analysis with immunohistochemistry, we discovered novel, spatiotemporal accumulation of PA metabolites in both the Drosophila embryonic trachea and differentiating eye imaginal discs. These results suggest that spatiotemporal patterning of polyamines is required during organogenesis.

Chapters 4-5 shift focus toward innovative applications of screening methods to advance preclinical research for breast cancer brain metastasis through the identification of genetic mediators of metastatic outgrowth (Chapter 4) and characterization of potential blood-brain permeable and cytotoxic drugs (Chapter 5). Current treatment options for patients with brain metastases include chemotherapy, radiotherapy, and HER2 targeted therapies where appropriate. Although these treatments initially improve overall survival (chemotherapy vs. no chemotherapy, 16.4 vs. 3.7 months; surgery vs. no surgery, 20.3 vs. 11.3 months)\(^\d\), there are no effective treatment strategies for patients with refractory disease. A primary cause for the lack of treatment options is our limited understanding of the genetic basis of brain metastasis. During metastasis, cells undergo a dramatic evolution that includes active adaptation to the distinct metastatic microenvironment. In Chapter 4, we use an integrative method building off of RNA-seq data describing single-cell, dormant brain metastases versus large overt metastases. Candidate genes from RNA-seq data were functionally validated in a high throughput genetic screen that functionally tested their effect on
tumorigenesis and metastasis in vivo using a genetic model of cancer in Drosophila. This approach enabled the identification of several genes whose RNAi-mediated knockdown resulted in significant decreases in primary tumor formation or metastasis. This integrative screening approach allows us to identify genes that have a functional role in primary tumor formation and metastatic development or progression.

Another cause for the lack of treatment options is the drug-impermeable nature of the blood brain barrier (BBB), which hinders the development of novel adjuvant therapies for brain metastases. Based on physiochemical properties indicative of BBB permeability, our collaborators have developed a method for synthesizing diverse oxindole members of the Strychnos family of indole alkaloid, which have been shown to inhibit mitosis, reduce proliferation of glioma and neuroblastoma cell lines, and cross the barrier created by brain endothelial cells in an in vitro model of the BBB model. In Chapter 6, I describe our integrated method combining screening with RNA-seq and protein-protein network interaction data for the rapid evaluation and mechanism of action prediction for these novel spirooxindole drugs. Using two brain trophic breast cancer cell lines, we are able to assay for the cytotoxic efficacy of novel compounds in breast cancer brain metastasis. We rapidly identified the mechanism of action of the most promising drug to be ribosome biogenesis by analyzing induced network dysregulation from RNA-seq data. We have further begun implementation of a high-content screening strategy for structure-activity analysis of compounds to ultimately
facilitate prediction of the structural and synthesis compound properties conducive for compound efficacy.
2.1 Background

Chemical genetics has recently emerged as a complimentary method to traditional genetics where the central theme is the use of small molecules for studying biological systems\textsuperscript{37–40}. One especially promising form of chemical genetics is inverse drug screening, where known bioactive compounds are screened for phenotypes\textsuperscript{37,39–42}. This inverse approach is analogous to RNA interference (RNAi) screens in that the compounds all have known or putative targets, and thus compound “hits” provide insights into the biological pathways involved in the process of interest\textsuperscript{37,39–42}. Some advantages of the small molecule approach are speed, reversibility, wide applicability across species, and efficiency (one compound can probe multiple putative targets)\textsuperscript{40}. These screens have been harnessed to investigate multiple aspects of biology including mitosis, pigmentation, development, insulin signaling, and wound healing\textsuperscript{119–123}. In

\* Reproduced from Molecular Biosystems, Ref. 105 with permission from the Royal Society of Chemistry; copyright 2014.
particular, *Drosophila melanogaster* has been used for whole organism scale small molecule screens to study various biological processes\(^7\),\(^1\)\(^2\).4.

*Drosophila* is a versatile model system used to understand the development and physiology of multiple tissue types\(^5\),\(^7\),\(^1\)\(^2\),\(^5\)\(^7\),\(^1\)\(^0\)\(^7\),\(^1\)\(^2\)\(^5\),\(^1\)\(^2\)\(^6\),\(^1\)\(^2\)\(^8\). Traditionally, the unsurpassed genetic and molecular tools available for *in vivo* studies has relegated the development of *in vitro* tools to a secondary role\(^1\)\(^0\)\(^6\). However, for increased throughput and ease, *Drosophila* cell and organ culture is becoming more widespread, especially in genetic and chemical screens (Figure 2.1 A-B)\(^1\)\(^0\)\(^7\),\(^1\)\(^0\)\(^6\),\(^1\)\(^0\)\(^8\)–\(^1\)\(^1\)\(^0\). Still, *Drosophila* cell culture tools are relatively undeveloped, limiting the utility of *Drosophila* cell culture as a model system. In particular, there is no chemically defined medium (CDM) available for the culture of *Drosophila* cell lines; they all either include undefined extracts (e.g., yeast extract) or require supplementation with undefined and highly variable serum such as fetal bovine serum (FBS) or fly extract (FEX)\(^1\)\(^1\)\(^1\). These undefined supplements limit the control and reproducibility of cell culture experiments, and due to their complex nature hinder proteomic analyses. While several companies offer serum-free media for the growth of insect cells, these formulations are proprietary.

The most recent attempts to rationally design chemically defined media for *Drosophila* cell culture were made over 30 years ago, before the development and spread of high-throughput screening techniques\(^5\),\(^1\)\(^1\)\(^1\). Wyss’ systematic attempt was based on the requirements of two embryonic cell lines, Kc and C\(\)\(^a\)\(^1\)\(^1\)\(^1\),\(^1\)\(^2\)\(^7\)–\(^1\)\(^2\)\(^9\). The resulting formulation, ZO media, could support Kc cell growth upon inclusion of fly extract, and
was also used to create an epithelial-like cell line from *Chironomus tentans* upon supplementation with FBS, yeast extract, and insulin\textsuperscript{111,130}.
Figure 2.1: (A) Advantages and limitations of *in vivo* and *in vitro* experimentation. The precise control offered by *in vitro* culture is abrogated by the required medium supplementation with undefined extracts. (B) Comparison of mammalian versus *Drosophila* cell culture. (C) Comparison of Cl.8 growth kinetics in complete serum containing media, ZO media unsupplemented, and “ZO Fortified.” ZO Fortified supports initial attachment and proliferation of Cl.8 cells whereas ZO unsupplemented does not (Figure A.1). Single and double asterisks denote $p < 0.05$ and $p < 0.01$, respectively, for two-tailed t-test for unmatched pairs. (D) Screening pipeline. Small molecule libraries are screened on two cell lines for their ability to promote proliferation in a minimal, serum-free medium. Putative hit compounds are then compared between cell lines to minimize false positives, and analysis on putative targets is conducted. The screen can be iterated upon to identify compound synergies by incorporating hit compounds into the background media and rescreening. (E) Target scoring pipeline. Compound-protein interaction scores ($s$) are identified from a database and linked to compound z-scores. Non-*Drosophila* protein targets are “translated” to their *Drosophila* orthologs retaining their percent identity match score, $q_i$. *Drosophila* protein target p-values are then calculated by summing for each interaction with a compound its compound-protein interaction score, percent identity match scores, and squared compound z-scores (which follows a chi-squared distribution). This analysis yields a list of significantly targeted proteins for a process of interest. Protein target lists can be also converted to their encoding genes and gene ontology enrichment can be performed. More details can be found in the materials and methods.
A. 

In Vivo vs. In Vitro

- Increased physiological relevance
- Lower throughput
- Low resolution
- In vivo environment: Hemolymph (>700 Proteins)

In Vivo

- High degree of control
- High throughput
- High resolution
- Artificial environment:
  - Culture Media: (UNDEFINED extracts)
  - Fetal bovine serum, fly extract, yeast extract

B. 

Mammalian vs. Drosophila

- Increased human relevance
- Require CO₂ incubator
- Harbor human pathogens

Mammalian

- Low genetic redundancy
- Easy to maintain
  - Does not require CO₂ incubator
  - Room temperature incubation
- Does not harbor human pathogens

C. 

Clone 8

<table>
<thead>
<tr>
<th>Intensity Fold Change</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>dH/d0</td>
<td>NS</td>
<td>**</td>
<td>*</td>
<td>**</td>
<td>**</td>
<td>*</td>
</tr>
</tbody>
</table>

D. 

Compound Libraries

Proliferation Assay Image Acquisition

Hit Compounds

Top Shared Hit

Target Analysis

E. 

Compound Interaction Score $s_i$

Drosophila ortholog targets ($q_i$)

Target p-values

$\sum_{i=1}^{n} s_i \cdot q_i \cdot z_i$

Significant biological targets, ontologies

27
Applied to culture media design, small molecule screens have previously been used to identify media supplements enabling self-renewal of embryonic stem cells and to find inducers of B-cell expansion\textsuperscript{131,132}. Similarly, a compound cocktail was optimized to support long-term growth of human embryonic stem cells using five predefined candidate molecules associated with known pathways\textsuperscript{133}. However, to our knowledge no systematic attempt has been made to both identify growth promoters and optimize \textit{Drosophila} culture media in a high-throughput fashion. We hypothesized that by performing a pilot inverse small molecule screen on \textit{Drosophila} cells we could identify 1) novel compounds and compound synergies stimulating cell proliferation, 2) genetic targets and biological pathways important for growth, and 3) a combination of compounds sufficient for long-term growth and maintenance of \textit{Drosophila} cell lines in a chemically defined medium.

Here, we demonstrate a high-throughput inverse drug-screening platform to identify novel compounds and genetic targets important for proliferation of \textit{Drosophila} cells. By systematically identifying and scoring protein targets of the screened compounds, we can identify genes and pathways in addition to compounds important for growth. We have developed an approach that harnesses a chemical-protein interaction database to “translate” cumulative small molecule scores to gene target scores to elucidate targets with small effects (e.g. small effects from multiple compounds with same target) (Figure 2.1 E). This approach is an improvement over traditional methods where only “hit” compounds’ targets are investigated, and can be
applied to screens in other model organisms for which databases are available. The pipeline can also be expanded to identify compound synergies that can be exploited to design CDM capable of supporting long-term growth of multiple Drosophila cell lines (Figure 2.1 D). The protocol can be used as a template for the rational design of media, to identify growth-promoting factors, and implicate signaling pathways important for growth.

In this proof-of-principle screen, we focus on two standard Drosophila cell lines, the adherent Clone 8 (Cl.8), which has previously been used to identify novel insect-specific growth factors, and S2-DRSC (S2), which is frequently used for recombinant protein production and grows in suspension. The screen led to the identification of multiple candidate molecules relevant for stimulating growth and viability of both cell lines. In particular, the pilot screen revealed polyamines as the critical missing component of a CDM for Drosophila cells, and sufficient for enabling long-term growth of Cl.8 and Kc167 cells without requiring any weaning of the cells from sera. To our knowledge this is the first successful attempt to harness a small molecule screen to systematically define the minimal requirements for long-term Drosophila cell growth in a chemically defined environment.
2.2 Materials and Methods

2.2.1 Cell culture

Cl.8, S2, and Kc167 (Kc) cells were expanded in optimized serum-containing media, Cl.8, S2, and Kc167 media, respectively. Cl.8 media contains M3 media (Sigma-Aldrich®) supplemented with FBS (2%), insulin (5 μg/m), and fly extract (2.5%). S2 media is based on Schneider’s media (Gibco® Life Technologies) supplemented with FBS (10%). Kc media containing M3 media supplemented with yeast extract (10 mg/mL), bactopeptone (25 mg/mL), and 5% FBS. All cell lines were obtained and cultured according to instructions from the Drosophila Genomics Resource Center (DGRC). Fly extract was prepared from adult yw flies as described by the DGRC. For culture in chemically defined media, cells were rinsed three times in PBS to remove residual serum and seeded at around 70% confluency in ZO Fortified or ZB Media (ZO Fortified with 1 μM spermidine added at time of passage). Upon reaching confluency, cells were passaged 1:2 retaining half the spent media, similar to routine maintenance in complete media, with cells passaged in ZB Media receiving fresh doses of 1 μM spermidine.

2.2.2 Basal media

While Wyss’ ZO medium was never widely adopted, we selected it as the starting point for designing a completely chemically defined basal medium as it represents the most recent and thorough systematic effort to create a chemically defined medium for Drosophila. ZO media was initially acquired commercially (Sweden National
Veterinary Institute) but is currently prepared in our lab in small batches as described by Wyss with similar results\textsuperscript{128,130}. Preliminary efforts focused on testing compound candidates from the literature for their effect on proliferation (data not shown). Various proliferation assays were tested, with CyQUANT\textsuperscript{®} Direct Proliferation Assay (Life Technologies) yielding the best calibration between fluorescent intensity and cell number. CyQUANT is a DNA content-based assay that uses a background suppressing dye that is selectively permeable to dead cells (lacking membrane integrity), enabling specific labeling of live cells\textsuperscript{107}. Conveniently, this assay yields good calibrations between cell number and fluorescent intensity ($r^2 > 0.95$) even when used at 0.25x the suggested working concentration (data not shown).

Using this approach we developed an intermediate “ZO Fortified” medium, made up of ZO media supplemented with insulin (5 μg/mL), trehalose (26.4 mM), L-alanyl-L-glutamine (ala-gln, 12 μM), and L-ascorbic acid 2-phosphate sesquimagnesium salt hydrate (A2P, 0.08 μM), with pH adjusted to 6.75. Insulin is a growth stimulator through insulin receptor signaling that is frequently used as a media supplement; trehalose is a disaccharide present in insect hemolymph that has roles in protection from environmental stresses such as temperature and oxidation; ala-gln is a stabilized dipeptide version of the essential amino acid L-glutamine; A2P is a stable form of L-ascorbic acid which is essential for many insects\textsuperscript{128,130,138–141}. Whereas ZO unsupplemented does not support proliferation of Cl.8 cells, ZO Fortified supports short-
term cell growth and initial attachment (Figure 2.1 C, Figure A.1). ZO Fortified is capable of supporting slow growth of Cl.8 cells through approximately 3 passages.

2.2.3 Pilot screen

To more efficiently identify compounds important for cell growth in serum-free media, we selected five small molecule libraries for pilot screening: the Wnt Pathway Library (75 compounds), Autophagy Library (97 compounds), Kinase Inhibitor Library (80 compounds), Phosphatase Inhibitor Library (33 compounds), and Ion Channel Ligand Library (72 compounds) (Enzo® Life Sciences). The Wnt pathway is important for regulating cell proliferation and differentiation\textsuperscript{142}. The Autophagy Library encompasses the target of rapamycin (TOR) pathway, involved in the regulation of growth and apoptosis with respect to nutrition, and is important in cancer\textsuperscript{138,143–145}. The Phosphatase and Kinase Libraries target tumor suppressors and oncogenes involved in TOR signaling, as well as other kinases and phosphatases involved in cell growth, proliferation, and survival\textsuperscript{138,143}. Ion channels have well-known roles in cell proliferation and cancer\textsuperscript{146,147}. These libraries come in a convenient 96-well format in dimethyl sulfoxide (DMSO) at a concentration of 10 μM.

Because most compound targets are identified using mammalian models, their efficacy and specificity to \textit{Drosophila} are largely unknown. Further, many of the compounds tested are likely growth inhibitors rather than stimulators. To supplement the commercial libraries, we thus created a custom library consisting of supplements we
hypothesized to be important for cell growth (39 unique compounds/supplements, 53 total, electronic supplementary information). Preliminary investigations yielded working concentrations for some of these compounds; therefore, custom library stocks were prepared either at 10 μM (for un-tested compounds) or 10,000x their working concentration (based on preliminary tests). The vast majority of these compounds stocks were prepared in DMSO, but some were prepared in water or ethanol based on solubility (electronic supplementary information). Based on the screening concentrations selected, the final DMSO dilution is 1:1000 or greater, and thus not expected to impact results. These six libraries served as the basis for the pilot screen. The combined contents of the tested libraries along with raw scores from the screening experiments are listed in the electronic supplementary information.

Five stock plates of randomized compounds in singlet (410 total) were prepared at 75 μM in water (1:133 dilution) using an automated liquid handler (Eppendorf epMotion® 5075). Ten strategically located negative controls were incorporated into each stock plate by dosing wells with DMSO (1:133, carrier only control). Four randomly placed positive control wells were also included on each plate by dosing wells with DMSO (1:133) for consistency, but for flexibility positive control treatments were prepared in the working stock. Working stocks at 2 and 20 μM (2x) were then prepared in ZO Fortified by diluting the 5 stock plates 1:3.75 (high concentration working stock) and then serially diluting 1:10 (low concentration working stock). Positive control wells were prepared by replacing 10% of the media with 100% fly extract (FEX)107,111,128.
Cells were rinsed three times in phosphate buffered saline (PBS) and seeded at 50,000 cells/well (50 μL) in ZO Fortified. Working stocks, 50 μL, were then immediately added on top of the cells (day 0). Plates were incubated at 25° C for four days, and then imaged at 10x magnification on a Nikon microscope using automated screen acquisition in MetaMorph® software. Plates were then treated with CyQUANT® (0.25x final concentration) for one hour at 25°C, and fluorescent intensity was measured with a plate reader (Biotek Synergy H2).

2.2.4 Screen analysis and hit selection

Strictly standardized mean difference (SSMD) scores for classifying assay quality were calculated with respect to negative controls on a plate-wise basis, with the average assay quality being excellent for both cell lines and screened concentrations51. We investigated the effect of performing positional corrections (median polishing, etc.) to account for plate-location effects and found that screen quality, as assessed by the SSMD, were overall best when no positional corrections were made51,148. As a hit selection metric we used z-scores, which measure the number of standard deviations from the negative control that a small molecule treatment causes. Z-scores were calculated on a plate-wise basis, and hit cutoffs of z = ±3 were selected51. Visual inspection of images was used to remove false-positive hits, which generally occurred due to autofluorescence of a few of the compounds.
2.2.5 Growth kinetics experiments

Cells were rinsed three times in PBS and seeded in 96-well plates at a concentration of 50,000 cell/well (by plating 5 μL of a 1x10^7 cell/mL suspension in PBS on top of 95 μL of media). Three replicates per media were assayed with CyQUANT (0.25x final concentration) daily, starting immediately after seeding (day 0). After addition of CyQUANT, cells were incubated for one hour at 25°C, and fluorescent intensity was measured with a plate reader (Biotek Synergy H2) at a gain of 70 to facilitate day-by-day comparisons in intensity. A fold change in intensity of less than one indicates proliferation of cells.

2.2.6 Polyamine-depleted spermidine dose-response measurements

Cl.8 cells were rinsed three times in PBS and seeded in ZO Fortified at 1x10^6 cell/mL. After nine days of culture in ZO Fortified, cells were harvested and seeded in tissue culture flasks in ZO Fortified supplemented with 0, 0.1, 1, or 10 μM spermidine (in triplicate). After six days of culture, cells were manually counted.

2.2.7 Cell immunostaining and western blots

For EdU incorporation, Cl.8 and S2 cells were rinsed three times in PBS and seeded in optical grade 96-well plates at a concentration of 50,000 cell/well (by plating 5 μL of a 1x10^7 cell/mL suspension in PBS on top of 95 μL of media). After four days of culture, cells were assayed with Invitrogen’s Click-iT® EdU Alexa Fluor® 647 Imaging Kit according to their instructions, with a three hour EdU incorporation time. Cells were
then stained with DAPI (1:1000) and two positions per well were imaged at 40x magnification on an EVOS® fluorescent microscope (AMG). CellProfiler, a customizable image analysis package was used to quantify total number of cells (from DAPI images) and number of EdU positive cells (from channel 647 images) within each image$^{149,150}$. A minimum of 375 cells was analyzed for each condition tested.

For western blots, Cl.8 and S2 cells were rinsed three times in PBS and seeded in tissue culture flasks at a concentration of 1,000,000 cell/mL in the various media. Cells were cultured until positive control samples (complete media) reached ~90% confluency, at which point all cell samples were lysed for 30 minutes on ice in a buffer containing 50 mM Tris, 150 mM NaCl, 1% Nonidet P 40 substitute (Sigma 74385), and 1% Sigma protease inhibitor cocktail (P8340). Nucleic acid and cell debris were cleared by centrifugation at 12,000 rpm for 20 minutes at 4°C. Because polyamines are known to play roles in initiation of transcription and translation, “housekeeping” genes routinely used as loading controls for western blots are likely inappropriate for this application$^{151-154}$. Thus samples were normalized by total loaded protein. Sample protein concentrations were determined three independent times per sample using Coomassie Plus (Bradford) Assay (Thermo Scientific 23238). ANOVA comparison of loading concentrations across all samples found no significant variation between samples ($p > 0.34$). Proteins were resolved by SDS-PAGE in reducing conditions, electroblotted to PVDF membranes (Hybone 10600087), and probed with either 1:1000 anti-rabbit p44/p42 MAPK (Erk1/2) (Cell Signaling) or 1:1500 anti-mouse MAPK
(dpERK1/2) (Sigma). Blots were developed using WesternBreeze chromogenic western blot immunodetection kit (Invitrogen) and quantification was done in Fiji\textsuperscript{155}.

2.2.8 Compound target analysis

The Search Tool for Interactions of Chemicals (STITCH) database was used to identify protein targets of the tested compounds\textsuperscript{102–104}. STITCH integrates data from multiple databases and catalogs the interactions between more than 300,000 compounds and 2.6 million proteins in 1,133 different species\textsuperscript{104}. The STITCH download file of chemical-protein links was queried for all of the screened compounds, and interactions and combined interaction scores for human (\textit{Homo sapiens}), mouse (\textit{Mus musculus}), and \textit{Drosophila melanogaster} were compiled, resulting in a list of 23,158 uniquely targeted proteins. To assign one score per interaction and penalize inconsistent results, \textit{z}-scores at high and low concentration were averaged. Drugs yielding no STITCH results were removed from the analysis (78 compounds total). \textit{Drosophila} orthologs and \% identity match scores for the human and mouse proteins were then found using Ensembl\textsuperscript{156}. Proteins yielding no orthologs were removed from analysis (9,491 unique proteins).

This analysis resulted in a list of 6,090 proteins (~22\% of the \textit{Drosophila} protein-coding genome) that were potentially targeted by the small molecule screen\textsuperscript{156,157}. In order to “translate” the compound scores to a scored protein list, we began by linking the absolute value of each average \textit{z}-score to each compound-protein interaction. The
average of the two concentration z-scores was used to penalize compounds that yielded inconsistent results. Because the database (STITCH) is rooted in text-mining, interaction scores do not provide any information about whether a given compound acts as an inhibitor or activator of its target protein. Thus absolute values were used. To generate protein lists of targets important for stimulating versus inhibiting proliferation, compounds were separated into potential proliferation agonists or antagonists (average z-score positive or negative, respectively), and targets were scored for these two sets of compounds separately.

Specifically, for each \( n \) compound related to \( d \) Drosophila proteins, each \( i^{th} \) relation has an interaction score \( (s_i) \) and % identity match score \( (q_i) \) between 0 and 1, and z-score \( z_i \). Approximately then, the protein score, \( \sum (s_i q_i z_i^2) \), follows a chi-squared distribution with degrees of freedom \( n \), and p-values can be calculated accordingly. These protein p-values are thus based on: 1) the strength with which their interactor compounds affected proliferation (absolute z-score), 2) the confidence level of the compound-protein interaction \( (s_i) \), 3) the number of times a protein was targeted by unique compounds (summing scores per targeted protein), 4) the degree of conservation between originally targeted mammalian proteins and their Drosophila orthologs \( (q_i) \).

To generate significantly targeted protein lists and account for multiple hypothesis testing, we used the standard Bonferroni correction and implemented a p-value cutoff of \( (0.05/d) \) for each set of scores. Drosophila protein IDs were converted to
gene IDs using FlyBase\textsuperscript{157}. We then used the Database for Annotation, Visualization, and Integrated Discover (DAVID) to identify enriched Gene Ontology and KEGG annotations\textsuperscript{158,159}. Importantly, because our library does not target the entire proteome, this resource enables definition of the background facilitating accurate determination of pathway enrichment. Finally, REVIGO was used to reduce redundancy in ontology annotations by grouping them into terms, which are visualized in semantic similarity plots\textsuperscript{160}.

2.3 Results

2.3.1 Small molecule screen in ZO Fortified

From the screen, 12 and 20 unique positive hit compounds that increase cell numbers were identified for the Cl.8 and S2 cell lines, respectively, and 73 (for Cl.8) and 33 (for S2) unique negative hit compounds decreased cell numbers (Figure 2.2, electronic supplementary information). Of the positive hits, three compounds were identified in both cell lines. All of the repeated positive hits were polyamines (spermine, spermidine, and putrescine), with spermine ($z_{Cl.8} = 10.4$, $z_{S2} = 17.7$) and spermidine ($z_{Cl.8} = 10.0$, $z_{S2} = 8.7$) representing the two strongest positive hits in both cell lines (Figure 2.2 E).

Of the negative hits, 21 compounds were found in both cell lines (Figure 2.2 F, electronic supplementary information). To narrow the analysis, another threshold was applied, yielding four compounds with $z$-scores $< -10$ in both cell lines: pyrvinium
pamoate ($z_{Cl.8} = -17.7, z_{S2} = -13.0$), staurosporine ($z_{Cl.8} = -18.6, z_{S2} = -17.3$), AG-879 ($z_{Cl.8} = -17.3, z_{S2} = -13.9$), and PKC-412 ($z_{Cl.8} = -17.1, z_{S2} = -11.0$) (Figure 2.2 F). Pyrvinium pamoate is an androgen receptor inhibitor and anticancer agent\textsuperscript{161,162}. Staurosporine and PKC-412 are inhibitors of protein kinases, with staurosporine being the precursor for PKC-412 development\textsuperscript{163,164}. AG-879 is tyrosine kinase inhibitor and suppressor of malignant transformation\textsuperscript{165,166}. Staurosporine is a known autophagy inducer, indicating that our approach does detect cell number changes resulting from compound treatments.
Figure 2.2: (A-B) Z-scores for all screened compounds on (A) Cl.8 cells and (B) S2 cells. (C-D) Representative images of hit compounds on Cl.8 (C) and S2 (D) cells. (E-F) Venn diagrams for unique positive (E) and negative (F) hit compounds.
**C** Spermidine, Cl.8
- \( z = 10.0 \)

**D** Spermidine, S2
- \( z = 8.4 \)

- Neg. Ctrl, Cl.8
  - \( z = -0.03 \)

- Staurosporine, Cl.8
  - \( z = -6.6 \)

- Neg. Ctrl, S2
  - \( z = -0.50 \)

- Staurosporine, S2
  - \( z = -17.3 \)

**E**
- Spermidine
- Spermine
- Putrescine

- Positive
  - \( z > 3.0 \)
  - CI8
  - S2
  - 20

- Negative
  - \( z < -10.0 \)
  - Staurosporine
  - Pyrvinium pamoate
  - AG-879
  - PKC-412

**F**
- CI8
  - 52
  - 21

- S2
  - 12
2.3.2 Growth properties in chemically defined medium

Due to the lower cost of spermidine and our desire for an economical CDM we incorporated 1 μM spermidine into ZO Fortified, newly termed “ZB Media,” and conducted growth kinetics experiments to characterize the effect of supplementing ZO Fortified with spermidine. To determine the extensibility of this medium, the growth of Kc167 cells, an isolate of the Kc167 embryonic cell line, was also investigated. For all three cell lines tested, ZB Media yields higher proliferation rates than ZO Fortified, indicating that spermidine does indeed promote cell growth and attachment (Figure 2.3 A-C, G-I). Kc167 cells in particular proliferate in ZB Media at comparable rates to in the complete serum-containing medium. Interestingly, ZO Fortified, which we developed for Cl.8 cells, does not significantly improve growth compared to ZO unsupplemented for S2 or Kc167 cells (Figure 2.3 D-F).
Figure 2.3: Growth kinetics of Clone 8, Schneider 2, and Kc167 cells in ZO unsupplemented, ZO Fortified, ZB Media, and their respective complete, serum-containing media. Error bars represent standard deviations. (A-C) Changes in averaged fluorescent intensity measurements indicate changes in DNA content from viable cells. Error bars represent standard deviations. (D-F) Fold changes in intensity from day 0 are taken to represent fold changes in cell number. Each of the four cell lines proliferate significantly better in ZB Media (with spermidine supplementation) than ZO Fortified by the fifth day. Kc167 cells in particular proliferate at comparable rates to the complete, serum-containing medium by day 5. Error bars represent standard deviations for which propagation of error is accounted. Single and double asterisks denote $p < 0.05$ and $p < 0.01$, respectively, for two-tailed t-test for unmatched pairs. Corresponding p-values can be found in the electronic supplementary information. (G-I) Cells from day five of the same experiment. The apparent low confluency of Kc167 cells (I) is due to a high number of suspension cells that cannot be seen in one plane of view.
2.3.3 Long-term culture in chemically defined medium

To determine ZB Media’s ability to support long-term cell growth, we cultured Cl.8, S2, and Kc167 cells in ZB Media versus ZO Fortified. After 3 passages in the two chemically defined media, Cl.8 cells continue to proliferate in ZB media (ZO Fortified supplemented with spermidine) whereas those cultured in ZO Fortified (no spermidine supplementation) stall in growth (Figure A.2). Inclusion of spermidine has enabled Cl.8 growth through 98 passages (thus far) with no signs of decreasing growth rates. Importantly, cell morphology in ZB Media is consistent with the serum containing media. Surprisingly, Cl.8 cells do not require a weaning from serum to adapt to the CDM, they adapt quickly to sustained growth in ZB Media. Adapted Cl.8 cells (passaged 10 times in ZB Media) passaged 1:2 become confluent after 1-2 days, comparable to Cl.8 cells cultured in Cl.8 medium. Adapted Cl.8 cells can also be passaged at higher dilutions (1:6) and reach confluency within three to four days. Further, an investigation into the growth kinetics of Cl.8 cells at various seeding densities shows that spermidine improves growth of cells in ZO Fortified even at low concentrations, and that Cl.8 cells proliferate in ZB Media even when seeded at low concentrations, albeit much more slowly (down to 250,000 cell/mL, Figure A.3, Figure A.4). To determine if there is any dependence of proliferation on potential secreted factors, we have performed spent media titrations, which fail to show any benefit on growth induced by spent media (Figure A.5). ZB media
is also capable of supporting growth of Kc167 cells through at least 14 passages without any signs of stalling in growth (so far; experiments with Kc167 having been initiated subsequent to the initial studies).

S2 cells grown in ZB Media reach passage four before growth stalls whereas those grown in ZO Fortified are unable to progress through passage 1 (Figure A.2). Thus while ZB Media enables additional growth of S2-DSRC cells, ZB Media requires further improvement to support long term growth of S2 cells.

Therefore, we have developed a medium capable of supporting long-term growth of Cl.8 and Kc167 cells by identifying a supplement cocktail of five components. Cl.8 cells adapted to ZB media are able to recover after storage in liquid nitrogen in ZB Media supplemented with 0.2 M trehalose and 10% DMSO. We have found that inclusion of trehalose in freezing media significantly improves cryopreservation in multiple media/cell types, consistent with findings for mammalian cells\textsuperscript{169–171}.

2.3.4 Biological effect of spermidine

Short-term dose-response experiments (data not shown) failed to show strong dose-dependence of proliferation on spermidine concentration. However, when polyamine-depleted Cl.8 cells were used, a clear dependence of proliferation on spermidine concentration was observed (Figure 2.4 A-B). Supplementation with 1 μM spermidine caused a significant increase in proliferation ($p < 5 \times 10^{-10}$) of Cl.8 cells
compared to 0 or 0.1 μM spermidine supplementation. A higher dose, however, failed to produce any further increase in cell number after six days.

To more specifically show that spermidine does indeed promote proliferation, we measured DNA synthesis by 5-ethynyl-2′-deoxyuridine (EdU) incorporation experiments on Cl.8 and S2 cells. ZB Media yields a significantly higher percentage of EdU positive cells than ZO Fortified for both Cl.8 (p < 0.005) and S2 (p < 0.025) cells (Figure 2.4 C-E).
Figure 2.4: (A-B) Proliferation of polyamine-depleted Cl.8 cells increases in a dose-dependent manner with spermidine supplementation. Error bars represent standard deviations with for which propagation of error is accounted. Single and double asterisks denote p < 0.05 and p < 0.01, respectively, for two-tailed t-test for unmatched pairs. (C) Spermidine supplementation (ZB Media) results in increased EdU incorporation in both Cl.8 and S2 cells. Error bars represent standard deviations. Single and double asterisks denote p < 0.05 and p < 0.01, respectively, for two-tailed t-test for unmatched pairs. (D-E) Representative images of EdU (magenta) and DAPI (blue) stained Cl.8 (D) and S2 (E) cells in each culture medium tested.
Because spermidine was previously found to stimulate phosphorylation of tyrosine kinases and ERK1/2 in the Ras/MAPK signaling cascade, we also investigated MAPK activity of Cl.8 and S2 cells in our spermidine supplemented medium\textsuperscript{172,173}. Western blots indicate that spermidine supplementation does in fact increase ERK double phosphorylation for Cl.8 cells (\( p < 0.012 \)), but not for S2 cells (\( p > 0.75 \)) (Figure 2.5 A-B). Total ERK levels are not significantly influenced by spermidine supplementation for either cell line (\( p > 0.45 \)), although there does seem to be a trend of increasing ERK levels across the iterating improvements to Wyss’ original ZO Media (Figure 2.5 C-D, \( p \)-values in electronic supplementary information).
Figure 2.5: (A-B) Spermidine supplementation (ZB Media) results in increases in ERK phosphorylation (dpERK) for Cl.8 cells ($p < 0.012$), but not for S2 cells ($p > 0.75$). (B-C) Although ERK expression does not change significantly with spermidine supplementation for either Cl.8 ($p > 0.55$) or S2 ($p < 0.45$) cells, there is a trend of increasing ERK expression across ZO Media (unsupplemented) improvements. Error bars represent standard deviations across three replicates. Single and double asterisks denote $p < 0.05$ and $p < 0.01$, respectively, for two-tailed t-test for unmatched pairs. Corresponding p-values can be found in the electronic supplementary information.
A

<table>
<thead>
<tr>
<th>Clone.8 Cells</th>
<th>Schneider 2 Cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZO</td>
<td>ZO Fort</td>
</tr>
<tr>
<td>dpERK</td>
<td></td>
</tr>
</tbody>
</table>

B

![Bar graph showing normalized optical density of dpERK for Clone.8 and Schneider 2 cells under different conditions.](image)

C

<table>
<thead>
<tr>
<th>Clone.8 Cells</th>
<th>Schneider 2 Cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZO</td>
<td>ZO Fort</td>
</tr>
<tr>
<td>ERK</td>
<td></td>
</tr>
</tbody>
</table>

D

![Bar graph showing normalized optical density of ERK for Clone.8 and Schneider 2 cells under different conditions.](image)
2.3.5 Compound target analysis

Target analysis on compounds with positive average z-scores yielded 111 gene product candidates significantly targeted for the Cl.8 cell line and 53 gene product candidates significantly targeted for the S2 cell line. All of the gene products that were targeted in S2 were also targeted in Cl.8. Some of the strongest genes targeted by positive-scoring compounds for both cell lines were glycogen phosphorylase (GlyP), ornithine decarboxylase 1 (ODC1), S-adenosylmethionine decarboxylase (SamDC), casein kinase IIα (CkIIα), and ornithine aminotransferase precursor (Oat). These findings are consistent with the importance of polyamines, with ODC1 and SamDC both being central upstream enzymes in the polyamine biosynthesis pathway\textsuperscript{174–176}.

Ontology enrichment on scored protein lists yielded multiple annotations that were visualized using REVIGO scatter plots, which group similar annotations into broader “terms” and plot them on semantic axis where similar terms are closer together; points on scatter plots are sized according to the number of ontology annotations per term and are colored based on their p-values (Figure 2.6). Among enriched ontology terms for targets of positive z-score compounds, polyamine metabolism, ornithine metabolism, and cellular modified amino acid biosynthesis pathways emerged for both cell lines.
Figure 2.6: REVIGO scatter plots of enriched ontology terms for targets of (A) positive scoring compounds for Cl.8 cells, (B) positive scoring compound for S2 cells, (C) negative scoring compounds for Cl.8 cells, and (D) negative scoring compounds for S2 cells. Enriched ontology annotations are grouped into terms by semantic similarity and plotted on semantic axes, where similar terms are clustered closely on the plot. Circles representing terms are colour-coded according to p-value and sized according to the number of ontology annotations per term.
KEGG pathway enrichment was also conducted for targets of positive-scoring compounds, with glutathione metabolism, arginine and proline metabolism, and cysteine and methionine metabolism all significantly enriched terms for Cl.8 cells, and arginine/proline metabolism and cysteine/methionine metabolism being significantly enriched for S2 cells as well. This is consistent with the importance of polyamine metabolism for cells cultured in ZO Fortified, as polyamines are synthesized from both arginine and methionine. The first step in polyamine metabolism is the production of ornithine from arginine; at the same time, L-methionine is used to create decarboxylated S-adenosyl-L-methionine (DcAdoMet) which acts as an aminopropyl group donor to either putrescine or spermidine to produce either spermidine or spermine, respectively\textsuperscript{174,177}.

Target analysis on compounds with negative average \( z \)-scores yielded 266 gene product candidates significantly targeted for the Cl.8 cell line and 166 gene product candidates significantly targeted for S2 cell line. Some of the strongest genes targeted by negative-scoring compounds in both cell lines were phosphorylase kinase \( \gamma \) (PhK\( \gamma \)), calmodulin (Cam), and downstream of raf1 (Dsor1). Among enriched ontology terms, protein amino acid phosphorylation, protein kinase activity, phosphorus metabolic process, and ATP binding emerged. The target of rapamycin (TOR) pathway and progesterone-mediated oocyte maturation were found as significantly enriched KEGG pathways for Cl.8 cells, with ribosome as well as again the progesterone-mediated oocyte maturation being enriched for S2 cells. The TOR pathway is strongly involved in
the regulation of cell growth, proliferation, and survival, especially in the context of coupling growth with nutrition\textsuperscript{138,143–145}.

A complete list of compounds, average z-scores, target proteins, as well as lists of target proteins and their associated p-values and enriched ontology annotations are provided in the electronic supplementary information. Validation of candidates and whether the targeting is definitively antagonistic or agonistic remains to be determined in future investigations and is outside the scope of the present study.

2.4 Discussion

In this study, we have identified five compound additives that enable Wyss’ ZO medium to support long-term growth and maintenance of \textit{Drosophila} Cl.8 and Kc167 cells\textsuperscript{128,130}. In particular, the polyamine spermidine was found through an inverse drug screen to be the critical component missing from ZO Fortified, our supplemented version of ZO medium that enables this long-term growth. The other two central polyamines, spermine and putrescine, were also identified from our compound screen to be significant growth promoters for both Cl.8 and S2 cell lines. Consistent with this finding, polyamines were found to be required in a chemically defined medium formulation for the flesh fly, \textit{Boettcherisca peregrina}, potentially generalizing the need for polyamine supplementation in chemically defined media for invertebrate cell culture\textsuperscript{178}. While many of the supplements used in culture media are undefined, it is known that polyamines are typically present in high levels in fermented foods such as
yeast extract, which is routinely included in many *Drosophila* culture media\textsuperscript{179}. It is possible that *Drosophila* cells may be missing upstream components for polyamine synthesis, or lack the signals to synthesize polyamines, and thus required exogenous polyamine supplementation.

Although polyamines are known to be essential for proliferation, the exact biological functionalities for these molecules are both pleiotropic and incompletely defined\textsuperscript{151,174}. Here we show that spermidine supplementation significantly increases phosphorylation levels and, potentially, expression levels of ERK. This is consistent with findings that spermidine specifically stimulates the phosphorylation of tyrosine kinases and ERK1/2 in the Ras/MAPK activated signaling cascade\textsuperscript{172,173}.

Importantly, the magnitude of the effects of spermidine on proliferation is relatively small, at least as measured in assays of short duration (four days or less), suggesting that polyamines become limiting only after significant cellular depletion. This assertion is also supported by the finding that dose-dependence of proliferation on spermidine concentration is stronger for cells that have been depleted of polyamines compared to normal cells. It appears that *Drosophila* cells may be unable to synthesize sufficient polyamines from the available nitrogen sources in ZO Fortified; therefore, polyamine supplementation is required and transport is likely crucial. This hypothesis is supported by the gene target analysis and ontology enrichment implicating the importance of arginine and methionine metabolism, both of which are required upstream for polyamine synthesis. In normal rat kidney cells, pharmacological knockout
of polyamine synthesis required several days before polyamine levels were significantly reduced, potentially explaining why exogenous polyamine supplementation would have small effects in the short term but be required in the long term\textsuperscript{180}. Due to the suboptimal growth conditions of the screening media (ZO Fortified), use of two cell lines to increase screening resolution, and our protein-target analysis, we were able to detect the relatively small short-term effects of polyamine supplementation.

Our media development pipeline can be expanded for the rational design of media to improve conditions for industrial production of recombinant proteins from insect cells, increase success rates for creating new cell lines from primary cultures, or define the minimal essential factors required for the proliferation of other types of insect cells. However, one drawback of our pilot screen was the low genome coverage; STITCH target analysis indicated that only 22\% of the proteome was targeted. This is due to our screening of a limited number of targeted compound libraries. This library selection and coverage issue is another promising application of our target analysis approach, which could be extended to develop rationally designed compound libraries that target the maximum possible genome coverage based on known interactions cataloged in STITCH. However, it is important to note that the target analysis approach is heavily dependent on the quality and quantity of data cataloged by interaction databases like STITCH.

One drawback of our method is the single-factor basis of our screening pipeline, which does not specifically identify compound synergies promoting proliferation.
Pooling strategies specifically used for synergy identification, however, are not compatible with our target scoring approach\textsuperscript{181}. As an alternative method to identify synergies, we propose iterative expansion of the screen, where validated factors would be incorporated into the media background before compound rescreening. This methodology could elucidate compound synergies with previously identified compounds, and our biological target identification technique coupled with this iterative approach could potentially lead to important clues about novel biological mechanisms and crosstalk between pathways and proteins important for growth. Our cumulative gene target scoring approach is an improvement over traditional screens that only consider biological targets of individual hit compounds, and can be applied to query a wide range of biological processes in other model systems.

2.5 Acknowledgments

We would like to thank C. Wyss and S. Restrepo for helpful and informative discussions before the initiation of this research project, J. Chen, E. Lee, M. Connelly, and E. Smith for their contributions to cell maintenance and media preparations. The authors are also very grateful for helpful feedback from P. Cherbas, L. Cherbas, and members of the Zartman lab. Cell lines were obtained from the \textit{Drosophila} Genomics Resource Center, supported by NIH grant 2P40OD010949-10A1. This work was supported by the University of Notre Dame.
3.1 Introduction

Polyamines (PAs)—putrescine, spermidine and spermine—are ubiquitous organic molecules influencing multiple cellular processes, including gene regulation, signal transduction and ion channel gating\textsuperscript{182–184}. In particular, PAs have been implicated in the regulation of a variety of developmental processes. In retina development of rabbits, chicks, and rats, PAs were found to be spatially localized\textsuperscript{185–189}. In rabbits, PA depletion disrupts cone-differentiation\textsuperscript{189}. Polyamines also function in the growth and differentiation of the mammary epithelium\textsuperscript{190–192}. Dynamic levels of PAs in embryogenesis of chicks and frogs has linked them with embryonic gastrulation\textsuperscript{193–197}. Various enzymes within the metabolic pathway have been found to be required for embryonic viability in mice\textsuperscript{198,199}.

In \textit{Drosophila}, dynamic PA levels have been observed throughout development, with increased PA levels detected during periods of rapid growth\textsuperscript{200–204}. Interestingly,
spermidine increases the net rate of RNA synthesis in larvae\textsuperscript{152}, suggesting a possible mechanism of broad transcriptome regulation. Although several studies have investigated specific PA metabolism enzymatic activities during development\textsuperscript{205,206}, there has not been a comprehensive analysis. Outside of development, \textit{Drosophila} has been used as a pharmacological model for polyamine-conjugate chemotherapeutics\textsuperscript{207,208}, for mechanistic studies in signaling pathways\textsuperscript{209}, and in the context of aging and autophagy\textsuperscript{210–212}. We have previously shown that polyamines are required for long-term growth of \textit{Drosophila} cell lines in a chemically defined medium\textsuperscript{105}. In the current work, we combine integrative bioinformatic analysis of transcriptional data with an immunohistochemical investigation of spermidine and spermine (Spd/Spn) levels during development. We identify novel dynamic spatiotemporal PA accumulation profiles in both the larval eye imaginal disc and the embryonic trachea. To our knowledge, this is the first report of dynamic spatially restricted accumulation of PAs during specific stages of organogenesis in \textit{Drosophila}.

3.2 Results

To characterize PA regulation in development, we used publicly available expression data across 30 stages of \textit{Drosophila} development\textsuperscript{213}. Hierarchical clustering and principal component analysis (PCA) of putative \textit{Drosophila} PA metabolism genes resulted in four clusters correlated with the enzymatic flow of PA metabolism: upstream
biosynthesis (red), downstream biosynthesis (green), interconversion (blue), inhibition (yellow) (Figure 3.1, Figure A.6).
Figure 3.1: (A) Polyamine metabolic pathway, adapted\textsuperscript{214}. Metabolites in white boxes, enzymes in ovals—human gene top, \textit{Drosophila} ortholog bottom (bold). Enzymes are color-coded to indicate observed regulatory clusters. (B) Principle component analysis (PCA) of PA genes across development reveals four clusters that follow the regulatory logic of the metabolic pathway (black arrow): upstream (red) and downstream (blue) biosynthesis, interconversion (blue), and inhibition (yellow).
To link regulatory clusters to metabolite accumulation, *in vivo* PA levels were extracted from previous studies quantifying developmental organismal levels\(^202,203\). Average regulatory cluster dynamics align to predict some basic dynamics of PA accumulation through development (Figure 3.2: Normalized gene cluster dynamics and measured PA accumulation profiles (left and right axes, respectively) across stages of *Drosophila* embryonic (top) and larval (bottom) development (gaps represent data gaps from the literature)\(^213,202,203\). Average cluster dynamics align to predict basic dynamics of PA accumulation through development importantly, large shifts in transcriptional and metabolic profiles coincide with significant developmental events: the shift from proliferation to differentiation during embryonic development occurs as the interconversion cluster reaches a diminished quasi-steady-state (Figure 3.2, highlighted grey); the shift from exponential growth in preparation for pupariation coincides with an increase in the inhibition cluster and subsequent decreases in PA levels (Figure 3.2).
Figure 3.2: Normalized gene cluster dynamics and measured PA accumulation profiles (left and right axes, respectively) across stages of *Drosophila* embryonic (top) and larval (bottom) development (gaps represent data gaps from the literature) $^{213,202,203}$. Average cluster dynamics align to predict basic dynamics of PA accumulation through development.
Embryogenesis (0-24 hours)

Larval and Pupal Development

Avg. Normalized Expression pmol/50 embryos

Avg. Normalized Expression µmol/animal

Upstream Interconversion Inhibition

Upstream

Downstream

Interconversion

Inhibition

Putrescine

Spermidine

Spermine

Embryogenesis (0-24 hours) Larval and Pupal Development
Analysis of Spd/Spn in the developing embryos revealed PA accumulation within the lumen of tubes formed by tracheal cells, visualized with expression of actin-GFP under breathless (Figure 3.3 C-D). The embryonic trachea is an excellent model for tubulogenesis\textsuperscript{215,216}, and develops in the latter half of embryogenesis by invagination of epithelial cells from the outer wall of the organism to form precursor sacs (Figure 3.3 A-B)\textsuperscript{217}. Breathless (btl) expressing tracheal cells migrate and intercalate toward sources of Branchless, and tubes are formed by cellular secretion and arrangement around a chitinous matrix that causes luminal expansion to the final diameter\textsuperscript{218}. Specifically, PAs are accumulated in the lumen of the developing embryonic trachea with the same dynamics as the luminal chitin-binding protein (Figure 3.3 C-D).
Figure 3.3: (A) The embryonic trachea is a branched tubular organ that develops in the latter half of embryogenesis\textsuperscript{215,216}. (B) Tube maturation depends on the formation of a chitinous matrix (shown in green) that expands the lumen diameter. This matrix is generated when btl-expressing tracheal cells (orange) experience a “secretion pulse”, inducing luminal deposition of chitin-binding proteins that expand the tube diameter\textsuperscript{218}. A subsequent endocytosis pulse rapidly clears the luminal matrix. Figure adapted\textsuperscript{219–221}. (C-D) Analysis of PA accumulation in developing embryos revealed PA deposition within the lumen of expanding trachea. Polyamines were accumulated with the same dynamics of 2A12 (magenta), a luminal marker for GASP, a chitin-binding protein\textsuperscript{222}. (E-F) Differentiation of the larval eye-antennal imaginal disc is driven by the morphogenetic furrow (orange), which initiates at the posterior part of the disc and proceeds to the anterior side. The progression of the morphogenetic furrow is dependent on well characterized morphogens including Decapentaplegic (Dpp) and Hedgehog (Hh) (blue) as well as and Wingless (Wg) (red), Notch (purple), and Epidermal Growth Factor Receptor (EGFR) (green)\textsuperscript{223–227}. Figure adapted\textsuperscript{227,228}. (G-I) PAs (red) are accumulated along the morphogenetic furrow in eye discs expressing E-cadherin::GFP (green, apical cell boundaries), with differentiated ommatidia present to the posterior (g-h, day 4; i-j, day six after egg laying). The PA pattern is initially more widespread (G-H), and becomes progressively more confined to the furrow (I-J). Scale bars 20 µm (G,I), 5 µm (H,J)
Along with embryogenesis, we also discovered spatial accumulation of polyamines during morphogenesis of the eye imaginal disc. Spd/Spn is accumulated along the morphogenetic furrow, visualized with E-cadherin-green fluorescent protein (GFP) fusion protein (Figure 3.3 G-J). In Drosophila, the eye-antennal imaginal disc is the precursor structure to the adult compound eye and antenna (Figure 3.3 E-F), and its transformation from an undifferentiated epithelial sac to a patterned organ of photoreceptor complexes called ommatidia has long been used to study cell proliferation, differentiation, and patterning. Prior to differentiation, a wave of mitosis determines the size of the eye and the number of ommatidia. Following this wave, the morphogenetic furrow initiates at the posterior end of the disc, and rapidly differentiates the epithelia into light receptors and support cells comprising the ommatidia. Specifically, PAs localize to the morphogenetic furrow during retinal patterning of the eye imaginal disc (Figure 3.3 G-J), and are located apically to E-cadherin (Figure A.7). Interestingly, the PA pattern is initially intense and widespread (Figure 3.3 G-H), but becomes progressively more confined to the furrow (Figure 3.3 I-J). Interestingly, GASP was found to accumulate in the same pattern along the morphogenetic furrow in eye discs (results not shown).

3.3 Discussion

While no previous research has demonstrated a link between PAs and tubulogenesis, yeast cells with depleted amines had severely affected intercellular chitin
structures. Because *Drosophila* tracheal tube maturation depends on cellular secretion and arrangement around a chitinous matrix, the increased level of PAs observed in the tracheal lumen correlates with chitin and extracellular matrix deposition. In chicks, PAs are necessary to sustain tissue proliferation in the retina and are dynamically accumulated during normal development. Our observation of dynamic PA accumulation with differentiating cells along the morphogenetic furrow (Figure 3.3 H-K) suggests that these roles for PAs in eye development are likely conserved between *Drosophila* and mammalian retinal development. Further, MAPK signaling, known to be important for eye differentiation, was recently been linked to polyamine metabolism in *Drosophila*, providing a putative function for PAs in this organ.

Here we demonstrate that the developmental-stage-specific regulation of the PA metabolic pathway results in dynamic accumulation of PAs in particular organs during development. We have discovered novel, spatiotemporal accumulation of PA metabolites in developing embryonic trachea and differentiating eye imaginal discs. While future research will be required to elucidate specific developmental roles of PAs in these organs, this study highlights the utility of *Drosophila* as a powerful model organism for investigating PA biology. Further, the observation of spatiotemporal accumulation of PA levels as measured by immunohistochemistry can lead to novel assays for studying PA transport.
3.4 Methods

Publicly available whole genome microarray expression data accessed through ModMine (http://intermine.modencode.org/) was used\(^{213,233}\). First, *Drosophila* orthologs to human PA metabolism genes were identified using Basic Local Alignment Search Tool (BLAST)\(^{234}\). Gene expression data for these orthologs was then isolated, mean-centered and standard deviation normalized. Polyamine metabolic genes were hierarchically clustered in Matlab using the Spearman average distance\(^{235}\). Principal component analysis (PCA) was conducted in Matlab. Metabolite concentrations were extracted from published values measured in the embryo\(^{203}\) and larval/pupa\(^{202}\) using an online plot digitizer (http://arohatgi.info/WebPlotDigitizer/).

The fly strains used were: GAL4-btl, UAS-Act:GFP (Bloomington Stock Center) and E-cad::GFP\(^{236}\). Antibody stainings of eye imaginal discs and embryos were conducted as previously described\(^{237,238}\). Polyamines were visualized using an antibody to spermidine/spermine (Spd/Spn) (Abcam). Tracheal lumen was visualized using 2A12 (Developmental Studies Hybridoma Bank). Isotype controls of mouse and rabbit IgG (Vector Labs) yielded no signal (data not shown). Secondary antibodies were goat anti-rabbit IgG 561 (Invitrogen), and goat anti-mouse IgG 647 (Invitrogen).

Imaging was performed on a Nikon Eclipse Ti confocal microscope (Nikon Instruments Inc.) with a Yokogawa spinning disc (Andor Technology). Image data was collected on an iXonEM+ cooled CCD camera (Andor Technology) using MetaMorph\(^{\circledR}\) v7.7.9 software (Molecular Devices). Shown are maximum intensity z-projections.
3.5 Acknowledgments

Stocks obtained from the Bloomington *Drosophila* Stock Center were used in this study. We are grateful for feedback from Erin Howe, Cody Narciso, Qinfeng Wu, and Pavel Brodskiy.
CHAPTER 4:

RAPID TRANSCRIPTOMICS WITH HIGH-THROUGHPUT FUNCTIONAL VALIDATION
ELUCIDATES NOVEL THERAPEUTIC TARGETS FOR BREAST CANCER BRAIN METASTASIS

4.1 Introduction

Breast cancer remains the most commonly diagnosed carcinoma and an estimated 231,840 American women will be diagnosed in 2015\textsuperscript{239}. Of these women, an estimated 40,290 will succumb to their disease\textsuperscript{239}. There have been a number of studies aimed at identifying diagnostic and prognostic markers that might serve as therapeutic targets for breast cancer initiation and progression\textsuperscript{240–249}. One of these studies revealed that breast cancer can be classified into four distinct subgroups based on expression of the estrogen and progesterone hormone receptors (ER and PR, respectively) and amplification of human epidermal growth factor receptor (HER2)\textsuperscript{249}. These subtypes now form the basis of breast cancer molecular classification: luminal A (ER/PR positive, HER2 negative; 30-60%); luminal B (ER/PR positive, HER2 positive; 10-20%); triple negative or basal-like (ER/PR negative, HER2 negative; 15-20%); and HER2 (ER/PR negative, HER2 positive; 10-20%).

\footnote{Collaborative work: Erin N. Howe, Xuejuan Tan, Patricia Skallos, Brandon Ashfeld, Jun Li, Siyuan Zhang}
negative, HER2 positive; 5-15%)⁹⁷–¹⁰³. The most common treatments target ER signaling (tamoxifen and aromatase inhibitors), and as such are only effective for patients with ER positive disease²⁵⁶–²⁵⁸. There are currently four approved treatments targeting HER2: the monoclonal antibodies, trastuzamab²⁵⁹–²⁶² and pertuzumab²⁶³; the antibody-microtubule inhibitor conjugate T-DM1²⁶⁴,²⁶⁵ (trastuzemab emastine); and the dual HER1/HER2 kinase inhibitor, lapatanib²⁶⁶,²⁶⁷. Chemotherapy is the other treatment option, but even cancers that initially respond well to treatment can eventually become drug resistant through a variety of mechanisms²⁶⁸–²⁷⁰. Further, around 30% of women diagnosed with early-stage breast cancer will progress to metastatic breast cancer (MBC), for which treatment strategies are extremely limited, particularly due to limited responses to first-line therapies (targeted or chemo-based)²⁶⁹–²⁷¹.

Despite the development of these targeted treatments for primary ER⁺ or HER2⁺ breast tumors, the overall incidence of patients relapsing with brain metastasis is increasing²⁷²,²⁷³. Further, 30% of breast cancer mortality is attributable to brain metastasis²⁷⁴. Although women diagnosed with localized breast cancer have an expected 99% five-year survival, when metastases to the brain or central nervous system (CNS) are present, the five-year survival drops precipitously to 25%, with a median survival of less than one year²³⁹,²⁷⁵. Patients with triple negative/basal-like breast cancer have the highest rate of brain metastases (35-50%), followed by patients with HER2 over-expressing disease (HER2⁺, 25-35%); however, even patients with the low-grade disease have a 5% chance of developing brain metastases²⁷⁶–²⁸².
Further, the incidence of brain metastasis is increasing with better control of systemic disease. In breast cancer, 40% of patients with HER2-amplified primary breast tumors who received first line therapy ultimately developed brain metastases even when the primary disease was under control\textsuperscript{281}. Many women who have stable systemic disease, or respond to initial treatment, ultimately develop brain metastasis. The greatest challenge in the clinical management of patients who relapse with brain metastasis is the extremely limited treatment options. Systemic treatments, such as chemotherapies or targeted therapies, cannot effectively treat micrometastatic brain lesions or prevent brain relapse, largely due to their inability to penetrate the blood-brain barrier. Currently, no clinically approved drug shows promising efficacy for brain metastases\textsuperscript{283}. Therefore, it is critical that we expand our mechanistic understanding of breast cancer brain metastases so that we can develop effective systemic targeted therapeutics.

Brain metastasis is an inefficient but highly regulated process\textsuperscript{284}. The cellular characteristics of the tumor cell--both its inherent genetic traits and poorly defined non-genetic changes influenced by the microenvironment--dictate successful passage through the metastatic cascade. While millions of tumor cells diseminate from a primary tumor, only a few will form metastatic tumors clinically\textsuperscript{285}. These can in turn form multiple inoperable metastatic lesions that are phenotypically different and innately resistant to first-line therapies designed for primary tumors\textsuperscript{286}. During the early stages of spreading, metastatic cells undergo a dramatic evolution that includes both
the passive selection of suitable cells for the specific organ and active adaptation to the distinct metastatic microenvironment. Here we used RNA-seq data that was generated in the lab of our collaborator, Professor Siyuan Zhang (Department of Biological Sciences, University of Notre Dame) comparing dormant, small single-cell brain metastases to large, actively proliferating brain metastases (Figure 4.1). Differential expression analysis\textsuperscript{96} of these datasets led to the identification of a BrainMets signature (S. Zhang), which contains 537 genes that are significantly up-regulated between single-cell and proliferative brain metastases.
Figure 4.1: Schematic of transcriptome deep-sequencing experiment identifying BrainMets signature. Single (7 dpi) or actively proliferating (40 dpi) RFP labeled MDA-231 tumor cells were isolated from the brain and subjected to RNA-seq. Actively proliferating metastases were classified as either small, medium, or large by visual inspection. Differential expression analysis revealed 537 genes that were significantly different between d7 and d40 small, medium, or large metastases. These 537 genes were taken to be the “BrainMets signature.” Figure graciously provided by Dr. Siyuan Zhang.
Intracarotid injection of MDA-231 cells → Single cell mets → Time (days) → Overt mets

Transcriptome Changes (RNA-seq)

BrainMets Signature:
537 significantly upregulated genes
To functionally evaluate these BrianMets signature genes, we harnessed a whole organism genetic model in *Drosophila* and directly assessed their impact on tumorigenesis and metastasis in a high-throughput screen\textsuperscript{287}. Recent advances in the *Drosophila* cancer models represent an intriguing alternative to traditional time-consuming and high cost genetic work in mice\textsuperscript{288,289,77,290}. The *Drosophila* model system enables the efficient, rapid and inexpensive simultaneous manipulation of multiple genes and analysis of complex genotypes in a defined cell population\textsuperscript{69,71}. It has been demonstrated that *Drosophila* cancer models are effective at revealing both fundamental aspects of human gliomas\textsuperscript{291} and interactions among cancer cells and the tumor microenvironment\textsuperscript{77,289,288,290,292}. This strategy has also been used as a high throughput drug screen to evaluate whole-animal compound efficacy of candidate cancer therapeutics\textsuperscript{293}. In particular, a similar high-throughput approach in *Drosophila* led to the development of Vandetanib (ZD6474), the first FDA approved drug for the treatment of metastatic medullary thyroid cancer\textsuperscript{294,295}. Remarkably, while *Drosophila* has an open circulatory system that lacks blood, its functional equivalent to blood (hemolymph) is separated from the brain by a BBB that is similar in physiology and structure to that of humans, contains critical and genetically highly conserved genes that regulate BBB integrity\textsuperscript{296,297,78,79,298}.

While massive exploration of sequencing data reveals extensive genetic and epigenetic changes during metastasis, functionally testing the importance of changes observed from -omic analyses in a tissue context requires a genetic model system. In
this context, *Drosophila* provides several advantages, particularly its suitability for the rapid screening of genes correlated with a biological process in an *in vivo* context. Our approach of building off of RNA-seq with RNAi functional screening in *Drosophila* enables high-throughput identification of key genetic mediators of brain metastasis that would not have been feasible using traditional murine models.

4.2 Results

To identify functional drivers of metastatic progression, we took advantage of a well-established *Drosophila* tumorigenesis model, which allows for the rapid screening of hundreds of RNAi constructs\(^{299}\). We utilized an epithelial tumor model based in the larval eye imaginal disc, the precursor to the adult eye, that metastasizes into the adjacent brain tissue\(^{287}\) (Figure 4.2). The model is based on expression of UAS-transgenes under control of a FLP-out GAL4 transgene that stably expresses GAL4 after FLP-induced recombination\(^{71,69}\). FLP-induced recombination, and thus transgene expression, is restricted to the eye discs using an eyeless promoter-driven FLP transgene\(^{287}\). Spatial extent of transgene expression is visualized using UAS-GFP to mark GAL4-expressing cells (green, bottom left). Concurrent expression of Ras\(^{V12}\) (an oncogenic form of Ras) and dlg\(^{\text{RNAi}}\) (the epithelial polarity gene *discs large*) results in neoplastic overgrowth of the eye discs and metastatic invasion into the neighboring brain and ventral nerve chord\(^{287,292}\) (bottom right). The tumor subsequently migrates and metastasizes into the larval brain and CNS, where GFP expression allows for
identification of tumor growth and metastasis. A single genetic cross yields larva bearing an RNAi construct under control of the same exogenous promoter that drives expression of GFP, Ras$^{V12}$ and $dlg^{RNAi}$, allowing for specific expression in the tumor-forming cells. Thus a candidate gene’s impact on tumorigenesis and metastasis can be assayed rapidly using this model. The line was first introduced in 2011, where it was used to characterize the requirement of phosphoinositide 3-kinase signaling, specifically through phosphatase and tensin homolog (PTEN), for Ras$^{V12}$, $dlg^{RNAi}$ dependent tumors.
Figure 4.2: *Drosophila* tumor screening model. Schematic of a *Drosophila* larva and relevant tissues (top). The model we utilized expresses UAS-transgenes under control of a FLP-out GAL4 transgene that stably expresses GAL4 after FLP-induced recombination. FLP-induced recombination, and thus transgene expression, is restricted to the eye discs using an eyeless promoter-driven FLP transgene. Spatial extent of transgene expression is visualized using UAS-GFP to mark GAL4-expressing cells (green, bottom left). Additional expression of oncogenic Ras<sup>V12</sup> results in neoplastic overgrowth of the tissue but no metastatic outgrowth. Expression of shRNA against the epithelial polarity marker discs large<sup>301,302</sup> (dlg) results in elimination of those cells by neighboring cells without FLP-induced transgene expression. Concurrent expression of Ras<sup>V12</sup> and dlg<sup>RNAi</sup> results in neoplastic overgrowth of the eye discs and metastatic invasion into the neighboring brain and ventral nerve chord<sup>287,292</sup> (bottom right). Figure adapted from Halder and Mills<sup>289</sup>.
Developing eye tissue
Brain
Ventral nerve chord

+Ras$^{v12}$  +dlg$^{RNAi}$  +Ras$^{v12}$, dlg$^{RNAi}$

eyFLP;
act>CD2>Gal4;
UAS-GFP
To identify genetic drivers of brain metastasis, we used the Zhang lab’s BrainMets signature to identify candidate genes for functional screening in our Drosophila model. Their BrainMets (Figure 4.1) signature of 537 genes incorporated any gene found to be differentially regulated between the d7 tumors and the d40 small, medium, or large tumors, with differential expression analysis conducted independently for each tumor size. In order to identify a generalized metastasis signature, we took the differential expression analysis conducted by the Zhang lab and combined significance on a gene-wise basis using Fisher’s test (electronic supplementary information). This approach treats each tumor size as a separate experiment testing the same hypothesis. After combining significance from each 40 dpi tumor size, our BrainMets signature was reduced to only 126 genes with differential expression between dormant and actively proliferating metastases.

We further narrowed our search to include only genes that were up-regulated in the overt metastases, since these would be the most therapeutically targetable; this narrowed our gene list to 108 genes that were significantly and consistently up-regulated in the overt metastases. From these 108 upregulated genes, Drosophila homologs were identified using Basic Local Alignment Search Tool (BLAST), resulting in 241 Drosophila genes to represent 81 of the 108 up-regulated human genes\textsuperscript{234}. Stocks were identified and obtained from the Transgenic RNAi Project at Harvard Medical School (TRiP) through the Bloomington Stock Center and from the Vienna Drosophila Resource Center (VDRC)\textsuperscript{72,73}. This resulted in almost 400 RNAi Drosophila lines
representing 79 genes that could be rapidly screened for their functional role in tumorigenesis and brain metastasis.
Figure 4.3: Pipeline for identification of screening reagents. Briefly, the 537 differentially regulated genes between small, medium, and large overt metastases were reduced to 126 genes using Fisher's exact test to combine significance. Conversion of human genes to Drosophila genes by homology resulted in 241 Drosophila genes representing 81 of the 126 signature genes. Finally, we acquired 368 RNAi reagents to functionally test those 126 human genes in our Drosophila tumor screen.
BrainMets Signature 537 Genes
325 from small
89 from medium
149 from big

Combine Significance Fisher's Test

126 Signature Genes
small/medium/large combined

Filter Targetable Genes
upregulated in overt metastases

108 Upregulated Human Genes

Identify Homologs

241 Drosophila Genes
representing 81 human genes

Identify RNAi Lines

368 RNAi Reagents
All 368 Drosophila RNAi lines were crossed to the tumor tester line to generate tumors with shRNA expression. At least 15 tumor-bearing larvae were imaged and analyzed for each RNAi construct, resulting in over 6,000 images of larvae to be analyzed. Total GFP integrated intensity was used as an assay readout and yielded good screen quality, as determined by the strictly standardized mean difference for screen quality (SSMD$_{QC} = 1.8$) based on the negative control (yw, no RNAi) and positive control (shRNA against the tumor suppressor PTEN)$^{51,287}$. Of the 368 RNAi constructs screened, 66 RNAi constructs representing 51 unique human genes significantly decreased tumor growth and metastasis to at least a moderate effect (Figure 4.2). Of those, 39 RNAi constructs representing 14 unique human genes strongly decreased tumor growth and metastasis. Data are presented as the SSMD of the integrated intensity of GFP, with the negative control in green normalized to zero, and knockdown of the tumor suppressor PTEN serving as a positive control in purple.
Figure 4.4: SSMD plot from tumor screen. Flies from the GFP expressing tumor line were crossed to flies bearing RNAi constructs under control of UAS, enabling spatially restricted expression of shRNA in the tumor region. Tumors were given six days to develop in larva before larva were harvested and imaged. Integrated intensity of GFP was calculated for at least 15 larvae per cross, and SSMD scores were calculated with respect to the negative control (yw, no additional RNAi). shRNA against the tumor suppressor PTEN was used as a positive control. Hits are characterized as moderate or strong negatives as indicated.
Figure 4.5: (A-D) Integrated GFP intensity for SERPINF1, RAB11B, MRPL37, and SNRPD2 RNAi lines and controls. Error bars represent standard deviation. (E) Representative images of larvae for controls (YW, negative; shPTEN, positive) and shSERPINF1, shMRPL37, shSNRPD2, and shRAB11b.
After identifying our hit *Drosophila* genes that decreased tumor growth and metastasis, we translated these genes back to their human homologs to validate the RNA-seq expression data. Although the original BrainMets signature gene list was found from the comparison of single-cell versus overtly proliferating brain metastases, acquisition of these samples is time, cost, and labor intensive. Therefore we chose to use primary breast tumors and overt brain metastases for initial validation of expression level changes. Because we model these tumors in a murine background using human breast cancer cells (MDA-231-Br), primers must be species-specific as well as target-specific. Primers targeting hit genes were designed and validated by gel electrophoresis before quantitative reverse-transcription polymerase chain reaction (qRT-PCR) was conducted. So far only 13 out of the 51 genes implicated to decrease tumor growth from the *Drosophila* screen have been tested by qRT-PCR (Figure 4.6 A). Of those, only two, Rab11B and SERPINF1, were found to be significantly up-regulated in the brain metastases compared to the primary tumor (Figure 4.6 B).
Figure 4.6: (A) qRT-PCR validation of tumor mediating gene mRNA levels in primary mammary gland tumors versus intracranially-injected brain metastases. Error bars represent standard deviation; asterisk denotes p<0.05 (t-test). (B) Rab11b and SERPINF1 are up-regulated in brain metastases.
4.3 Methods

4.3.1 Identification of *Drosophila* screening reagents

Zhang and collaborators previously conducted differential expression analysis comparing 7 dpi versus 40 dpi for small, medium, and large tumors independently using cuffdiff\textsuperscript{96}. To identify candidate functional mediator genes, we took the approach of treating all 40 dpi tumors as independent tests of the same hypothesis, and combined significance on a gene-wise basis using Fisher’s combined probability test. We further narrowed our search to only include genes up-regulated in the overt metastases. *Drosophila* homologs were identified using BLAST\textsuperscript{234}. Stocks were identified and obtained from TRiP through the Bloomington Stock Center or from VDRC\textsuperscript{72,73}.

4.3.2 Tumor screen

A tumor tester line of the following genotype was used: eyFLP; UAS-Ras\textsuperscript{V12}, UAS-dlg\textsuperscript{RNAi}, UAS-GFP/Cyo, Gal80\textsuperscript{T6}; act>CD2>GAL4, UAS-GFP\textsuperscript{287}. For each RNAi line, 15 tumor tester line virgins were collected and crossed to eight males from the RNAi line. Flies were given overnight to mate and lay eggs before being passaged to a new vial. Progeny larvae from the cross were collected the sixth day after egg laying, when larvae were wandering but not yet pupariating, for imaging. Larvae were collected in 50% glycerol and placed at -20°C for approximately 15 minutes for immobilization before being imaged on an EVOS\textsuperscript{TM} FL cell imaging system using GFP filter cube (50% brightness) and
transmitted light (approximately 45% brightness). At least 15 larvae were assayed for each cross.

Transmitted light images were thresholded and used as regions of interest for calculation of integrated GFP intensity that represented the tumors in each larva. SSMD scores were calculated for each line with respect to the negative control (yw, no additional RNAi)\(^51\). shRNA against the tumor suppressor PTEN was used as a positive control\(^287\). Image and statistical analysis was conducted using custom MATLAB scripts (electronic supplementary information).

4.3.3 PCR

Primers were designed using Primer-BLAST\(^303\). Target and species-specificity were verified to produce both single products and species-specific amplification by gel electrophoresis using cDNA from HEK-293T and NIH-3T3 cells. RNA was harvested from cells using the PureLink RNA kit (Invitrogen) and reverse transcribed into cDNA using the Verso cDNA kit (Thermo). GoTaq (Promega) was used to validate primers. cDNA from primary mammary gland tumors and intracranially injected brain metastasis samples was provided by Dr. Siyuan Zhang’s lab. For validation of the RNA-seq data, we performed SYBR Green qRT-PCR using validated primers on five biological replicates per sample. Relative mRNA levels were calculated using the comparative cycle threshold (C\(_T\)) method \(2^{-\Delta\Delta C_T}\)\(^{304}\). C\(_T\) values for actin were subtracted from C\(_T\) values of the target gene
to calculate $\Delta C_T$. Relative mRNA levels ($\Delta\Delta C_T$) were then calculated from $2^{-\Delta C_T}$ for each sample, divided by the average of the control (primary tumor) sample.

### 4.4 Discussion

Here we used RNA-seq data to identify candidate genes required during the progression from small cell aggregates to actively proliferating brain metastases, and developed an integrative high-throughput approach to functionally test the genes’ involvement in metastatic progression. Traditional techniques, such as the injection of human cancer cells expressing shRNA into mice, are both expensive and time-consuming. To overcome these limitations, we used an established *Drosophila* tumor model amenable to HTS. Our methodology led to the identification of 51 unique human genes necessary for tumor growth in our model, 14 of which resulted in dramatically smaller tumors. Of these genes, increased expression in brain metastases has been validated for two genes, Rab11b and SERPINF1.

Rab11b is a member of the Ras superfamily of small GTP-binding proteins. The Rab11 family is involved in membrane trafficking, directing receptors and adhesion proteins to the cell surface$^{305-307}$. There are three Rab11 proteins, Rab11a, with ubiquitous expression; Rab11b, primarily expressed in the brain, testes, and heart; and Rab11c, expressed in the lung, kidney, and gastric tract$^{308-311}$. The most evidence connecting Rab11 with cancer is through Rab11c, also known as Rab25, with evidence supporting its role as both a tumor agonist and antagonist$^{312-314}$. The effect of
deregulating Rab11 will be highly dependent on the cargo that it is directing. A central collaborator in this project, Dr. Erin Howe, is conducting follow-up studies to mechanistically probe the role of Rab11b in brain metastases.

Serpin family F member one (SERPINF1), also known as pigment epithelium-derived factor (PEDF), is a non-inhibitory member of the serine protease inhibitor family that was originally identified from conditioned media to be a secreted inducer of neuronal differentiation in retinoblastoma cells\textsuperscript{315–317}. The fact that SERPINF1 acts as a neurotrophic factor, combined with our finding that its down-regulation results in decreases tumorigenesis, supports the idea that dormant tumor cells undergo reprogramming to become more “brain-like,” enabling them to survive and thrive in the brain and ultimately resulting in aggressive overt metastases. However, SERPINF1 is generally thought to be a tumor suppressing gene that acts as a strong inhibitor of angiogenesis and whose expression has actually been correlated with lower grade, less metastatic phenotypes in liver and prostate cancers as well as glioma; further it has been shown to induce endothelial cell apoptosis via p53\textsuperscript{318–322}. In fact, the work of Fitzgerald \textit{et al.} contradicts our findings and reports down-regulation of PEDF in brain mets compared to the primary tumor and PEDF inhibition of brain metastasis\textsuperscript{323}. However, anti-plasminogen activator members of the serpin family, serpin B2 and neuroserpin, have been implicated to protect brain metastatic cells from lethality induced by stromal secreted plasmin\textsuperscript{324}. 

103
Two more promising hits, due to the consistency of their effect on *Drosophila* tumor phenotypes, are SNRPD2 and MRPL37. SNRPD2, or small nuclear ribonucleoprotein Sm D2, is required for nuclear pre-mRNA splicing and small nuclear ribonucleoprotein biogenesis. SNRPD2 is one of the seven Sm proteins that form the small nuclear ribonucleoprotein particles (snRNPs) that assemble to splice pre-mRNA into mature mRNA. While splicing of pre-mRNA to generate functional transcripts is a well-regulated process, aberrant splicing can result in tumor-specific splice variants. Thus, targeting the spliceosome is an attractive therapeutic target for cancer. Two drugs targeting SF3b to impair splicing have been shown to have antitumor activity both *in vitro* and *in vivo*.

The mitochondrial ribosomal protein MRPL37, along with other mitochondrial proteins and mitochondrially-transcribed rRNA, is a structural component of mitochondrial ribosomes responsible for translation of proteins in the mitochondria that are required for oxidative phosphorylation. While malignant transformation is hallmarked by metabolic reprogramming, it is primarily through increased glycolysis as opposed to oxidative phosphorylation. However, MRPL44 transcription was recently shown to be a predictor of poor prognosis in papillary thyroid carcinoma, specifically for lymph node metastasis, through its correlation with levels of oxidative phosphorylation. Further, inhibition of mitochondrial translation through peptide deformylase (HsPDF) was found to induce apoptosis in hematopoietic cancers and myeloid leukemia. However, there is also recent evidence that the
antiproliferative effects of diminished mitochondrial transcription are not attributable to changes in respiration, but to a second “ribosome quality-control pathway” whose activation impairs cell proliferation\textsuperscript{347}.

4.5 Acknowledgments

We thank the TRiP at Harvard Medical School (NIH/NIGMS R01-GM084947) and the Vienna \textit{Drosophila} Resource Center for providing the transgenic RNAi fly stocks used in this study. We thank the Walther Cancer Foundation Advancing Basic Cancer Research Program and the Mike and Josie Harper Cancer Research Institute for financial support of this research. M. Burnette was supported in part by the American Cancer Society Institutional Research Grant.
5.1 Introduction

Despite clinical advances in early tumor detection and, where applicable, the development of targeted treatment regimens for primary breast tumors, the overall incidence of patients relapsing with brain metastasis is increasing. Forty percent of patients with human epidermal growth factor receptor 2 amplified (HER2+) primary breast tumors who received first line therapy (trastuzumab or Herceptin) ultimately developed brain metastases even when the primary disease was under control. There are several known risk factors for the development of brain or central nervous system (CNS) metastases, including tumor size less than five cm, and less than four positive lymph nodes; however, women at the greatest risk of developing brain metastases are those with HER2+ or triple negative breast cancer (TNBC, no over expression of estrogen receptor, progesterone receptor, or HER2). Current treatment options for patients

5 Collaborative work with Erin N. Howe, Kevin X. Rodriguez, Jennifer L. Meloche, Jayda Meisel, Patricia Skallos, Xeujuan Tan, Mayland Chang, Brandon L. Ashfeld, and Siyuan Zhang
with brain metastases include chemotherapy, radiotherapy, and HER2-targeted therapies where appropriate. Although these treatments initially improve overall survival (chemotherapy vs. no chemotherapy, 16.4 vs. 3.7 months; surgery vs. no surgery, 20.3 vs. 11.3 months)\textsuperscript{113}, there are no effective treatment strategies for patients with refractory disease. Although there is an urgent clinical need, the development of new treatment options for patients with brain metastases has been slow. There are two main barriers in the development of therapies: 1) a limited understanding of the genetics of brain metastases, and how breast cancer cells change over time in the brain; and 2) the protective nature of the blood brain barrier (BBB); over 98% of potential CNS-active anticancer compounds fail during development due to inadequate BBB penetration\textsuperscript{114}. Unfortunately, the current process for designing and testing drugs for brain metastases is slow, and most drugs are discarded due to lack of efficacy or inability to cross the BBB\textsuperscript{114}.

Further, the genetic heterogeneity of cancer hinders the efficacy of therapeutics designed to target specific proteins or signaling pathways. Despite this, the rational architectural design of new small molecules based on a structure-activity correlation with a targeted signaling pathway remains one of the most widely utilized approaches in drug design\textsuperscript{348–351}. The majority of cancer therapeutics are based on natural compound scaffolds, and recent advances in organic synthesis techniques have allowed for the creation of libraries of small molecules based on these compounds\textsuperscript{352–355}. Although the design and synthesis of these libraries can be rapidly accomplished,
mechanism of action (MoA) of these novel compounds in human cells remains a major obstacle in the field of drug discovery\textsuperscript{118,356,357}.

Here we present our approach for characterizing novel compounds synthesized by our collaborators in the Ashfeld lab. They have developed a method for the rapid synthesis of novel bis-spirooxindoles, which exhibit anti-mitotic activity in breast cancer cells and readily cross the barrier created by brain epithelial cells in an \textit{in vitro} BBB model (Figure 5.1)\textsuperscript{115–117}. Using this synthesis route, the Ashfeld lab has synthesized 30 structurally diverse compounds (Table 5.1).

![Figure 5.1](image)

Figure 5.1: General molecular framework of spirooxindoles (A) and novel bis-spirooxindoles (B) with highlighted points of structural diversity (Table 5.1).
TABLE 5.1

STRUCTURES OF NOVEL COMPOUNDS

<table>
<thead>
<tr>
<th>Entry</th>
<th>PG</th>
<th>R₁</th>
<th>R²</th>
<th>Ar</th>
<th>11</th>
<th>NDXXXX</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Me</td>
<td>H</td>
<td>H</td>
<td>Ph</td>
<td>11a</td>
<td>ND1000</td>
</tr>
<tr>
<td>2</td>
<td>Me</td>
<td>H</td>
<td>p-Cl</td>
<td>Ph</td>
<td>11b</td>
<td>ND1002</td>
</tr>
<tr>
<td>3</td>
<td>Me</td>
<td>H</td>
<td>p-CF₃</td>
<td>Ph</td>
<td>11c</td>
<td>ND1003</td>
</tr>
<tr>
<td>4</td>
<td>Me</td>
<td>H</td>
<td>p-OMe</td>
<td>Ph</td>
<td>11d</td>
<td>ND4008</td>
</tr>
<tr>
<td>5</td>
<td>Me</td>
<td>H</td>
<td>o-Br</td>
<td>Ph</td>
<td>11e</td>
<td>ND1007</td>
</tr>
<tr>
<td>6</td>
<td>Bn</td>
<td>H</td>
<td>o-Br</td>
<td>Ph</td>
<td>11f</td>
<td>ND1011</td>
</tr>
<tr>
<td>7</td>
<td>Me</td>
<td>Me</td>
<td>o-Br</td>
<td>Ph</td>
<td>11g</td>
<td>ND4026</td>
</tr>
<tr>
<td>8</td>
<td>Me</td>
<td>Br</td>
<td>o-Br</td>
<td>Ph</td>
<td>11h</td>
<td>ND4027</td>
</tr>
<tr>
<td>9</td>
<td>Me</td>
<td>OMe</td>
<td>o-Br</td>
<td>Ph</td>
<td>11i</td>
<td>ND4032</td>
</tr>
<tr>
<td>10</td>
<td>Me</td>
<td>H</td>
<td>o-CF₃</td>
<td>Ph</td>
<td>11j</td>
<td>ND1012</td>
</tr>
<tr>
<td>11</td>
<td>Me</td>
<td>H</td>
<td>o-Me</td>
<td>Ph</td>
<td>11k</td>
<td>ND1013</td>
</tr>
<tr>
<td>12</td>
<td>Me</td>
<td>H</td>
<td>o-Br</td>
<td>p-Cl</td>
<td>11l</td>
<td>ND1014</td>
</tr>
<tr>
<td>13</td>
<td>Me</td>
<td>H</td>
<td>o-Br</td>
<td>p-OMe</td>
<td>11m</td>
<td>ND1015</td>
</tr>
<tr>
<td>14</td>
<td>Boc</td>
<td>H</td>
<td>o-Br</td>
<td>Ph</td>
<td>11n</td>
<td>ND1016</td>
</tr>
<tr>
<td>15</td>
<td>Ts</td>
<td>H</td>
<td>o-Br</td>
<td>Ph</td>
<td>11o</td>
<td>ND4036</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Entry</th>
<th>R₁</th>
<th>R²</th>
<th>R³</th>
<th>12</th>
<th>NDXXXX</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ts</td>
<td>H</td>
<td>o-Br</td>
<td>12a</td>
<td>ND4012</td>
</tr>
<tr>
<td>2</td>
<td>Me</td>
<td>H</td>
<td>p-Cl</td>
<td>12b</td>
<td>ND4028</td>
</tr>
<tr>
<td>3</td>
<td>Me</td>
<td>Br</td>
<td>p-Cl</td>
<td>12c</td>
<td>ND3004</td>
</tr>
<tr>
<td>4</td>
<td>Boc</td>
<td>H</td>
<td>p-Cl</td>
<td>12d</td>
<td>ND4033</td>
</tr>
<tr>
<td>5</td>
<td>Me</td>
<td>Br</td>
<td>o-Br</td>
<td>12e</td>
<td>ND4038</td>
</tr>
<tr>
<td>6</td>
<td>Me</td>
<td>OMe</td>
<td>p-F</td>
<td>12f</td>
<td>ND4039</td>
</tr>
<tr>
<td>7</td>
<td>Boc</td>
<td>H</td>
<td>p-F</td>
<td>12g</td>
<td>ND4040</td>
</tr>
<tr>
<td>8</td>
<td>Me</td>
<td>Me</td>
<td>p-Cl</td>
<td>12h</td>
<td>ND3005</td>
</tr>
<tr>
<td>9</td>
<td>Me</td>
<td>OMe</td>
<td>p-Cl</td>
<td>12i</td>
<td>ND3006</td>
</tr>
</tbody>
</table>
Here we present our method to evaluate efficacy of these spirooxindoles and rapidly identify their MoA using a transcriptional approach. We demonstrate that a novel bis-spirooxindole effectively kills TNBC cells through inhibition of protein translation by targeting the ribosome. This compound holds promise for a sub-type of breast cancer that has no available targeted therapeutics. We have also laid the groundwork for a high-content screening strategy to facilitate prediction of the structural and synthesis compound properties conducive for compound efficacy.

5.2 Results

5.2.1 Assessing half-maximal effective concentration (EC$_{50}$)

Compound efficacy was tested against a highly aggressive TNBC cell line, MDA-231, as well as the isogenic brain-trophic cell line, MDA-231-Br. Cells were treated for four days with stock plates containing randomly arrayed compounds at concentrations spanning from 100 $\mu$M to 10 pM (log scale), then fixed and stained using a sulforhodamine B assay. EC$_{50}$ values were calculated from a non-linear four-parameter curve fit to measured absorbances.
**TABLE 5.2**

**COMPOUND EC50 VALUES**

<table>
<thead>
<tr>
<th>Entry</th>
<th>MDA-231 (µM)</th>
<th>MDA-231-Br (µM)</th>
<th>Entry</th>
<th>MDA-231 (µM)</th>
<th>MDA-231-Br (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11a</td>
<td>52.79</td>
<td>30.24</td>
<td>12a</td>
<td>15.43</td>
<td>82.02</td>
</tr>
<tr>
<td>11b</td>
<td>25.63</td>
<td>25.14</td>
<td>12b</td>
<td>19.17</td>
<td>15.43</td>
</tr>
<tr>
<td>11c</td>
<td>19.15</td>
<td>17.74</td>
<td>12c</td>
<td>61.68</td>
<td>171.9</td>
</tr>
<tr>
<td>11d</td>
<td>11.09</td>
<td>24.4</td>
<td>12d</td>
<td>112.8</td>
<td>179.2</td>
</tr>
<tr>
<td>11e</td>
<td>91.73</td>
<td>32.15</td>
<td>12e</td>
<td>27.89</td>
<td>56.36</td>
</tr>
<tr>
<td>11f</td>
<td>145</td>
<td>129</td>
<td>12f</td>
<td>33.95</td>
<td>66.1</td>
</tr>
<tr>
<td>11g</td>
<td>34.81</td>
<td>27.26</td>
<td>12g</td>
<td>20.91</td>
<td>30.4</td>
</tr>
<tr>
<td>11h</td>
<td>33.29</td>
<td>19.72</td>
<td>12h</td>
<td>66.78</td>
<td>30.04</td>
</tr>
<tr>
<td>11i</td>
<td>20</td>
<td>31.96</td>
<td>12i</td>
<td>18.18</td>
<td>15.77</td>
</tr>
<tr>
<td>11j</td>
<td>12.51</td>
<td>23.8</td>
<td>13a*</td>
<td>3.823</td>
<td>17.72</td>
</tr>
<tr>
<td>11k</td>
<td>18.99</td>
<td>31.55</td>
<td>13b*</td>
<td>22.62</td>
<td>12.46</td>
</tr>
<tr>
<td>11l</td>
<td>61.62</td>
<td>117.3</td>
<td>13a*</td>
<td>3.823</td>
<td>17.72</td>
</tr>
<tr>
<td>11m</td>
<td>6.517</td>
<td>40.75</td>
<td>13b*</td>
<td>22.62</td>
<td>12.46</td>
</tr>
<tr>
<td>11n</td>
<td>8.677</td>
<td>157.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11o</td>
<td>544.7</td>
<td>170.2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*In an effort to increase water solubility, 12d and 12g were converted to their free N-H counterparts to yield compounds 13a and 13b respectively.

While compounds exhibit a diverse range of efficacy, 20 and 15 compounds were found to have EC50 values less than 50 µM for MDA-231 and MDA-231-Br cells, respectively (Table 5.2). Three compounds had EC50 values less than 10 µM in MDA-231 cells. Interestingly, some compounds did exhibit selectivity for one cell line over the
other. While none of the compounds were as effective as the DNA-intercalating chemotherapeutic doxorubicin, many were more effective than temozolamide, a semi-BBB-permeable therapeutic currently in clinical trials for treating breast cancer brain metastases. 13b was the most effective cytotoxic compound in MDA-231-Br cells, with an EC₅₀ of 12.46 μM (Figure 5.2). EC₅₀ curves for the two most effective compounds against MDA-231 (Figure 5.2 A-B) and MDA-231-Br (Figure 5.2 C-D) can be found in Figure 5.2.
Figure 5.2: EC$_{50}$ curves for top performing compounds. The top two most efficacious compounds against MDA-231 cells—13b (A) and 12b (C)—demonstrated comparable efficacy to the tyrosine kinase inhibitor, lapatanib. MDA-231-Br cells are, in general, more resistant to our compounds than their parental line, but 13b (B) and 12i (D) were more effective than temozolamide.
5.2.2 Detecting mechanism of action by network dysregulation (DeMAND)

Identifying the MoA for a novel compound is critical for determining both the on- and off-target effects of the compound. On-target effects determine the efficacy of a compound, as well as give insight into appropriate patient populations, while off-target effects determine potential toxicity. Efficacy and toxicity are the two main challenges of drug development. Transcriptomics analyses, such as microarray or RNA-seq, are the most time- and cost-effective ways to access the effect of a compound. However, a compound’s target may not be differentially regulated at the transcriptional level; for example, the drug paclitaxel functions by stabilizing microtubules. To begin elucidating the MoA of our compounds, we utilized a recently developed hybrid experimental and computational method, detecting mechanism of action by network dysregulation (DeMAND). DeMAND takes as input both expression data and a network of known interactions for each gene, known as that gene’s regulon. Each regulon is composed of connected genes, or edges. In Figure 5.3 A, although G₀ expression is not affected by treatment, the regulon components G₂, G₅, and G₇ are down-regulated, while G₃ and G₄ are up-regulated. While differential expression analysis would not identify G₀ as a target, given the degree of dysregulation of the G₀ regulon, DeMAND would identify G₀ as a target.
Figure 5.3: Experimental basis for DeMAND\textsuperscript{118}. (A) DeMAND determines targeted genes by network analysis. In this example, while $G_0$ expression is not affected by treatment, the regulon components $G_2$, $G_5$, and $G_7$ are down-regulated, while $G_3$ and $G_4$ are up-regulated. Although differential expression analysis would not identify $G_0$ as a potential target, given the degree of dysregulation of the $G_0$ regulon, DeMAND identifies $G_0$ as a target. (B) Experimental pipeline. (C) Cytoscape visualization of targets of 13b implicated from DeMAND with FDR q-value < 0.001.

13b is the most effective cytotoxic compound in MDA-231-Br cells, with an EC\textsubscript{50} of 12.46 $\mu$M. To predict the MoA for this compound, we treated MDA-231-Br cells at approximately five times the EC\textsubscript{25} of 62.30 $\mu$M and harvested RNA 6, 12, and 24 hours post-treatment (Figure 5.3 B). To capture the state of the transcriptome with a high level of fidelity, we performed RNA-seq on three replicates for each timepoint, and a DMSO vehicle control. Sequencing was performed on the MiSeq (Illumina), with paired end 150 basepair reads. Sample level normalization was performed using cuffnorm prior
to DeMAND analysis. DeMAND identified 3,400 genes as potential mediators of 13b activity with a predicted MoA false discovery rate (FDR) less than 0.001, 1,706 and 1,694 of which had an average decrease or increase in expression, respectively (Figure 5.4 A).

The ideal timepoint to identify the MoA for a compound depends on how quickly the effector genes are regulated by the compound. As 13b is a novel compound, we examined a range of timepoints (Figure 5.3 B) to determine the appropriate timepoint for analysis. The genes that are most differentially regulated at 24 hours begin changing as early as six hours, as the highly-expressed blue genes and lowly-expressed yellow genes both move toward gray (Figure 5.4 A). Although the differential gene expression is most dramatic within 24 hours of treatment (Figure 5.4 A), we found that the genes with the most dysregulated regulons, as identified by DeMAND were very similar for all individual timepoints, or all timepoints considered together (Figure 5.4 B). DeMAND outputs a ranked list of genes, with the most highly-ranked gene being the gene most significantly predicted to be an effector. For the top 30 genes identified by DeMAND as putative mediators of 13b MoA, the rank of each gene does not change significantly whether all timepoints are considered together, or each timepoint is considered separately (Figure 5.4). This suggests that 13b begins dysregulating the networks of its effector genes quickly. To simplify further analysis, we focused on all timepoints combined.
Figure 5.4: DeMAND\textsuperscript{118}. (A) Normalized expression of top 100 DeMAND implicated genes at 6, 12, and 24 hours post-treatment. (B) Rank in DeMAND output of top 30 DeMAND implicated genes using combined versus individual timepoint expression.
To further explore the common function of DeMAND-predicted effector genes, and thus the MoA of 13b, we performed Gene Set Enrichment Analysis (GSEA)\textsuperscript{366,367}. Genes were ranked by their inverse p-value from the timepoint-combined DeMAND results multiplied by the sign of the change in expression. Thus, genes were weighted by their DeMAND-predicted significance and assigned directionality by their average change in direction. GSEA identified 236 negatively-enriched gene sets and 130 positively-enriched gene sets with FDR q-value < 0.001 (Table A.1, Table A.2). The 15 most significant positively-regulated gene sets are shown in blue, while the 15 most negatively-regulated gene sets are shown in yellow (Figure 5.5 A). From these gene enrichment sets, we identified the KEGG Ribosome pathway as having the highest enrichment score (Figure 5.5 A). Indeed, the top 30 DeMAND-predicted genes encode ribosomal proteins (Figure 5.4 B), and 75% of the top 100 genes encode ribosomal proteins (Table A.3). Further supporting the common functional role of these genes, there is considerable overlap between 14 of the 15 most negatively-enriched gene sets (Figure 5.5 B), with 15 ribosomal proteins appearing in 14 of the gene sets. This suggests that 13b negatively-regulates ribosomal function.
Figure 5.5: DeMAND implicates ribosomes as target of 13b. (A) Top 15 most significant negatively-(yellow) and positively-(blue) regulated gene sets from GSEA. (B) Overlap between genes in the top 15 most negatively-enriched gene sets. (C) Relative rank of predicted effector genes across custom classes of common MoAs. Genes predicted by DeMAND to most significantly contribute to 13b MoA cluster strongly in the ribosome regulation class.
To clarify the MoA of 13b, we created custom gene sets that represent the broad classes of therapeutic MoA—apoptosis or cell cycle regulation, regulation of the ribosome, disruption of microtubule function, DNA damage, kinase activation, G-protein coupled receptor (GPCR) signaling, and ion channel signaling\textsuperscript{368,369}. We analyzed the relative rank of predicted effector genes across these MoA classes (Figure 5.5 C). We found that the genes predicted by DeMAND to most significantly contribute to 13b MoA cluster strongly in the ribosome regulation class. None of the other common drug MoA classes contain genes that ranked in the top 100 predicted 13b effector genes (Figure 5.5 C). Taken together, this analysis strongly suggests that 13b negatively effects ribosomal proteins to down-regulate ribosomal function to exert its cytotoxic effect. Compound 13b begins to exert this effect at six hours post-treatment, suggesting rapid dysregulation of ribosomal function.

The ribosome is a complex consisting of ribosomal RNA (rRNA) and ribosomal proteins, and is responsible for translating messenger RNA (mRNA) into protein. Translation of mRNA into protein is the most energy-consumptive process in a cell, and inhibition of this process is under investigation as a cancer therapeutic strategy\textsuperscript{370,371}. To validate the effect of 13b on ribosomal function, we examined protein translation. MDA-231-Br cells were treated with DMSO, doxorubicin (dox), paclitaxel (taxol), or 13b at the previously determined EC\textsubscript{25} concentrations (Table 5.2). Dox is a DNA-intercalating agent, while taxol stabilizes microtubules\textsuperscript{364}. Although both compounds are toxic, neither directly interferes with protein synthesis. Following 15 hours of treatment, total cellular
protein levels were examined (Figure 5.6 A). We found that 13b significantly decreased the total amount of protein, while neither dox nor taxol exhibited this effect. The actin cytoskeleton was stained with phalloidin to visualize cellular morphology (Figure 5.6 B). None of the treatment groups caused a dramatic morphological change consistent with apoptosis, indicating that the 15 hour timepoint is early enough to observe MoA-induced cellular changes. Cells treated with taxol exhibited cellular rounding, consistent with the known MoA. Taken together, this strongly supports our bioinformatically-predicted MoA for compound 13b, and suggests that 13b functions by inhibiting protein translation.
Figure 5.6: 13b inhibits protein translation. (A) Total protein levels following 15 hours of treatment with DMSO (negative), dox, taxol, or 13b. Compound 13b significantly decreases total protein levels whereas dox (DNA intercalation) and taxol (microtubule stabilization) do not. (B) Phalloidin staining to visualize acting cytoskeleton following 15 hours of treatment with DMSO (negative), dox, taxol, or 13b. No treatments caused dramatic morphological changes indicative of apoptosis, indicating that 15 hours is early enough to observe MoA and not cell-death-induced changes. Taken together, this strongly supports the predicted MoA that 13b functions by inhibiting protein translation.
5.2.3 Analyzing structure-activity relationships by uHCS

With our collaborators in the Ashfeld lab having developed a route to rapidly synthesize novel, structurally diverse bis-spirooxindoles, we aimed to build on this work to analyze structure-activity relationships between drugs based on the differences in their induced cell responses. A central paradigm in drug discovery relies on functional drug HTS using a biological assay for the given process of interest, for example a biosensors readout for a genetic biomarker\(^6,372,373\). Unfortunately, this target-based drug discovery is a highly specific approach with limited utility for elucidating putative applications for novel drugs, yielding only a yes/no answer for a given biomarker, and thus necessitating subsequent screens to elucidate drug functionality or alternative applications for drug repurposing\(^46\). An alternative approach that is becoming increasingly utilized is HCS, which relies on image-based cell morphological data, providing an extremely rich, high-dimensional data source describing cell responses to pharmacological perturbations\(^43,46,374,375\). In a recent proof-of-principle paper, Di et al. coined the term uHCS to describe their approach of both extracting and using more than 500 cell-morphological parameters to describe cell response paradigms to drug perturbations\(^47\). The basic principle is extraction of multiparametric cell and neighborhood level statistical data from images of cell morphology for use in machine learning algorithms to separate drug-induced cell response paradigms into a \(n\)-dimensional space, where drugs with similar MoA can be grouped based on similar
dose-response behavior\textsuperscript{47}. Here we aimed to build off of this approach to use uHCS for structure-activity analysis of our novel structurally diverse compounds.

The first step and primary bottleneck to developing uHCS methodology is optimization of image collection and cell segmentation. As a proof-of-principle, we used rudimentary high-content analysis to distinguish between two cell lines based on principal component analysis (PCA) of differences in nuclear morphology (Figure 5.7 A-B). We used two brain metastatic breast cancer cell lines, MDA-231-Br and BT474-Br2. The BT474-Br2 cell line basally overexpresses HER2 and was selected for reliable brain metastasis through \textit{in vivo} passaging\textsuperscript{376}. Cells were seeded and stained with Hoescht to visualize nuclei (Figure 5.7 A, left). Nuclei were segmented using CellProfiler with the IdentifyPrimaryObjects module\textsuperscript{377,378}. Each nucleus is presented as a separate color in the representative images shown (Figure 5.7 A, right). Custom MATLAB scripts were used to identify 44 cell-level features, including area, eccentricity, and surface-area-to-volume ratio. PCA was used to condense cell morphological features to the two primary components (Figure 5.7 B). The two cell types can be clearly differentiated in that principle component space (Figure 5.7 B).
Figure 5.7: HCS approach. (A) Representative images of MDA-231-Br and BT474-Br2 and segmented nuclei. Fluorescent images represent maximum z-projections. (B) PCA on summary statistics of the nuclear segmentation revealed a clear separation between MDA-231-Br and BT474-Br2 cell lines along the first two principle components (representing 80.1% of the data variability). Scale bars represent 100 μm. (C) General approach: drug-induced cell characteristics will be quantified using a range of features from image analysis of segmented cell nuclear areas. Briefly, drugs are screened in dose-response on MDA-231-Br cells, which are then fixed, stained, and imaged. A custom image processing pipeline is employed to segment cells and extract statistical features describing the images. Machine learning is used to reduce data dimensionality to visualize dose-response trajectories in n-dimensional space (first three components shown here). These response modalities can then be correlated based on their corresponding drug’s structural or synthetic properties, enabling us to generate a “map” of generalized cell responses that describe specific drug features.
Synthetic Spiroxindole Library Drugs
Brain Trophic Breast Cancer Cell Line MDA-231-Br
Dose-response information
Segmentation
Cell & Neighborhood Response Statistics
Dose-Response Phenotype Trajectories
Feedback into evaluation and synthesis
Having implemented rudimentary high-content analysis to differentiate cell types, we aimed to analyze structure-activity relationships by characterizing the differential cell changes induced by our structurally diverse novel compounds. To do this, we developed a simplified version of the Di et al. method to succinctly describe cell responses induced by novel drugs, enabling us to generate a “map” of generalized cell responses describing specific drug features (Figure 5.7 C). We applied this method to perform an uHCS of our novel compounds versus a panel of control drugs: dox, taxol, temozolamide, lapatanib, and nocodazole. Eight wells of the drug carrier alone (DMSO) was used as a negative control. Cells were seeded at 10,000 cells/well and treated with the same randomized drug plates used for EC\textsubscript{50} determination. After 18 hours of treatment, cells were fixed and stained with CellTracker and DAPI for cytoplasm and nuclei visualization, respectively (Figure 5.8 A). A custom MATLAB script was used to segment cell and nuclear areas, extract 96 features describing segmented regions, and calculate summary statistics on an image basis. These summary statistics were then mean-centered and standard deviation normalized (based on the negative controls) before PCA was used to visualize drug-induced dose-response trajectories in cell shape changes (Figure 5.8 B). Dose-response trajectories were described quantitatively using a second order polynomial fit for each drug. Finally, the coefficient of determination, as described by Di et al., was used to quantify how well a trajectory for a given compound predicts cell responses induced by another compound\textsuperscript{47}. Drugs were then clustered
based on their coefficient of determination in order to group drugs based on similarity in their induced cell change dose-response trajectories (Figure 5.8 C).
Figure 5.8: (A) Cell segmentation. Representative images and segmentation for DMSO control cells versus dox, taxol, and 13b treated cells (10 µM). Scale bar represents 50 µm. (B) Differential dose-response trajectories of induced cells' morphological responses in principle component space. (C) Drug clustering based on induced dose-response trajectories, as measured by complete linkage of the coefficient of determination\(^1\), which quantifies how well a trajectory for a given compound predicts cell responses induced by another compound. Drug labels are colored by the drug’s EC\(_{50}\), green to red being low to high.
A

Segmented

Cytoplasm

Nuclei

CellTracker/DAPI

10 µM

Doxorubicin

Paclitaxel

ND4041

B

Size by dose (µM):

0.00001 0.1 100

DMSO

Doxorubicin

Paclitaxel

Nocodazole

ND4041

ND1008

C

Complete Linkage Coefficient of Determination (1-R²)
Structurally and mechanistically different drugs are expected to induce distinct dose-response trajectories. Indeed, some induced trajectories for mechanistically-distinct control drugs—the microtubule stabilizer taxol versus the DNA-intercalating agent dox—are very different (Figure 5.8 B). However, others are similar despite being dissimilar drugs—the microtubule inhibitor nocodazole and the DNA-intercalating agent dox. Further, the negative control trajectory of DMSO, which was not in fact tested in dose-response but at a constant 1% and thus simply represents variability, suggests that the pipeline is not robust against noise and requires further development. Hierarchical clustering of compounds based on their induced trajectories succinctly describes those differences visualized in principle component space (Figure 5.8 C), and it becomes clear that similar target drugs do not cluster together, nor do drugs with similar efficacy, providing further evidence that the pipeline is not yet adequate for analyzing structure-activity relationships. However, this analysis was based on only one image per condition and a single replicate experiment. Analysis of additional replicates to test conditions in quadruplicate would likely strongly improve assay quality, based on the results of Di et al. Further, subdivision of cells into subpopulations (for large cell aggregates versus single cells) could provide additional statistical power.
5.3 Methods

5.3.1 Cell culture

Human breast cancer cell lines MDA-231 and MDA-231-Br were cultured in DME/F12 with 10% fetal bovine serum (FBS), 2.5 mM L-Glutamine, 15 mM HEPES and penicillin (100 unit/mL)/streptomycin (100 µg/mL). All cell lines were maintained and grown in a 37°C incubator with 5% CO₂.

5.3.2 Randomized compound stock plates

To generate a randomized plate for compound screening, all compounds were first solubilized in DMSO at 10 mM. 10 mM compounds were serially diluted 1:10 into DMSO to span a concentration range from 10 mM to 1 nM. An Eppendorf EpMotion 5075 robotic pipettor followed a randomization program to dilute compounds 1:10 into DME/F12 medium without FBS. The randomized plate contained compounds in 10% DMSO spanning a concentration range from 1 mM to 100 pM.

5.3.3 Determination of EC₅₀

Cells were plated in 96-well tissue culture plates at a concentration of 1,000 cells per well in 90 µL of medium, and given overnight to adhere. 10 µL of compound was pipetted from the randomized plates to the cell plates, yielding a working solution of 1% DMSO, with compound concentrations ranging from 100 µM to 10 pM. Plates were incubated for four days, then fixed and stained using a sulforhodamine B assay. Briefly, medium was removed and cells were fixed in 10% trichloracetic acid (Sigma-
Aldrich) for one hour. Plates were rinsed repeatedly in deionized water and dried prior to staining with a solution of 0.4% sulforhodamine B (Sigma-Aldrich) in 1% acetic acid (Sigma-Aldrich). Following one hour of staining, plates were rinsed five times in 1% acetic acid and dried. 100 µL of 10 mM tris base was used to solubilize each well. Plates were incubated on an orbital shaker for 10 to 20 minutes to ensure full solubilization, and absorption at 554 nm was read on a BioTek Synergy H1 plate reader. Deconvolution of the randomization was performed in MATLAB (MathWorks), and EC50 values were calculated by performing a non-linear four-parameter curve fit in Prism 6 (GraphPad Software). Three technical replicates were performed for each experiment, and all experiments were repeated three times.

5.3.4 Prediction of mechanism of action (DeMAND)

MDA-231-Br cells were plated in 6-well plates such that they would be 70% confluent and allowed to adhere overnight. Cells were treated with 13b at the EC25 value, 62.5 µM. Cells were collected in triplicate at 6, 12, and 24 hours post-treatment, and RNA was extracted with the PureLink kit (Invitrogen). Libraries for RNA-seq were prepared according to the SmartSeq2 protocol379. RNA-seq was performed on the MiSeq (Illumina), with paired end 150 basepair reads. Following sequencing, transcripts were aligned and normalized using cuffnorm96,365. Sample level normalization was performed using cuffnorm prior to DeMAND analysis96,365.
R was used for heatmap visualizations and the Cleveland plot. DeMAND was run using the breast cancer regulatory network, BRCA-MCF7, and cuffnorm normalized sequencing data\textsuperscript{118,380}. For genes with 0 fpkm reads across all samples, 0.001 was added to the first replicate of each timepoint sample. To further explore the common function of DeMAND-predicted effector genes, and thus the MoA of 13b, we performed GSEA\textsuperscript{366,367}. GSEA was conducted using a custom ranked list with genes ranked by their inverse p-value from the timepoint-combined DeMAND results multiplied by the sign of their timepoint-combined change in expression. Cytoscape was used for network visualizations\textsuperscript{381,382}. For visualization of GSEA in the Cleveland plot, gene set names were truncated as follows: transferase activity transferring phosphorus containing groups—transferase activity; phosphotransferase activity alcohol group as acceptor—phosphotransferase activity; reactome formation of the ternary complex and subsequently the 43s complex—formation of the ternary complex; reactome activation of the mRNA upon binding of the cap binding complex and EIFs and subsequent binding to 43s—reactome activation of the mRNA; reactome SRP dependent cotranslational protein targeting to membrane—protein targeting to membrane; reactome nonsense mediated decay enhanced by the exon junction complex—nonsense mediated decay; reactome three UTR mediated translational regulation—three UTR translational regulation; reactome influenza viral RNA transcription and replication—influenza RNA transcription; post translational protein modification—post translational modification; and protein serine threonine kinase activity—serine threonine kinase activity.
For analysis of broad drug MoA classes (Figure 5.5 C), multiple gene annotation sets were combined: Apoptosis (GO biological process—Regulation of apoptosis, GO biological process—Apoptotic program); Cell cycle (GO biological process—Regulation of the cell cycle, Biocarta—Cell cycle pathway); Ribosome (KEGG—Ribosome, GO biological process—Ribosome biogenesis and assembly); Microtubule disruption (GO biological processes—Microtubule polymerization or depolymerization, Reactome—Formation of tubulin folding intermediates, GO biological processes—Microtubule cytoskeleton organization and biogenesis); DNA damage (KEGG mismatch repair, KEGG non-homologous end joining, KEGG nucleotide excision repair); Kinase activation (GO biological processes—Activation of protein kinase activity, GO molecular function—Kinase regulator activity); GPCR activation (Biocarta—GPCR pathway, Reactome—Signaling by GPCR); and Ion channel signaling (GO molecular function—Extracellular ligand gated ion channel activity, GO molecular function—Ion channel activity, GO molecular function—Ligand gated channel activity. The genes predicted by DeMAND to be affected by 13b treatment were ranked, with one being the gene whose regulon was most significantly altered by treatment. The rank of each gene in the combined gene sets was plotted.

5.3.5 Protein assay

MDA-231-Br cells were plated in 96-well tissue culture plates at a concentration of 10,000 cells/well in 90 µL of medium, and given overnight to adhere. Treatments
were prepared in medium at 10x EC\textsubscript{25} concentrations and 10 µL was pipetted to the cell plates, yielding a working solution of 1% DMSO, with EC\textsubscript{25} concentrations—dox (1.14 µM), taxol (0.05 µM), 13b (62.3 µM). Cells were lysed in RIPA lysis buffer (ThermoFisher, 89900), with the addition of 2% SDS, and protease and phosphatase inhibitor (Life Technologies, 88668) according to the manufacturer’s instructions. Lysate was assayed for protein concentration using the Pierce BCA protein assay kit (ThermoFisher, 23225) according to the manufacturer’s instructions.

5.3.6 Phalloidin staining

At the protein assay endpoint, cells were stained with Alexa Fluor 594 Phalloidin (ThermoFisher, A12381) according to the manufacturer’s instructions. Briefly, cells were washed twice with PBS, fixed in 4% PFA for ten minutes, washed twice with PBS, and permeabilized with 0.1% Triton-X for five minutes. Cells were washed twice with PBS, blocked with 1% BSA in TBST for five minutes, and stained with phalloidin in 1% BSA in TBST for 20 minutes. Cells were counterstained with DAPI to show nuclei, washed twice with PBS and imaged in PBS. Imaging was performed on an EVOS FL system (Life Technologies) with a 20x objective.

5.3.7 Classification of MDA-231-Br and BT-474-Br2 cell lines

Cells were plated on a glass bottom 96-well tissue culture plate at a concentration of 5,000 cells per well. Cells were given 24 hours to adhere to the plate. Following confirmation of cellular adhesion, cells were treated with 1 µg/mL Hoechst
stain and incubated for 15 minutes. Cells were then placed in an environmental chamber (37°C, 5% CO₂) and imaging was performed on a Nikon Eclipse Ti confocal microscope with a Yokogawa spinning disc. Image data were collected on an iXonEM+ cooled CCD camera (Andor Technology) using MetaMorph® software (Molecular Devices). The center of each well was imaged in a z-stack (54 slices with 1.5 μm spacing) at 10x magnification.

To segment nuclei, images were loaded into CellProfiler and the IdentifyPrimaryObjects module was used. After generating a maximum intensity projection from the z-stack data in FIJI, automatic thresholding was used and nuclei were defined as objects with a diameter between 2 and 40 pixels. Each nucleus is presented as a separate color in the representative images shown (Figure 5.7, left). Custom MATLAB scripts were then used to identify 44 cell-level features, which were then mean-centered and standard deviation normalized before PCA was conducted in MATLAB. For each cell line, one image was analyzed.

5.3.8 Ultra-high-content screen

Cells were plated in 96-well tissue culture plates at a concentration of 10,000 cells/well in 90 μL of medium, and given overnight to adhere. 10 μL of compound was pipetted from the randomized plates to the cell plates, yielding a working solution of 1% DMSO, with compound concentrations ranging from 100 μM to 10 pM. Plates were incubated for roughly 12 hours before being fixed in 4% PFA, rinsed with PBS, and
stained with 1 µM CellTracker® Deep Red (ThermoFisher C34565) and DAPI for 15 minutes. Imaging was performed on a Nikon Eclipse Ti confocal microscope with a Yokogawa spinning disc. Image data were collected on an iXonEM+ cooled CCD camera (Andor Technology) using MetaMorph® software (Molecular Devices). The center of each well was imaged using automated screen acquisition with autofocus at 10x magnification. Custom MATLAB (MathWorks) scripts were used for cell and nuclear segmentation during all subsequent analysis (electronic supplementary information). After cell and nuclear segmentation, 96 image-based features were extracted from images to describe the drug-induced cell responses in each well. Features were subsequently mean-centered and standard deviation normalized to the negative controls on a plate-wise basis before PCA was conducted in MATLAB. Drug-induced dose-response trajectories were quantitatively described by fitting second order polynomials to describe the second principle component score as a function of the first principle component score for each drug. As in Di et al., the coefficient of determination was used to quantify the differences in compounds-induced dose-response trajectories. Hierarchical clustering was then conducted in using the complete linkage of the coefficient of determination as a distance metric to visualize drugs inducing similar cell responses. In this proof-of-principle screen, we tested over 20 novel compounds synthesized by the Ashfeld lab along with control drugs with known targets: dox, temozolamide, nocodazole, and taxol. DMSO was used as a negative control.
5.4 Discussion

Our inter-disciplinary approach for the rapid evaluation and MoA identification of novel bis-spirooxindoles led to the identification of a bis-spirooxindole that effectively kills TNBC cells, holding promise as a treatment for a sub-type of breast cancer that has no available targeted therapeutics. Interestingly, 13b appears to affect its cytotoxicity by inhibiting the ribosome and thus protein translation. Ribosome biogenesis is well-established as correlating with rates of cell proliferation and growth, and many believe that cancer cells are indeed “addicted to increases in ribosome biogenesis and number”\textsuperscript{383–389}. Despite this, inhibition of ribosome biogenesis is a relatively new therapeutic target, with the only two ribosome biogenesis inhibiting drugs just recently being developed\textsuperscript{390–393}. CX-5461, which inhibits rRNA synthesis through selective inhibition of RNA polymerase I, entered phase I clinical trials for hematological malignancies less than five years after its discovery, and patients are currently being recruited for a phase I/II trial to test its efficacy against TNBC\textsuperscript{394}. 13b is thus particularly promising due to its demonstrated efficacy against brain trophic breast cancer cells and its relative novelty as one of only a few drugs with a new and promising MoA.

Our implementation of a HCS strategy to facilitate prediction of compound properties conducive for cytotoxic efficacy is another promising avenue upon which we can build. Unlike a standard high-throughput cell culture based assay, ultra high-content image analysis provides a large quantity of information (big data) to describe drug-induced responses. Future work will build upon this foundation to relate ultra high-
content cellular readouts to drug structure, by quantitatively describing drug structures using established methods and then incorporating those descriptions as features for machine learning\textsuperscript{395–397}. This approach will ultimately enable us to incorporate synthesis information about novel compounds into our pipeline to identify synthesis routes that impact drug efficacy. Such a method of iterating between the synthesis of chemical compounds with desired characteristics such as BBB-penetration or cytotoxic efficiency and uHCS will provide a mechanism for more efficiently searching the large chemical parameter space for efficacious drugs to treat breast cancer brain metastases.

5.5 Acknowledgments

We thank TRiP at Harvard Medical School (NIH/NIGMS R01-GM084947) and the Vienna Drosophila Resource Center for providing transgenic RNAi fly stocks used in this study. We thank the Walther Cancer Foundation Advancing Basic Cancer Research Program and the Mike and Josie Harper Cancer Research Institute for financial support of this research. M. Burnette was supported by the American Cancer Society Institutional Research Grant.
CHAPTER 6:
CONCLUSIONS AND DISCUSSION

6.1 Summary of findings

We have integrated a variety of advanced technologies, including RNA-seq and HTS, to generate a CDM for insect cell culture (biotechnology, Chapter 2), identify a novel role for metabolites in development (developmental biology, Chapter 3), and characterize potential therapeutics and targets for breast cancer brain metastasis (disease, Chapters 4 and 5).

To characterize the minimal requirements for long-term maintenance of Drosophila cell lines, we developed an inverse screening strategy to identify small molecules and synergies stimulating proliferation in a CDM. In this chemical-genetics approach, a compound-protein interaction database is used to systematically score genetic targets on a screen-wide scale to extract further information about cell growth. This approach is a general strategy that can be applied to a wide-range of bioprocess optimization problems, including the development of novel cell culture media as well as drug cocktails in precision medicine applications.
Building on findings generated during our design of the CDM, we integrate previously published whole genome expression and PA metabolic data across *Drosophila* organismal development to show developmental-stage-specific regulation of the PA metabolic pathway\(^\text{112}\) (Chapter 3). Combining this transcriptomic analysis with immunohistochemistry, we discovered novel, spatiotemporal accumulation of PA metabolites in both the *Drosophila* embryonic trachea and differentiating eye imaginal discs. These results suggest that spatiotemporal patterning of polyamines is required during organogenesis, which is at odds with the current paradigm that polyamines and other metabolites primarily play roles in cell homeostasis and, when dysregulated, disease. Recent discoveries have also shown that traditional morphogens also have important metabolic roles\(^{398-400}\). As morphogens are spatiotemporally patterned during development, it follows that some metabolites may reflect that underlying pattern formation. We believe that these findings will prove of general importance toward advancing our knowledge of how embryos and tissues develop. The functional significance and regulation of extracellular polyamine accumulation in the developing trachea and eye disc is an open question that will require significant effort to investigate.

In Chapter 4:, we used an integrative method building off of RNA-seq data describing single-cell, dormant brain metastases versus large overt metastases. Candidate genes from the RNA-seq data underwent rapid functional validation in a HTS testing their effect on tumorigenesis and metastasis *in vivo* using a genetic model of
cancer in *Drosophila*. This approach enabled the identification of several genes whose knockdown via RNAi resulted in significant decreases in primary tumor formation or metastasis. This integrative screening approach allows us to identify genes that have a functional role in primary tumor formation, metastatic development, or progression. Our methodology led to the identification of 51 unique human genes necessary for tumor growth in our model, 14 of which resulted in dramatically smaller tumors in the *Drosophila* system. Of these genes, increased expression in brain metastases has been validated for two genes, Rab11b and SERPINF1. Future studies will validate the functional role of hit genes in human breast cancer cells in an immunocompromised murine background to further demonstrate preclinical efficacy.

In Chapter 5:, I described our integrated method combining screening with RNA-seq and protein-protein network interaction data for the rapid evaluation and mechanism of action prediction for novel spirooxindole drugs, and our initial implementation of a high-content screening strategy to facilitate prediction of the compound properties conducive for compound efficacy. Efficacy of spirooxindoles were evaluated *in vitro* and their mechanism of action rapidly identified using a transcriptional approach from RNA-seq data. Our inter-disciplinary approach lead to the identification of a bis-spirooxindole, 13b, that effectively kills triple negative breast cancer cells, holding promise as a treatment for a sub-type of breast cancer that has no available targeted therapeutics. Interestingly, compound 13b appears to affect cytotoxicity by inhibiting the ribosome and thus protein translation. While protein
translation and ribosome biogenesis have received attention as therapeutic targets for chemotherapy, there are currently only two cancer drugs that specifically target ribosomal biogenesis\(^{390-393,401}\). One of these, CX-5461, entered phase I clinical trials for hematological malignancies less than five years after its discovery, and patients are currently being recruited for a phase I/II trial to test its efficacy against triple negative breast cancer\(^{394}\). Compound 13b is particularly promising due to its demonstrated efficacy against brain trophic breast cancer cells and relative novelty as one of only a few drugs with a new and promising mechanism of action.

6.2 New questions and future directions

Our media development pipeline (Chapter 2) can be applied to a variety of biological processes, including developing novel cell culture media or drug cocktails for precision medicine. Expanding on our approach by including iterative rescreening, where validated factors would be incorporated into the media background prior to a secondary screen, could elucidate compound synergies with previously identified compounds. This approach, coupled with our biological target identification technique, could lead to important insight into crosstalk between pathways and proteins important for growth. Our cumulative gene scoring approach could be applied to any inverse drug screen for which a drug-protein interaction database is available, and is a vast improvement over traditional screens that only consider biological targets of individual hit compounds rather than cumulative effects.
The function of the observed spatiotemporal patterns of polyamine accumulation in the developing trachea and eye disc (Chapter 3) will require substantial work to investigate.

Our approach for rapid identification and functional evaluation of mediators of metastatic outgrowth (Chapter 4) led to the identification of 14 genes that resulted in dramatically smaller tumors in the Drosophila system. Of these, so far Rab11b and SERPINF1 were validated to be more highly expressed in brain metastases compared to the primary tumor. In the future, more of the candidate genes can be retested for their expression levels in primary versus secondary metastatic tumors, and mechanistic studies can be performed to validate their functional roles in metastatic outgrowth. The most promising hits can ultimately be knocked down in human breast cancer cells injected into an immunocompromised murine background to functionally test for their requirement in brain metastasis. Another promising avenue to build off of this approach is to conduct a second, complimentary screen. In addition to the RNA-seq data we harnessed to identify transcriptional changes in tumors relevant for metastatic outgrowth, the Zhang lab generated a complimentary dataset describing the corresponding transcriptional changes undergone by the brain microenvironment during tumor outgrowth. Thus, generating of a fly that would enable us to perform the complimentary screen to functional test for the requirement for transcriptional reprogramming of the brain microenvironment in successful metastatic outgrowth. Such
a tester line would require two binary expression systems to independently perturb expression in the tumor and in the brain.

Our implementation of a high-content screening strategy to facilitate prediction of compound properties conducive for cytotoxic efficacy (Chapter 5) is an area for continued future work. By relating ultra high-content cellular readouts to drug structure will ultimately enable us to incorporate synthesis information about novel compounds into our machine-learning pipeline to identify synthesis routes that impact drug efficacy. Functional data from these screens can be used to identify effective structural components, which in turn can be used to design the next round of chemical synthesis.
Figure A.1: Comparison of Cl.8 cell proliferation in complete serum containing media, ZO media unsupplemented, and “ZO Fortified.” Cells were seeded at 50,000 cells/well and imaged after three days. ZO Fortified supports initial attachment and proliferation of Cl.8 cells whereas ZO unsupplemented does not.
Figure A.2: Cl.8 and S2 cells during long term culture in ZO Fortified versus ZB Media; images at the point of discernable difference in cell growth between ZO Fortified and ZB Media. S2 cells passaged into ZO Fortified were unable to reach passage two (confluency) whereas those supplemented with spermidine (ZB Media) were able to reach confluency and undergo two population doublings before growth rates stalled (passage four). Cl.8 cells passaged into ZO Fortified became confluent and underwent one population doubling (passage three) before growth stopped. Cl.8 cells cultured in ZB Media are able to proliferate long-term (currently passage 35). Cl.8 cells after adaptation (ten passages in ZB Media) can be frozen and thawed successfully.
Figure A.3: CI8 growth kinetics are dependent on seeding density. Spermidine causes significant increase in growth for all seeding densities tested. Corresponding p-values across multiple days (data shown in Figure A.4) can be found in the electronic supplementary information.
Figure A.4: Cl.8 growth kinetics are dependent on seeding density. Even at 125,000 cell/mL seeding density, ZB Media promotes cell growth (d5/d0 intensity fold change of 1.12 versus 1.01 for ZO Fortified). However, growth at this low of seeding densities is low enough that cells will likely deplete culture medium before reaching confluency. Thus seeding density must be considered when using ZB Medium.
Figure A.5: Supplementing ZB Media with spent media ZB media from adapted or unadapted Cl.8 cells does not induce proliferation of Cl.8 or S2 cells. Adapted (passage 33 in ZB Media) and unadapted Cl.8 cells were seeded in ZB Media at 1x10^6 cell/mL and allowed to proliferate for five days. Spent media was then harvested and Cl.8 and S2 cells were seeded at 50,000 cell/well in 96-well plates with either 0, 25, 50, 75, or 100% spent media from adapted or unadapted cells. Cells were allowed to proliferate for eight days and then assayed with CyQUANT.
Figure A.6: (A) Hierarchical clustering of known PA pathway genes over developmental time reveals that genes at similar levels of metabolic flow show similar expression profiles (cluster colors consistent with coloring in Figure 3.1, Figure 3.2). Clusters can be interpreted to represent different roles in the metabolic pathway: pink is upstream biosynthesis, green is downstream biosynthesis, blue is interconversion, and yellow is inhibition. (B) Violin plots depicting the time-course expression of the average mean-centered expression of each of the four gene clusters across developmental stages. Each violin plot shows the frequency distribution of the mean expression level of all genes per cluster.
Figure A.7: Orthogonal z-projection reveals that PAs are located apical to E-cadherin (five days after egg laying, same disc as Figure 3.3 G). Scale bar 5 μm.
**TABLE A.1**

**TOP 50 NEGATIVELY ENRICHED TERMS FROM GSEA**

<table>
<thead>
<tr>
<th>Term</th>
<th>Size</th>
<th>NES</th>
</tr>
</thead>
<tbody>
<tr>
<td>KEGG RIBOSOME</td>
<td>85</td>
<td>-9.49</td>
</tr>
<tr>
<td>REACTOME PEPTIDE CHAIN ELONGATION</td>
<td>84</td>
<td>-9.39</td>
</tr>
<tr>
<td>REACTOME INFLUENZA VIRAL RNA TRANSCRIPTION AND REPLICATION</td>
<td>100</td>
<td>-9.07</td>
</tr>
<tr>
<td>REACTOME METABOLISM OF MRNA</td>
<td>197</td>
<td>-9.05</td>
</tr>
<tr>
<td>REACTOME METABOLISM OF RNA</td>
<td>235</td>
<td>-8.86</td>
</tr>
<tr>
<td>REACTOME 3 UTR MEDIATED TRANSLATIONAL REGULATION</td>
<td>103</td>
<td>-8.82</td>
</tr>
<tr>
<td>REACTOME NONSENSE MEDIATED DECAY ENHANCED BY THE EXON JUNCTION COMPLEX</td>
<td>103</td>
<td>-8.51</td>
</tr>
<tr>
<td>REACTOME SRP DEPENDENT COTRANSLATIONAL PROTEIN TARGETING TO MEMBRANE</td>
<td>103</td>
<td>-8.36</td>
</tr>
<tr>
<td>STRUCTURAL CONSTITUENT OF RIBOSOME</td>
<td>75</td>
<td>-8.31</td>
</tr>
<tr>
<td>REACTOME INFLUENZA LIFE CYCLE</td>
<td>132</td>
<td>-8.00</td>
</tr>
<tr>
<td>REACTOME TRANSLATION</td>
<td>139</td>
<td>-7.92</td>
</tr>
<tr>
<td>REACTOME METABOLISM OF PROTEINS</td>
<td>339</td>
<td>-6.19</td>
</tr>
<tr>
<td>REACTOME CELL CYCLE</td>
<td>343</td>
<td>-6.00</td>
</tr>
<tr>
<td>Term</td>
<td>Size</td>
<td>NES</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>REACTOME ACTIVATION OF THE MRNA UPON BINDING OF THE CAP BINDING COMPLEX AND EIFS AND SUBSEQUENT BINDING TO 43S</td>
<td>55</td>
<td>-5.79</td>
</tr>
<tr>
<td>REACTOME FORMATION OF THE TERNARY COMPLEX AND SUBSEQUENTLY THE 43S COMPLEX</td>
<td>47</td>
<td>-5.78</td>
</tr>
<tr>
<td>REACTOME DNA REPLICATION</td>
<td>169</td>
<td>-5.55</td>
</tr>
<tr>
<td>REACTOME CELL CYCLE MITOTIC</td>
<td>273</td>
<td>-5.48</td>
</tr>
<tr>
<td>RNA BINDING</td>
<td>231</td>
<td>-5.27</td>
</tr>
<tr>
<td>REACTOME CELL CYCLE CHECKPOINTS</td>
<td>103</td>
<td>-5.23</td>
</tr>
<tr>
<td>STRUCTURAL MOLECULE ACTIVITY</td>
<td>223</td>
<td>-5.23</td>
</tr>
<tr>
<td>REACTOME ADAPTIVE IMMUNE SYSTEM</td>
<td>439</td>
<td>-5.20</td>
</tr>
<tr>
<td>REACTOME MITOTIC G1 G1 S PHASES</td>
<td>126</td>
<td>-5.18</td>
</tr>
<tr>
<td>REACTOME S PHASE</td>
<td>103</td>
<td>-5.16</td>
</tr>
<tr>
<td>REACTOME MITOTIC M M G1 PHASES</td>
<td>150</td>
<td>-5.11</td>
</tr>
<tr>
<td>REACTOME G1 S TRANSITION</td>
<td>105</td>
<td>-5.04</td>
</tr>
<tr>
<td>KEGG REGULATION OF ACTIN CYTOSKELETON</td>
<td>195</td>
<td>-4.96</td>
</tr>
<tr>
<td>REACTOME SYNTHESIS OF DNA</td>
<td>87</td>
<td>-4.94</td>
</tr>
<tr>
<td>REACTOME VIF MEDIATED DEGRADATION OF APOBEC3G</td>
<td>48</td>
<td>-4.91</td>
</tr>
<tr>
<td>REACTOME PLATELET ACTIVATION SIGNALING AND AGGREGATION</td>
<td>183</td>
<td>-4.89</td>
</tr>
<tr>
<td>REACTOME ER PHAGOSOME PATHWAY</td>
<td>56</td>
<td>-4.84</td>
</tr>
<tr>
<td>REACTOME HOST INTERACTIONS OF HIV FACTORS</td>
<td>116</td>
<td>-4.83</td>
</tr>
<tr>
<td>Term</td>
<td>Size</td>
<td>NES</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>REACTOME SCFSKP2 MEDIATED DEGRADATION OF P27 P21</td>
<td>53</td>
<td>-4.81</td>
</tr>
<tr>
<td>REACTOME HIV INFECTION</td>
<td>185</td>
<td>-4.76</td>
</tr>
<tr>
<td>REACTOME AUTODEGRADATION OF THE E3 UBIQUITIN LIGASE COP1</td>
<td>45</td>
<td>-4.71</td>
</tr>
<tr>
<td>REACTOME CDK MEDIATED PHOSPHORYLATION AND REMOVAL OF CDC6</td>
<td>45</td>
<td>-4.71</td>
</tr>
<tr>
<td>KEGG PATHOGENIC ESCHERICHIA COLI INFECTION</td>
<td>51</td>
<td>-4.68</td>
</tr>
<tr>
<td>REACTOME P53 INDEPENDENT G1 S DNA DAMAGE CHECKPOINT</td>
<td>47</td>
<td>-4.65</td>
</tr>
<tr>
<td>REACTOME HEMOSTASIS</td>
<td>399</td>
<td>-4.65</td>
</tr>
<tr>
<td>REACTOME M G1 TRANSITION</td>
<td>76</td>
<td>-4.63</td>
</tr>
<tr>
<td>REACTOME DESTABILIZATION OF MRNA BY AUF1 HNRNP D0</td>
<td>48</td>
<td>-4.62</td>
</tr>
<tr>
<td>REACTOME CROSS PRESENTATION OF SOLUBLE EXOGENOUS ANTIGENS ENDOSOMES</td>
<td>43</td>
<td>-4.61</td>
</tr>
<tr>
<td>REACTOME ASSEMBLY OF THE PRE REPLICATIVE COMPLEX</td>
<td>61</td>
<td>-4.60</td>
</tr>
<tr>
<td>REACTOME REGULATION OF MITOTIC CELL CYCLE</td>
<td>72</td>
<td>-4.59</td>
</tr>
<tr>
<td>REACTOME REGULATION OF ORNITHINE DECARBOXYLASE ODC</td>
<td>47</td>
<td>-4.58</td>
</tr>
<tr>
<td>REACTOME ANTIGEN PROCESSING CROSS PRESENTATION</td>
<td>67</td>
<td>-4.56</td>
</tr>
<tr>
<td>REACTOME CDT1 ASSOCIATION WITH THE CDC6 ORC ORIGIN COMPLEX</td>
<td>52</td>
<td>-4.55</td>
</tr>
<tr>
<td>REACTOME SCF BETA TRCP MEDIATED DEGRADATION OF EMI1</td>
<td>48</td>
<td>-4.53</td>
</tr>
<tr>
<td>Term</td>
<td>Size</td>
<td>NES</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>REACTOME P53 DEPENDENT G1 DNA DAMAGE RESPONSE</td>
<td>51</td>
<td>-4.50</td>
</tr>
<tr>
<td>REACTOME APC C CDH1 MEDIATED DEGRADATION OF CDC20 AND OTHER</td>
<td>59</td>
<td>-4.49</td>
</tr>
<tr>
<td>APC C CDH1 TARGETED PROTEINS IN LATE MITOSIS EARLY G1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE A.2
TOP 50 POSITIVELY ENRICHED TERMS FROM GSEA

<table>
<thead>
<tr>
<th>Term</th>
<th>Size</th>
<th>NES</th>
</tr>
</thead>
<tbody>
<tr>
<td>POST TRANSLATIONAL PROTEIN MODIFICATION</td>
<td>407</td>
<td>4.38</td>
</tr>
<tr>
<td>PROTEIN SERINE THREONINE KINASE ACTIVITY</td>
<td>178</td>
<td>4.30</td>
</tr>
<tr>
<td>NUCLEAR PART</td>
<td>495</td>
<td>4.24</td>
</tr>
<tr>
<td>PHOSPHOTRANSFERASE ACTIVITY ALCOHOL GROUP AS ACCEPTOR</td>
<td>299</td>
<td>4.15</td>
</tr>
<tr>
<td>TRANSCRIPTION FACTOR BINDING</td>
<td>279</td>
<td>4.13</td>
</tr>
<tr>
<td>KINASE ACTIVITY</td>
<td>331</td>
<td>4.04</td>
</tr>
<tr>
<td>TRANSFERASE ACTIVITY TRANSFERRING PHOSPHORUS CONTAINING</td>
<td>378</td>
<td>3.99</td>
</tr>
<tr>
<td>GROUPS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROTEIN KINASE ACTIVITY</td>
<td>254</td>
<td>3.98</td>
</tr>
<tr>
<td>REACTOME UNFOLDED PROTEIN RESPONSE</td>
<td>71</td>
<td>3.89</td>
</tr>
<tr>
<td>TRANSCRIPTION COFACTOR ACTIVITY</td>
<td>212</td>
<td>3.87</td>
</tr>
<tr>
<td>NUCLEAR LUMEN</td>
<td>328</td>
<td>3.84</td>
</tr>
<tr>
<td>LIGASE ACTIVITY</td>
<td>85</td>
<td>3.81</td>
</tr>
<tr>
<td>MEMBRANE ENCLOSED LUMEN</td>
<td>383</td>
<td>3.80</td>
</tr>
<tr>
<td>Term</td>
<td>Size</td>
<td>NES</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>ORGANELLE LUMEN</td>
<td>383</td>
<td>3.78</td>
</tr>
<tr>
<td>MACROMOLECULAR COMPLEX ASSEMBLY</td>
<td>244</td>
<td>3.77</td>
</tr>
<tr>
<td>NUCLEOPLASM</td>
<td>242</td>
<td>3.75</td>
</tr>
<tr>
<td>PROTEIN AMINO ACID PHOSPHORYLATION</td>
<td>237</td>
<td>3.73</td>
</tr>
<tr>
<td>PHOSPHORYLATION</td>
<td>269</td>
<td>3.71</td>
</tr>
<tr>
<td>KEGG UBIQUITIN MEDIATED PROTEOLYSIS</td>
<td>116</td>
<td>3.69</td>
</tr>
<tr>
<td>CELLULAR COMPONENT ASSEMBLY</td>
<td>260</td>
<td>3.68</td>
</tr>
<tr>
<td>TRANSCRIPTION FROM RNA POLYMERASE II PROMOTER</td>
<td>426</td>
<td>3.67</td>
</tr>
<tr>
<td>RESPONSE TO STRESS</td>
<td>452</td>
<td>3.59</td>
</tr>
<tr>
<td>REGULATION OF TRANSCRIPTION</td>
<td>492</td>
<td>3.53</td>
</tr>
<tr>
<td>INTRACELLULAR TRANSPORT</td>
<td>248</td>
<td>3.51</td>
</tr>
<tr>
<td>STK33 UP</td>
<td>191</td>
<td>3.51</td>
</tr>
<tr>
<td>TRANSCRIPTION ACTIVATOR ACTIVITY</td>
<td>157</td>
<td>3.51</td>
</tr>
<tr>
<td>DNA METABOLIC PROCESS</td>
<td>234</td>
<td>3.50</td>
</tr>
<tr>
<td>KEGG BASAL TRANSCRIPTION FACTORS</td>
<td>31</td>
<td>3.49</td>
</tr>
<tr>
<td>REACTOME DIABETES PATHWAYS</td>
<td>112</td>
<td>3.49</td>
</tr>
<tr>
<td>RESPONSE TO ENDOGENOUS STIMULUS</td>
<td>175</td>
<td>3.45</td>
</tr>
<tr>
<td>DNA REPAIR</td>
<td>116</td>
<td>3.43</td>
</tr>
<tr>
<td>Term</td>
<td>Size</td>
<td>NES</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
<td>-----</td>
</tr>
<tr>
<td>RESPONSE TO DNA DAMAGE STIMULUS</td>
<td>147</td>
<td>3.42</td>
</tr>
<tr>
<td>REACTOME SIGNALLING BY NGF</td>
<td>194</td>
<td>3.38</td>
</tr>
<tr>
<td>REGULATION OF CELL CYCLE</td>
<td>164</td>
<td>3.38</td>
</tr>
<tr>
<td>PROGRAMMED CELL DEATH</td>
<td>389</td>
<td>3.34</td>
</tr>
<tr>
<td>TRANSCRIPTION COACTIVATOR ACTIVITY</td>
<td>116</td>
<td>3.34</td>
</tr>
<tr>
<td>APOPTOSIS GO</td>
<td>388</td>
<td>3.32</td>
</tr>
<tr>
<td>REACTOME RNA POL II TRANSCRIPTION PRE INITIATION AND PROMOTER OPENING</td>
<td>38</td>
<td>3.30</td>
</tr>
<tr>
<td>REGULATION OF KINASE ACTIVITY</td>
<td>136</td>
<td>3.23</td>
</tr>
<tr>
<td>DNA DIRECTED RNA POLYMERASEII HOLOENZYME</td>
<td>57</td>
<td>3.19</td>
</tr>
<tr>
<td>PIGF UP.V1 UP</td>
<td>171</td>
<td>3.18</td>
</tr>
<tr>
<td>REGULATION OF RNA METABOLIC PROCESS</td>
<td>415</td>
<td>3.17</td>
</tr>
<tr>
<td>REACTOME PERK REGULATED GENE EXPRESSION</td>
<td>25</td>
<td>3.15</td>
</tr>
<tr>
<td>NUCLEOCYTOPLASMIC TRANSPORT</td>
<td>76</td>
<td>3.13</td>
</tr>
<tr>
<td>REGULATION OF PROTEIN KINASE ACTIVITY</td>
<td>134</td>
<td>3.13</td>
</tr>
<tr>
<td>PROTEIN KINASE CASCADE</td>
<td>241</td>
<td>3.12</td>
</tr>
<tr>
<td>NUCLEAR TRANSPORT</td>
<td>76</td>
<td>3.12</td>
</tr>
<tr>
<td>NUCLEOPLASM PART</td>
<td>180</td>
<td>3.11</td>
</tr>
<tr>
<td>REGULATION OF TRANSFERASE ACTIVITY</td>
<td>138</td>
<td>3.11</td>
</tr>
</tbody>
</table>
### TABLE A.3

**TOP 100 DEMAND GENES FROM TIME-POINTS COMBINED**

<table>
<thead>
<tr>
<th>MoA Rank</th>
<th>Gene</th>
<th>Pvalue</th>
<th>MoA Rank</th>
<th>Gene</th>
<th>Pvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RPS17</td>
<td>3.27E-13</td>
<td>51</td>
<td>RPL4</td>
<td>1.09E-12</td>
</tr>
<tr>
<td>2</td>
<td>RPS28</td>
<td>4.73E-13</td>
<td>52</td>
<td>RPS9</td>
<td>1.10E-12</td>
</tr>
<tr>
<td>3</td>
<td>RPS13</td>
<td>5.16E-13</td>
<td>53</td>
<td>RPS3A</td>
<td>1.10E-12</td>
</tr>
<tr>
<td>4</td>
<td>RPS20</td>
<td>5.21E-13</td>
<td>54</td>
<td>RPL26</td>
<td>1.11E-12</td>
</tr>
<tr>
<td>5</td>
<td>RPS23</td>
<td>5.65E-13</td>
<td>55</td>
<td>RPL35</td>
<td>1.12E-12</td>
</tr>
<tr>
<td>6</td>
<td>RPL19</td>
<td>5.91E-13</td>
<td>56</td>
<td>RPL14</td>
<td>1.14E-12</td>
</tr>
<tr>
<td>7</td>
<td>RPS5</td>
<td>6.05E-13</td>
<td>57</td>
<td>RPL13A</td>
<td>1.14E-12</td>
</tr>
<tr>
<td>8</td>
<td>RPS2</td>
<td>6.12E-13</td>
<td>58</td>
<td>RPS15</td>
<td>1.15E-12</td>
</tr>
<tr>
<td>9</td>
<td>RPL9</td>
<td>6.53E-13</td>
<td>59</td>
<td>RPS16</td>
<td>1.19E-12</td>
</tr>
<tr>
<td>10</td>
<td>RPL11</td>
<td>6.77E-13</td>
<td>60</td>
<td>RPS6</td>
<td>1.21E-12</td>
</tr>
<tr>
<td>11</td>
<td>FAU</td>
<td>7.18E-13</td>
<td>61</td>
<td>RPL21</td>
<td>1.24E-12</td>
</tr>
<tr>
<td>12</td>
<td>RPS7</td>
<td>7.21E-13</td>
<td>62</td>
<td>RPS10</td>
<td>1.25E-12</td>
</tr>
<tr>
<td>13</td>
<td>RPL30</td>
<td>7.35E-13</td>
<td>63</td>
<td>RPL37A</td>
<td>1.27E-12</td>
</tr>
<tr>
<td>14</td>
<td>RPL6</td>
<td>7.41E-13</td>
<td>64</td>
<td>RPL10</td>
<td>1.40E-12</td>
</tr>
<tr>
<td>15</td>
<td>RPS11</td>
<td>7.47E-13</td>
<td>65</td>
<td>RPL26L1</td>
<td>1.52E-12</td>
</tr>
<tr>
<td>16</td>
<td>RPL18</td>
<td>7.65E-13</td>
<td>66</td>
<td>RPS14</td>
<td>1.59E-12</td>
</tr>
<tr>
<td>17</td>
<td>RPL3L</td>
<td>7.66E-13</td>
<td>67</td>
<td>RPS29</td>
<td>1.72E-12</td>
</tr>
<tr>
<td>18</td>
<td>RPS4Y1</td>
<td>7.66E-13</td>
<td>68</td>
<td>RPL29</td>
<td>1.90E-12</td>
</tr>
<tr>
<td>19</td>
<td>RPL36AL</td>
<td>7.70E-13</td>
<td>69</td>
<td>RPL23</td>
<td>2.37E-12</td>
</tr>
<tr>
<td>20</td>
<td>RPL39</td>
<td>7.93E-13</td>
<td>70</td>
<td>RPL36</td>
<td>2.52E-12</td>
</tr>
<tr>
<td>21</td>
<td>RPS24</td>
<td>8.05E-13</td>
<td>71</td>
<td>RPS19</td>
<td>2.61E-12</td>
</tr>
<tr>
<td>22</td>
<td>RPL41</td>
<td>8.14E-13</td>
<td>72</td>
<td>NOP14</td>
<td>2.95E-12</td>
</tr>
<tr>
<td>23</td>
<td>RPL27</td>
<td>8.16E-13</td>
<td>73</td>
<td>MAGEB2</td>
<td>3.78E-12</td>
</tr>
<tr>
<td>24</td>
<td>RPS3</td>
<td>8.17E-13</td>
<td>74</td>
<td>DDX56</td>
<td>3.86E-12</td>
</tr>
<tr>
<td>25</td>
<td>RPL38</td>
<td>8.27E-13</td>
<td>75</td>
<td>RPLP0</td>
<td>4.27E-12</td>
</tr>
<tr>
<td>26</td>
<td>RPSA</td>
<td>8.37E-13</td>
<td>76</td>
<td>VARS</td>
<td>4.27E-12</td>
</tr>
<tr>
<td>27</td>
<td>MRPL13</td>
<td>8.44E-13</td>
<td>77</td>
<td>RPL23A</td>
<td>5.69E-12</td>
</tr>
<tr>
<td>28</td>
<td>RPS15A</td>
<td>8.50E-13</td>
<td>78</td>
<td>MYOC</td>
<td>7.42E-12</td>
</tr>
<tr>
<td>29</td>
<td>RPL37</td>
<td>8.50E-13</td>
<td>79</td>
<td>RPS27</td>
<td>7.77E-12</td>
</tr>
<tr>
<td>30</td>
<td>RPL8</td>
<td>8.52E-13</td>
<td>80</td>
<td>FGF5</td>
<td>9.32E-12</td>
</tr>
<tr>
<td>31</td>
<td>RPS26</td>
<td>8.88E-13</td>
<td>81</td>
<td>PTN</td>
<td>9.76E-12</td>
</tr>
<tr>
<td>MoA Rank</td>
<td>Gene</td>
<td>Pvalue</td>
<td>MoA Rank</td>
<td>Gene</td>
<td>Pvalue</td>
</tr>
<tr>
<td>----------</td>
<td>--------</td>
<td>---------</td>
<td>----------</td>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>32</td>
<td>RPS25</td>
<td>8.92E-13</td>
<td>82</td>
<td>RPLP1</td>
<td>1.12E-11</td>
</tr>
<tr>
<td>33</td>
<td>RPL7A</td>
<td>8.95E-13</td>
<td>83</td>
<td>RPS27A</td>
<td>1.25E-11</td>
</tr>
<tr>
<td>34</td>
<td>CORO7</td>
<td>8.98E-13</td>
<td>84</td>
<td>IMP4</td>
<td>1.64E-11</td>
</tr>
<tr>
<td>35</td>
<td>RPL35A</td>
<td>9.08E-13</td>
<td>85</td>
<td>EIF1B</td>
<td>1.81E-11</td>
</tr>
<tr>
<td>36</td>
<td>RPL34</td>
<td>9.36E-13</td>
<td>86</td>
<td>MPG</td>
<td>1.91E-11</td>
</tr>
<tr>
<td>37</td>
<td>RPLP2</td>
<td>9.40E-13</td>
<td>87</td>
<td>ATG12</td>
<td>1.96E-11</td>
</tr>
<tr>
<td>38</td>
<td>RPS21</td>
<td>9.49E-13</td>
<td>88</td>
<td>MAP3K1</td>
<td>2.30E-11</td>
</tr>
<tr>
<td>39</td>
<td>RPS12</td>
<td>9.50E-13</td>
<td>89</td>
<td>BLMH</td>
<td>2.60E-11</td>
</tr>
<tr>
<td>40</td>
<td>RPS18</td>
<td>9.53E-13</td>
<td>90</td>
<td>PRODH</td>
<td>2.67E-11</td>
</tr>
<tr>
<td>41</td>
<td>RPL27A</td>
<td>9.55E-13</td>
<td>91</td>
<td>STAU1</td>
<td>3.15E-11</td>
</tr>
<tr>
<td>42</td>
<td>RPS8</td>
<td>9.60E-13</td>
<td>92</td>
<td>UCHL3</td>
<td>3.29E-11</td>
</tr>
<tr>
<td>43</td>
<td>RPL10A</td>
<td>1.01E-12</td>
<td>93</td>
<td>MPHOSPH6</td>
<td>3.81E-11</td>
</tr>
<tr>
<td>44</td>
<td>RPL13</td>
<td>1.01E-12</td>
<td>94</td>
<td>NFKB2</td>
<td>4.53E-11</td>
</tr>
<tr>
<td>45</td>
<td>RPL24</td>
<td>1.01E-12</td>
<td>95</td>
<td>PA2G4</td>
<td>4.55E-11</td>
</tr>
<tr>
<td>46</td>
<td>RPL12</td>
<td>1.01E-12</td>
<td>96</td>
<td>KRR1</td>
<td>4.91E-11</td>
</tr>
<tr>
<td>47</td>
<td>RPL28</td>
<td>1.02E-12</td>
<td>97</td>
<td>PSTPIP1</td>
<td>5.33E-11</td>
</tr>
<tr>
<td>48</td>
<td>RPL32</td>
<td>1.04E-12</td>
<td>98</td>
<td>PRDX1</td>
<td>5.74E-11</td>
</tr>
<tr>
<td>49</td>
<td>RPL7</td>
<td>1.04E-12</td>
<td>99</td>
<td>C11orf49</td>
<td>5.91E-11</td>
</tr>
<tr>
<td>50</td>
<td>RPL31</td>
<td>1.06E-12</td>
<td>100</td>
<td>B3GNT1</td>
<td>6.14E-11</td>
</tr>
</tbody>
</table>
REFERENCES


115. García Prado, E., García Gimenez, M. D., De la Puerta Vázquez, R., Espartero Sánchez, J. L. & Sáenz Rodríguez, M. T. Antiproliferative effects of mitraphylline, a pentacyclic oxindole alkaloid of Uncaria tomentosa on human glioma and


188


398. Fabre, B. *et al.* Analysis of the Drosophila melanogaster proteome dynamics during the embryo early development by a combination of label-free proteomics approaches. *PROTEOMICS* n/a–n/a (2016). doi:10.1002/pmic.201500482
