

Spring 2021

# Automating and Implementing Complex, Multistep Sample Preparation Processes to Generate Reproducible Proteomic and Phosphoproteomic Data

Brian Todd Mullis

Follow this and additional works at: <https://scholarcommons.sc.edu/etd>

 Part of the [Chemistry Commons](#)

---

## Recommended Citation

Mullis, B. T.(2021). *Automating and Implementing Complex, Multistep Sample Preparation Processes to Generate Reproducible Proteomic and Phosphoproteomic Data*. (Doctoral dissertation). Retrieved from <https://scholarcommons.sc.edu/etd/6303>

This Open Access Dissertation is brought to you by Scholar Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact [digres@mailbox.sc.edu](mailto:digres@mailbox.sc.edu).

AUTOMATING AND IMPLEMENTING COMPLEX, MULTISTEP SAMPLE  
PREPARATION PROCESSES TO GENERATE REPRODUCIBLE PROTEOMIC  
AND PHOSPHOPROTEOMIC DATA

by

Brian Todd Mullis

Bachelor of Science  
University of South Carolina, 2014

---

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Chemistry

College of Arts and Sciences

University of South Carolina

2021

Accepted by:

Ken Shimizu, Major Professor

Qian Wang, Major Professor

Norma Frizzell, Committee Member

John Lavigne, Committee Member

Susan Richardson, Committee Member

Tracey L. Weldon, Interim Vice Provost and Dean of the Graduate School

© Copyright by Brian Todd Mullis, 2021  
All Rights Reserved.

## ACKNOWLEDGEMENTS

I would first like to thank Dr. Qian Wang for his guidance throughout my graduate studies. He has helped me to think critically about my research questions and helped to further develop my understanding of how to conduct effective research. His patience and ability to thoughtfully challenge me has pushed me to become a better scientist. The lessons I learned in my graduate career will stick with me for life. I would like to thank Dr. Frank Berger for his advice and insight in my project working on thymidylate synthase inhibitor treatments. I would like to thank Dr. Johannes Stratmann who collaborated with us on the GLV project. Dr. Stratmann's thorough understanding of the biology of GLVs and plant signaling allowed us to make meaningful contributions on this research and he was extremely patient with me throughout our work together.

I would also like to thank my doctoral committee members, Dr. Ken Shimizu, Dr. Norma Frizzell, Dr. John Lavigne, and Dr. Susan Richardson, for their time and interests in my work. Thank you to my fellow lab members in the Wang lab, Foam, Jingyu, Nannie, and Libo, for their thoughtful questions and support.

Thank you to my family for your loving support. Thanks to Jonathan, Scott, Hannah, Mark, Kim, and the rest of my Midtown family. You all have made Columbia home for me. Finally, I would like to thank my wife, Lauren, for your love, friendship, constant patience, and for always believing in me.

## ABSTRACT

Phosphorylation is the most common post-translational modification that occurs within the cell and is intricately involved in many biological processes. It is important to understand the roles that protein kinases and phosphatases play by mapping protein phosphorylation as many disease states stem from aberrant signaling events. With the advent of modern mass spectrometry (MS), researchers are now provided with more complex phosphorylation patterns from thousands of proteins in each mass spectrometric analysis. Further developments in quantification techniques, such as stable isotope labeling with amino acids in cell culture (SILAC) and label-free quantitation, have allowed researchers to reliably quantify changes to these proteins in response to different treatments. A main challenge in MS-based phosphoproteomics is the need for phosphopeptide enrichment prior to MS analysis which can be labor-intensive and is plagued by irreproducibility. In the Wang lab, we have developed an automated phosphopeptide enrichment protocol that addresses many of the issues normally associated with conventional phosphopeptide enrichment workflows. Our established method outperforms conventional manual phosphopeptide enrichment methods in terms of phosphopeptide identifications and specificity. We have applied our enrichment method to two different projects, one focused on understanding thymidylate synthase inhibition in colorectal cancer cells and the second focused on understanding how tomato cells respond to green leaf volatiles.

Additionally, extracellular vesicles are a promising biomarker discovery tool that over the past 10 years researchers have used proteomic studies to identify potential biomarkers for disease diagnosis. One major problem is that the current extraction techniques are cumbersome and time consuming. To address these issues, we applied an IMCStip-based automated workflow for the purification of extracellular vesicles from urine samples. Further work is needed to optimize the extraction process, but this work highlights the potential for automating the extraction workflow aiding in EV-based biomarker discovery.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	iii
ABSTRACT .....	iv
LIST OF TABLES .....	viii
LIST OF FIGURES .....	x
CHAPTER 1: DEVELOPMENT OF AN AUTOMATED PHOSPHOPEPTIDE ENRICHMENT METHOD FOR GLOBAL AND TARGETED PHOSPHOPROTEOMIC STUDIES .....	1
1.1 INTRODUCTION .....	1
1.2 RESULTS AND DISCUSSION .....	5
1.3 CONCLUSION .....	54
1.4 MATERIALS AND METHODS .....	55
CHAPTER 2: USING PHOSPHOPROTEOMICS TO INVESTIGATE THYMIDYLATE SYNTHASE INHIBITOR TREATMENTS OF COLORECTAL CANCER CELL LINES. ....	61
2.1 INTRODUCTION .....	61
2.2 RESULTS & DISCUSSION .....	66
2.3 CONCLUSION .....	105
2.4 MATERIALS & METHODS .....	107
CHAPTER 3: PHOSPHOPROTEOMIC ANALYSIS OF TOMATO CELLS TREATED WITH GREEN LEAF VOLATILES. ....	112
3.1 INTRODUCTION .....	112
3.2 RESULTS & DISCUSSION .....	119

3.3 CONCLUSION .....	140
3.4 MATERIALS AND METHODS .....	141
CHAPTER 4: GENERATING AN AUTOMATED EXTRACTION OF EXTRACELLULAR VESICLES FROM URINE SAMPLES.....	146
4.1 INTRODUCTION.....	146
4.2 RESULTS & DISCUSSION .....	154
4.3 CONCLUSION .....	178
4.4 MATERIALS & METHODS.....	182
REFERENCES .....	185
APPENDIX A: REPRINT PERMISSION FOR CHAPTER 1 .....	196
APPENDIX B: REPRINT PERMISSION FOR FIGURE 2.1 .....	197
APPENDIX C: REPRINT PERMISSION FOR FIGURE 3.1 .....	198
APPENDIX D: REPRINT PERMISSION FIGURE 3.2.....	199
APPENDIX E: REPRINT PERMISSION FIGURE 4.1 .....	200
APPENDIX F: REPRINT PERMISSION FIGURE 4.2 .....	201

## LIST OF TABLES

Table 1.1. Desalting protocol used for testing peptide desalting resins in the IMCStip.....	8
Table 1.2. Peptides and associated MS information used for MRM Analysis. ....	9
Table 1.3. Initial phosphopeptide enrichment process with the inclusion of glycolic acid. Referred to as buffer set 1.....	15
Table 1.4. Resin testing for determining resin capacity.....	22
Table 1.5. Phosphopeptide enrichment process and buffer set.....	23
Table 1.6 Comparison of phosphopeptide recoveries and associated CVs for each phosphopeptide from different resin types (n = 8 per resin type) when the eluted samples are not dried versus when dried. ....	27
Table 1.7. Summary of phosphopeptide recoveries and associated CVs for each phosphopeptide from different resin types (n = 4 per resin type) when using elution buffer 1 or elution buffer 2.....	28
Table 1.8. Optimized automated desalting and phosphopeptide enrichment workflow. ....	32
Table 1.9. Phosphopeptide enrichment process using the Spin Tips from GL Sciences and Thermo Scientific. ....	33
Table 1.10. Summary of data from phosphopeptide enrichment of 200 µg of HCT116 cell lysate.....	39
Table 1.11. Summary of data from phosphopeptide enrichment of 200 µg of untreated HCT116 cell lysate using 10 mg of PolyTi resin. ....	40
Table 1.12. Summary of phosphopeptide recoveries (% Recov.) and associated CVs (% CV) for each phosphopeptide from different resins for the 3-week experiment (n = 8 per resin type per week). ....	43
Table 1.13. Summary of average inter-week phosphopeptide recoveries and associated CVs for each phosphopeptide from different resins for the 3-week experiment (n = 24 per resin type).....	44

Table 1.14. A summary table for the comparison between two different resins used in the IMCStip.....	51
Table 2.1. The relative growth inhibition of HCT116 using various concentrations of 5-FU based on the CTB assay. ....	67
Table 2.2. MS data from 8 HCT116 cell digest samples.....	70
Table 2.3. List of proteins that were identified across all analyzed phosphopeptide samples and were calculated to have a fold change of greater than 2 or less than 0.5.....	81
Table 2.4. Top 20 phosphorylated proteins from CDK1, CDK2, or MAPK family substrate groups that show differential responses to the 5-FdUMP/LV treatment at the 0.5-hour time point. ....	106
Table 3.1. Summary data from the tomato cells with different treatments and analyzed by different MS systems.....	123
Table 3.2. List of phosphoproteins whose abundance ratios were significantly increased in both the <i>cis</i> -3-hexenol-treated samples and the systemin-treated samples.....	130
Table 3.3. List of phosphoproteins from the <i>cis</i> -3-hexenol triplicate dataset that are linked to DAMP associated processes or pathways. ....	136
Table 4.1. Characteristics of exosomes and MVs that can be used to differentiate between these EV subtypes.....	148
Table 4.2. Extraction workflow for EVTRAP resin in the 1 mL Integra IMCStip. ....	155
Table 4.3. Extraction workflow for DEAE resin in the 1 mL Integra IMCStip. ....	156
Table 4.4. Band intensities from samples from experiments in Figure 4.12. ....	174
Table 4.5. Summary of protein, peptide, and peptide spectral matches from EVTRAP purified peptide samples. ....	177
Table 4.6. List of proteins identified by samples 1 and 2 from Table 5. ....	179

## LIST OF FIGURES

Figure 1.1. Standard workflow for sample preparation of phosphopeptide samples for MS analysis.....	6
Figure 1.2. Peptide recovery data from MRM assay from peptide samples subjected to peptide desalting extraction using different resins (C18A1, C18A3, RP) and different extraction techniques (Spin column based or IMCStip-based). ....	10
Figure 1.3. Peptide desalting for global peptide analysis. ....	11
Figure 1.4. Percent recovery of peptides was established using trypsin digested peptides purchased from Pierce. The starting sample volume was 200 $\mu$ L. ....	13
Figure 1.5. Summary of phosphopeptide enrichment data using four different resin types (n = 2 per resin type). ....	17
Figure 1.6. Phosphopeptide recovery data from MRM assay from peptide samples subjected to phosphopeptide enrichment extraction using different resins (ZrO <sub>2</sub> and PolyTi, N=8 per resin type).....	19
Figure 1.7. Comparison of number of phosphopeptides and non-phosphopeptides identified by two different workflows using ZrO <sub>2</sub> (n = 2, data are mean $\pm$ s.d.). ....	24
Figure 1.8. Comparison of number of phosphopeptides and non-phosphopeptides identified by two different workflows using PolyTi (n = 2, data are mean $\pm$ s.d.). ....	25
Figure 1.9. Comparison of phosphopeptide specificities between Wash Condition 1 and Wash Condition 2 for both PolyTi and ZrO <sub>2</sub> (n = 2 per resin type, data are mean $\pm$ s.d.). ....	29
Figure 1.10. Comparison of number of phosphopeptides and non-phosphopeptides identified by two different workflows (n = 2, data are mean $\pm$ s.d.). ....	30
Figure 1.11. Comparison of phosphopeptide enrichment methods. ....	38

Figure 1.12. The average recovery for each phosphopeptide from three different extractions performed once a week over 3-weeks using two different resins (n = 8 per resin type). .....	42
Figure 1.13. The distribution of singly-, doubly-, triply- phosphorylated peptides identified from 200 µg of HCT116 digest from the different extractions.....	45
Figure 1.14. Number of singly-, doubly-, and triply-, phosphorylated peptides identified by four different extractions with glycolic acid containing buffer set (n = 2 per resin type, data are mean ± s.d.). .....	47
Figure 1.15. Comparison of types of phosphopeptides enriched by PolyTi when using two different buffer sets (n = 2, data are mean ± s.d.). .....	48
Figure 1.16. Comparison of phosphoprofile from different resin types. ....	50
Figure 1.17. Summary of results from phosphopeptide enriched samples using either the Hamilton STAR or Integra Assist for the IMCStip-based extraction and a centrifuge for the spin column-based extraction.....	53
Figure 2.1. 5-FU is a thymidylate synthase inhibitor used to treat cancer.....	63
Figure 2.2. Cell viability plot of HCT116 cells after 5-FU treatment at each concentration. ....	68
Figure 2.3. A comparison of the total amount of phosphopeptides and non-phosphopeptides identified by either PolyTi or ZrO2 resin. ....	72
Figure 2.4. A comparison of peptide fold change ratios (log2) between two biochemical replicates of 200 µg of 1:1 mixed SILAC samples.....	73
Figure 2.5. Volcano plots of quantified proteins from each sample analyzed.....	75
Figure 2.6. Canonical pathways identified using the proteins from the phosphopeptide enriched data set. ....	77
Figure 2.7. EIF2 signaling pathway with proteins identified and quantified within the phosphopeptide enriched sample set. ....	78
Figure 2.8. Tox list generated by IPA software based on molecules from the phosphopeptide enriched dataset.....	79
Figure 2.9. CTB assay results from HCT116 and HCT116/200 cells untreated and treated for 24 hours with 10 µM 5-FdUMP/LV or 10 µM 5-FdUMP/LV/THY (n = 3, data are mean ± s.d.). ....	83

Figure 2.10. Schematic illustration of SILAC-based time course thymidylate synthase inhibitor treatment experiment. ....	85
Figure 2.11. Summary from the phosphoproteomic analysis of samples from TS inhibitor time course samples. ....	87
Figure 2.12. CLUE analysis of HCT116 5-FdUMP/LV treated dataset for all time points. ....	92
Figure 2.13. CLUE analysis of HCT116/200 5-FdUMP/LV treated dataset for all time points. ....	94
Figure 2.14. CLUE analysis of HCT116 5-FdUMP/LV/THY treated dataset for all time points. ....	96
Figure 2.15. Summary data from 0.5-hour 5-FdUMP/LV treated HCT116 samples (n = 3). ....	98
Figure 2.16. KSEA analysis of data from 0.5-hour 5-FdUMP/LV treated HCT116 samples (n = 3). ....	100
Figure 2.17. Heatmaps generated using data from 0.5-hour 5-FdUMP/LV treated HCT116, HCT116/200, and HCT116+THY samples. ....	103
Figure 3.1. Biosynthetic pathway of GLVs. ....	113
Figure 3.2. Schematic of different stressors that induce GLVs (red) and their associated roles they play in plant defense (black). ....	114
Figure 3.3. Western blot showing changes in phosphorylated MAPK1 in tomato cell lysate in response to different treatments at different time points. ....	121
Figure 3.4. Summary data from the treated tomato cells. Three samples per treatment type were collected. A sample from each treatment type was analyzed by the Orbitrap LTQ Velos, the QE-HFX, and the Orbitrap Fusion system. ....	124
Figure 3.5. Summary of the data generated from phosphopeptide enriched samples from tomato cell lysis (n = 3 per treatment, data are mean $\pm$ s.d.). ....	126
Figure 3.6. Bar graph of the number of proteins with abundance ratios greater than 2, less than 2 and greater than 0.5, and less than 0.5 for both <i>cis</i> -3-hexenol treated and systemin treated samples (n = 3 per treatment type). ....	128

Figure 3.7. Mass spectrum showing the fragmentation pattern from an MS2 scan of a phosphopeptide with the sequence VTSETDFMTEyVVTR with a charge state of 2+.....	134
Figure 3.8. DAMP/Hexenol signal perception and signal integration pathway based on phosphoproteomic dataset. ....	138
Figure 4.1. Diagram of EV subpopulations. ....	147
Figure 4.2. Network analysis of proteomic data generated by S. Wang et al. ....	151
Figure 4.3. Western blot of EV sample from 1 mL of urine.....	158
Figure 4.4. Western blot of EVTRAP purified samples from 1 mL of urine. ....	160
Figure 4.5. TEM images of purified EV samples. ....	161
Figure 4.6. Western blot of EVTRAP purified samples from urine. ....	163
Figure 4.7. Western blot of EVTRAP purified samples from urine. ....	165
Figure 4.8. Western blot of EVTRAP purified samples from urine. ....	167
Figure 4.9. Western blot of EVTRAP purified samples from 0.5 ml of urine.....	169
Figure 4.10. Western blot of EVTRAP purified samples from 0.25 mL of urine. ....	170
Figure 4.11. Western blot of EVTRAP purified samples from 0.5 mL of urine. ....	172
Figure 4.12. Western blot of EVTRAP purified samples from 0.5 mL of urine. ....	173
Figure 4.13. Total ion chromatograms (TICs) from LC-MS/MS analysis of EVTRAP purified peptide samples. ....	176

## CHAPTER 1

### DEVELOPMENT OF AN AUTOMATED PHOSHOPEPTIDE ENRICHMENT METHOD FOR GLOBAL AND TARGETED PHOSPHOPROTEOMIC STUDIES

#### 1.1 INTRODUCTION

##### *1.1.1 Importance of protein phosphorylation & phosphoproteomics*

Protein phosphorylation is an important post-translational modification used to communicate vital signals throughout the cell.<sup>1-5</sup> Kinases and phosphatases are used within the cell to turn “off” or “on” signaling cascades, leading to a variety of critical intracellular events, such as cell proliferation and apoptosis.<sup>3,4,6</sup> Understanding the roles that protein kinases and phosphatases play by mapping protein phosphorylation is of particular interest as many disease states, such as cancer and neurodegenerative diseases, stem from aberrant signaling events.<sup>7</sup> Such interest is exemplified by the number of kinase inhibitors approved by the United States Food and Drug Administration (FDA) for treating cancers and other diseases.<sup>8</sup> With the advent of modern mass spectrometry (MS), researchers are now provided more complex phosphorylation patterns from thousands of proteins in each mass spectrometric analysis.<sup>9-11</sup> Further developments in labeling techniques, such as stable isotope labeling with amino acids in cell culture (SILAC) and Tandem-Mass Tag (TMT) labeling, have allowed researchers to reliably quantify changes to these proteins in response to different stress, treatments, or novel compounds.<sup>12-16</sup>

### *1.1.2 Overview of sample preparation techniques for phosphoproteomics*

While MS-based phosphoproteomics has advanced over the past 10 years, the field is still plagued by a labor-intensive sample preparation process, limiting its translational impact. Phosphopeptide analysis requires a multi-step sample preparation process prior to mass spectrometric analysis starting from cell lysis, protein digestion, pre-enrichment desalting, and phosphopeptide enrichment. Each step can introduce variations that can obfuscate mass spectrometric results. This challenge to provide a consistent and efficient automated workflow needs to be addressed before this powerful analytical process can be further expanded and translated to more routine clinical and pharmaceutical applications.

Current sample preparation methods for generating phosphopeptide samples are segmented and laborious. Developing robust processes for these complex sample preparations that can provide reliable results is still needed. Alternatively, simplifying the workflow by eliminating the more laborious enrichment step is not feasible due to the low abundance of phosphopeptides. Recent advances in resin chemistries to specifically enrich phosphorylated peptides have been developed to address this low stoichiometry issue.<sup>17-19</sup> Two approaches are typically employed, with the first approach using antibodies to perform immunoaffinity extractions. This approach can be used to isolate specific proteins or peptides, or it can be generalized to target all peptides that have a phosphorylated tyrosine residue. Immunoaffinity extractions have been widely used, although they are limited by a lack of cost-effective antibodies that can consistently enrich the target proteins or peptides.<sup>17, 20</sup> The second approach uses metal-oxide or immobilized metal-ion affinity (IMAC) resins to extract phosphopeptides by leveraging the affinity of phosphates to iron, titania, or zirconia.<sup>18, 19</sup> These resins are more cost-

effective and have been used as the primary approach for enriching a broad spectrum of phosphorylated peptides.<sup>21</sup>

Automated metal-oxide and IMAC -based phosphopeptide enrichment methods are widely reported using automated liquid handling systems.<sup>17-21</sup> Many of these reported methods utilize automated liquid handling systems that are designed for single function operations, such as peptide desalting or phosphopeptide enrichment.<sup>22</sup> Since the focus of this study is to streamline multiple processes onto an integrated automated workflow, a more advanced platform such as Hamilton Microlab® STAR was selected. Additionally, the reported automated methods utilize either functionalize magnetic beads or spin columns containing packed resin. While magnetic bead and spin column -based methods can be automated on the STAR system, these required additional hardware such as a robotic arm to move the sample plate to and from the magnet stand or a centrifuge to carry out the spin column workflow. This additional hardware would increase capital costs and would reduce the overall workspace on the robotic platform. Hence, we selected a tip-based extraction technique that utilizes IMCStips to automate the extraction process.

### ***1.1.3 Tip-based sample preparation techniques***

Tip-based extractions have been popularized over the last two decades, mostly involving applications ranging from food and forensic analysis and more recently in protein purification.<sup>23</sup> One popular tip-based SPE is a monolith-based pipette tip that is used for desalting peptides prior to spotting on matrix-assisted laser desorption ionization time of flight mass spectrometer (MALDI TOF MS).<sup>24</sup> Tip-based extractions employ pipette tips containing resins that perform SPE which is an extraction technique by which analytes of

interest (AOI) suspended in solution are separated from contaminants based on chemical or physical properties. The resin inside of the tip is exposed to the sample either via unidirectional or bidirectional flow through the pipette tip. Tip-based extractions follow a stepwise procedure in which the resin is conditioned using buffers to prepare the resin to bind the AOI, the resin is then exposed to the sample containing the AOI, molecules that are non-specifically bound to resin are then removed using wash buffers, and the AOI is then eluted using elution buffer. Conventionally, the resin inside the tip used for the tip-based extraction is tightly restricted by two filters which results in high back pressure and clogging. In contrast, IMCStips contain loosely packed resin and operate as dispersive solid-phase extraction (dSPE). dSPE allows the resin to fully disperse throughout the liquid matrix, leading to faster binding of the AOI and limiting back pressure. The IMCStips are compatible with automated liquid handling platforms, including the Hamilton Microlab® STAR, which will be used to automate and standardize the desalting and phosphopeptide enrichment process.

This study is the first to report this novel dSPE in comparison to a traditional Spin Tip approach in our pursuit for a standardized phosphopeptide analysis. The efficiency of the workflow is benchmarked against the traditional spin format which requires intermittent manual interventions. Our integrated workflow couples the peptide desalting and the phosphopeptide enrichment using the dSPE in a pipette tip. This automated workflow demonstrated higher specificities and greater number of phosphopeptide identifications (IDs) than the conventional approaches (such as the Spin Tip method). In addition, a tandem mass spectrometric analysis performed over a 3-week period

demonstrated the consistency of the method. Such reliability is critical for the further applications of this method.

## 1.2 RESULTS AND DISCUSSION

### *1.2.1 Overview of method development*

The overall aim of this study was to standardize and automate the desalting and phosphopeptide enrichment portion of the phosphopeptide sample preparation process (Figure 1.1). The overall goal of this study was to establish an automated desalting and phosphopeptide enrichment method to improve reproducibility and the overall phosphopeptide enrichment efficiency in terms of phosphopeptide IDs and specificity. The general workflow for MS-based phosphoproteomics begins with the lysis of a cell or tissue sample using specific lysis buffer. The released proteins are treated with a reducing agent to reduce disulfide bonds, and then are irreversibly alkylated. The reduced and alkylated proteins are then digested using trypsin, a protease that cleaves proteins at the carboxyl side of arginine and lysine residues. Each of these steps needed to be optimized to ensure proper protein digestion. For example, the ratio of trypsin to peptide used for the trypsin digestion portion of the protocol needs to be established to ensure complete digestion. Our in-house protocol yields complete alkylation of cysteine residues and good peptide digest based on the number of missed cleavages observed in the resulting peptide samples. Once the lysis and digestion protocol was established, different peptide desalting resins were screened. C18 resins are commonly used for the purification of peptides to remove contaminating salts as the C18 chain has good affinity for peptides although has less affinity for more hydrophilic peptides.<sup>25</sup> Highly cross-linked poly(styrene/divinylbenzene) resins have also been used to desalt peptides and shown to

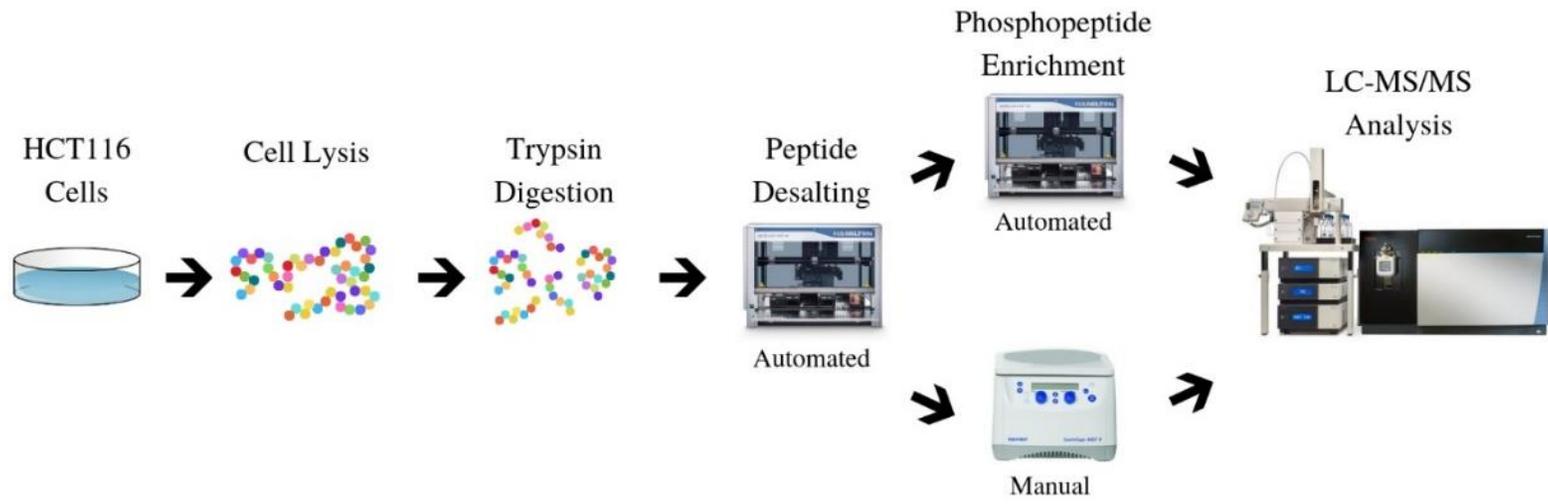


Figure 1.1. Standard workflow for sample preparation of phosphopeptide samples for MS analysis.

capture a wider range of peptides, including more hydrophilic peptides that are not captured by C18 resins.<sup>26</sup> Therefore, three different resins were evaluated in this work, including two C18 aqueous silica gel resins with bead sizes of 300 Å (C18AQ3) and 100 Å (C18AQ1) from Sorbent Technologies and one highly cross-linked divinylbenzene resin (RP) from Dow Chemical. These resins were tested in a spin column format and compared with an IMCStip format using the extraction protocol found in Table 1.1. For the IMCStip extraction, the process was automated on a Hamilton STAR system. For the spin column format, the same extraction protocol was performed using a rocker to mimic the IMCStip-based extraction.

### ***1.2.2 Establishment of MRM assay and comparison of peptide desalting resins***

To determine the ability of the different resins to recover peptides with these workflows, four peptides and three phosphopeptides were spiked in protein digest samples and a multiple reaction monitoring (MRM) assay was used to quantify the recovered peptides. The list of peptides and phosphopeptides used and their associated precursor and product ions are listed in Table 1.2. The peptide recovery data from the resulting peptide desalting extractions are found in Figure 1.2. From this data, the spin column format showed lower recovery of phosphopeptides in comparison to the IMCStip format. Additionally, the RP resin in the IMCStip consistently outperformed the other two resins in terms of overall recovery of both the peptides and phosphopeptides. To determine if different types of peptides were identified using the different resins, HEK cell digest samples were desalted using 10 mg of either C18A1, C18A3, or RP resin in the 1 mL IMCStip. The resulting samples were analyzed via mass spectrometry to determine the amount and type of peptides identified within each sample (Figure 1.3).

Table 1.1. Desalting protocol used for testing peptide desalting resins in the IMCStip (ACN: Acetonitrile, F.A.: Formic Acid, TFA: Trifluoroacetic Acid).

<b>Steps</b>	<b>Process</b>	<b>Solvent</b>	<b>Aspiration, uL</b>	<b>Volume, uL</b>	<b>Repeat #</b>	<b>Duration, min</b>
1	Activation	100% ACN	600	800	2	0.6
2	Condition	70% ACN, 0.1% F.A.	400	800	3	1.0
3	Equilibration	1% TFA	400	800	3	1.0
4	Binding	1% TFA	400	500	10	3.3
5	Washing I	0.1% TFA	400	800	3	1.0
6	Washing II	0.1% F.A.	400	800	3	1.0
7	Elution I	70% ACN, 0.1% F.A.	400	400	3	1.0
8	Elution II	70% ACN, 0.1% F.A.	400	400	3	1.0
					<b>Total</b>	<b>9.9</b>

Table 1.2. Peptides and associated MS information used for MRM Analysis.  
Phosphorylated amino acids are emboldened and underlined.

Peptides	Sequence	Ions	
		Precursors	Products
pAngiotensin2	DRV <b><u>Y</u></b> IHPF	376.1742 <sup>3+</sup>	Y [y5] - 756.3117 <sup>+</sup> , V [b3] - 371.2037 <sup>+</sup>
Aquapoin-2	RQSVELH <b><u>S</u></b> PQSLPR	571.9526 <sup>3+</sup>	P [y6] - 697.3991 <sup>+</sup> , Q [y5] - 600.3464 <sup>+</sup>
pp60(v-SRC)	RRLIEDNE <b><u>Y</u></b> TARG	558.2649 <sup>3+</sup>	N [y7] - 890.3404 <sup>+</sup> , Y [y5] - 647.2549 <sup>+</sup>
Bradykinin	RPPGFSPFR	354.1944 <sup>3+</sup>	S [y4] - 506.2722 <sup>+</sup> [1], P [y3] - 419.2401 <sup>+</sup> [2]
Angiotensin1	DRVYIHPFHL	432.8998 <sup>3+</sup>	I [b5] - 647.3511 <sup>+</sup> , I [a5] - 619.3562 <sup>+</sup>
Angiotensin2	DRVYIHPF	349.5188 <sup>3+</sup>	I [y4] - 513.2820 <sup>+</sup> , V [b3] - 371.2037 <sup>+</sup> [1]
Neurotensin	ELYENKPRRPYIL	558.3105 <sup>3+</sup>	Y [y11] - 483.6051 <sup>3+</sup> [2], L [a2] - 197.1285 <sup>+</sup> [1]

## Peptide Desalting Extraction- MRM Assay

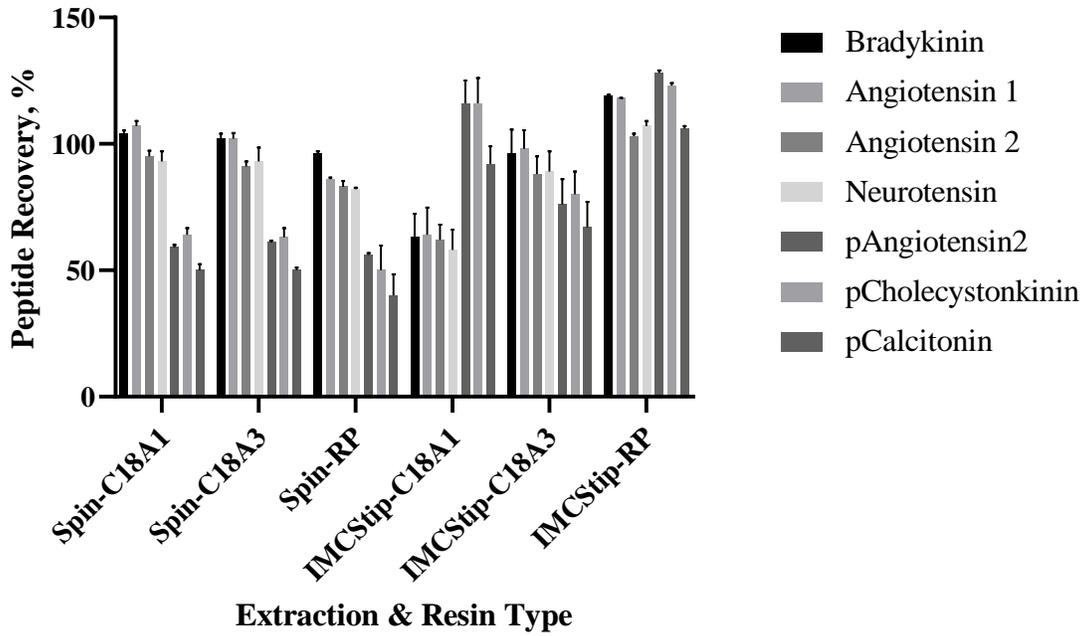
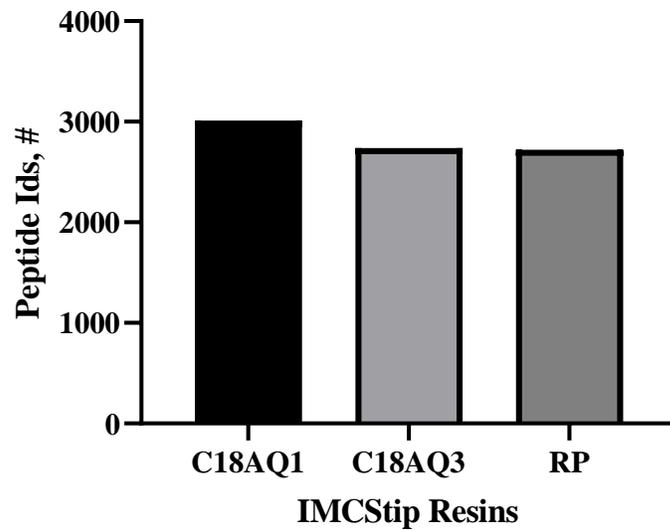


Figure 1.2. Peptide recovery data from MRM assay from peptide samples subjected to peptide desalting extraction using different resins (C18A1, C18A3, RP) and different extraction techniques (Spin column based or IMCStip-based).

## A Peptide Desalting Extraction- Global Peptide Identifications



## B

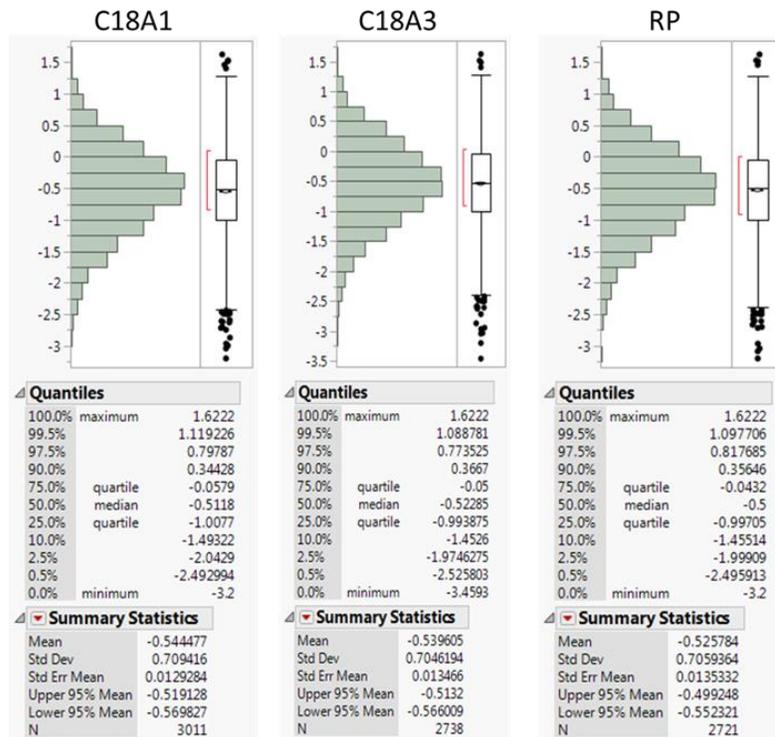


Figure 1.3. Peptide desalting for global peptide analysis. A) Total number of peptides identified for each of the different samples desalted using the specified resin. B) Histograms of GRAVY values for peptides identified within the samples desalted with the specified resin.

From this analysis the total number of peptides identified from samples desalted using either C18A1, C18A3, and RP was 3011, 2738, and 2721, respectively. This indicates that there is no significant difference in the number of peptides able to be identified using different resin types. Additionally, based on the grand average of hydrophobicity (GRAVY) values of the identified peptides, there is no significant differences in the GRAVY values of the peptides extracted using the different resin types. Based on these results, RP resin was selected and used for all future desalting extractions based on multiple different factors including phosphopeptide recovery data from the MRM assay and resin cost. After selection of the RP resin, the binding capacity of the resin was determined using trypsin digested peptides purchased from Pierce as the starting sample, and the peptide recovery was determined using the Pierce Quantitative Colorimetric Peptide Assay (Thermo Scientific) (Figure 1.4).

The peptide recovery data in Figure 1.4 highlights that peptide recovery is dependent on the amount of starting material and the peptide concentration of the starting sample. Samples with less than 200  $\mu\text{g}$  of starting material have a peptide recovery of less than 90%. This could be due to a concentration effect. In this experiment, the starting sample had a fixed volume of 200  $\mu\text{L}$ . Samples with less than 200  $\mu\text{g}$  of peptide have a lower the concentration of peptide and, therefore, more time may be needed for the peptides to bind to the resin. Samples with more than 200  $\mu\text{g}$  of starting material had a peptide recovery of less than 90%. This is most likely due to the binding sites of the resin being saturated, especially on the exterior of the resin where peptides can most readily access. This is consistent with the observation from Kates *et al.*, which showed that the binding kinetics for porous affinity resins are driven by “fast” and “slow” binding

## Recovery Post Desalting

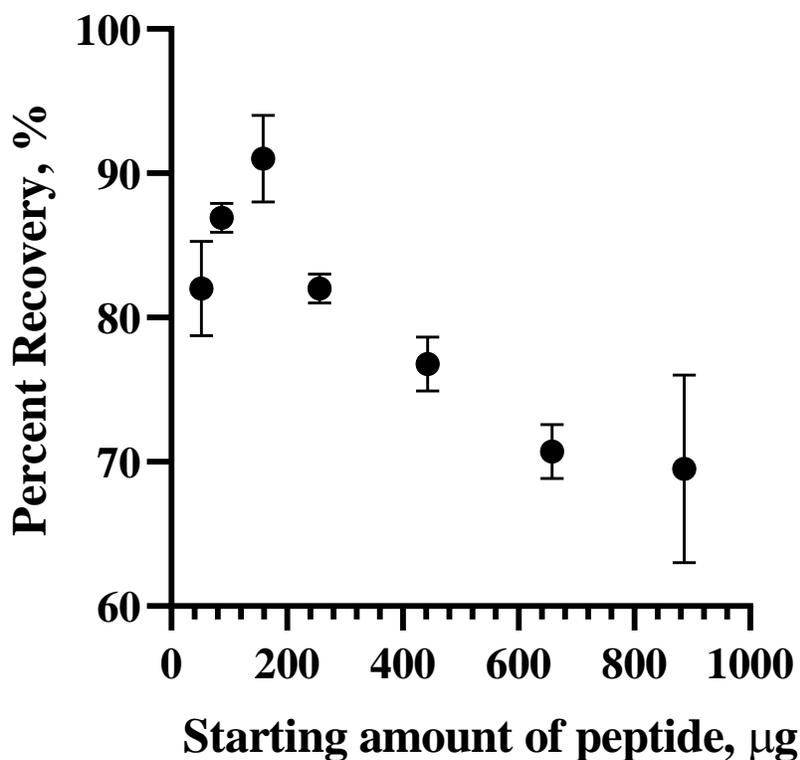


Figure 1.4. Percent recovery of peptides was established using trypsin digested peptides purchased from Pierce. The starting sample volume was 200 µL. The desalting program was preset on the automation platform using 5 mg of reverse phase resin (n = 3 per starting amount of peptide, data are mean ± s.d.). The starting peptide sample and eluted peptide samples were quantified using Pierce Quantitative Colorimetric Peptide Assay (Thermo Scientific).

events.<sup>27</sup> Based on the above results we set the upper binding capacity for the 5 mg RP IMCStip to 200 µg. This is due to analytes being able to readily bind to the exterior of resins while binding sites on the interior of the resin are rate limited by diffusion.

### ***1.2.3 Comparison of phosphopeptide enrichment resins***

For the phosphopeptide enrichment several different resins were surveyed. As previously mentioned, metal oxide resins or immobilized metal-ion affinity resins have been widely used for phosphopeptide enrichment. This is due to the affinity of phosphate groups for the metal groups within the resin. One stipulation in our search was that the resin needed to be compatible with IMCStips to facilitate the automation of the enrichment process. Thus, the resin needed to have a pore size larger than 30 microns to be retained by the frit of the IMCStip. Four different resins were initially tested for use in the phosphopeptide enrichment protocol. A TiO<sub>2</sub> resin and a ZrO<sub>2</sub> resin from Cospheric, as metal oxide resins have been shown to efficiently enrich phosphopeptides. Iliuk *et al.* previously showed that their polymer-based metal ion affinity capture (PolyMAC) method obtained highly specific isolation of phosphopeptides from cell lysate samples, and PolyMAC extraction kits are commercially available (Tymora Analytical).<sup>28, 29</sup> The PolyMAC method utilized metal ion-functionalized magnetic beads, which are not compatible with IMCStips. In working with Tymora Analytical, they were able to perform the conjugation on silica beads to generate a metal ion-functionalized silica bead (PolyTi). Additionally, an agarose-based Fe-NTA resin from Cube Biotech was tested. As a starting point for the phosphopeptide enrichment workflow, the workflow published by Iliuk *et al.* was adapted for the IMCStips (Table 1.3). The phosphopeptide enrichment workflow from

Table 1.3. Initial phosphopeptide enrichment process with the inclusion of glycolic acid. Referred to as buffer set 1.

<b>Phosphopeptide Enrichment Workflow</b>			
<b>Steps</b>	<b>Process</b>	<b>Solvent</b>	<b>Sample, <math>\mu\text{L}</math></b>
1	Activation	100% ACN	400
2	Condition	80% ACN	400
3	Equilibration	100 mM glycolic acid, 1% TFA, 50% ACN	400
4	Sample	100 mM glycolic acid, 1% TFA, 50% ACN	200
5	Wash I	100 mM glycolic acid, 1% TFA, 50% ACN	400
6	Wash II	100 mM Acetic Acid, 1% TFA, 80% ACN	400
7	Wash III	80% ACN	400
8	Elution I	1.5% $\text{NH}_4\text{OH}$	100
9	Elution II	1.5% $\text{NH}_4\text{OH}$	100

Table 1.3 and buffer set (buffer set 1), which included glycolic acid and acetic acid was used, as it was shown that these chemicals helped improve phosphopeptide specificities.<sup>30</sup> The workflow from Table 1.3 was automated on the Hamilton STAR system. For the initial experiment, 200  $\mu\text{g}$  of HCT116 cell digest was used as the starting material. Samples using the automated workflow from Table 1.3 were collected after phosphopeptide enrichment and dried completely. The dried samples were then resuspended and then analyzed via MS (Figure 1.5).

Data from our initial studies with the four different resin types showed that  $\text{ZrO}_2$  and PolyTi achieved the highest number of phosphopeptide identifications with  $2517 \pm 130$  and  $2337 \pm 115$ , respectively, compared to  $1330 \pm 13$  phosphopeptide identifications for  $\text{TiO}_2$  and  $1428 \pm 73$  phosphopeptide identifications for Fe-NTA. While the PolyTi resin achieved higher phosphopeptide identifications, the phosphopeptide specificity of  $64\% \pm 2\%$  was lower than  $\text{ZrO}_2$  and Fe-NTA which had phosphopeptide specificities of  $79\% \pm 2\%$  and  $71\% \pm 2\%$ . In subsequent analysis, the phosphopeptide specificity of  $\text{TiO}_2$  resin decreased even further and samples from Fe-NTA yielded samples that cause considerable clogging of the LC system. Thus, the  $\text{TiO}_2$  and Fe-NTA resins were not explored further for the phosphopeptide enrichment. One of the overall goals of the phosphopeptide enrichment was to achieve an in-depth phosphopeptide profile, while limiting the amount of non-phosphopeptides in the enriched samples. Obtaining the highest number of phosphopeptides gives a better view of which kinase substrates are highly phosphorylated, indicating which kinases are most highly active within the dataset. Non-phosphopeptides invariably occlude the identifications of phosphopeptides within the sample and usually bind non-specifically to the resin, thus decreasing the

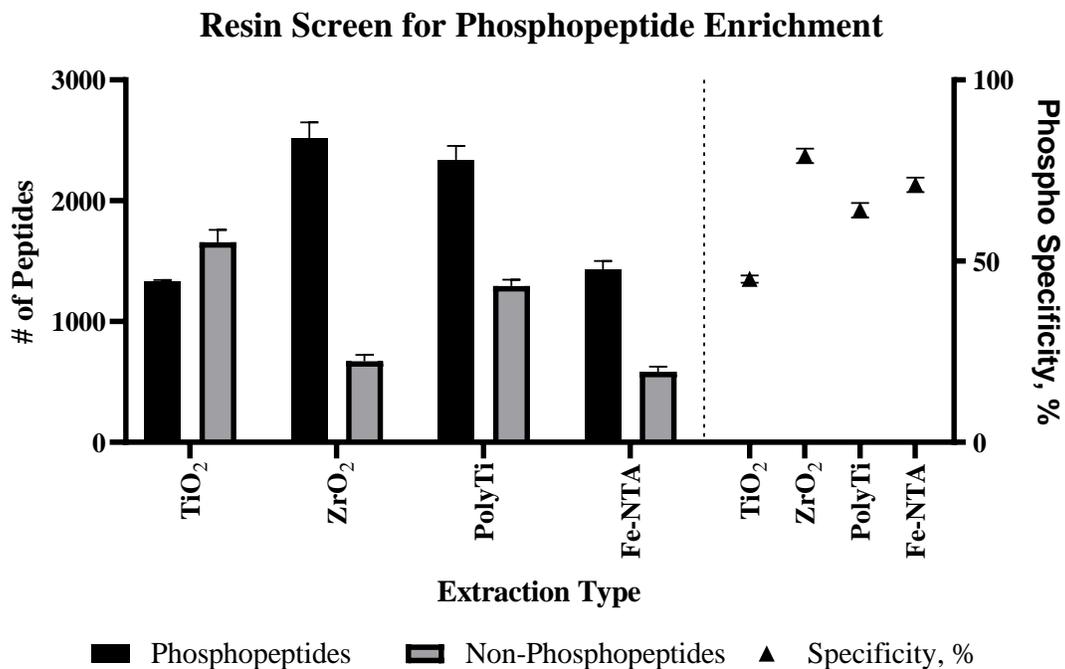


Figure 1.5. Summary of phosphopeptide enrichment data using four different resin types (n = 2 per resin type). These samples were processed using buffer set 1 prior to optimization.

reproducibility of the enrichment and complicating downstream interpretation of the resulting dataset.

The LC-MS system used for the analysis of the sample significantly influenced the number of phosphopeptide identifications. For instance, samples run on a LC column that was 100 mm in length compared to a column that was 250 mm in length generally achieved better separation of the peptides enabling higher identifications by the MS.<sup>31</sup> Additionally, one of the major advancements in MS over the past 10 years is in sensitivity, which is due to a host of different achievements including faster scan rates and optimizations in almost all facets of the instrument including ion storage, ion routing, and fragmentation. For phosphoproteomics specifically, there is a balance needed in terms of scan rate during the MS analysis to ensure the highest number of phosphopeptides, but also the phosphorylation sites need to be localized to a specific site, requiring high-quality fragmentation spectra, which generally requires slower scan rates, compromising the number of phosphopeptide identifications.<sup>10, 32</sup>

To determine the ability of the phosphopeptide enrichment method to recovery specific phosphopeptides, the MRM assay that we established for the 4 non-phosphopeptides and the 3 phosphopeptides was used to quantify the recovery of phosphopeptides and the lack of recovery of non-phosphopeptides. The 4 non-phosphopeptides and 3 phosphopeptides were spiked into a peptide sample and subjected to phosphopeptide enrichment. After the phosphopeptide enrichment, the samples were dried. The resulting samples were resuspended, and the phosphopeptides and peptides were quantified using the established MRM assay (Figure 1.6).

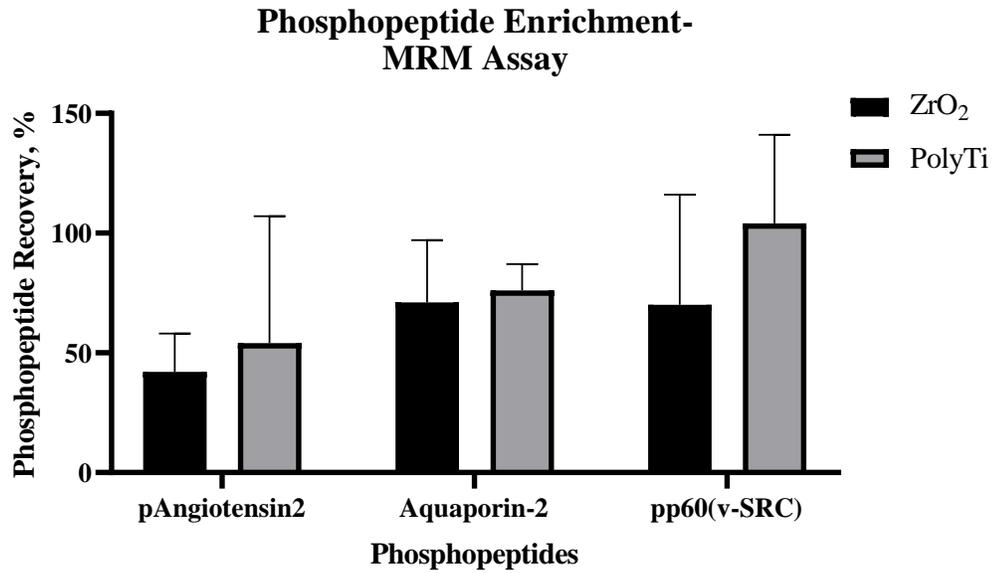


Figure 1.6. Phosphopeptide recovery data from MRM assay from peptide samples subjected to phosphopeptide enrichment extraction using different resins (ZrO<sub>2</sub> and PolyTi, N=8 per resin type).

Using the original phosphopeptide enrichment workflow, there was considerable variation in the amount of recovered phosphopeptides for both resin types with an average %CV of 29% for ZrO<sub>2</sub> and 34% for PolyTi. Based on these results, significant optimization in terms of reproducibility was needed to ensure a robust phosphopeptide enrichment protocol that can be applied for the MRM assay. Using the MRM assay, none of the four non-phosphopeptides were detected after the phosphopeptide enrichment using either resin types and none were not detected in any phosphopeptide enriched samples in subsequent experiments.

#### ***1.2.4 Optimization of phosphopeptide enrichment***

The overall goal of this work was to establish a phosphopeptide enrichment that was robust and reliable. For global phosphoproteomic analysis, robustness and reliability are measured by number of phosphopeptide IDs, phosphopeptide specificity, and sample-to-sample reproducibility. For targeted phosphoproteomic analysis, robustness and reliability are measured by phosphopeptide recovery and sample-to-sample reproducibility. The progress of the optimization of the phosphopeptide enrichment was measured using both global and target phosphoproteomic analyses. The optimization of the phosphopeptide enrichment was done by altering the following parameters: resin:cell lysate ratio, sample, washing and elution buffers, aspiration/dispense cycle number for each extraction step, and aspiration/dispense speeds at each extraction step. Each of these parameters was optimized for the phosphopeptide enrichment and only a summary of this optimization process is presented here. To begin the optimization for the phosphopeptide enrichment, a fixed quantity of resin was used with varying amounts of cell lysate to

determine if we achieved higher phosphopeptide identifications as we increased the starting material (Table 1.4).

The data showed no significant improvement across different amounts of protein digest starting material for a set amount of resin. This is probably due to the saturation of the binding sites on the resin. The upper range of digest amount (400-1600  $\mu\text{g}$ ) would be impractical in clinical settings, therefore for all subsequent experiments, 100-200  $\mu\text{g}$  of peptide starting material were used. During the optimization portion of this study, clogging issues occasionally occurred for both the  $\text{ZrO}_2$  and PolyTi samples. We hypothesized that trace amounts of glycolic acid were being carried over into the eluted samples, which clogged the LC column. This problem was remedied by an additional desalting step after phosphopeptide enrichment; however, this would introduce an additional step to the overall workflow, increasing the overall processing time and the potential for the loss of material. An alternative phosphopeptide enrichment buffer set was developed in place of buffer set 1, here referred to as buffer set 2 (Table 1.5). Buffer set 2 was based on previous work by Zhou *et al.* that showed that high concentrations of TFA (5-10%) in the sample binding buffer leads to high phosphopeptide specificity and the inclusion of 200 mM NaCl during the wash steps helped remove the non-specifically bound non-phosphopeptides.<sup>33, 34</sup> The two different buffer sets were compared for overall phosphopeptide identifications and phosphopeptide specificity (Figures 1.7-1.8). Samples generated using buffer set 2 outperformed samples generated using buffer set 1 in terms of phosphopeptide specificity and phosphopeptide identifications. Also, samples prepared using buffer set 2 had no issues with clogging whereas samples prepared with buffer set 1

Table 1.4. Resin testing for determining resin capacity. 5 mg of resin was used.

<b>Resin Type</b>	<b>Protein Digest Amount</b>	<b>Protein Ids, #</b>	<b>Phosphopeptide Ids, #</b>	<b>Phosphopeptide Specificity, %</b>	<b>MS/MS Scans</b>
ZrO <sub>2</sub>	200 µg	1070	1902	97%	8155
ZrO <sub>2</sub>	400 µg	1220	2074	89%	8634
ZrO <sub>2</sub>	800 µg	1308	2126	84%	9733
ZrO <sub>2</sub>	1600 µg	1180	2263	96%	10103
PolyTi	200 µg	883	1389	89%	7859
PolyTi	400 µg	829	1473	97%	7632
PolyTi	800 µg	958	1615	88%	8748
PolyTi	1600 µg	911	1378	86%	8900

Table 1.5. Phosphopeptide enrichment process and buffer set. Referred to as buffer set 2.

<b>Phosphopeptide Enrichment Workflow</b>			
<b>Steps</b>	<b>Process</b>	<b>Solvent</b>	<b>Sample, <math>\mu\text{L}</math></b>
1	Activation	100% ACN	400
2	Condition	80% ACN	400
3	Equilibration	80% ACN, 6% TFA	400
4	Sample	80% ACN, 6% TFA	200
5	Wash I	80% ACN, 6% TFA	400
6	Wash II	50% ACN, 6% TFA, 200 mM NaCl	400
7	Wash III	30% ACN, 0.1% TFA	400
8	Elution I	1.5% $\text{NH}_4\text{OH}$	100
9	Elution II	1.5% $\text{NH}_4\text{OH}$	100

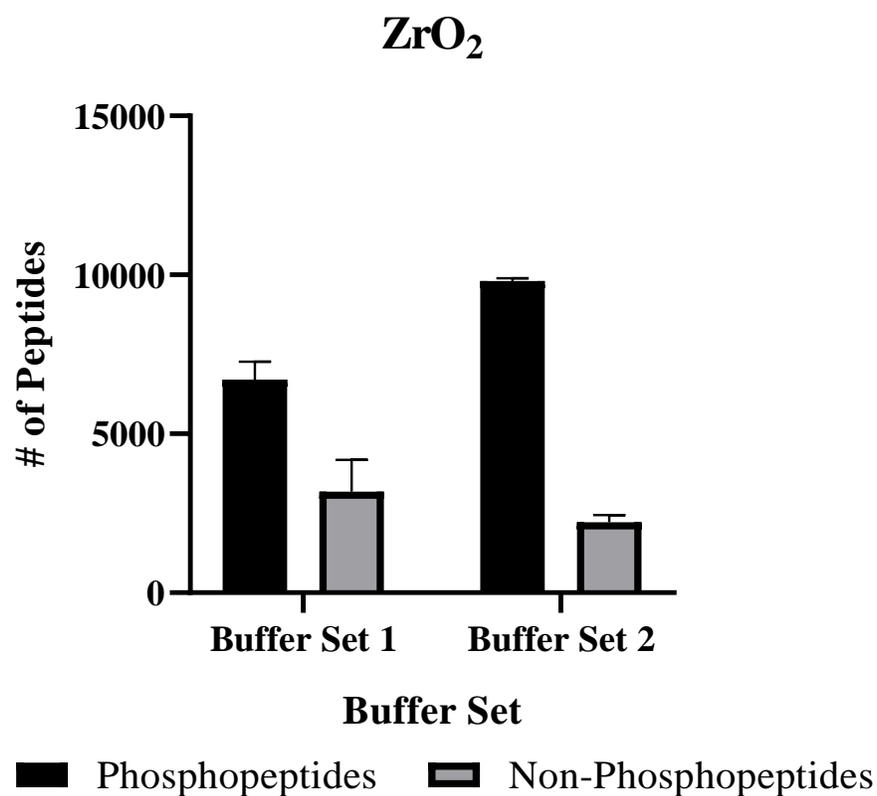


Figure 1.7. Comparison of number of phosphopeptides and non-phosphopeptides identified by two different workflows using ZrO<sub>2</sub> (n = 2, data are mean ± s.d.). The buffer set 1 is the glycolic acid containing buffer set and the buffer set 2 is the current buffer set. Sample analysis was done using the Orbitrap Fusion.

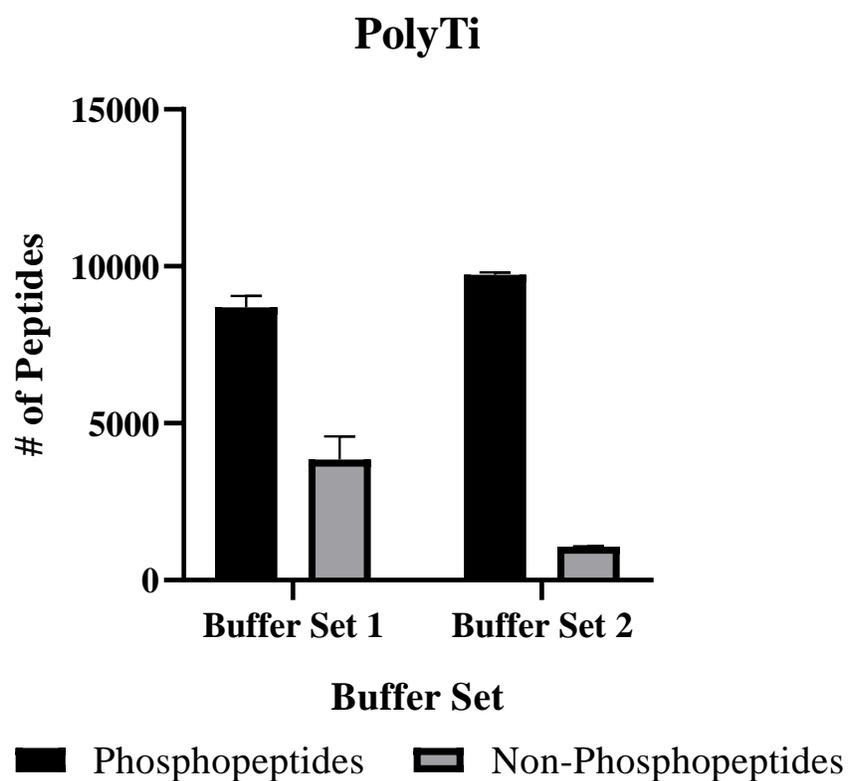


Figure 1.8. Comparison of number of phosphopeptides and non-phosphopeptides identified by two different workflows using PolyTi (n = 2, data are mean  $\pm$  s.d.). The buffer set 1 is the glycolic acid containing buffer set and the buffer set 2 is the current buffer set. Sample analysis was done using the Orbitrap Fusion.

consistently caused higher pressure when injected onto the LC or clogged the LC altogether. Buffer set 2 was used for all subsequent experiments.

One parameter that affected recovery in the MRM assay was neutralizing the sample using formic acid and injecting directly instead of vacuum drying the sample after elution. When the samples were dried for the MRM assay, the recoveries were highly variable (CVs greater than 30%), while neutralizing the sample and skipping the dry down step decreased the CVs to less than 20% (Table 1.6). Loss of peptides during the drying step could be due to the peptides having increased time to bind to the plastics and not being recovered during the resuspension process or by simply eliminating other variables associated with drying and reconstituting the sample. Additionally, the inclusion of 50% acetonitrile in the elution buffer increased phosphopeptide recovery for the MRM assay (Table 1.7).

Next, the mixing steps for each step in the enrichment were altered. Increasing the aspiration speed from 30  $\mu\text{L/s}$  to 60  $\mu\text{L/s}$ , adjusting dispense speeds from 5  $\mu\text{L/s}$  to 15  $\mu\text{L/s}$ , and increasing the cycle number from 3 to 5 during the wash steps led to an average increase of phosphopeptide specificities by 27% for both PolyTi and  $\text{ZrO}_2$  (Figure 1.9). To determine if a second elution step was necessary, a comparison between the phosphopeptide samples eluted with one elution step versus samples eluted with two elution steps was performed. This experiment used buffer set 2 and used 1.5% Ammonium Hydroxide, 50% ACN as the elution buffer (Figures 1.10). A second elution step increased the number of phosphopeptide identifications by 160% and increased the overall phosphopeptide specificity for global analysis. Together the optimization process led to significant improvements in the phosphopeptide enrichment process in terms of

Table 1.6 Comparison of phosphopeptide recoveries and associated CVs for each phosphopeptide from different resin types (n = 8 per resin type) when the eluted samples are not dried versus when dried.

		<b>pAngiotensin2</b>		<b>Aquaporin-2</b>		<b>pp60(v-SRC)</b>	
		<b>Average Recovery</b>	<b>% CV</b>	<b>Average Recovery</b>	<b>% CV</b>	<b>Average Recovery</b>	<b>% CV</b>
<b>Resin Type</b>							
<b>No Dry Down</b>	ZrO <sub>2</sub>	29%	28%	70%	6%	52%	10%
	PolyTi	62%	6%	73%	3%	61%	7%
<b>Dry Down</b>	ZrO <sub>2</sub>	42%	16%	71%	26%	70%	46%
	PolyTi	54%	53%	76%	11%	104%	37%

Table 1.7. Summary of phosphopeptide recoveries and associated CVs for each phosphopeptide from different resin types (n = 4 per resin type) when using elution buffer 1 or elution buffer 2. Elution buffer 1 is 1.5% Ammonium Hydroxide. Elution buffer 2 is 1.5% Ammonium Hydroxide, 50% ACN.

Phosphopeptides	Elution Buffer 1				Elution Buffer 2			
	PolyTi		ZrO <sub>2</sub>		PolyTi		ZrO <sub>2</sub>	
	% Recovery	% CV						
pAngiotensin2	2%	82%	40%	14%	41%	10%	85%	6%
Aquaporin-2	1%	25%	56%	3%	59%	8%	80%	4%
pp60(v-SRC)	5%	26%	54%	6%	61%	12%	81%	7%

## Difference in Phosphopeptide Specificity: Wash Condition 1 vs Wash Condition 2

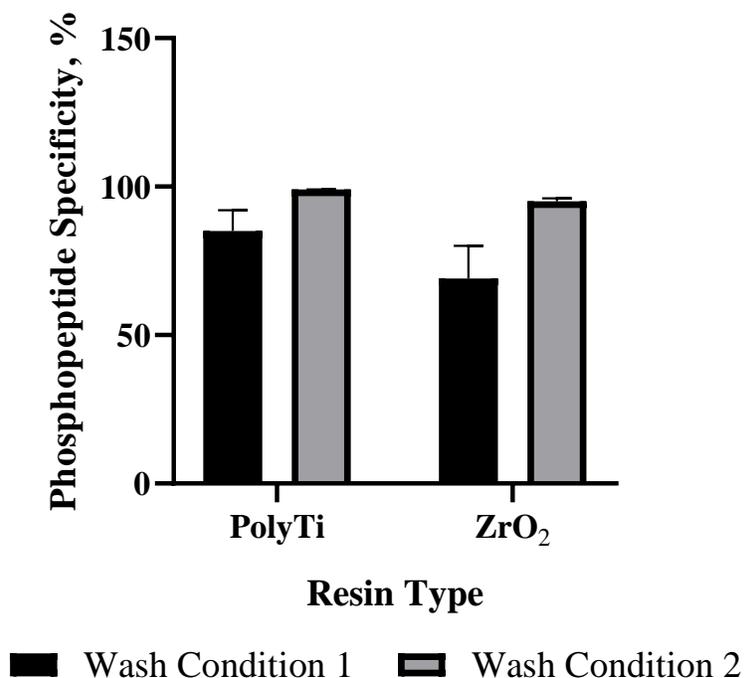


Figure 1.9. Comparison of phosphopeptide specificities between Wash Condition 1 and Wash Condition 2 for both PolyTi and ZrO<sub>2</sub> (n = 2 per resin type, data are mean  $\pm$  s.d.). Wash Condition 1: 3 cycles at 30  $\mu$ L/s aspiration speed and 5  $\mu$ L/s dispense speed for each wash solution; Wash Condition 2: 5 cycles at 60  $\mu$ L/s aspiration speed and 15  $\mu$ L/s dispense speed for each wash solution. Sample analysis was done using the LTQ-Orbitrap Velos Pro.

### Comparison of Extraction Features: 1 Elution vs 2 Elutions

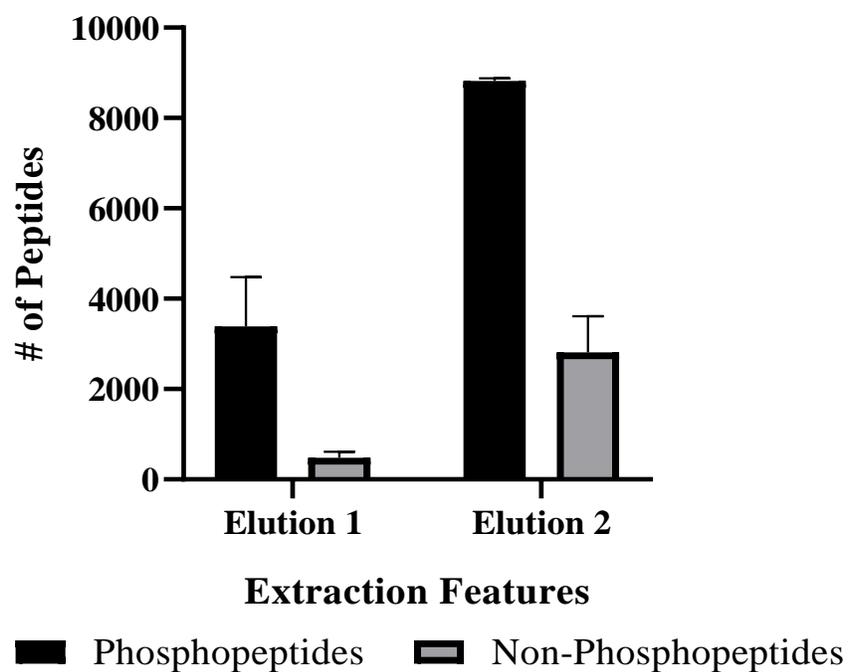


Figure 1.10. Comparison of number of phosphopeptides and non-phosphopeptides identified by two different workflows ( $n = 2$ , data are mean  $\pm$  s.d.). The first workflow contained only one elution step and the second workflow contained two elution steps. Sample analysis was done using the Orbitrap Fusion.

phosphopeptide identifications and phosphopeptide specificity for global analysis and increased recovery and reproducibility for targeted analysis. The optimization process also highlighted the benefits of the IMCStips as control of specific aspiration and dispense speeds have significant effects on the efficiency of the phosphopeptide enrichment. The final optimized automated desalting and phosphopeptide enrichment workflow is detailed in Table 1.8 with specific aspiration/dispense speeds, volumes, buffers, and the time associated with each step of the extraction process.

This final automated desalting and phosphopeptide enrichment process was then compared to a standard Spin Tip workflow (Table 1.9). As part of the standardization process, the entire workflow was unified with one single cell type, human colon cancer cell line (HCT116). The cell culture, lysis, protein precipitation, and digested were all performed identically in bulk prior to splitting the samples into multiple aliquots for comparison of Spin Tips and IMCStips. The selection of colon cancer cell line as the initial starting point is due to the relatively well-established characterization of the cell line and the need for more effective colorectal cancer screens.<sup>35</sup> A recent report highlighted that higher incidence rates of colon cancer for adults age 20 to 39 than adults aged 40-54 years old raises concerns for the well-being of the general public.<sup>36</sup> Current novel biomarker research focuses on genetic, epigenetic, and protein markers as new screening modalities for early detection. This study is envisioned as an additional toolkit to expand that search for biomarkers, especially focusing on aberrant post-translational modifications involving phosphorylation events.

Table 1.8. Optimized automated desalting and phosphopeptide enrichment workflow.

IMCStip	Process	Solvent	Aspirate, $\mu\text{L}$	Cycle Number	Aspirate/Dispense Speed, $\mu\text{L}/\text{sec}$	Sample, $\mu\text{L}$	Time, mins
RP	Activation	100% ACN	250	1	30/15	400	1
RP	Equilibration	1% TFA	250	3	30/15	400	2
RP	Sample	1% TFA	200	10	30/15	250	10
RP	Wash 1	1% TFA	250	3	30/15	400	2
RP	Wash 2	0.5% Formic Acid	250	3	30/15	400	2
RP	Elution	80% ACN, 6% TFA	200	5	30/15	200	3
Phospho.	Activation	100% ACN	250	1	30/15	400	1
Phospho.	Condition	80% ACN	200	3	30/15	400	3
Phospho.	Equilibration	80% ACN, 6% TFA	200	5	30/15	400	3
Phospho.	Sample	80% ACN, 6% TFA	180	40	30/10	200	30
Phospho.	Wash I	80% ACN, 6% TFA	200	3	60/15	400	3
Phospho.	Wash II	50% ACN, 6% TFA, 200 mM NaCl	200	3	60/15	400	3
Phospho.	Wash III	30% ACN, 0.1% TFA	200	3	60/15	400	3
Phospho.	Elution I	1.5% $\text{NH}_4\text{OH}$ , 50% ACN	80	5	30/5	100	3
Phospho.	Elution II	1.5% $\text{NH}_4\text{OH}$ , 50% ACN	80	5	30/5	100	3

Table 1.9. Phosphopeptide enrichment process using the Spin Tips from GL Sciences and Thermo Scientific. For each step the specified solvent and volume was added to each Spin Tip. The Spin Tips were then centrifuged and the flow through was discarded. Elution 1 and elution 2 were collected, combined, and lyophilized. The dried samples were then stored at -80 °C until LC-MS/MS analysis.

<b>Process</b>	<b>Solvent</b>	<b>Volume, <math>\mu</math>L</b>	<b>Centrifuge Speed, <math>g</math></b>	<b>Time, mins</b>
Condition	100% ACN	200	3000	2
Equilibration	80% ACN, 6% TFA	200	3000	2
Sample (First Pass)	80% ACN, 6% TFA	200	1000	10
Sample (Second Pass)	80% ACN, 6% TFA	200	1000	10
Wash I	80% ACN, 6% TFA	200	3000	2
Wash II	80% ACN, 6% TFA	200	3000	2
Wash III	50% ACN, 6% TFA, 200 mM NaCl	200	3000	2
Wash IV	50% ACN, 6% TFA, 200 mM NaCl	200	3000	2
Wash V	30% ACN, 0.1% TFA	200	3000	2
Wash VI	30% ACN, 0.1% TFA	200	3000	2
Elution I	1.5% NH <sub>4</sub> OH, 50% ACN	100	1000	5
Elution II	1.5% NH <sub>4</sub> OH, 50% ACN	100	1000	5

### *1.2.5 Comparison of automated phosphopeptide enrichment to conventional method*

The automated workflow which coupled desalting and phosphopeptide enrichment required less than 10 minutes of hands-on time to place reagents on the system. After 2 hours, the eluates were ready for injection. Spin Tips required more manual interventions with several centrifugation steps, and the entire processing time was completed in similar time frames for approximately 8 samples. Increasing the sample number increased the processing time, whereas the time commitment for automated method was the same for 1 to 96 samples.

Each sample aliquot for analysis by Orbitrap Fusion comprised of 200 µg of peptide from the HCT116 digests. For the PolyTi and ZrO<sub>2</sub> resins in IMCStip, the average phosphopeptide identifications were 10,191 and 9,796, respectively (Figure 1.11A). The average phosphopeptide identifications from Thermo and GL Biosciences Spin Tips were 7,408 and 5,902, respectively (Figure 1.11A). The average specificities for PolyTi ranked highest at 93%, ZrO<sub>2</sub> at 86%, and the two Spin Tips were significantly lower at 58% for Thermo Scientific and 32% for GL Biosciences (Figure 1.11B). The differences may be attributed to resin chemistries and/or the difference between dSPE in tips versus the Spin Tips. Some of these differences could not be examined as resins from Thermo and GL passed through the bottom filter of IMCStips. Hence, the PolyTi resin from IMCStips were transferred to a spin column format. The results from the PolyTi resins filled in spin column and IMCStips were comparable in terms of phosphopeptide identifications, but the specificities from the spin column were significantly lower than tip extraction (Figure 1.11C, D). This result suggests that PolyTi resins are superior to

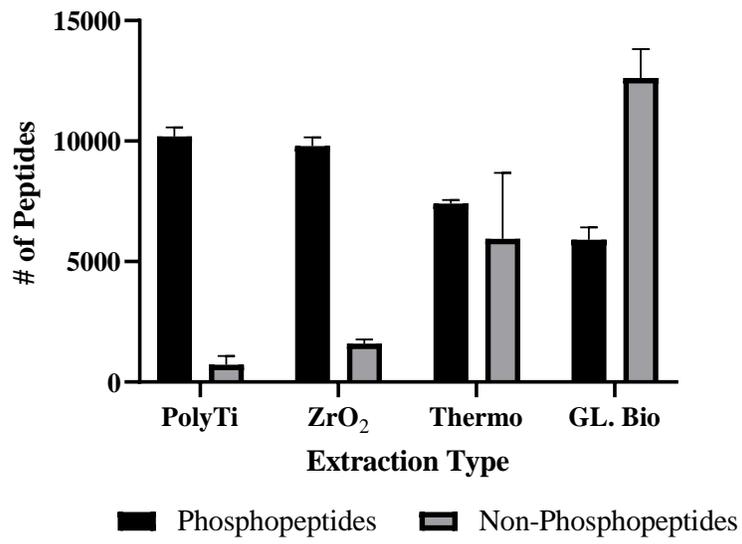
TiO<sub>2</sub> as the spin column containing PolyTi had higher phosphopeptide IDs compared to the Spin Tip products from GL and Thermo. The extraction using dSPE in the IMCStip also outperform the spin column format within the scope of the current study. Such a difference in specificities between dSPE and spin column is intriguing and warrants further study in the future using expanded portfolio of resins and different analytes. A summary of the results from each can be found in Table 1.10-1.11.

Reproducibility of phosphopeptide enrichment methods was assessed using two biochemical replicates. Technical reproducibility was obtained by injecting a single sample that was divided in half and injected sequentially. The peptide overlap for the technical replicate from this “split” sample was 63% (Figure 1.11E). This overlap percentage is the maximum overlap, where a significant portion (~40%) of the sample-to-sample differences are due to the limitations of the data dependent acquisition (DDA) mass spectrometry. The peptide overlap between biochemical replicates for the two resin types was 60.4% and 55.2% for PolyTi and ZrO<sub>2</sub>, respectively (Figure 1.11E). For comparison, the peptide overlap between Thermo Spin Tip biochemical replicates and GL Spin Tip biochemical replicates was 44.1% and 60.2%, respectively (Figure 1.11F).

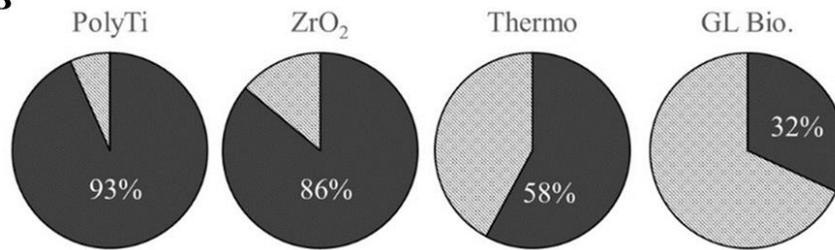
### ***1.2.6 Phosphopeptide enrichment for MRM assay applications***

Given the intrinsic limitation of DDA, the reproducibility of the established automated phosphopeptide enrichment was further assessed using an in-house MRM assay with three phosphopeptides and three non-phosphopeptides. For this initial work, several commercially available phosphopeptides were screened, but these three phosphorylated peptides provided sufficient MS signal intensities.

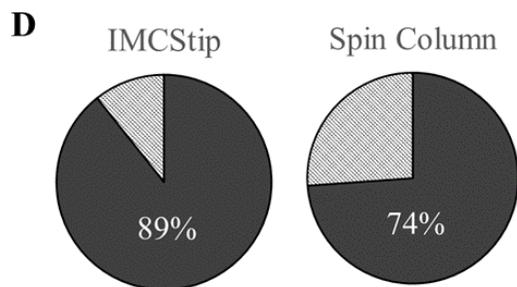
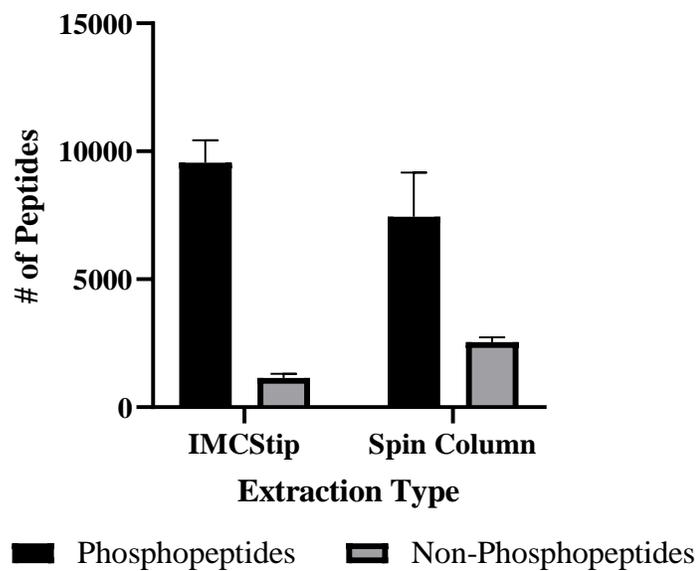
### A Phosphopeptide Specificity- IMCS vs Spin Tip



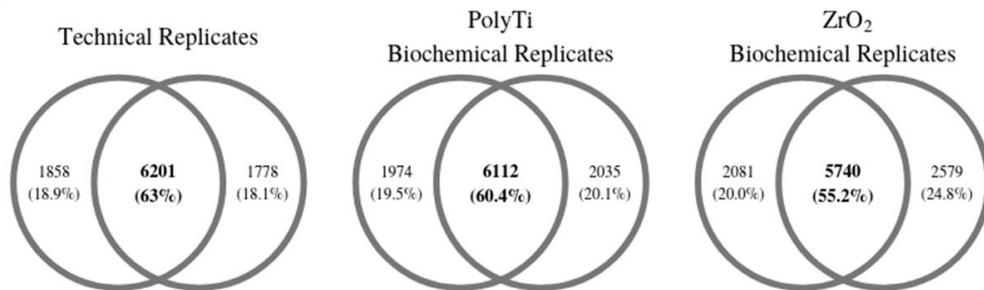
### B



### C Efficiency of Phosphopeptide Enrichment



### E



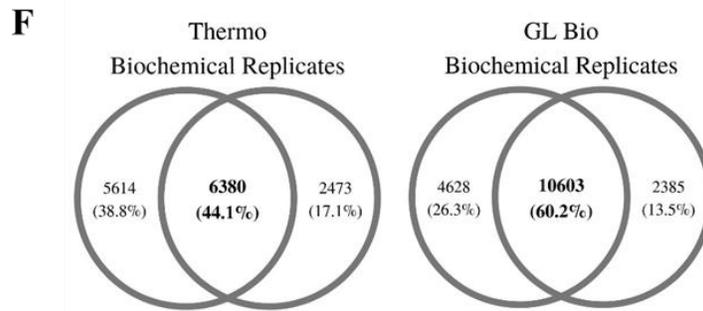


Figure 1.11. Comparison of phosphopeptide enrichment methods. Two biochemical replicates were processed for each extraction type. A) Average number of phosphopeptides and non-phosphopeptides identified by PolyTi and ZrO<sub>2</sub> represent automated extractions using IMCStips. Thermo and GL Bio are two different vendors using TiO<sub>2</sub> in Spin Tips. B) Average specificities by resin type, where higher specificities are indicative of more efficient extraction as MS scans are dedicated to identifying phosphopeptides instead of non-phosphorylated peptides. C) Average number of phosphopeptides and non-phosphopeptides identified using IMCStips and Spin Columns both using PolyTi resin. For this experiment 200 μg of untreated HCT116 cell digest was used as the starting material for both the IMCStip and the spin column samples. D) Average phosphopeptide specificities for the IMCStip and the Spin Column method, both of which used PolyTi resin. E) Peptide overlap between samples analyzed by the orbitrap fusion. Peptide overlap between technical replicates of a phosphopeptide sample. Peptide overlap between biochemical replicates that were enriched for phosphopeptides using PolyTi and ZrO<sub>2</sub>. F) Peptide overlap between samples processed using the Spin Tips and analyzed by the orbitrap fusion. Peptide overlap between biochemical replicates that were enriched for phosphopeptides using the Thermo Spin Tip and using the GL Bio Spin Tip.

Table 1.10. Summary of data from phosphopeptide enrichment of 200  $\mu\text{g}$  of HCT116 cell lysate. Sample analysis was done using the Orbitrap Fusion.

<b>Resin Type</b>	<b>IMCStip or Spin Tip</b>	<b>Protein Ids, #</b>	<b>Peptide Ids, #</b>	<b>Phosphopeptide Ids, #</b>	<b>Phosphopeptide Specificity, %</b>	<b>MS/MS Scans</b>
PolyTi	IMCStip	2976	10941	10570	97%	62803
PolyTi	IMCStip	3088	10890	9812	90%	64310
ZrO <sub>2</sub>	IMCStip	3197	11928	10160	85%	64514
ZrO <sub>2</sub>	IMCStip	3046	10855	9432	87%	68148
Thermo	Spin Tip	3738	10759	7557	70%	55413
Thermo	Spin Tip	4314	15936	7260	46%	74705
GL Bio	Spin Tip	4780	20235	6413	32%	76860
GL Bio	Spin Tip	4407	16794	5391	32%	71133

Table 1.11. Summary of data from phosphopeptide enrichment of 200 µg of untreated HCT116 cell lysate using 10 mg of PolyTi resin. Sample analysis was done using the Orbitrap Fusion.

<b>IMCStip or Spin Column</b>	<b>Protein Ids, #</b>	<b>Peptide Ids, #</b>	<b>Phosphopeptide Ids, #</b>	<b>Phosphopeptide Specificity, %</b>	<b>MS/MS Scans</b>
IMCStip	3029	9675	8687	90%	60671
IMCStip	3244	11736	10425	89%	61195
Spin Column	2858	8079	5706	71%	51512
Spin Column	3466	11897	9172	77%	60400

For three peptides, PKC substrate 4 from Anaspec, MAPK1 from New England Peptide, and MAPK14 from New England Peptide, the signal intensities were unreliable and were therefore not used in the assay. The sequence of the six peptides that were used for the MRM assay are listed in Table 1.2 along with the precursor and daughter ions. These peptides were spiked into a matrix of digested bovine serum albumin and subjected to automated phosphopeptide enrichment. The enrichment was done using the two different resins (n = 8 per resin) once a week over a 3-week period. From these experiments, the average recoveries for all phosphopeptides from both resin types were greater than 50% (Figure 1.12). The inter-plate and intra-plate averages and associated CVs for each resin type can be found in the supplemental information (Table 1.12-1.13). Among the two resins, ZrO<sub>2</sub> provided higher recoveries than PolyTi with inter-plate averages of 77%, 73%, and 63% for pAngiotensin2, Aquaporin-2, and pp60 (v-SRC), respectively (Table 1.13). The %CVs for the resins were below 20% (Table 1.13). The higher recoveries with the ZrO<sub>2</sub> resin for the MRM assay may be due to the resin's preferential recovery of singly-phosphorylated peptides compared to the PolyTi resin as all three phosphopeptides used in the MRM-assay were singly-phosphorylated. This observation is in line with the global analysis data, where the ZrO<sub>2</sub> contained 10-15% more singly-phosphorylated phosphopeptide identifications compared to the PolyTi resin (Figure 1.13).

### ***1.2.7 Comparison of phosphopeptide profiles from different resins***

Zirconia, titania and iron are common metals used for phosphopeptide enrichment. Since zirconia and titania-based chemistries were available for dSPE in tips, the differences between the two resins using the same automated workflow were further investigated.

## Phosphopeptide Recovery From MRM Assay

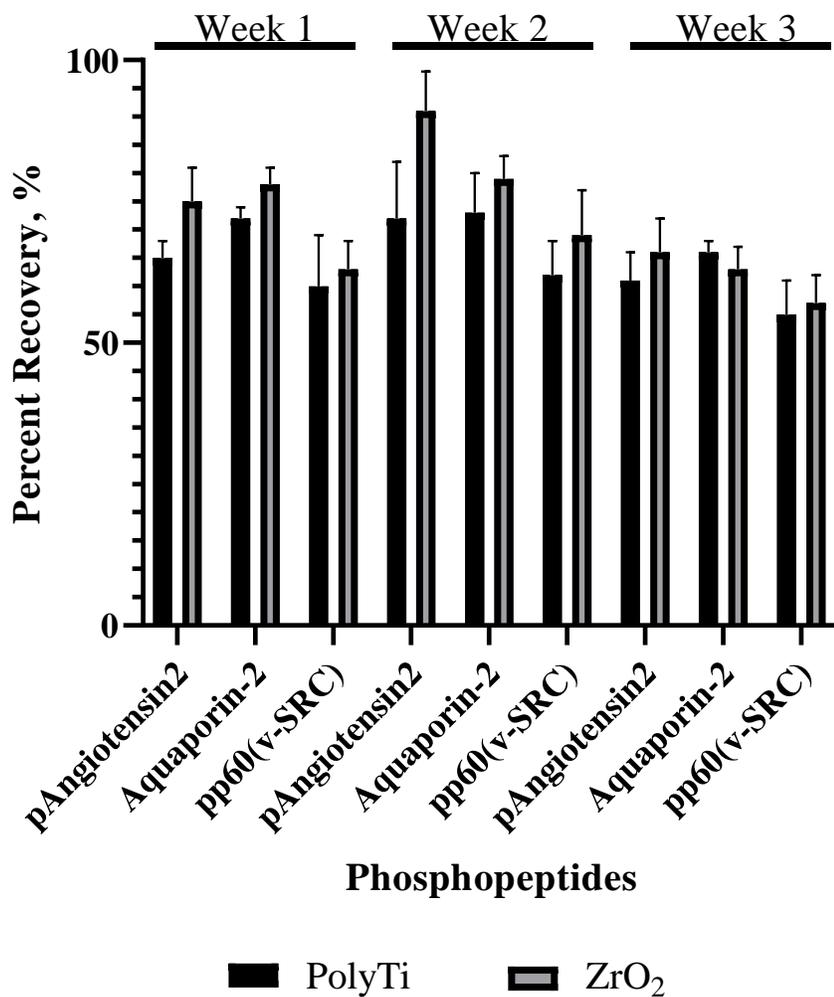


Figure 1.12. The average recovery for each phosphopeptide from three different extractions performed once a week over 3-weeks using two different resins (n = 8 per resin type).

Table 1.12. Summary of phosphopeptide recoveries (% Recov.) and associated CVs (% CV) for each phosphopeptide from different resins for the 3-week experiment (n = 8 per resin type per week).

Phosphopeptides	Week 1				Week 2				Week 3			
	PolyTi		ZrO <sub>2</sub>		PolyTi		ZrO <sub>2</sub>		PolyTi		ZrO <sub>2</sub>	
	% Recov.	% CV	% Recov.	% CV	% Recov.	% CV	% Recov.	% CV	% Recov.	% CV	% Recov.	% CV
pAngiotensin2	65%	4%	75%	7%	72%	14%	91%	7%	61%	8%	66%	9%
Aquaporin-2	72%	3%	78%	4%	73%	9%	79%	5%	66%	4%	63%	6%
pp60(v-SRC)	60%	16%	63%	8%	62%	9%	69%	12%	55%	11%	57%	9%

Table 1.13. Summary of average inter-week phosphopeptide recoveries and associated CVs for each phosphopeptide from different resins for the 3-week experiment (n = 24 per resin type).

Phosphopeptides	Phosphopeptide Recovery: Inter-week Average Recovery		Phosphopeptide Recovery: Inter-week CVs	
	PolyTi	ZrO <sub>2</sub>	PolyTi	ZrO <sub>2</sub>
pAngiotensin2	66%	77%	12%	16%
Aquaporin-2	70%	73%	8%	11%
pp60(v-SRC)	59%	63%	13%	13%

## Singly- and Multi- Phosphorylated Peptides

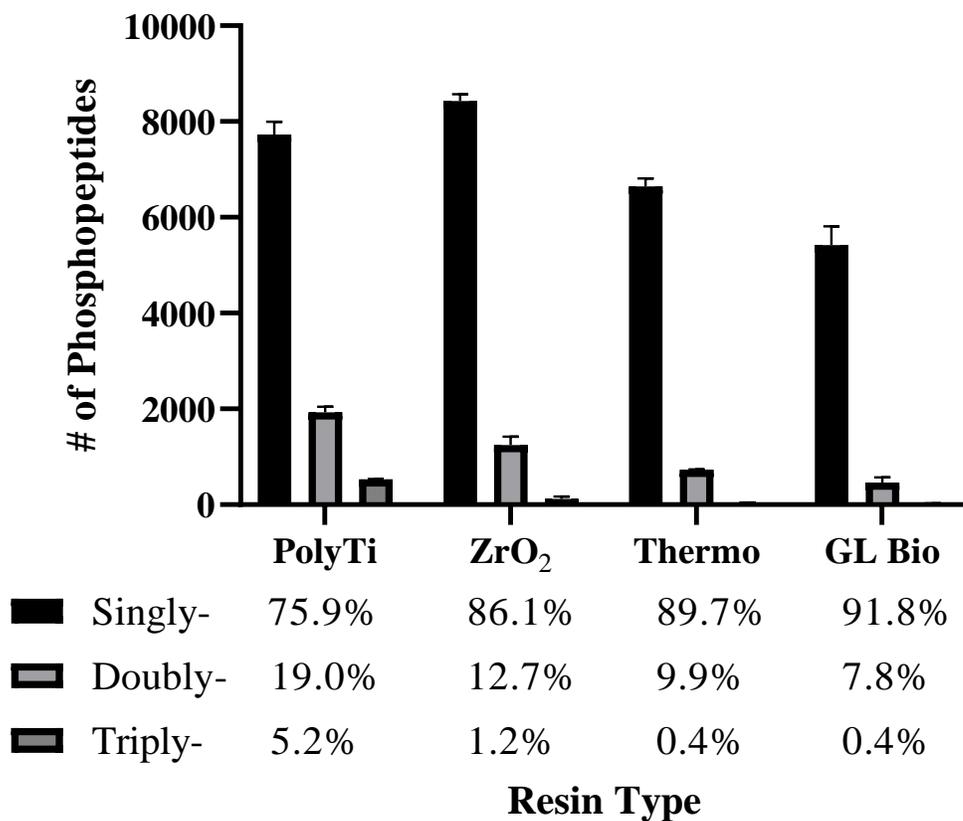


Figure 1.13. The distribution of singly-, doubly-, triply- phosphorylated peptides identified from 200 µg of HCT116 digest from the different extractions.

The pool of singly, doubly and triply phosphorylated peptides from the two different resins were minor, where 86% of phosphopeptides identified with the zirconia resin were singly-phosphorylated compared to 75% for PolyTi resin. For PolyTi resin, on average, 19% of the phosphorylated peptides identified were multiply-phosphorylated compared to 13% for ZrO<sub>2</sub> and less than 10% for both TiO<sub>2</sub>-based Spin Tips (Figure 1.13). Prior studies have also reported a selective bias of zirconia towards singly phosphorylated peptides over the doubly and multiply-phosphorylated peptides, likely due to the strong interaction between zirconia and phosphates.<sup>37</sup> From previous data using a glycolic acid containing buffer set and the LTQ-Orbitrap Velos Pro for sample analysis, the phosphopeptides enriched using PolyTi showed nearly equal enrichment of singly- and doubly- phosphorylated peptides (Figure 1.14). Additionally, a comparison in the number of singly-, doubly-, and triply-phosphorylated peptides identified using PolyTi with the current buffer set and the previous glycolic acid containing buffer set was done using the Orbitrap Fusion for sample analysis (Figure 1.15). From this comparison, a slight decrease in the average number of doubly-phosphorylated peptides was observed, from 37% to 19%, when using the current buffer set. This decrease in recovery of doubly phosphorylated could be attributed to a decrease in pH as the current buffer set uses a higher concentration of TFA creating an environment that could be more favorable toward binding singly-phosphorylated peptides to the resin. The peptide overlap between PolyTi and ZrO<sub>2</sub> (pooled peptide identifications from two biological replicates for both resin types), showed a slight decrease in peptide overlap compared to the biological replicates (Figure 1.16A). However, these differences in overlap could be due to a variety of factors, one being the inherent limitations of data-dependent acquisitions.

### Comparison of Average number of singly-, doubly-, triply-phosphorlated peptides

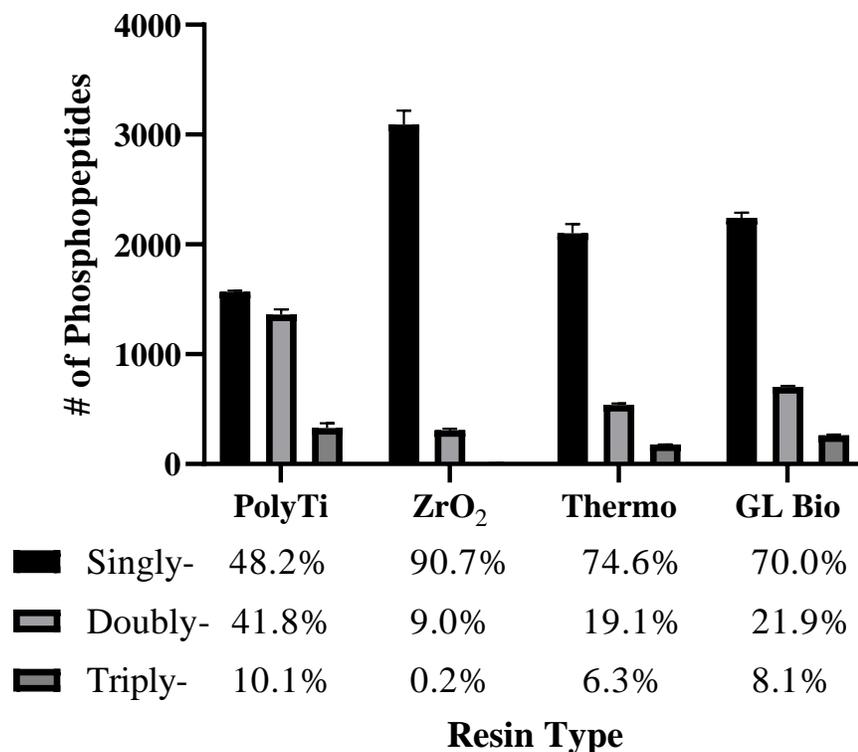


Figure 1.14. Number of singly-, doubly-, and triply-, phosphorylated peptides identified by four different extractions with glycolic acid containing buffer set (n = 2 per resin type, data are mean ± s.d.). Sample analysis was done using the LTQ-Orbitrap Velos Pro.

## Differentially Enriched Phosphopeptides When Using Different Buffers

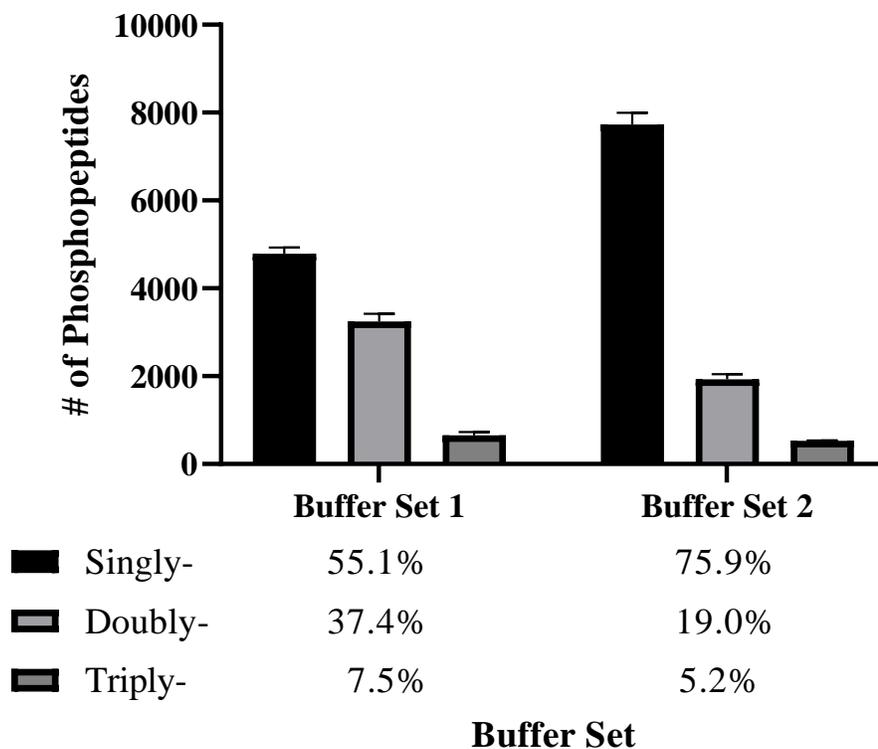


Figure 1.15. Comparison of types of phosphopeptides enriched by PolyTi when using two different buffer sets (n = 2, data are mean ± s.d.). The first buffer set contains glycolic acid while the second buffer set contains a higher concentration of TFA. Sample analysis was done using the Orbitrap Fusion.

The hydrophobicity (GRAVY score) and the isoelectric point (pI value) for all identified phosphopeptides were determined to further detect any differences between the two resins. Previous reports have indicated that IMAC based resins enrich more hydrophilic phosphopeptides and metal oxide-based resins enrich more acidic phosphopeptides.<sup>38</sup> However, this analysis showed no significant differences between PolyTi and ZrO<sub>2</sub>, in which the distribution of GRAVY scores and pI values of identified phosphopeptides are similarly aligned (Figure 1.16B, C). In general, positive (+) GRAVY scores are associated with peptides with overall hydrophobic character, whereas negative (−) GRAVY scores indicate hydrophilic character, suggesting that both resins preferentially enrich hydrophilic phosphopeptides. Both resins also show majority (> 50%) of the identified phosphopeptides are acidic (pI < 6), but still ~ 30% of the identified phosphopeptides have basic isoelectric values > 8.

As summary of the results with the two phosphopeptide enrichment resins is presented in Table 1.14. Based on the analysis, we find that the only differences in the phosphopeptides identified by PolyTi and ZrO<sub>2</sub> resins are differences in numbers of singly- versus multiply- phosphorylated peptides. However, this observation could be highly dependent on variations in the sample binding environment as well as the chromatography and ionization conditions used during the LC-MS/MS analysis. These differences in types of phosphopeptides between the resins are intriguing, and further systematic investigation is needed to understand these differences and how such subtle differences could be exploited for future studies.

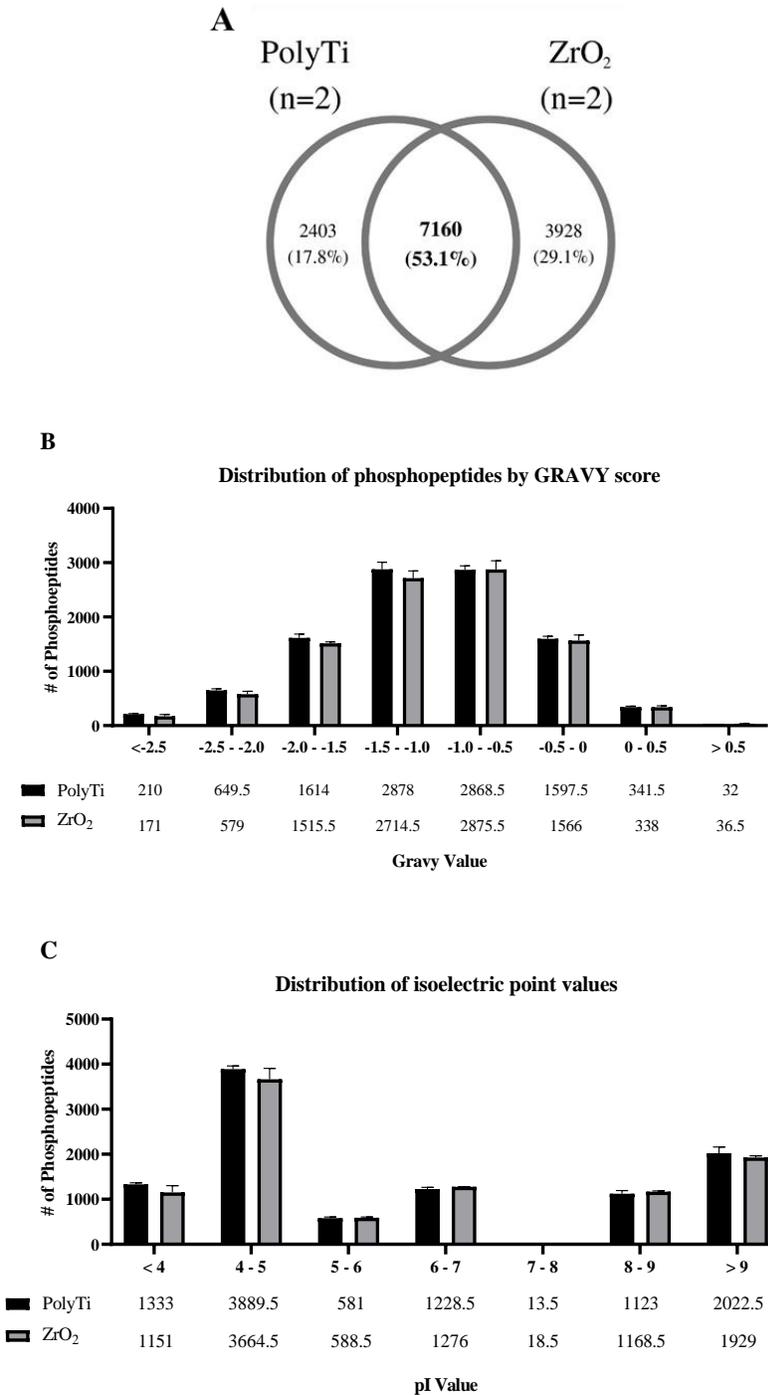


Figure 1.16. Comparison of phosphoprofile from different resin types. A) Peptide overlap between biochemical replicates (n = 2 per resin type) that were enriched for phosphopeptides using PolyTi or ZrO<sub>2</sub>. B) The average distribution of gravity values for all phosphorylated peptides identified by both PolyTi and ZrO<sub>2</sub>. C) The average distribution of pI values for all phosphorylated peptides identified by both PolyTi and ZrO<sub>2</sub>.

Table 1.14. A summary table for the comparison between two different resins used in the IMCStip. The data shown was generated from phosphopeptide enrichments of 200 µg of HCT116 protein digest using two different resin types in the IMCStip using the optimized automated extraction protocol (n = 2 per resin type, data are mean ± s.d.).

	<b>IMCStip Resins</b>	
	<b>PolyTi</b>	<b>ZrO<sub>2</sub></b>
<b>Average Phosphopeptide Ids, #</b>	10191 ± 379	9796 ± 364
<b>Average Phosphopeptide Specificity, %</b>	93.3 ± 3.3 %	86.0 ± 0.9 %
<b>Sample to Sample Peptide Overlap, %</b>	60.4 %	55.2 %
<b>Average Multi-phosphorylated peptide, %</b>	24.1 ± 0.3 %	13.9 ± 1.8 %
<b>Average Peptide pI, #</b>	6.33 ± 2.58	6.43 ± 2.56
<b>Average Peptide Gravy Score, #</b>	-1.05 ± 0.66	-1.05 ± 0.65

### *1.2.8 IMCStip-based extraction method transferability*

One of the advantages of the IMCStip-based workflow is that the method can be applied to several different types of automated liquid handling systems. This allows for more board adoption of the workflow as labs with different automated liquid handling systems could apply the same workflow. To demonstrate the transferability of the automated desalting and phosphopeptide enrichment workflow, programmed our method on an Integra Assist system and compared the desalting and phosphopeptide enrichment to the Hamilton STAR system. For comparison, a spin column method was used using RP resin for the desalting portion and PolyTi resin for the phosphopeptide enrichment (Figure 1.17).

The samples generated using the IMCStips with either the Integra Assist or the Hamilton STAR showed high phosphopeptide identifications with  $11395 \pm 220$  phosphopeptide identifications from samples generated using the Integra Assist and  $10126 \pm 237$  phosphopeptide identifications from samples generated using the Hamilton STAR. Additionally, the phosphopeptide specificity for IMCStip-based extractions did not differ significantly and had an average specificity of 90%. In contrast, the phosphopeptide samples generated using spin column-based extraction showed lower phosphopeptide identifications compared to the IMCStip-based extraction with an average of  $7439 \pm 1733$  and showed lower phosphopeptide specificity with an average of 74%. Together these results highlight the efficiency of the IMCStip-based extraction in comparison to the spin column-based method and the transferability of the IMCStip-based extraction across different automated liquid handling systems.

**Phosphopeptide Enrichment using 10  $\mu$ L PolyTi Resin:  
IMCStip v Spin Column**

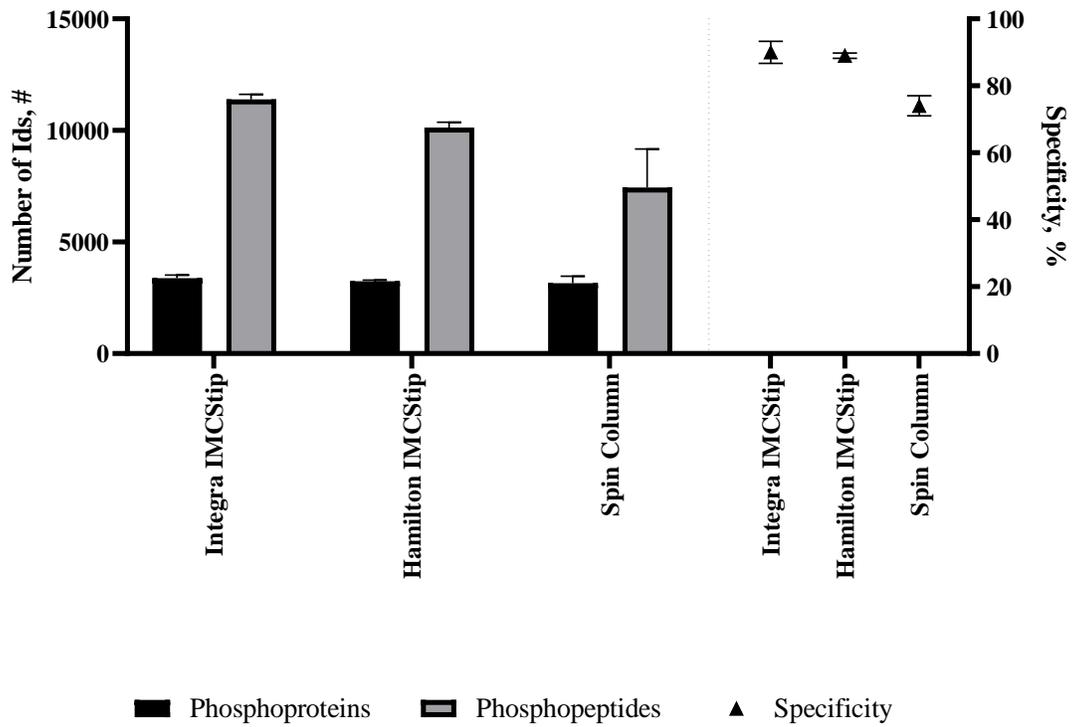


Figure 1.17. Summary of results from phosphopeptide enriched samples using either the Hamilton STAR or Integra Assist for the IMCStip-based extraction and a centrifuge for the spin column-based extraction. For the phosphopeptide enrichment, 10  $\mu$ L PolyTi resin was using for each extraction type (n = 3 per extraction type, data are mean  $\pm$  s.d.).

### 1.3 CONCLUSION

We established an automated workflow that couples desalting and phosphopeptide enrichment within 2 hours by leveraging dSPE that allows for reproducible desalting and phosphopeptide extraction. The total number of phosphopeptides identified reached ~10,000 per sample without fractionation. The program was integrated onto the Hamilton Microlab STAR without additional hardware. Greater than 90% phosphopeptide specificity was achieved on PolyTi, with biochemical replicates showing similar reproducibility as the technical replicates, indicating that extraction efficiencies are at the detection limits of the MS. Using the established MRM assay, greater than 50% phosphopeptide recovery for all three phosphopeptides with CVs below 20% for both resin types is reported over a 3-week period. Based on these results, this automated workflow should facilitate future studies for systematically screening different buffers, resins, and cell treatment conditions.

Automating sample preparation is a growing field that continues to make technological advances in terms of hardware and software. The progress is primarily pushed forward due to the emergence of precision medicine where patients are treated on an individual level. Precision medicine opens new and exciting opportunities to help treat diseases effectively however, it creates a major challenge in handling the sample throughput required to carry out true precision medicine. To address these throughput issues not only is there a need to automate sample preparation processes, but there is also a need to be able to standardize preparation workflows across different automation platforms. Establishing standardized methods is critical to allow researchers to obtain reproducible results across different laboratories and expand the capacity of the

throughput allowed. The automated workflow described here can be used to standardize sample preparation, leading to reproducible results to study phosphorylation events. In future studies, we plan to expand upon our automated method to include both cell lysis and trypsin digestion to generate a “cells to peptides” workflow that will lead to better standardization. Additionally, the established automated method can also be applied to many other workflows that require significant hands-on time or suffer from low throughput. In the next decade, the automated workflow that we describe here could help address sample throughput issues, especially in the clinic, to help drive precision medicine.

## 1.4 MATERIALS AND METHODS

### *1.4.1 Cell culture and treatment*

HCT116 cells were grown in Gibco™ DMEM/F-12, containing 15 mM of HEPES and supplemented with 10% FBS. Cells were passaged by trypsin/EDTA treatment at 80-90% confluence. The cells were passaged at least 5 times prior to treatment with 10 mM H<sub>2</sub>O<sub>2</sub>/Na<sub>3</sub>VO<sub>4</sub> for 30 minutes. After treatment, the cells were washed 3 times with cold 1x PBS. The cells were collected by scraping from the dish, and pelleted by centrifuging for 5 min at 500 × g. The supernatant was removed and the cell pellets were stored at -80°C until further processing.

### *1.4.2 Cell lysis*

For HCT116 cell lysis, ~1 × 10<sup>8</sup> cells were lysed with 1 mL of RIPA buffer (Thermo Scientific Pierce, Rockford, IL) containing 1x of Halt Protease and Phosphatase Inhibitor Cocktail (Thermo Scientific Pierce, Rockford, IL). Proteins were reduced with addition of 10 mM TCEP at 56 °C for 30 minutes, then alkylated with addition of 25 mM

iodoacetamide for 30 minutes in dark at room temperature (20-25°C). Next, 1.5 mL of ice-cold acetone was added to the sample to precipitate the protein. The mixture was placed at -80°C for 30 minutes followed by -20°C for 30 minutes. The sample was then centrifuged for 5 minutes at 2000 × g and the supernatant was discarded. The protein pellet was washed using 1 mL of ice-cold acetone and centrifuged for 5 minutes at 2000 × g. The supernatant was then discarded. The resulting pellet was suspended in 50 mM ammonium bicarbonate to a final concentration of 5 mg/mL followed by overnight trypsin digestion (1:50 enzyme to protein by mass ratio) at 37°C using MS Grade Trypsin Protease (Thermo Scientific Pierce, Rockford, IL). The sample was then acidified to 1% TFA and then centrifuged for 5 minutes at 5000 × g. The supernatant was then aliquoted and stored at -80°C until required for phosphopeptide enrichment.

#### ***1.4.3 Phosphopeptide enrichment of HCT116 cell digests***

Cell digest was thawed on ice for 30 minutes. Aliquots of 20 µL of cell digest was then diluted with 1% TFA to final volume of 200 µL and transferred to a 1 mL 96-well plate. The automated phosphopeptide enrichment used 300 µL IMCStips (IMCS, Irmo, SC) which contain loosely packed resins that were specific to the workflow. All reagents were poured into large reagent troughs, and the program was designed to aliquot solutions and buffers into specific positions used for the enrichment process. The peptide desalting portion of the extraction was performed using pipette tips containing 5 mg RP. IMCStips containing 5 mg ZrO<sub>2</sub> or 10 mg PolyTi were used to enrich phosphopeptides from the eluted peptide samples. Spin Tips containing TiO<sub>2</sub> resin were purchased from GL Biosciences and Thermo Scientific. All Spin Tip-based extractions were processed according to vendor specifications with slight modifications to the reagents. For the spin

column-based extraction using PolyTi resin, 10 mg of resin was loaded into a spin column and the extraction protocol followed the same protocol as the automated method, however, instead of mixing by pipetting on the automated liquid handler, the spin columns were placed on a tumbler at room temperature to facilitate proper mixing. For each step in the extraction, the spin columns were centrifuged at  $1000 \times g$  for 2 minutes. All wash buffers and sample loading buffers were prepared in bulk to standardize the comparisons.

For each experiment, the desalting and phosphopeptide enrichment method that was used will be specified. Once the sample was eluted, the eluted solutions were vacuum dried. The samples were then stored at  $-80^{\circ}\text{C}$  until ready for analysis.

#### ***1.4.4 LTQ-Orbitrap Velos Pro analysis for global phosphopeptide analysis***

This analysis was used during the optimization of the phosphopeptide enrichment protocol. Dried phosphopeptide samples were dissolved in  $4.8 \mu\text{L}$  of 0.25% formic acid with 3% (vol/vol) acetonitrile and  $4 \mu\text{L}$  of each sample was injected into Easy-nLC 1000 (Thermo Fisher Scientific). Peptides were separated on a 40-cm in-house packed column ( $360 \mu\text{m OD} \times 75 \mu\text{m ID}$ ) containing C18 resin ( $2.2 \mu\text{m}$ ,  $100 \text{ \AA}$ ; Michrom Bioresources). The mobile phase buffer consisted of 0.1% formic acid in ultrapure water (buffer A) with an eluting buffer of 0.1% formic acid in 80% (vol/vol) acetonitrile (buffer B) run with a linear 60-min gradient of 6–30% buffer B at flow rate of  $300 \text{ nL/min}$ . Easy-nLC 1000 was coupled online with a hybrid high-resolution LTQ-Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific). The mass spectrometer was operated in the data-dependent mode, in which a full-scan MS (from  $m/z$  300 to 1,500 with the resolution of 30,000 at  $m/z$  400) was followed by MS/MS of the 10 most intense ions [normalized

collision energy - 30%; automatic gain control (AGC) -  $3.0 \times 10^4$ ; maximum injection time (MS1) - 100 ms; 90 s exclusion; maximum injection time (MS2) – 150 ms].

#### ***1.4.5 Orbitrap Fusion analysis for global phosphopeptide analysis***

Thermo Scientific Orbitrap Fusion run under conditions established by the staff at the Nevada Proteomics Center was used for sample analysis. Liquid chromatography mass spectrometry (LC-MS) was performed on an UltiMate 3000 RSLCnano system (Thermo Scientific, San Jose, CA) on a self-packed UChrom C18 column (100  $\mu\text{m}$  x 35 cm). The gradient used consisted of Solvent B from 2-90% (Solvent A: 0.1% formic Acid, Solvent B: acetonitrile, 0.1% formic Acid) over 175 mins at 50°C using a digital Pico View nanospray source (New Objectives, Woburn, MA) that was modified with a custom-built column heater and an ABIRD background suppressor (ESI Source Solutions, Woburn, MA). The self-packed column tapered tip was pulled with a laser micropipette puller P-2000 (Sutter Instrument Co, Novato, CA) to an approximate id of 10  $\mu\text{m}$ . The column was then packed with 1-2 cm of 5  $\mu\text{m}$  Sepax GP-C18 (120 Å) (Sepax Technologies, Newark, DE) followed by 40 cm of 1.8  $\mu\text{m}$  Sepax GP-C18 (120Å) at 9000 psi using a nano LC column packing kit (nanoLCMS, Gold River, CA). The MS1 precursor selection range is from 400-1500 m/z at a resolution of 120K and an automatic gain control (AGC) target of  $2.0 \times 10^5$  with a maximum injection time of 100 ms. Quadrupole isolation window at 0.7 Th for MS2 analysis using CID fragmentation in the linear ion trap with a collision energy of 35%. The AGC was set to  $4.0 \times 10^3$  with a maximum injection time of 150 ms. The instrument was operated in a top speed data-dependent mode with a most intense precursor priority with dynamic exclusion set to an exclusion duration of 60 s with a 10 ppm tolerance. A technical replicate refers to multiple injections from a single

eluate to ensure system suitability, whereas biochemical replicate refer to single injection from a single sample to assess reproducibility of the extraction method.

#### ***1.4.6 Database search and analysis***

The raw data files obtained from LTQ-Orbitrap Velos Pro and Orbitrap Fusion were searched against a human database using SEQUEST on Proteome Discoverer (Version 2.2, Thermo Fisher). The peptide precursor mass tolerance was set to 10 ppm, and the fragment mass tolerance was set to 0.6 Da. Search criteria included a static modification of cysteine residues of +57.0214 Da and a variable modification of +15.995 Da to include potential oxidation of methionine and a modification of +79.966 Da on serine, threonine, or tyrosine for the identification of phosphorylation (three modifications allowed per peptide). Searches were performed with full tryptic digestion and allowed a maximum of two missed cleavages on the peptides analyzed from the sequence database. False discovery rates (FDR) were set to less than 1% for each analysis at the PSM level. Phosphorylation site localization from CID spectra was determined by PhosphoRS on Proteome Discoverer 2.2. All Venn diagrams for peptide overlap were generated using Venny 2.1 (<http://bioinfogp.cnb.csic.es/tools/venny/>). Confidence view was assigned a score of 0.4, indicating medium confidence. The GRAVY values of phosphopeptides were calculated using GRAVY calculator (<http://gravy-calculator.de/index.php>) and the pI values were calculated using the Compute pI/Mw tool ([https://web.expasy.org/compute\\_pi/](https://web.expasy.org/compute_pi/)).

#### ***1.4.7 Phosphopeptide enrichment for MRM assay***

Four purified phosphopeptides from Anaspec and two purified phosphopeptides from New England Peptide were screened for compatibility with the MRM assay based on

stability and ionization efficiencies. Three phosphopeptides were selected as optimal candidates for monitoring by MS/MS. Three additional non-phosphopeptides were added to the assay to determine recovery of non-phosphorylated peptides. The standard sample comprised of 2 pmol of 3 phosphorylated and 2 pmol of 3 non-phosphorylated peptides fortified with 50 µg of trypsin digested bovine serum albumin (BSA). Once the starting sample was generated, the samples were aliquoted into a 96-well plate and enriched for phosphopeptides with either a 5 mg ZrO<sub>2</sub> or 10 mg PolyTi using the automated phosphopeptide enrichment method. The desalting portion of the automated extraction was excluded in this assay. The eluates were neutralized 1:1 by volume with 20% formic acid. The neutralized eluates were then stored at -20°C until ready for analysis.

#### ***1.4.8 Multiple reaction monitoring (MRM) assay***

MRM assay was used to analyze phosphopeptide samples after the automated enrichment. For the MRM assay, a TSQ Endura (Thermo Fisher) coupled to a Vanquish UHPLC was used. Reverse-phase separation was performed using a Waters BEH C18 analytical column (100 x 2.1 mm, 1.7 µm) heated at 40°C with an LC gradient of 2%-35% B (0.1% formic Acid in acetonitrile) for 10 minutes. TSQ Endura was run under optimized conditions with a Q1 and Q3 resolution of 1.2 FWHM and 0.7 FWHM, respectively, and CID Gas was set to 1.5 mTorr. All raw files were processed using Skyline 4.0. All samples were compared and quantified based on a neat sample injected before and after each sample set.

## CHAPTER 2

### USING PHOSPHOPROTEOMICS TO INVESTIGATE THYMIDYLATE SYNTHASE INHIBITOR TREATMENTS OF COLORECTAL CANCER CELL LINES.

#### 2.1 INTRODUCTION

##### *2.1.1 Colorectal cancer and thymidylate synthase inhibitor treatments*

Colorectal Cancer (CRC) is one of the most deadly cancers in the United States and is expected to cause over 50,000 deaths in 2019 according to the American Cancer Society.<sup>39</sup> Small molecule chemotherapy drugs, mostly commonly thymidylate synthase (TS) inhibitors such as 5-fluorouracil (5-FU) and 5-fluoro-2'-deoxyuridine monophosphate (5-FdUMP) and folate analogs such as folinic acid (drug name leucovorin or LV) are used as a first line treatment of CRC. TS has two binding sites, a nucleotide binding site, and a folate binding site, which are used in the catalysis of the reductive methylation of 2'-deoxyuridine-5'-monophosphate (dUMP) by N<sup>5</sup>N<sup>10</sup>-methylenetetrahydrofolate (CH<sub>2</sub>THF), to form deoxythymidine monophosphate (dTMP) and dihydrofolate. TS is important for DNA synthesis as it is the primary generator of dTMP and thus, if inhibited, dTMP will be limited, leading to a disruption in DNA synthesis. This treatment is especially detrimental to highly proliferating cells (i.e. cancer cells) and thus TS inhibitors have been widely used as a first line of treatment for cancer patients. Currently, the standard of care for cancer patients includes a 5-FU and LV cotreatment used to inhibit TS by forming a stable ternary complex with metabolized forms of these drugs, namely 5-FdUMP and 5,10-methylenetetrahydrofolate (5,10-

CH<sub>2</sub>FH<sub>4</sub>) (Figure 2.1).<sup>40-42</sup> This process limits the availability of dTMP and disrupt normal DNA and RNA processes.

### ***2.1.2 Resistance to thymidylate synthase inhibitor treatments***

One of the main problems in cancer treatment is drug resistance. Innate or acquired drug resistance is thought to be the main cause of treatment failure in 90% of patients with metastatic cancer.<sup>43</sup> There have been significant efforts to develop predictive markers of TS inhibitor resistance and to find additional drug targets to help combat resistance. Several different hypotheses for TS inhibitor resistance have been presented: existence of “drug tolerant” cells prior to drug treatment, acquired genetic mutations during treatment, and alterations in signaling pathways which immediately respond to drug treatment.<sup>44</sup> Of these hypotheses for TS inhibitor resistance, the existences of “drug tolerant” cells prior to treatment and alterations in signaling pathways in response to treatment are most likely. This is based on cell culture studies and clinical studies where TS inhibitor resistance is established in a few weeks which is insufficient for de novo genetic alterations.<sup>44</sup>

To understand how existing mutations affect TS inhibitor resistance, particularly 5-FU resistance, a large-scale study in which 5-FU response in 77 CRC cell lines was monitored. This study found that replication error deficiency (RER), characterized by the existence of frequency mutations in (CA)<sub>n</sub> and other simple repeats, correlates well with 5-FU resistance. This correlation could allow clinicians to screen patients for their RER status (+ or -) to help determine the 5-FU dose mostly likely to lead to effective treatment.<sup>45</sup> As for acquired 5-FU resistance from alterations in signaling pathways, several studies have shown alterations in protein expression and post-translational

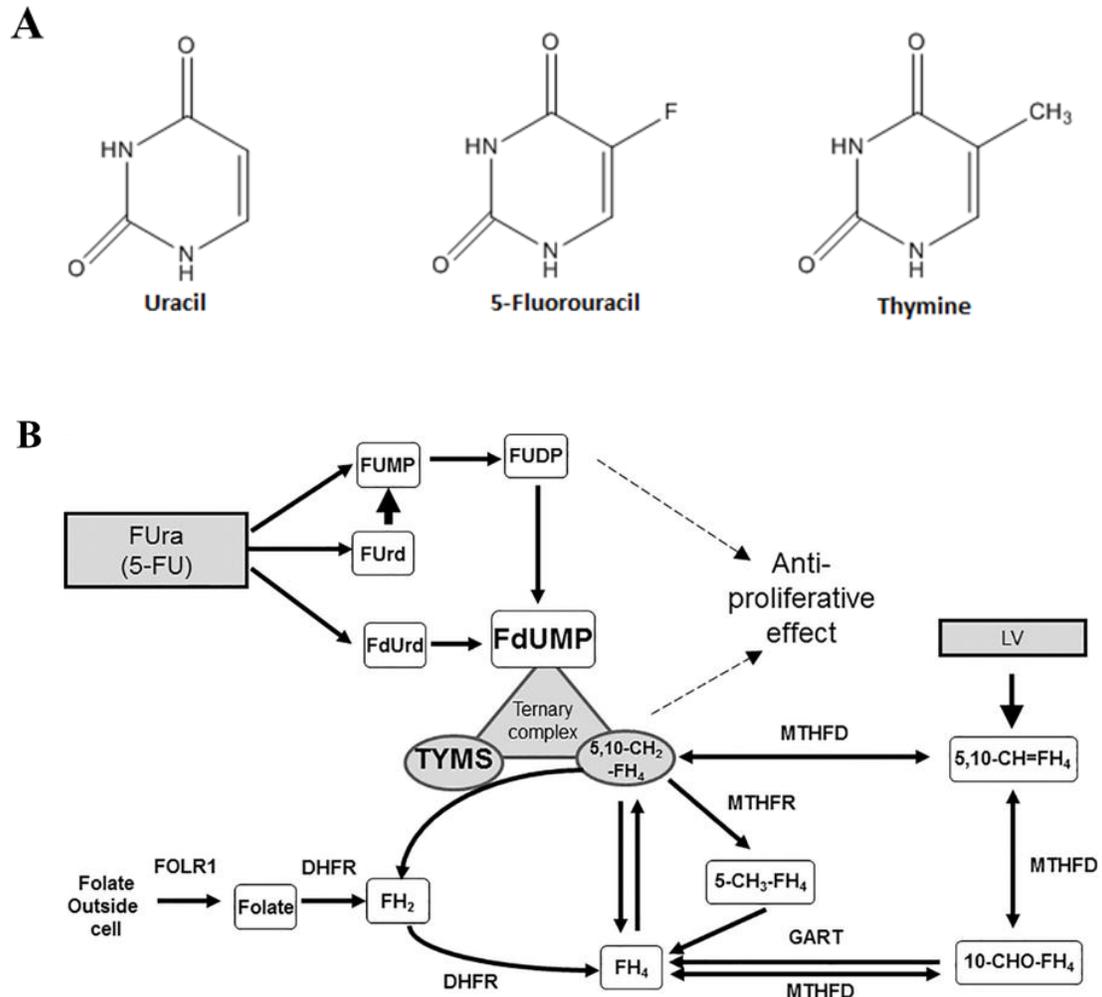


Figure 2.1. 5-FU is a thymidylate synthase inhibitor used to treat cancer. A) Chemical structure of Uracil, 5-Fluorouracil, and Thymine. B) Thymidylate synthase inhibition by 5-FU and LV. The original figure is from reference 41 and the legend has been modified from the original.<sup>41</sup>

modifications, especially phosphorylation, upon drug treatment.<sup>44, 46-52</sup> For example, a study published in 2018 showed that HCT-8 cells with acquired resistance had alterations in the p53-signaling pathway and the WNT-signaling pathway.<sup>46</sup> The p53-signaling pathway is responsible for inducing cell arrest and apoptosis in response to DNA damage. The study found that Serine/threonine-protein kinase (CHK1) total and its phosphorylated form, were reduced in HCT-8 cells with acquired resistance. CHK1 is upstream of tumor protein p53 (p53) and suppression of CHK1 led to suppression of p53 and the p53-signaling pathway, allowing the cells to escape cell arrest and apoptosis.<sup>45</sup> Furthermore, the study found that activation of proteins in the WNT-pathway,  $\beta$ -catenin and Transcription factor 4, caused the suppression of CHK1 and the p53-signaling pathway.<sup>45</sup> Together, this study suggests that proteins within the p53-signaling pathway and the WNT-signaling pathway could be potential therapeutic targets that could be explored to address 5-FU resistance.

Interestingly, cell cycle perturbation has also been observed in TS inhibitor resistant cells. These resistant cell lines have decreased expression levels of CDK2 and its phosphorylated form, leading to an increased delay in G<sub>1</sub> and G<sub>1</sub>/S which allows for longer DNA synthesis and DNA repair. Additionally, other proteins associated with cell cycle transition show decreased phosphorylation levels including retinoblastoma-associated protein (RB1), cyclin D3, and cyclin A.<sup>48, 49</sup>

The MAPK signaling pathway is also thought to play a role in response to TS inhibitor treatments as an anti-apoptotic response, however, the understanding of the pathway's role is complex and reports have been contradictory.<sup>53-55</sup> In 2005, Kim *et al.* showed that reduction of kinase suppressor of Ras1 (KSR1), a positive regulator of ERK

activation, lead to a decrease in ERK activation and reduced sensitivity to chemotherapy treatments.<sup>53</sup> However, in 2010, Yang *et al.* showed that inhibition of p38 MAPK significantly increased colorectal cancer cell sensitivity to 5-FU treatments.<sup>55</sup>

### ***2.1.3 Overview of aims and approach of this study***

The aim of this work is to further understand the phosphoproteomic changes in response to TS inhibitor drug treatments in sensitive cells as well as resistant cells. We use quantitative MS-based phosphoproteomics in this study to understand and quantify changes in phosphoproteins in response to the TS-inhibitor cotreatment. This was done using stable isotope labeling by amino acids in cell culture (SILAC) to specifically label cells with heavy isotope labelled lysine and arginine generating a “heavy” cell line and, by extension, a “light” cell line. SILAC was first demonstrated for proteomics studies in 2002 by Ong *et al.*<sup>13</sup> With this approach, one of the cell lines (conventionally the “light” cell line) can be subjected to a drug or stress treatment and then combined with the untreated cell line and analyzed by mass spectrometry. From this analysis, differences in protein abundance and post translational modifications (PTMs) can be detected between the “light” and “heavy” sample. These quantified differences give an indication of how the cell responds to the stress or drug treatment that can be used to better understand the mechanism of action of the drug, identify potential biomarkers for resistance/sensitivity to the drug treatment, and screen therapeutic targets that can increase the treatment efficacy. Studies have identified proteins that change in response to TS-inhibitor treatment, but these studies were done using whole-genome cDNA microarray analysis.<sup>41</sup> Although this technology is useful in understanding changes in expression levels of

proteins in response to treatments, it is not able to show changes in PTMs including changes in phosphorylation signaling pathways.

One of the limitations of SILAC-based analysis is that the samples that are analyzed only give a snapshot of the phosphoproteome at the time of collection of the cells. While this can be useful, quantifying changes in the phosphoproteome over time can give a better indication of the drug's mechanism of action and how the cells respond. Conventionally, phosphoproteome changes happen relatively quickly (5-120 mins) as phosphorylation operates similar to a switch to turn “off” or “on” particular signaling pathways and protein expression changes happen over several hours to help drive larger cellular response including cell senescence and apoptosis.<sup>56, 57</sup> Therefore, performing a time course study of the cells subjected to the cotreatment will allow us to obtain several snapshots at different time points showing how changes in the phosphoproteome are used by the cell to respond to the treatment over time.

## 2.2 RESULTS & DISCUSSION

### *2.2.1 Establishment of TS inhibitor treatment with 5-FU*

We established an in-house 5-FU treatment of HCT116 cells by determining the 50% growth inhibition concentration (GI<sub>50</sub>) for a 48-hour treatment. HCT116 cells were treated with various concentrations of 5-FU including, 1.56, 3.12, 6.25, 12.5, 25, 50, 100, 200, 400 and 800 μM for 48 hours. Using the CTB assay, relative cell viability was determined for each 5-FU concentration (Table 2.1). A Monod equation was generated to calculate the GI<sub>50</sub> based on the data points from the CTB assay (Figure 2.2).

$$\text{RFU}_{[5\text{-FU}]} = \text{RFU}_o \times \frac{1.064}{[5\text{-FU}]}^{0.187}$$

Table 2.1. The relative growth inhibition of HCT116 using various concentrations of 5-FU based on the CTB assay.

<b>[5-FU], <math>\mu</math>M</b>	<b>590 nm, RFU</b>	<b>Percent Inhibition, %</b>
0	3470.7	0%
1.5625	3379.3	3%
3.125	2689.8	22%
6.25	2488.2	28%
12.5	2012.6	42%
25	1660.7	52%
50	1586.9	54%
100	1434.2	59%
200	1469.0	58%
400	1491.6	57%
800	885.9	74%

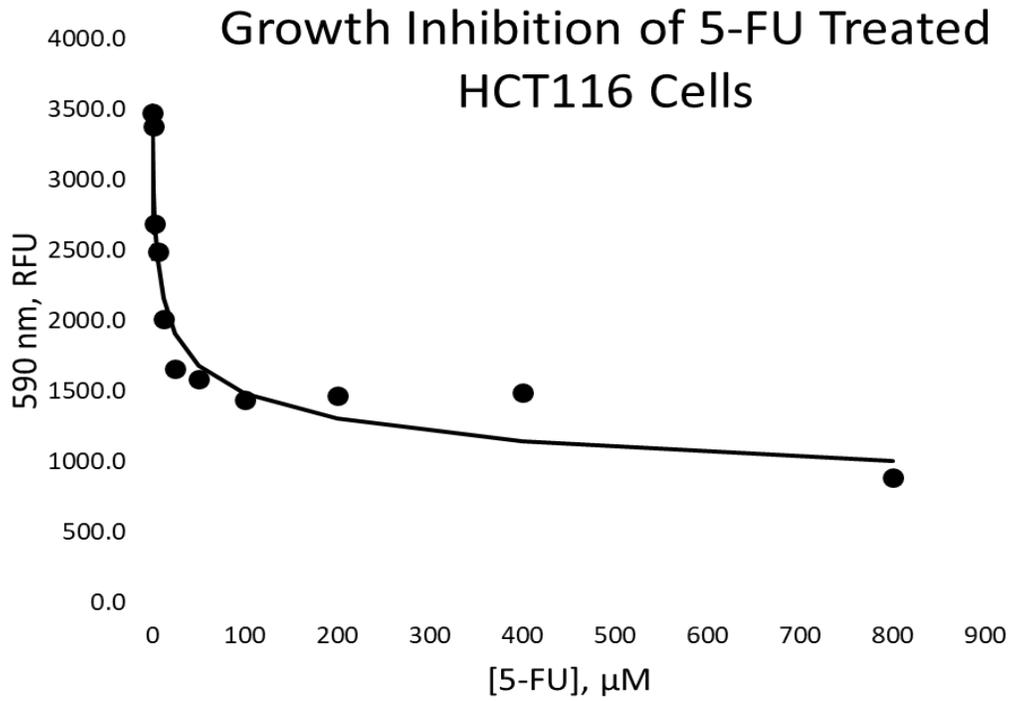


Figure 2.2. Cell viability plot of HCT116 cells after 5-FU treatment at each concentration. The cell viability was quantified by 590 nm RFU values after incubation with CTB reagent. The displayed trendline is based on the generated Monod equation using the 590 nm RFU values from the CTB assay where  $\text{RFU}_o$  is the RFU value at 0  $\mu\text{M}$  5-FU.

Based on this equation, the  $GI_{50}$  for our HCT116 cell line was 42  $\mu\text{M}$ , which was consistent with reported results that the  $GI_{50}$  5-FU concentration for the HCT116 cell line was reported to be between 0.5 and 77  $\mu\text{M}$ .<sup>45, 58</sup> Based on these results, 10  $\mu\text{M}$  5-FU was selected as the concentration to be used for the cell treatment with a calculated theoretical growth inhibition of 35%. This 5-FU treatment of HCT116 cells was then used to perform the SILAC-based quantitative phosphoproteomic analysis.

### ***2.2.2 Global phosphoproteomics of 5-FU treated HCT116 cells***

In previous studies, we established an automated desalting and phosphopeptide enrichment method using IMCStips on a Hamilton Microlab STAR™, which allowed us to identify over 10,000 phosphopeptides from 200  $\mu\text{g}$  of HCT116 cell lysate without fractionation. Here, we used this established automated method to desalt and enrich for phosphopeptides from 200  $\mu\text{g}$  of 1:1 mixed Heavy:Light cell lysate to quantitate protein fold changes in 5-FU treated HCT116 cells. For the analysis, 8 samples were desalted using RP IMCStips. Two of the desalted samples were collected and vacuum dried without further enrichment. Six samples were further enriched using phosphopeptide enrichment IMCStips, i.e., 3 using PolyTi IMCStips and 3 using  $\text{ZrO}_2$  IMCStips, respectively. Our previous studies showed a slight differential enrichment between the two resin types. Therefore, both PolyTi and  $\text{ZrO}_2$  resin types were used for enrichment to obtain a comprehensive phosphorylation profile (Table 2.2).

From the analysis, a total of 5,985 unique proteins and 67,250 unique peptides were identified from our analyzed samples. For the phosphopeptide enriched samples, the PolyTi IMCStips samples averaged a total of 10,465 phosphopeptides identified and an average specificity of 85%.  $\text{ZrO}_2$  IMCStips samples averaged 8,351 phosphopeptides

Table 2.2. MS data from 8 HCT116 cell digest samples.

Sample	Resin	Protein Ids, #	Peptide Ids, #	Phosphopeptide Ids, #	Phosphopeptide Specificity, %	Quantified Proteins, #	Quantified Peptides, #
1	RP	4200	38423	260	1%	3490	19785
2	RP	4132	36539	234	1%	2838	18911
3	PolyTi	2886	12598	11156	89%	2458	6788
4	PolyTi	2879	12388	10365	84%	2195	6678
5	PolyTi	2835	12085	9874	82%	2221	6422
6	ZrO <sub>2</sub>	2395	9012	8489	94%	1842	4919
7	ZrO <sub>2</sub>	2483	9589	7895	82%	1835	5175
8	ZrO <sub>2</sub>	2407	9138	8670	95%	1862	4921

identified with an average specificity of 90% (Figure 2.3). The unenriched samples processed using RP IMCStips samples showed very low phosphopeptide identifications, highlighting the need for the phosphopeptide enrichment.

### ***2.2.3 Reproducibility of SILAC-based quantification and phosphopeptide enrichment***

The reproducibility of the SILAC-based quantification was determined for unenriched and enriched samples by plotting the calculated  $\log_2$  fold change values for biochemical replicates (Figure 2.4). By Pearson correlation, the r-values for the comparison of calculated fold change ratios for two replicates from RP, PolyTi and ZrO<sub>2</sub> were obtained as 0.88, 0.80 and 0.78 respectively. Based on these high r-values, we concluded that the sample preparation and analysis process generated reproducible quantitative proteomic and phosphoproteomic data.

Using the calculated fold change ratios, we then determined the number of proteins “significantly” decreased or increased for each sample analyzed. A protein was considered “significantly” decreased or increased if the calculated fold change was above 2 or below 0.5 with a calculated p-value less than 0.05. For the unenriched samples, an average of  $685 \pm 6$  proteins were significantly decreased and an average of  $523 \pm 18$  proteins were significantly increased. For the phosphopeptide enriched samples, an average of  $659 \pm 20$  proteins were significantly decreased and an average of  $576 \pm 27$  proteins were significantly increased. Volcano plots for each sample were generated to show the overall distribution of protein fold changes (Figure 2.5).

## Efficiency of Phosphopeptide Enrichment

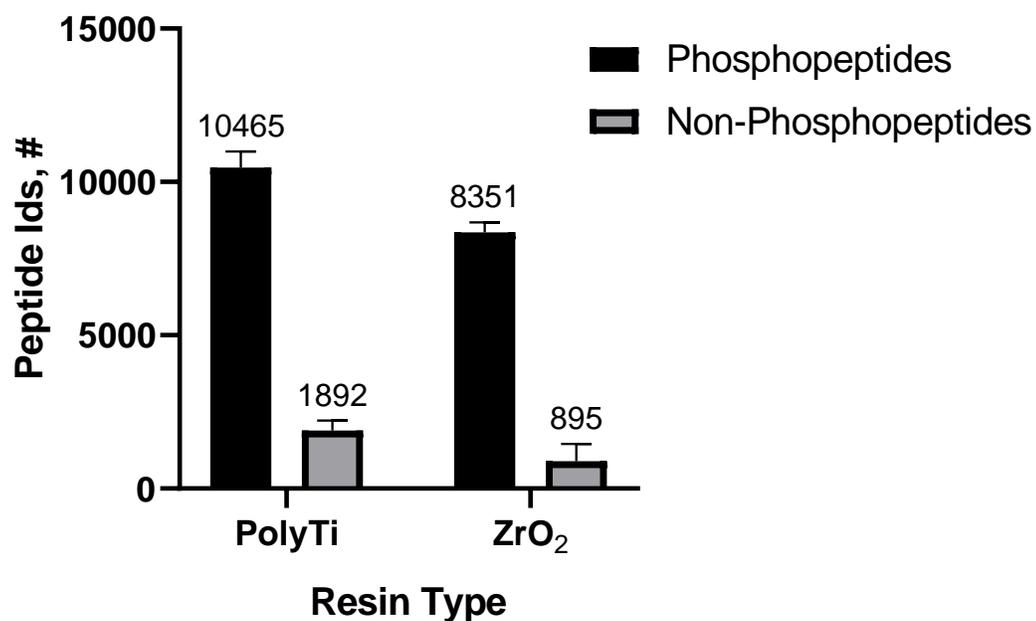


Figure 2.3. A comparison of the total amount of phosphopeptides and non-phosphopeptides identified by either PolyTi or ZrO<sub>2</sub> resin. All samples were processed with IMCStips using our fully automated phosphopeptide enrichment method (n = 3, data are mean ± s.d.).

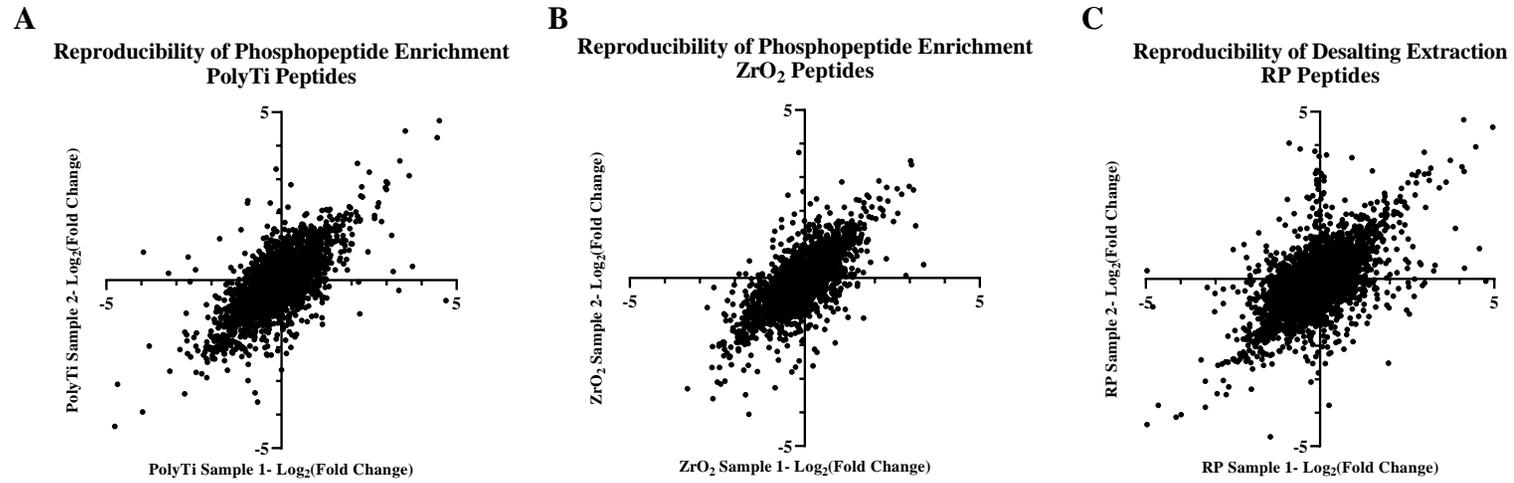
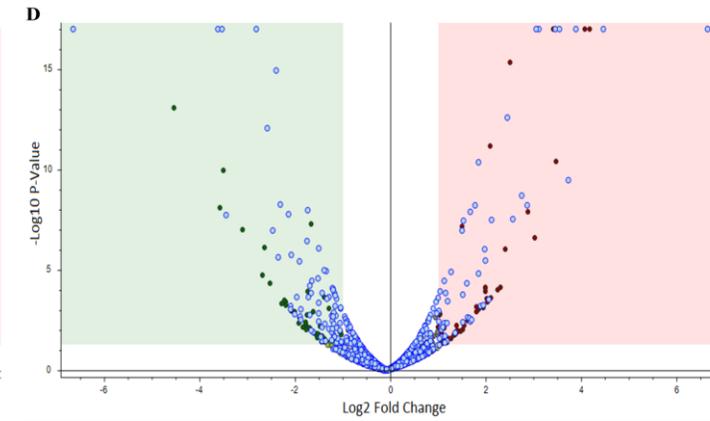
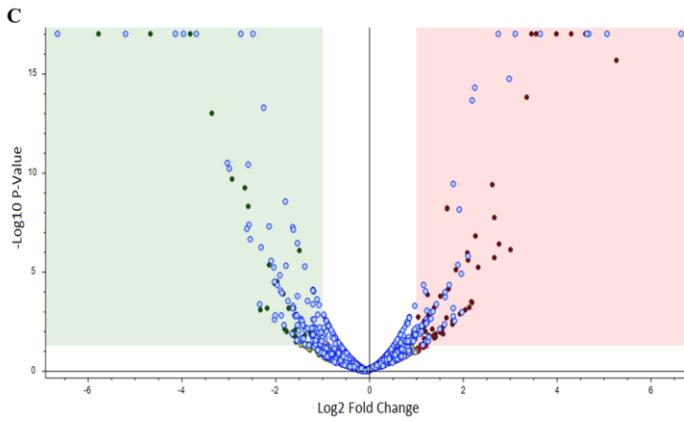
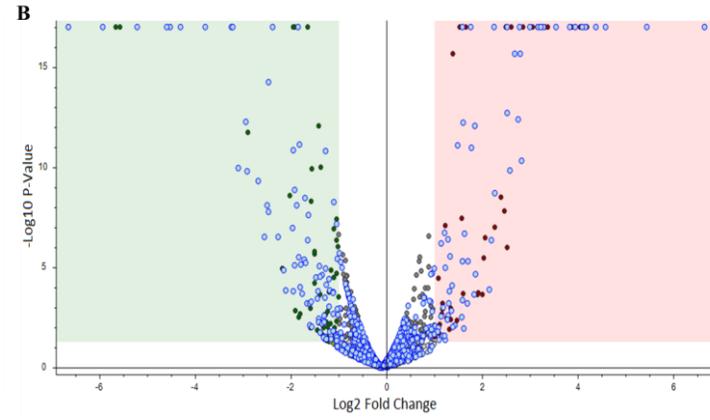
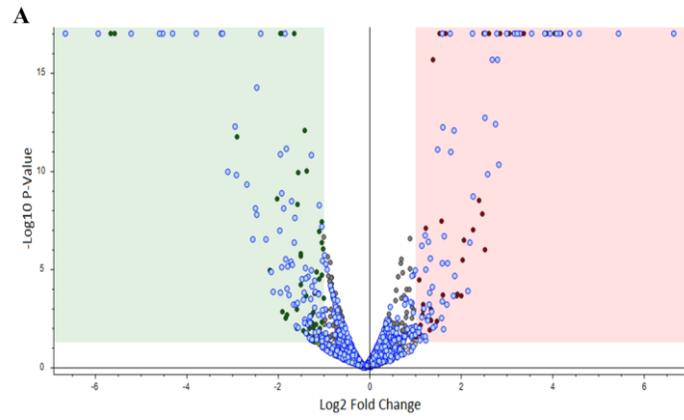


Figure 2.4. A comparison of peptide fold change ratios ( $\log_2$ ) between two biochemical replicates of 200  $\mu\text{g}$  of 1:1 mixed SILAC samples. Samples were prepared using A) the desalting extraction using RP IMCStips, B) the desalting extraction using RP IMCStips followed by the phosphopeptide enrichment using PolyTi IMCStips or C) the desalting extraction using RP IMCStips followed by the phosphopeptide enrichment using ZrO<sub>2</sub> IMCStips. R-values equal to 0.88, 0.80 and 0.78 respectively by Pearson correlation (P-Values < 0.00001).



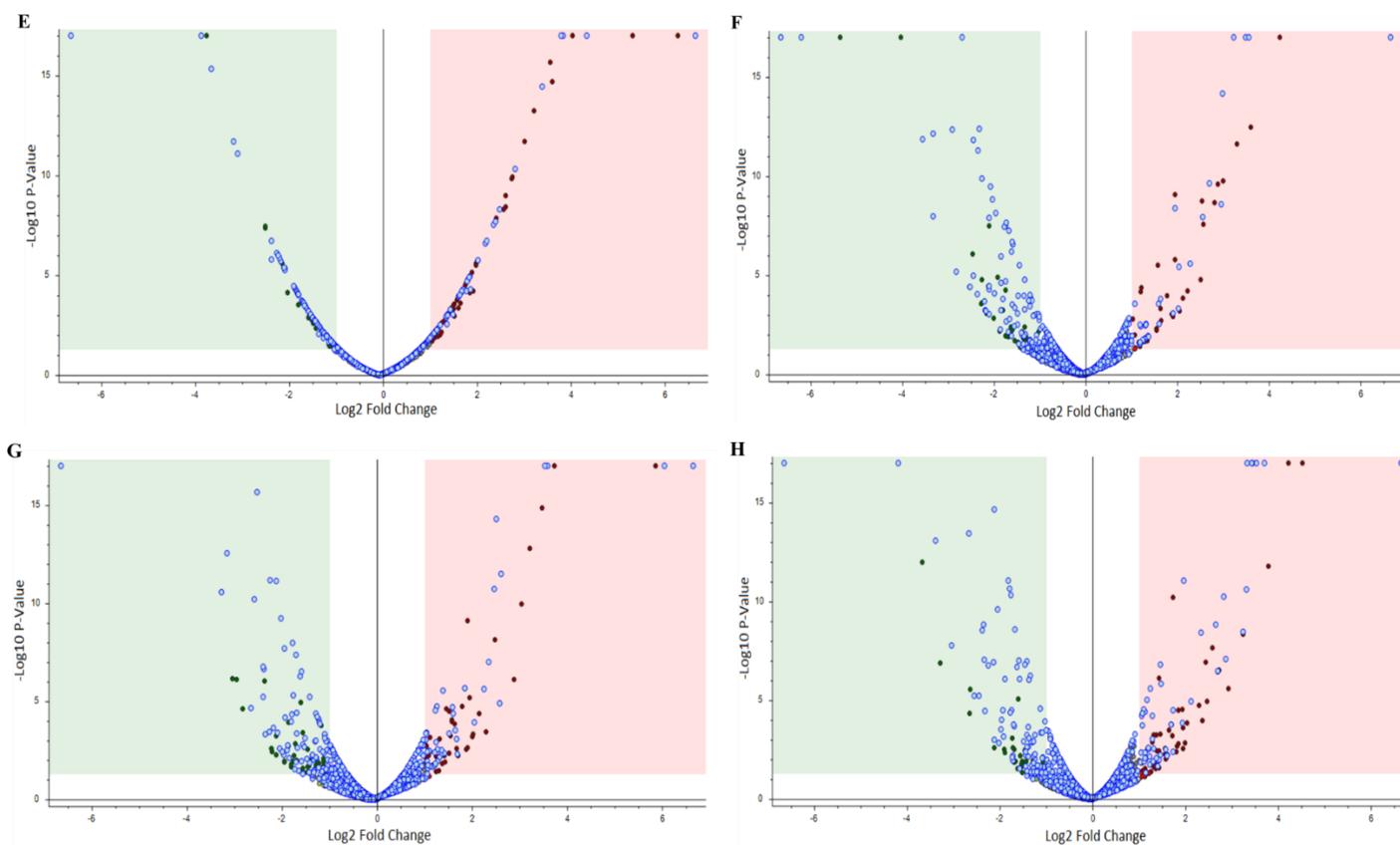


Figure 2.5. Volcano plots of quantified proteins from each sample analyzed. The volcano plot is based on the proteins fold change and its associated p-value. A) RP- Sample 1 Proteins: 679 Significantly Decreased, 505 Significantly Increased. B) RP- Sample 2: 690 Significantly Decreased, 541 Significantly Increased. C) PolyTi- Sample 1 Proteins: 657 Significantly Decreased, 576 Significantly Increased. D) PolyTi- Sample 2 Proteins: 645 Significantly Decreased, 605 Significantly Increased. E) PolyTi-Sample 3 Proteins: 701 Significantly Decreased, 606 Significantly Increased. F) ZrO<sub>2</sub>- Sample 1 Proteins: 660 Significantly Decreased, 531 Significantly Increased. G) ZrO<sub>2</sub>- Sample 2 Proteins: 640 Significantly Decreased, 553 Significantly Increased. H) ZrO<sub>2</sub>-Sample 3 Proteins: 652 Significantly Decreased, 584 Significantly Increased.

#### ***2.2.4 Pathway analysis of quantitative phosphoproteomic data***

To further determine the biological relevance of the data, Ingenuity Pathway Analysis (IPA) software was used. For this proof-of-concept study, we were primarily interested in analyzing the phosphopeptide enriched data set to understand the phosphorylation signaling pathways affected by the 5-FU treatment. The expression fold change data from the phosphopeptide enriched samples was analyzed using IPA to determine the affected signaling pathways (Figure 2.6). In our data set, we observed a collective decrease in kinases within the eukaryotic initiation factor-2 (EIF2) pathway, which are generally associated with growth factors such as PI3K, AKT, c-RAF, potentially in response cell stress (Figure 2.7). The EIF2 signaling pathway is generally associated with cellular growth and proliferation. It is known that in response to stress and amino acid starvation, phosphorylation of EIF2a can inhibit translational initiation and by association cellular growth.<sup>59-61</sup>

A tox list was generated to illustrate the pathways associated with toxicity that were enriched in the analyzed dataset (Figure 2.8). The NRF2-Mediated Oxidative stress response pathway was shown to be increased within our dataset. This pathway has been implicated as a primary response pathway to 5-FU treatment due to an increase in reactive oxygen species.<sup>42</sup> In previous studies, it has been shown that upregulation of NRF2 can lead to resistance to 5-FU treatment. Additionally, phosphorylation plays is thought to play a key role in the upregulation of NRF2 and NRF2-associated proteins.<sup>62</sup> The p53 signaling pathway was also shown to be affected in our dataset. It has recently been reported that p53 signaling pathway is crucial for initiating cell death in response to 5-FU treatment. In cells resistant to 5-FU treatment, the WNT-signaling pathway can

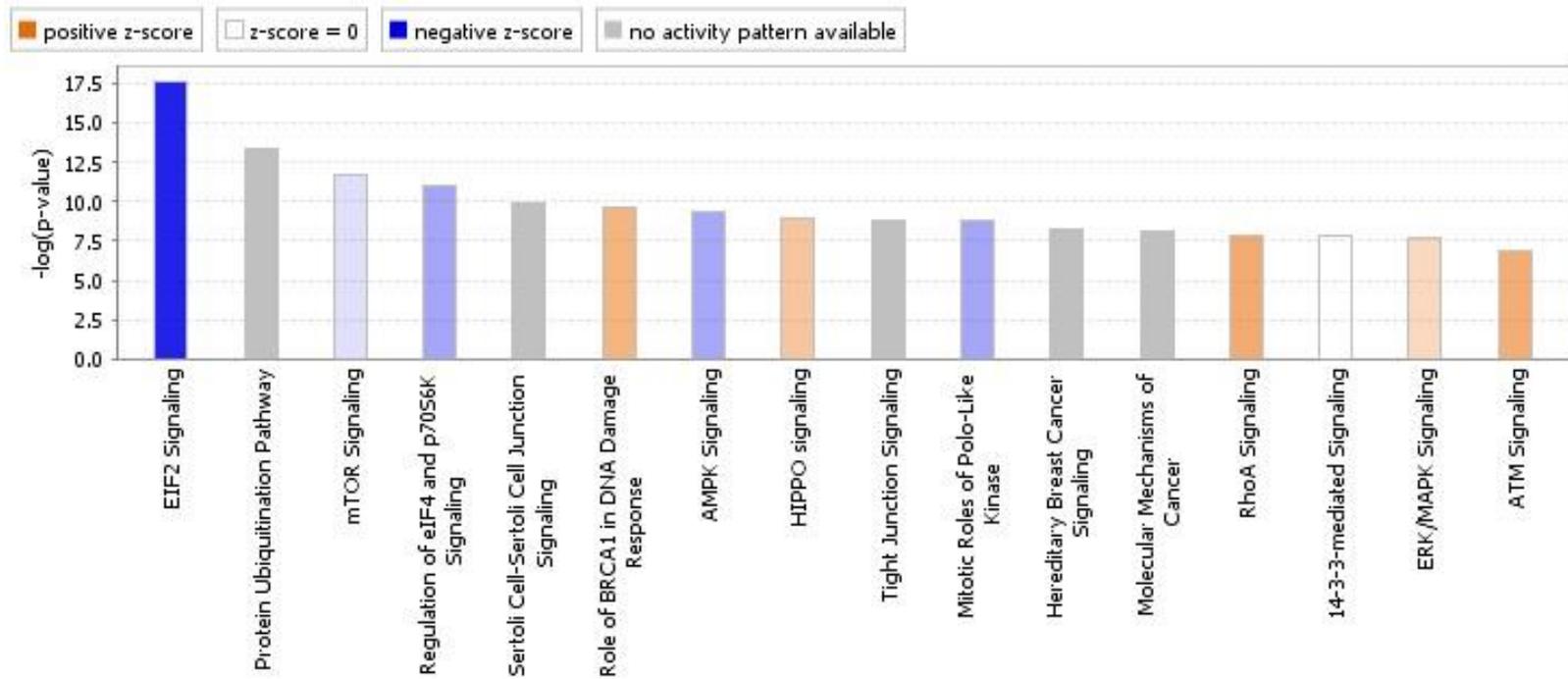


Figure 2.6. Canonical pathways identified using the proteins from the phosphopeptide enriched data set. The p-value represents the probability of random association of molecules with the identified pathway. A  $-\log(P\text{-value})$  threshold of 5 for this dataset. A positive z-score indicates an increase of the pathway, while a negative z-score indicates an inhibition of the pathway.

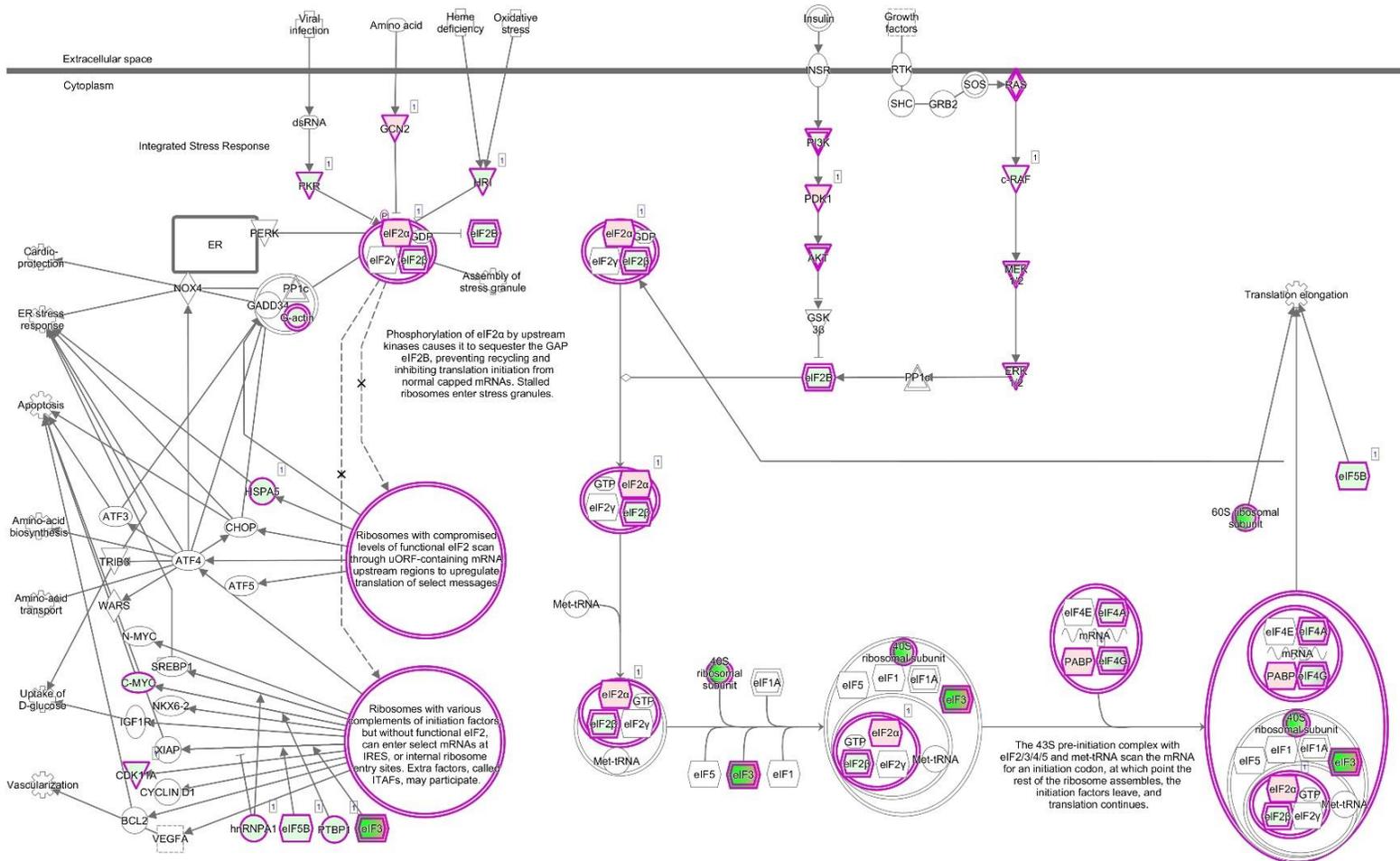


Figure 2.7. EIF2 signaling pathway with proteins identified and quantified within the phosphopeptide enriched sample set. The z-score for this pathway was calculated to be -2.46 indicating inhibition of the EIF2 signaling pathway. Molecules with purple bands indicate identification of the protein within the dataset.

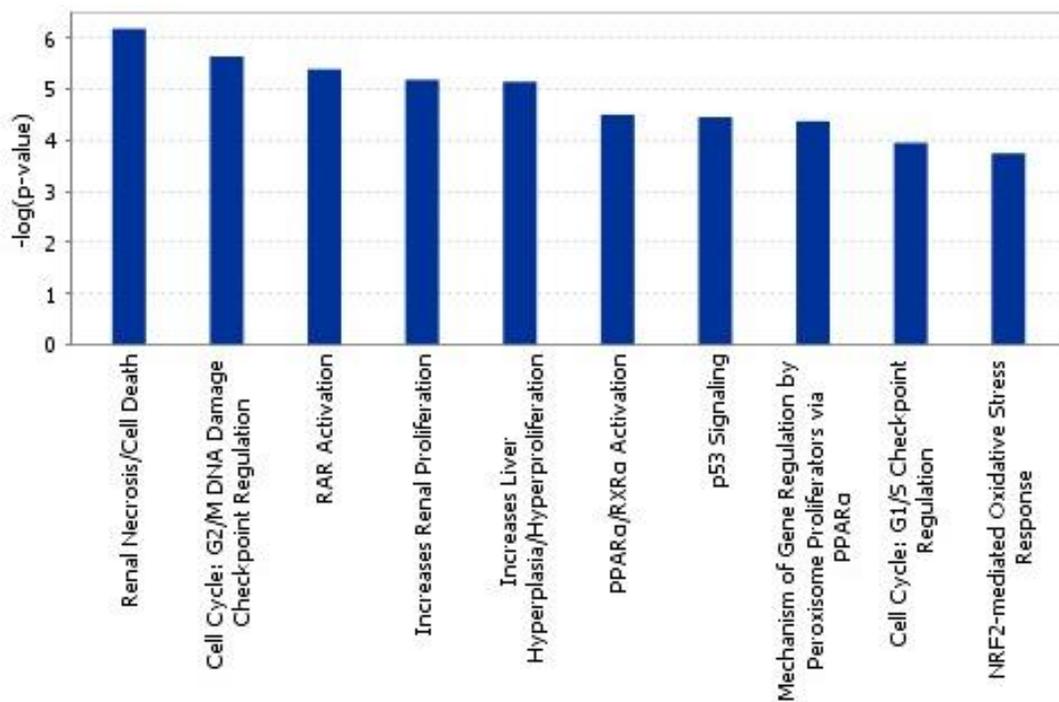


Figure 2.8. Tox list generated by IPA software based on molecules from the phosphopeptide enriched dataset. The p-value represents the probability of random association of molecules with the identified pathway. A  $-\log(P\text{-value})$  threshold of 3 was used for this dataset.

inactivate the p53 signaling pathway cell death response.<sup>46</sup> Together this analysis points to several different signaling pathways that could be responsible for drive the toxicity of the 5-FU treatment.

The phosphopeptide enriched data set was further filtered by selecting proteins that were identified across all analyzed phosphopeptide samples and were calculated to have a fold change of greater than 2 or less than 0.5. In total 22 proteins were identified, 11 proteins with a fold change greater than 2 and 11 proteins with a fold change less than 0.5 (Table 2.3). From this list Lamin-B1 is of particular interest as it was recently reported that the overexpression of the protein leads to worse clinical outcomes in patients with colon cancer due to an increase in mitotic catastrophe allowing cancers cells to avoid apoptosis.<sup>63</sup> Additionally, upregulation of deoxycytidine kinase, an important enzyme in DNA synthesis, has been shown to correlate well with 5-FU sensitivity and has been highlighted as a potential prognostic marker for the drug treatment.<sup>64</sup> The proteins in Table 2.3 highlight phosphoproteins that could be further investigated to understand their role in response to 5-FU.

### ***2.2.5 Establishment of TS inhibitor treatment with 5-FdUMP/LV***

Based on our initial study using 5-FU, it was determined that a more targeted TS inhibitor treatment may lead to a better understanding of the cell's response to TS inhibition. However, 5-FU is known to incorporate into both RNA and DNA as well as has many off-target affects that complicate the data analysis of the treated cells. To simplify our system, we established a more targeted TS inhibitor treatment using 10  $\mu$ M 5-FdUMP and 10  $\mu$ M folinic acid. This treatment has previously been used by Dr. Frank Berger *et al.* to study TS inhibition.<sup>42</sup> Furthermore, to understand how different colorectal

Table 2.3. List of proteins that were identified across all analyzed phosphopeptide samples and were calculated to have a fold change of greater than 2 or less than 0.5.

Uniprot Id	Protein Description	Average Fold Change	Phospho-Site(s)
Q8IV50	LysM and putative peptidoglycan-binding domain-containing protein 2	100.00	S24
Q9Y285	Phenylalanine--tRNA ligase alpha subunit	100.00	S301
Q8N6N3	UPF0690 protein C1orf52	11.91	T155
Q92878	DNA repair protein RAD50	9.77	S635; T690
Q92547	DNA topoisomerase 2-binding protein 1	5.46	S350; S888
P27707	Deoxycytidine kinase	4.35	S74
Q8N4S0	Coiled-coil domain-containing protein 82	4.03	S282
O14545	TRAF-type zinc finger domain-containing protein 1	3.50	S327; S415
Q8WWA1	Transmembrane protein 40	2.88	S141
Q08174	Protocadherin-1	2.78	S962; S984
Q9Y6W5	Wiskott-Aldrich syndrome protein family member 2	2.35	S293
Q5VTL8	Pre-mRNA-splicing factor 38B	0.44	S529
Q5JVS0	Intracellular hyaluronan-binding protein 4	0.36	S108
P22234	Multifunctional protein ADE2	0.31	S27
O43583	Density-regulated protein	0.30	S73
Q6AI08	HEAT repeat-containing protein 6	0.30	S337; S643
Q15717	ELAV-like protein 1	0.29	S202
Q9UKX7	Nuclear pore complex protein Nup50	0.28	S296
P20700	Lamin-B1	0.26	S23; S391; T575
Q9H910	Hematological and neurological expressed 1-like protein	0.24	S97
Q9UN79	Transcription factor SOX-13	0.01	S334
Q86U06	Probable RNA-binding protein 23	0.01	S149

cancer cell lines respond to TS inhibition, HCT116/200 cells, which was established by Dr. Frank Berger *et al*, was employed in our study. HCT116/200 cells overexpress TS and is resistant to conventional TS inhibitor treatments.<sup>42</sup> Additionally, to ensure that HCT116 cells were sensitive to the chemotherapy through the inhibition of TS, we performed a TS inhibition treatment with 10  $\mu$ M thymidine. Thymidine is phosphorylated by thymidylate kinase, thereby bypassing the need for TS and the cell is provided with dTTP for DNA synthesis, rescuing the cells from cell death.

Based on the CTB assay, HCT116 cells treated with 10  $\mu$ M 5-FdUMP and 10  $\mu$ M folinic acid for 24 hours decreased the cell viability to  $57.2 \pm 1.0$  %. As a comparison, HCT116/200 cells treated with 10  $\mu$ M 5-FdUMP and 10  $\mu$ M folinic acid for 24 hours did not have a significant change in the cell viability (Figure 2.9). Additionally, HCT116 cells and HCT116/200 cells treated with 10  $\mu$ M 5-FdUMP and 10  $\mu$ M folinic acid along with 10  $\mu$ M thymidine for 24 hours did not have a significant change in the cell viability. The data in Figure 2.9 indicates that HCT116 cells are sensitive to the 10  $\mu$ M 5-FdUMP/LV treatment due to TS inhibition as thymidine, which bypasses the need for TS, rescues the cells from the treatment. Additionally, the HCT116/200 cell line is resistant to the TS inhibitor treatment as no decrease in cell viability was detected after the 24-hour treatment (Figure 2.9). Based on these cell viability results, we used the 10  $\mu$ M 5-FdUMP and 10  $\mu$ M folinic acid treatment to study how the phosphoproteome is differentially affected for cells that are sensitive (HCT116) and resistant (HCT116/200) to the treatment as well as cells that are rescued from the treatment (HCT116 + thymidine).

### 5-FdUMP/LV- 24-hr Treatment

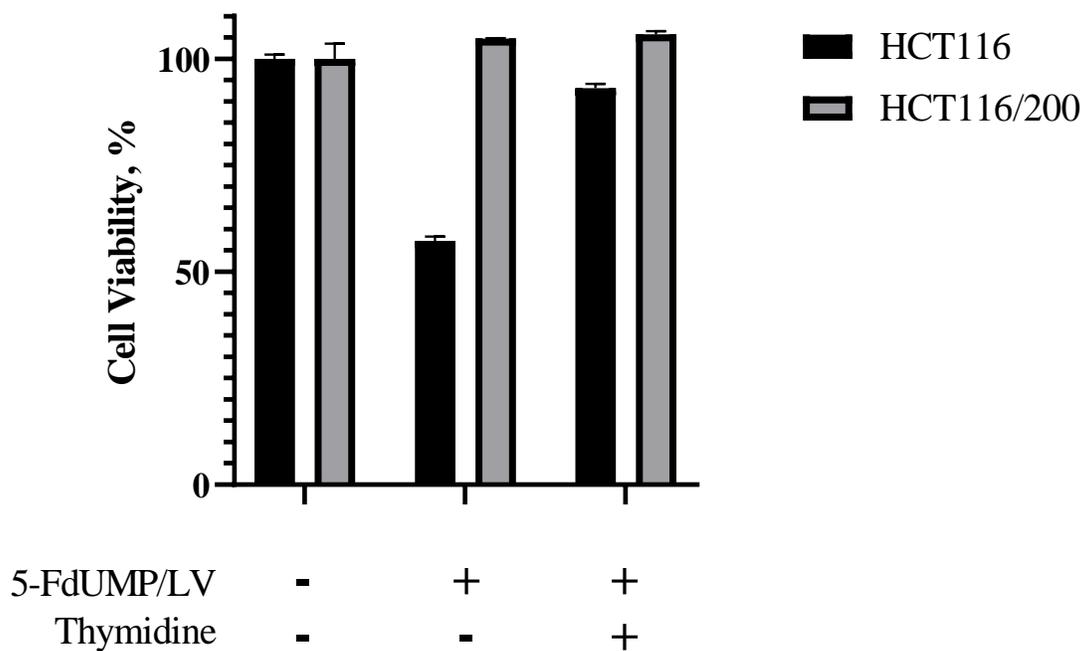


Figure 2.9. CTB assay results from HCT116 and HCT116/200 cells untreated and treated for 24 hours with 10  $\mu$ M 5-FdUMP/LV or 10  $\mu$ M 5-FdUMP/LV/THY (n = 3, data are mean  $\pm$  s.d.).

### ***2.2.6 Global phosphoproteomics of 5-FdUMP/LV treated HCT116, HCT116/200, HCT116 + THY cells over a set time-course***

Based on the initial 5-FU treated HCT116 phosphoproteomic data, it was seen that there were significant changes in the phosphoproteome in comparison to the untreated cells. However, this data only gave a snapshot of the cells at the 48-hour time point rather than an overall picture of how the phosphoproteome changes over the course of the treatment. To generate a more comprehensive dataset we performed a time-course experiment in which SILAC labeled HCT116 and HCT116/200 cells were treated with 10  $\mu$ M 5-FdUMP and 10  $\mu$ M folinic. Meanwhile, SILAC labeled HCT116 cells were treated with 10  $\mu$ M 5-FdUMP and 10  $\mu$ M folinic and 10  $\mu$ M thymidine. For the time-course experiment, cells were harvested after 0.5, 1, 2, 4, 8, 24, and 48 hours. The overall schematic of the experiment is detailed in Figure 2.10. Once the samples were analyzed by MS, the raw data was loaded to Proteome Discoverer to generate the SILAC-based quantitative data. From this analysis, it was determined that the total number of phosphoproteins identified from the samples was 2,319 and the total number of phosphopeptides was 15,167 with an overall phosphorylation specificity of 85% (Figure 2.11A). Furthermore, the total number of Class I phosphosites was 13,536. Class I phosphosites are defined as phosphosites that were localized to an amino acid with a probability of at least 75% and a probability localization score difference greater than or equal to 5. Class I phosphosites along with their associated fold changes is the data that is used in the downstream data analysis software, CLUster Evaluation (CLUE) and Kinase-Substrate Enrichment Analysis (KSEA), as to limit erroneous data interpretation that can come without this statistical cutoff.

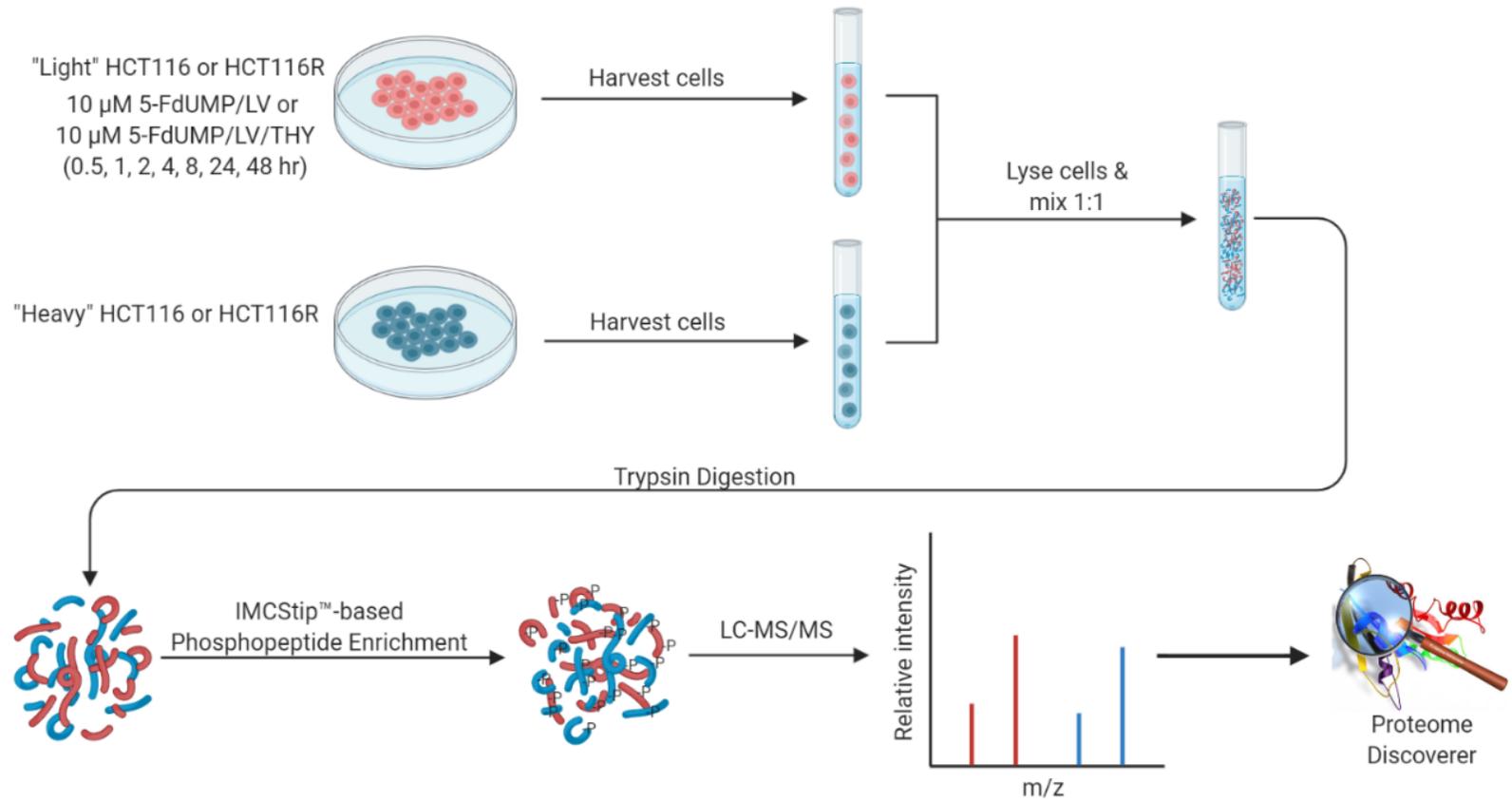


Figure 2.10. Schematic illustration of SILAC-based time course thymidylate synthase inhibitor treatment experiment.

To gain a better understanding of the phosphopeptides enriched during the phosphopeptide enrichment, the number of singly, doubly, and triply phosphorylated peptides was calculated. Of the phosphopeptides identified, 10,944 (72.2%) were singly phosphorylated, 3,351 (22.1%) were doubly phosphorylated, and 872 (5.7%) were triply phosphorylated (Figure 2.11B). This is in line with our previously published phosphoproteomic data for sample enriched with PolyTi resin.<sup>25</sup> Additionally, the number of phospho-serines, phospho-threonines, and phospho-tyrosines that were detected in the samples was calculated. Of the phosphosites identified, 11,919 (88.1%) were phospho-serines, 1,478 (10.9%) were phospho-threonines, and 139 (1.0%) were phospho-tyrosines (Figure 2.11B). This observed ratio of phospho-serines to phospho-threonines to phospho-tyrosines is in line with literature values and most likely represents the overall naturally occurring abundance of each phosphosite.<sup>65</sup> The low percentage of the phospho-tyrosines identified in our dataset highlights the need for specific enrichment of phosphorylated tyrosines as it is known that many tyrosine kinases, such as EGFR and VEGF, play a key role in cancer development.<sup>66</sup>

Using Proteome Discoverer, the fold change between the light and heavy (treated and untreated) proteins could be calculated for each of the samples for each time point. The cutoff for significance for the calculated fold change was set at greater than or equal to 2 and less than or equal to 0.5 which is the standard practice for quantitative proteomics. For each time point and each treatment type, the number of proteins that have a fold change greater than or equal to 2 or less than or equal to 0.5 was calculated (Figure 2.11C). Interestingly, at the 0.5-hour time point, the 5-FdUMP/LV treated HCT116 sample contained significantly more proteins with a fold change greater than or equal to 2

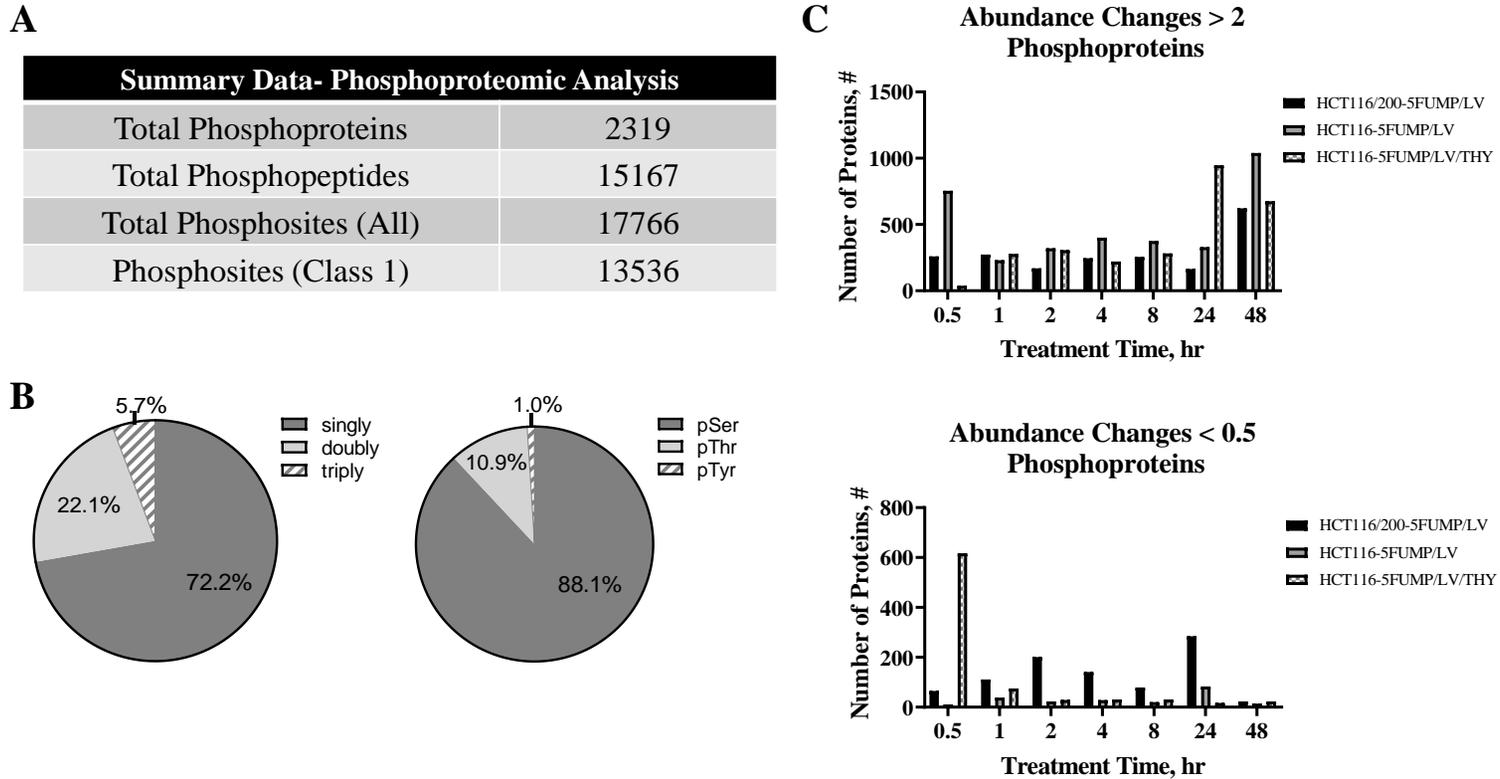


Figure 2.11. Summary from the phosphoproteomic analysis of samples from TS inhibitor time course samples. This summary data includes all samples from each different time point (0.5-, 1-, 2-, 8-, 24-, 48-hour) and from each different sample type (5-FdUMP/LV treated HCT116 cells, 5-FdUMP/LV treated HCT116/200 cells, and 5-FdUMP/LV/THY treated HCT116 cells) for a total of 18 different samples. A) Numbers of phosphoproteins, phosphopeptides, phosphosites, and phosphosites (Class 1) from the MS analysis. B) The phosphopeptide profiles including the percentage of singly, doubly, and triply phosphorylated peptides and the percentage of phosphoserines, phosphothreonines, and phosphotyrosines from all 18 samples. C) Bar graph of the number of phosphoproteins with abundance ratios greater than 2 or less than 0.5 for each of the different treated sample types and at different treatment times.

(754) in comparison to the 5-FdUMP/LV treated HCT116/200 cell sample (259) and to the 5-FdUMP/LV/THY treated HCT116 cell sample (38). Conversely, at the 0.5-hour time point, the 5-FdUMP/LV/THY treated HCT116 cell sample contained significantly more proteins with a fold change less than or equal to 0.5 (617) in comparison to the 5-FdUMP/LV treated HCT116/200 cell sample (65) and to the 5-FdUMP/LV/THY treated HCT116 cell sample (11). This implies that there is a significant early response through the phosphoproteome to the 5-FdUMP/LV treatment for 5-FdUMP/LV-sensitive cells that is observed to a lesser extent in resistant cells and is inverted for sensitive cells that are given thymidine alongside the 5-FdUMP/LV treatment. For the 1–8-hour time points, for 5-FdUMP/LV treated HCT116 cell samples, 5-FdUMP/LV treated HCT116/200 cell samples, and 5-FdUMP/LV/THY treated HCT116 cell samples, there are  $332 \pm 65$ ,  $236 \pm 40$ , and  $272 \pm 31$  proteins with fold changes greater than or equal to 2, respectively. For the 1–8-hour time points, for 5-FdUMP/LV treated HCT116 cell samples, 5-FdUMP/LV treated HCT116/200 cell samples, and 5-FdUMP/LV/THY treated HCT116 cell samples, there are of  $27 \pm 7$ ,  $133 \pm 45$ , and  $41.5 \pm 19$  proteins with fold changes less than or equal to 0.5, respectively. This data indicates that for 5-FdUMP/LV treated HCT116 cell samples and 5-FdUMP/LV/THY treated HCT116 cell samples, there is a relative leveling off of the phosphoproteome response in comparison to the 0.5-hour time point. In contrast, there is no significant difference, in terms of differential abundance of phosphoproteins, observed for the 5-FdUMP/LV treated HCT116/200 cells at the 0.5-hour and 1–8-hour time points. For the 24-hour time point, a significant increase in the number of proteins ( $n = 947$ ) with a fold change greater than or equal to 2 was observed

in the 5-FdUMP/LV/THY treated HCT116 cell sample, whereas the number for the 5-FdUMP/LV treated HCT116 cell sample and for the 5-FdUMP/LV treated HCT116/200 cell sample stayed relatively steady at, 329 and 164, respectively. For the 48-hour time point, for each sample type, there was a significant increase in the number of proteins with a fold change greater than or equal to 2 observed.

### ***2.2.7 Clue analysis of global phosphoproteomics data***

To better understand the changes in the phosphoproteome over time, CLUE was used to cluster phosphosites into different cluster groups and then further can implicate the kinase(s) likely involved in the temporal changes observed in the cluster plot.<sup>67</sup> The software first generates distinct phosphopeptide cluster groups using k-means clustering algorithms where the “k” is the number of clusters. The “k” with the highest enrichment score is the optimal cluster for the dataset.<sup>67</sup> For each cluster group, a temporal plot is generated showing the normalized fold change data for phosphopeptides within the cluster at each time point. The color gradient is formed based on an individual peptide’s membership score to the cluster where green signifies a low membership score and red signifies a high membership score. The “size” indicated for each cluster is the number of phosphosites that fit within the cluster. This program was previously applied to a human embryonic stem cell differentiation phosphoproteomic time course data set to determine which substrates and their associated kinases were involved in the differentiation process.<sup>67</sup> The CLUE analysis requires that the dataset includes only Class I phosphosites that were quantified across each time point. The time course datasets from each of the different treatment types, 5-FdUMP/LV treated HCT116, 5-FdUMP/LV/THY treated HCT116, and 5-FdUMP/LV treated HCT116/200, were restricted to include only Class I

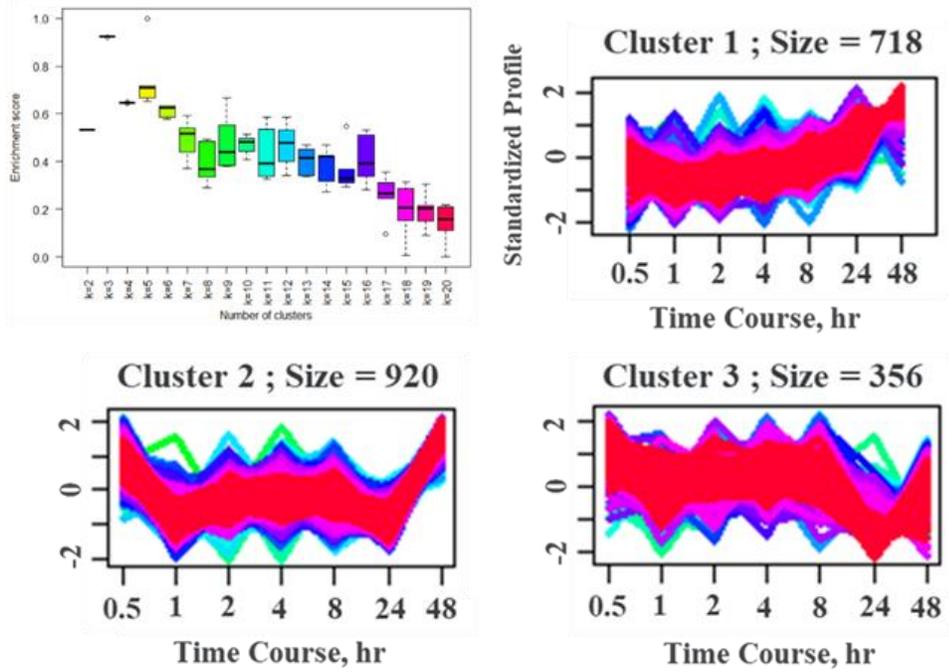
phosphosites that were quantified across each time point and the datasets were then analyzed using CLUE.

The CLUE analysis of the 5-FdUMP/LV treated HCT116 cells time course data resulted in three distinct phosphopeptide clusters based on the enrichment score generated from k-means clustering of the dataset. For each temporal plot generated, there is a slight decline in the normalized fold change from the 0.5-hour time point to the 1-hour time point. This slight decline is followed by general leveling off between 1 and 8 hours. For clusters 2 and 3, at the 24-hour time point there is a slight decrease in the plot followed by a sharp increase at the 48-hour time point, whereas for cluster 1 there is a gradual increase in the plot for both the 24-hour time point as well as the 48-hour time point (Figure 2.12A). For each of the clusters, the software was able to find one or more kinases whose substrates were enriched within the cluster, each with a p-value of less than 0.05.

For Cluster 1, Cyclin dependent kinase 2 (CDK2), Cyclin-dependent kinase 1 (CDK1), and Cyclin-dependent kinase 5 (CDK5) had substrates enriched within the cluster group (Figure 2.12B). In total, 39 phosphosites within the cluster could be linked to these kinases. The 39 phosphosites linked to CDK2, CDK1, and CDK5 were plotted based on the log base two of the fold change ( $\log_2(\text{FC})$ ) for each time point showing a similar plot to the temporal plot generated by CLUE (Figure 2.12C). The average  $\log_2(\text{FC})$  was calculated for the substrate group at each time point. From this data, time points 0.5-, 24-, and 48-hour each show an average  $\log_2(\text{FC})$  greater than 1, signifying a significant overall upregulation in phosphorylation of the substrates for those time points.

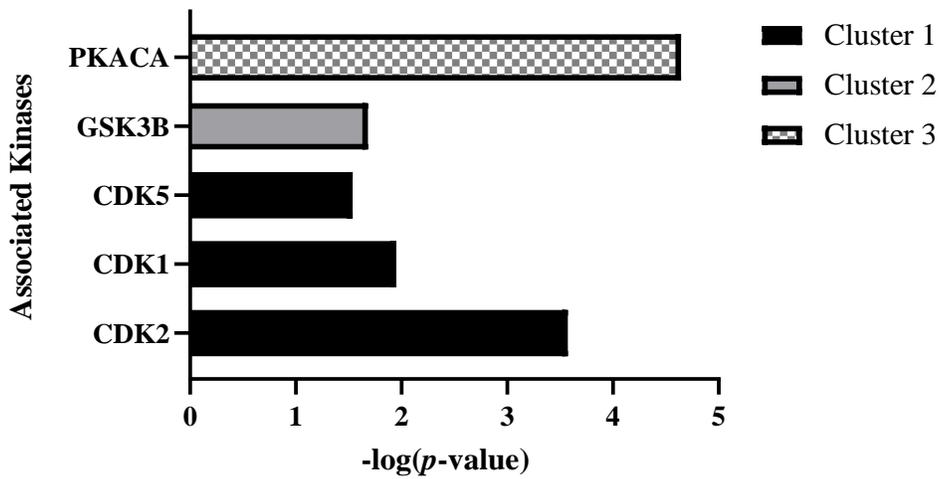
For Cluster 2, the CLUE analysis found glycogen synthase kinase 3 beta (GSK3 $\beta$ ) had substrates enriched within the cluster group. Five phosphosites within the cluster could

**A**



**B**

**Cluster Associated Kinases**



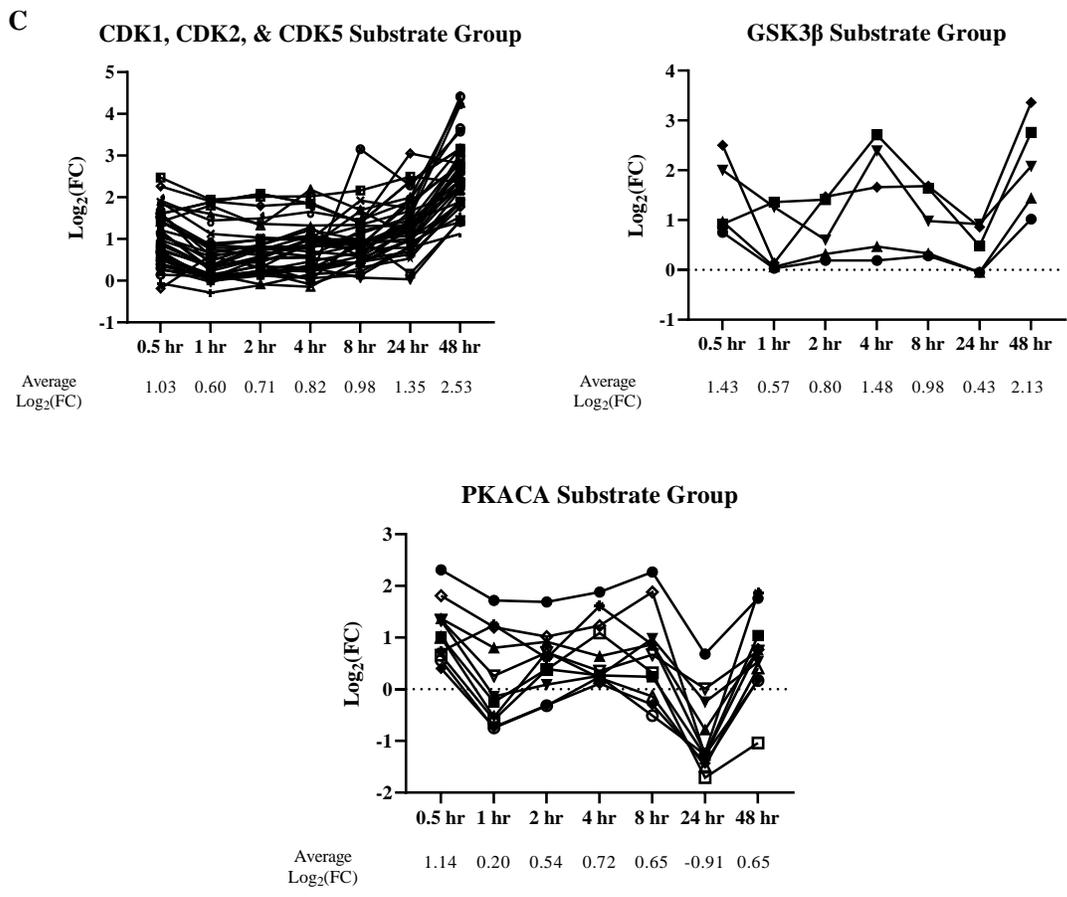


Figure 2.12. CLUE analysis of HCT116 5-FdUMP/LV treated dataset for all time points. This dataset only includes class 1 phosphosites and phosphosites that were quantified at all time points. A) HCT116 5-FdUMP/LV Clusters. B) Cluster Associated Kinases C)  $\text{Log}_2\text{FC}$  changes for specified substrate groups over time course treatment.

be linked to GSK3 $\beta$ . The 5 phosphosites linked to GSK3 $\beta$  were plotted based on the  $\log_2(\text{FC})$  for each time point showing a similar plot to the temporal plot generated by CLUE, although there is a slight deviation at the 4-hour time point (Figure 2.12C). From this data, the 0.5-hour, 4-hour, and 48-hour time points each show an average  $\log_2(\text{FC})$  greater than 1, indicating a significant overall upregulation in phosphorylation of the substrates for those time points.

For Cluster 3, the CLUE analysis found catalytic subunit  $\alpha$  of protein kinase A (PKACA) had substrates enriched within the cluster group. Eleven phosphosites within the cluster could be linked to PKACA. The 11 phosphosites linked to PKACA were plotted based on the  $\log_2(\text{FC})$  for each time point showing a similar plot to the temporal plot generated by CLUE (Figure 2.12C). From this data, only time point 0.5-hour shows an average  $\log_2(\text{FC})$  greater than 1, indicating a significant overall upregulation in phosphorylation of the substrates at that time point (Figure 2.12C).

The CLUE analysis of the 5-FdUMP/LV treated HCT116/200 cells time course data resulted in four distinct phosphopeptide clusters based on the enrichment score generate from k-means clustering of the dataset, however only one kinase, Casein Kinase 2 Alpha 1 (CK2A1), was able to be linked to a certain cluster, Cluster 2 (Figure 2.13A). In total, 19 phosphosites within cluster 2 could be linked to CK2A1. The 19 phosphosites linked to CK2A1 were plotted based on the log base two of the fold change ( $\log_2(\text{FC})$ ) for each time point showing a similar plot to the temporal plot generated by CLUE (Figure 2.13A). From this data, no time point had an average  $\log_2(\text{FC})$  greater than 1 or less than -1, indicating no significant overall upregulation or downregulation in phosphorylation of the substrates at these time points (Figure 2.13C).

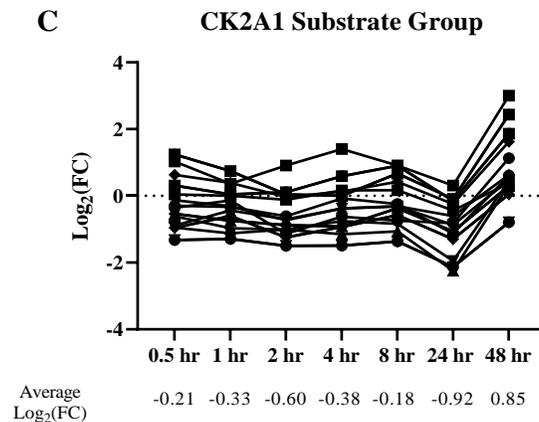
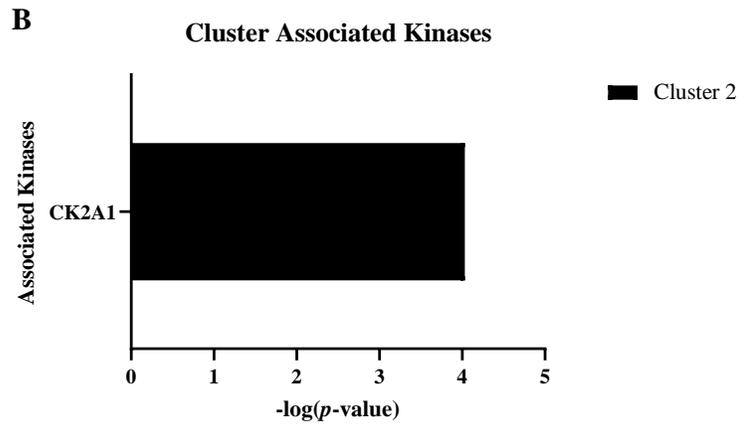
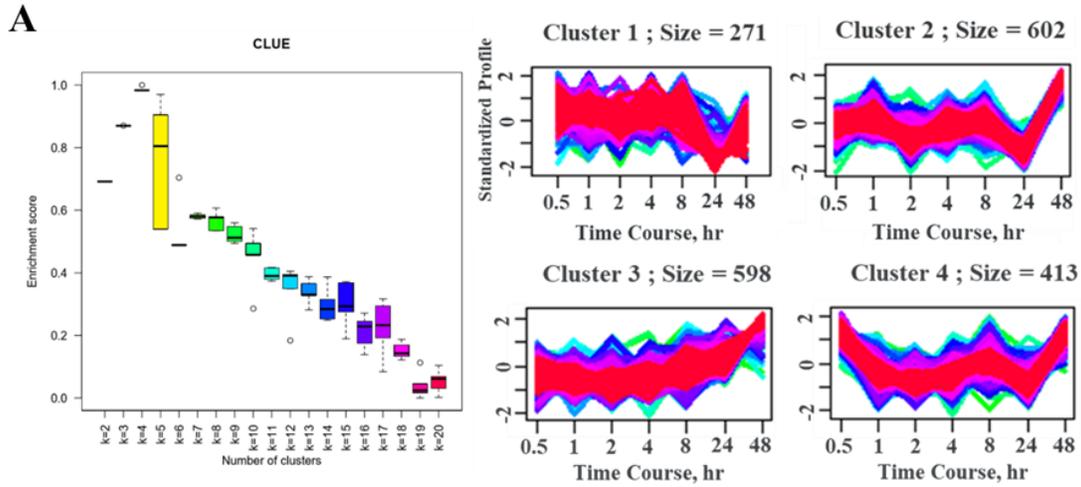


Figure 2.13. CLUE analysis of HCT116/200 5-FdUMP/LV treated dataset for all time points. This dataset only includes class 1 phosphosites and phosphosites that were quantified at all time points. A) HCT116/200 5-FdUMP/LV Clusters. B) Cluster Associated Kinases. C) Log<sub>2</sub>FC changes for specified substrate groups over time course treatment.

The CLUE analysis of the 5-FdUMP/LV/THY treated HCT116 cells time course data resulted in three distinct phosphopeptide clusters based on the enrichment score generate from k-means clustering of the dataset (Figure 2.14A). For cluster 1, two kinases, Cyclin Dependent Kinase 4 (CDK4) and CDK2, were linked to the phosphosites within the cluster.

In total, 29 phosphosites within cluster 1 could be linked to CDK4 and CDK2 (Figure 2.14B). The 29 phosphosites linked to CDK4 and CDK2 were plotted based on the  $\log_2(\text{FC})$  for each time point showing a similar plot to the temporal plot generated by CLUE where there is an overall increase in the average  $\log_2(\text{FC})$  over the course of the treatment. From this data, time points 24-hour and 48-hour have an average  $\log_2(\text{FC})$  greater than 1, signifying a significant overall upregulation in phosphorylation of the substrates at these time points (Figure 2.14C). For cluster 2, two kinases, Ribosomal protein S6 kinase beta-1 (p70S6K) and (cAMP-dependent protein kinase catalytic subunit alpha) PKACA, were linked to the phosphosites within the cluster. In total, 11 phosphosites within cluster 2 could be linked to p70S6K and PKACA. The 11 phosphosites linked to p70S6K and PKACA were plotted based on the  $\log_2(\text{FC})$  for each time point. From this plot, the  $\log_2(\text{FC})$  of the substrates vary significantly at some time points. For example, at the 1-hour time point the average  $\log_2(\text{FC})$  is  $0.69 \pm 1.32$  indicating significant differences in fold changes within this substrate group at that time point (Figure 2.14C). This could be due to some of the substrates within this cluster having low membership scores and thus be differentially regulated compared to substrates with high membership scores to the cluster. From this data, no time point had an average  $\log_2(\text{FC})$  greater than 1 or less than -1, indicating no significant overall upregulation or downregulation in phosphorylation of the substrates at

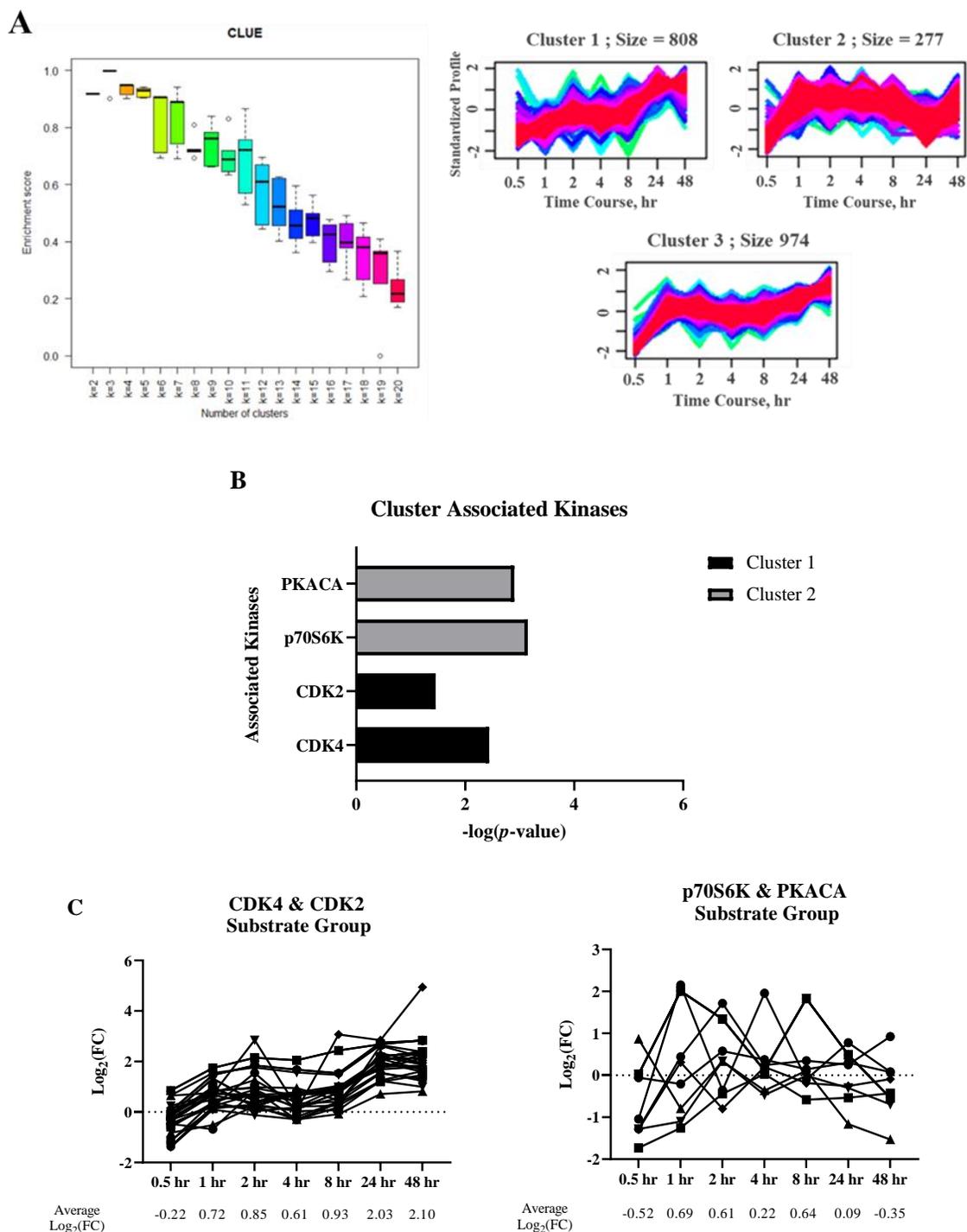


Figure 2.14. CLUE analysis of HCT116 5-FdUMP/LV/THY treated dataset for all time points. This dataset only includes class 1 phosphosites and phosphosites that were quantified at all time points. A) HCT116-5-FdUMP/LV/THY Clusters. B) Cluster Associated Kinases. C) Log<sub>2</sub>FC changes for specified substrate groups over time course treatment.

these time points. Based on these observations, we wanted to probe further into the early phosphorylation changes seen at the 0.5-hour time point for the 5-FdUMP/LV treated HCT116 cells. The early response of the phosphoproteome at the 0.5-hour time point has not been previously reported. To further confirm this data, we repeated the experiment in which HCT116 cells were treated with 5-FdUMP/LV for 0.5 hour in triplicate. The main objective was to confirm the early response seen in the original data was reproducible and, further, to probe which kinases may be involved in this early response to the drug treatment. The phosphopeptide samples were generated using the same workflow from the initial experiment and the data was processed using Proteome Discoverer.

From this analysis, it was determined that the average number of proteins identified from each sample was  $2,018 \pm 26$  and the average number of phosphopeptides was  $5334 \pm 249$  with an average phosphopeptide specificity of  $85 \pm 2\%$  (Figure 2.15A). Furthermore, the SILAC-based fold change calculation was performed for each protein where the fold change was calculated based on the data from each of the samples generated. Similar to the initial data set for the 0.5-hour 5-FdUMP/LV treated HCT116 cell sample, there were a high number of phosphoproteins with a fold change ratio greater than or equal to 2 (394) and a low number of phosphoproteins with a fold change ratio less than or equal to 0.5 (5) (Figure 2.15B).

### ***2.2.8 Kinase substrate enrichment analysis of global phosphoproteomics data***

For this dataset, KSEA was used to understand which kinases were most likely responsible for this early response to the drug treatment. KSEA is a software tool that draws on publicly available phosphosite databases such as PhosphoSitePlus and NetworKin to calculate kinase activity scores (z-scores) from quantitative phosphoproteomic data.<sup>68</sup> KSEA outputs

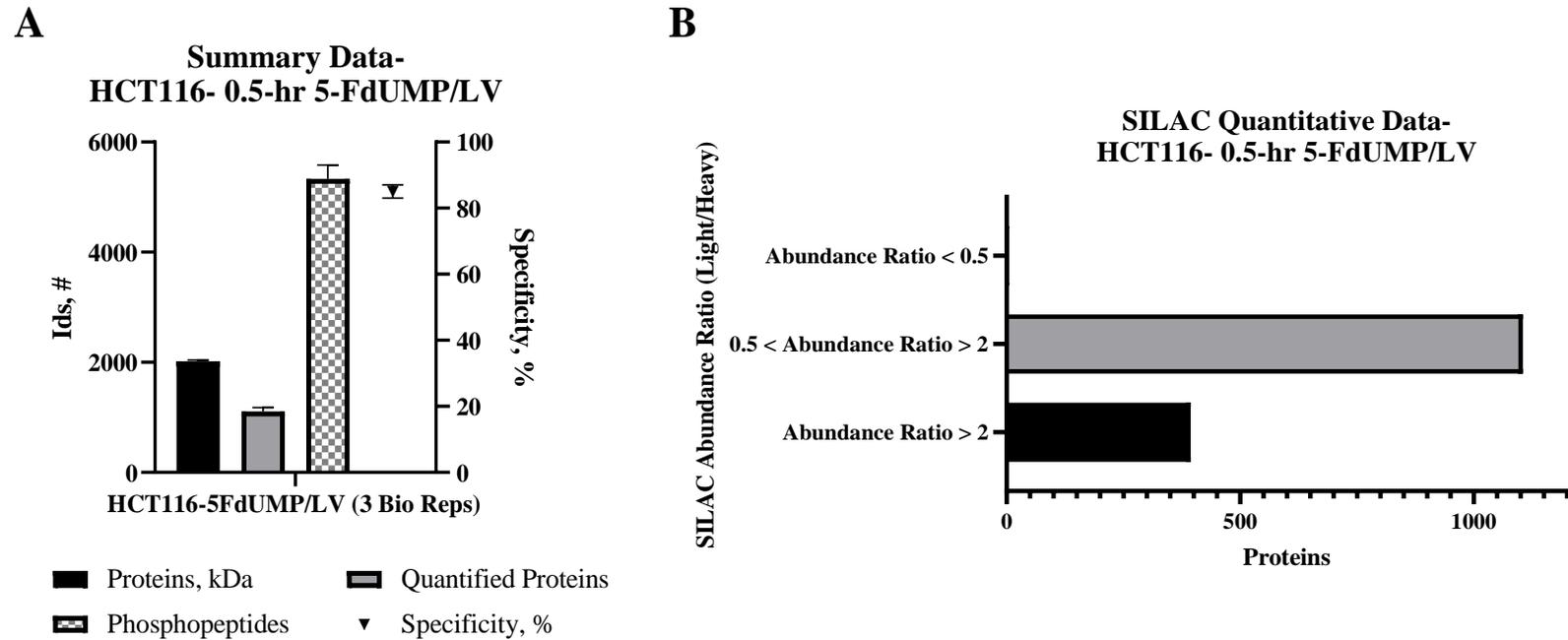


Figure 2.15. Summary data from 0.5-hour 5-FdUMP/LV treated HCT116 samples ( $n = 3$ ). A) Summary of phosphoproteomic data from 0.5-hour 5-FdUMP/LV treated HCT116 cell samples ( $n = 3$ ) including number of protein identifications, phosphopeptide identifications, quantified proteins, and phosphopeptide specificity. B) Summary of SILAC-based abundance ratios calculated from 0.5-hour 5-FdUMP/LV treated HCT116 cell samples ( $n = 3$ , data are mean  $\pm$  s.d.).

a bar plot of kinases based on their calculated z-score and their associated p-values as well as the substrate links the software was able to be determined within the dataset. For the analysis, both the PhosphoSitePlus database and the NetworKin database were used. The minimum number of substrates per kinase required (“m.cutoff”) was set to 5 and the p-value cutoff was set to 0.05.

From the KSEA analysis of the 0.5-hour 5-FdUMP/LV HCT116 cell sample triplicate dataset, it was determined that CDK2, CDK1, CDK4, and Mitogen-activated protein kinase 8 (MAPK8) had the highest calculated z-scores of 5.00, 3.84, 3.05, and 3.03, respectively, with associated p-values less than 0.05. Ribosomal Protein S6 Kinase B1 (RPS6KB1) and Ribosomal Protein S6 Kinase A1 (RPS6KA1) had the lowest calculated z-scores of -2.52 and -3.23, respectively, with associated p-value less than 0.05 (Figure 2.16A). Of note, the KSEA software normalizes the fold change data across the data set so that the resulting kinase z-scores are relative to the dataset. This is shown when plotting the substrate groups for the kinases with the highest and lowest calculated z-scores, where the average  $\log_2FC$  of substrates within the CDK2 and CDK1 substrate group is 1.14 and the average  $\log_2FC$  of substrates with the RPS6KB1 and RPS6KA1 substrate group is 0.13 (Figure 2.16B).

Using KSEA, we compared our triplicate data set to the previously generated 0.5-hour data for all three of the different sample types. Based on the kinase-substrate links determined by KSEA, we generated heatmaps to compare the  $\log_2FC$  of the specified substrates within different kinase substrate groups (Figure 2.17). Interestingly, for most highly active kinases, the CDK1 substrate group, the CDK2 substrate group, and the MAPK family substrate group, the 0.5-hour 5-FdUMP/LV treated HCT116 cell samples

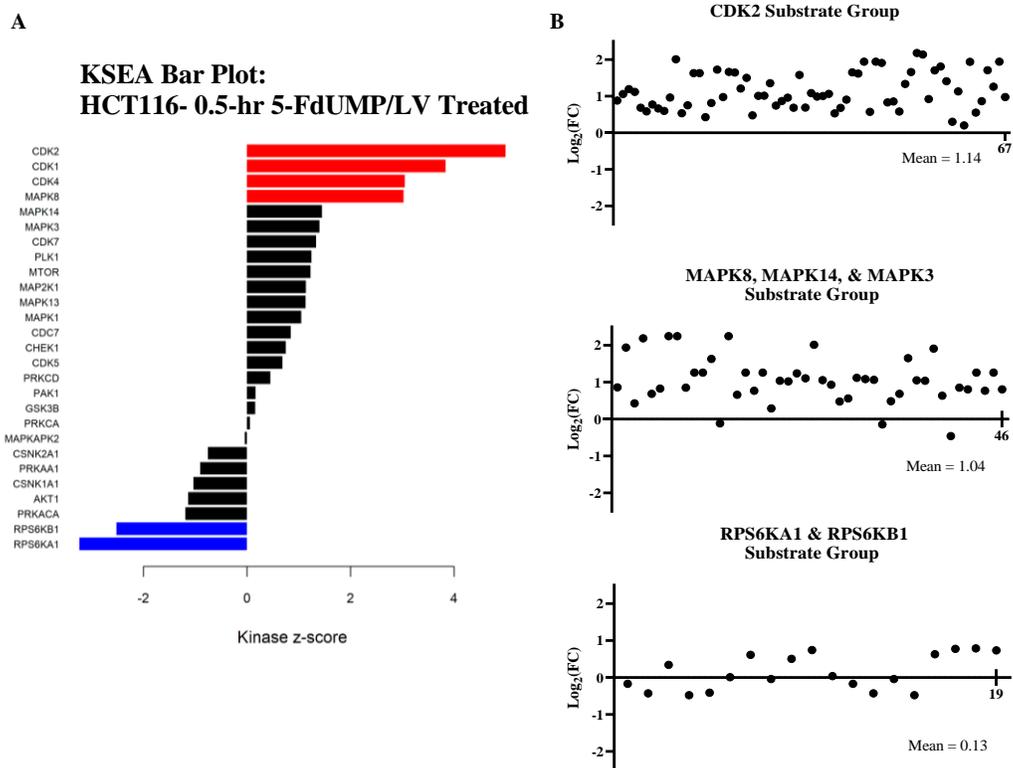
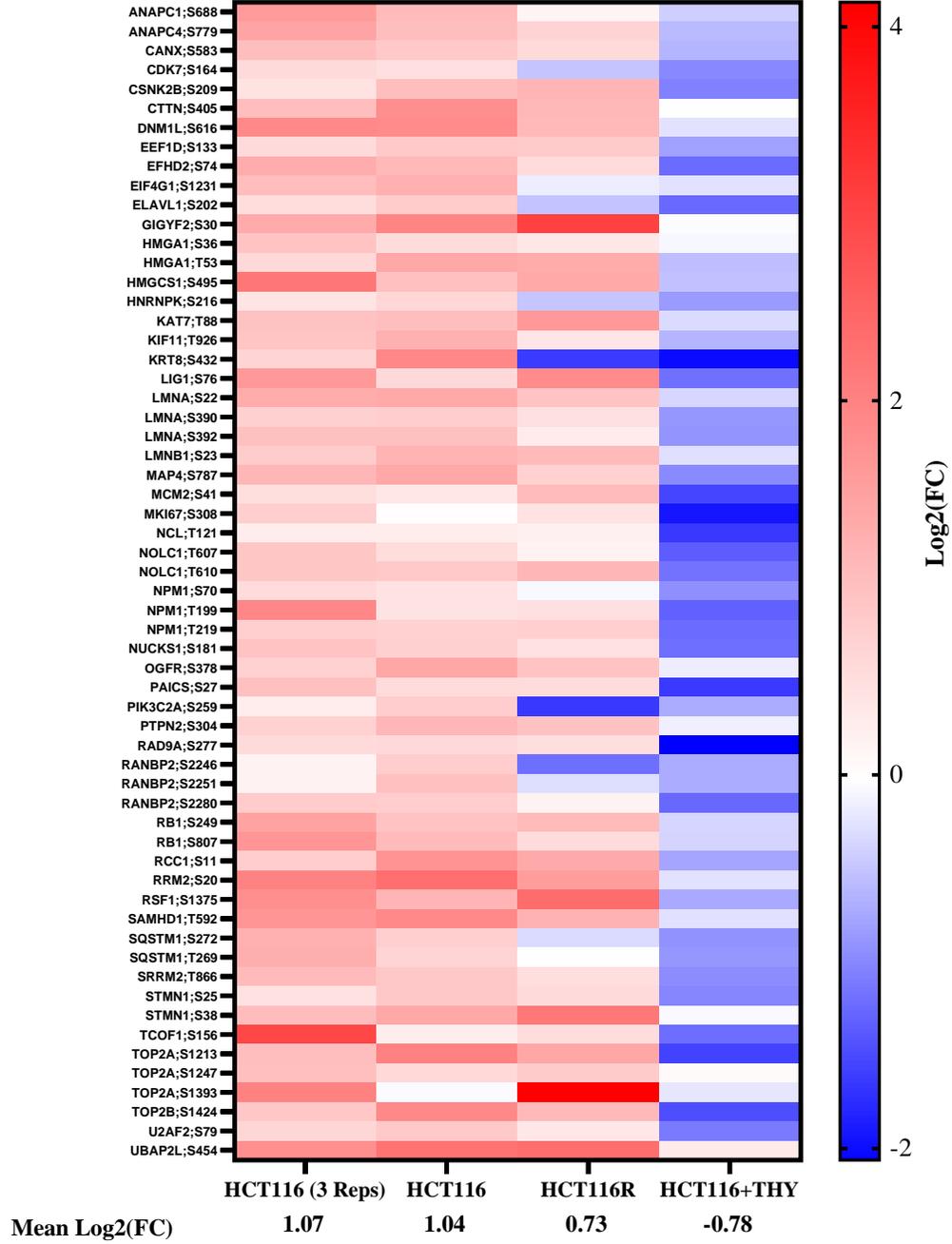


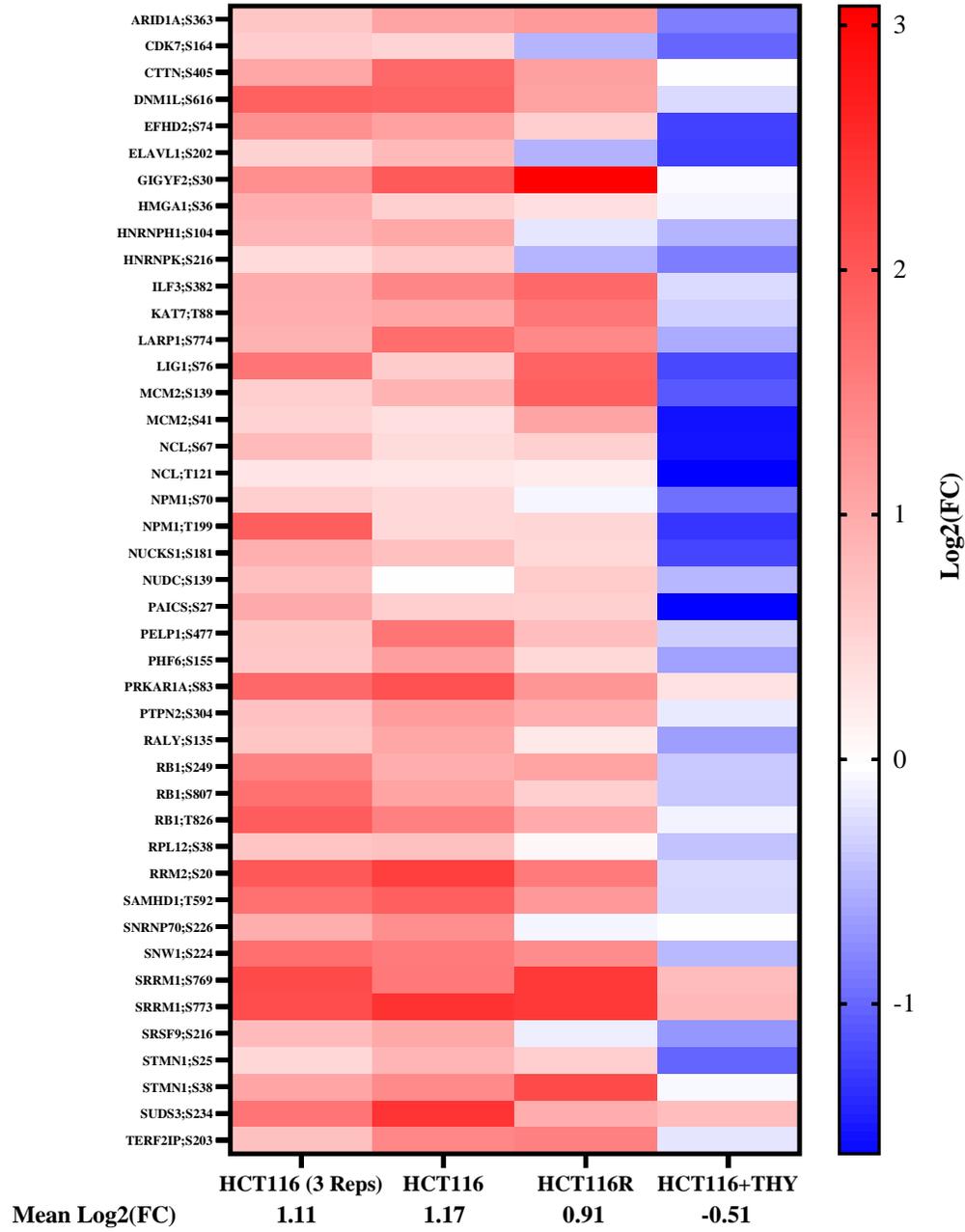
Figure 2.16. KSEA analysis of data from 0.5-hour 5-FdUMP/LV treated HCT116 samples ( $n = 3$ ). This dataset only includes class 1 phosphosites. A) KSEA bar plot from phosphoproteomic data from 0.5-hour 5-FdUMP/LV treated HCT116 cell samples ( $n = 3$ ). Kinases with red or blue bars have at least 5 linked substrates found within the dataset and the calculated z-score has a p-value that is less than 0.05. B) Dot plots of substrates'  $\log_2$  fold changes associated with the indicated kinase(s). The average  $\log_2$  fold change is indicated for each substrate group.

A

CDK1 Substrate Group Heatmap (m=60)



**B** CDK2 Substrate Group Heatmap (m = 43)



### C MAPK Family Substrate Group Heatmap (m=43)

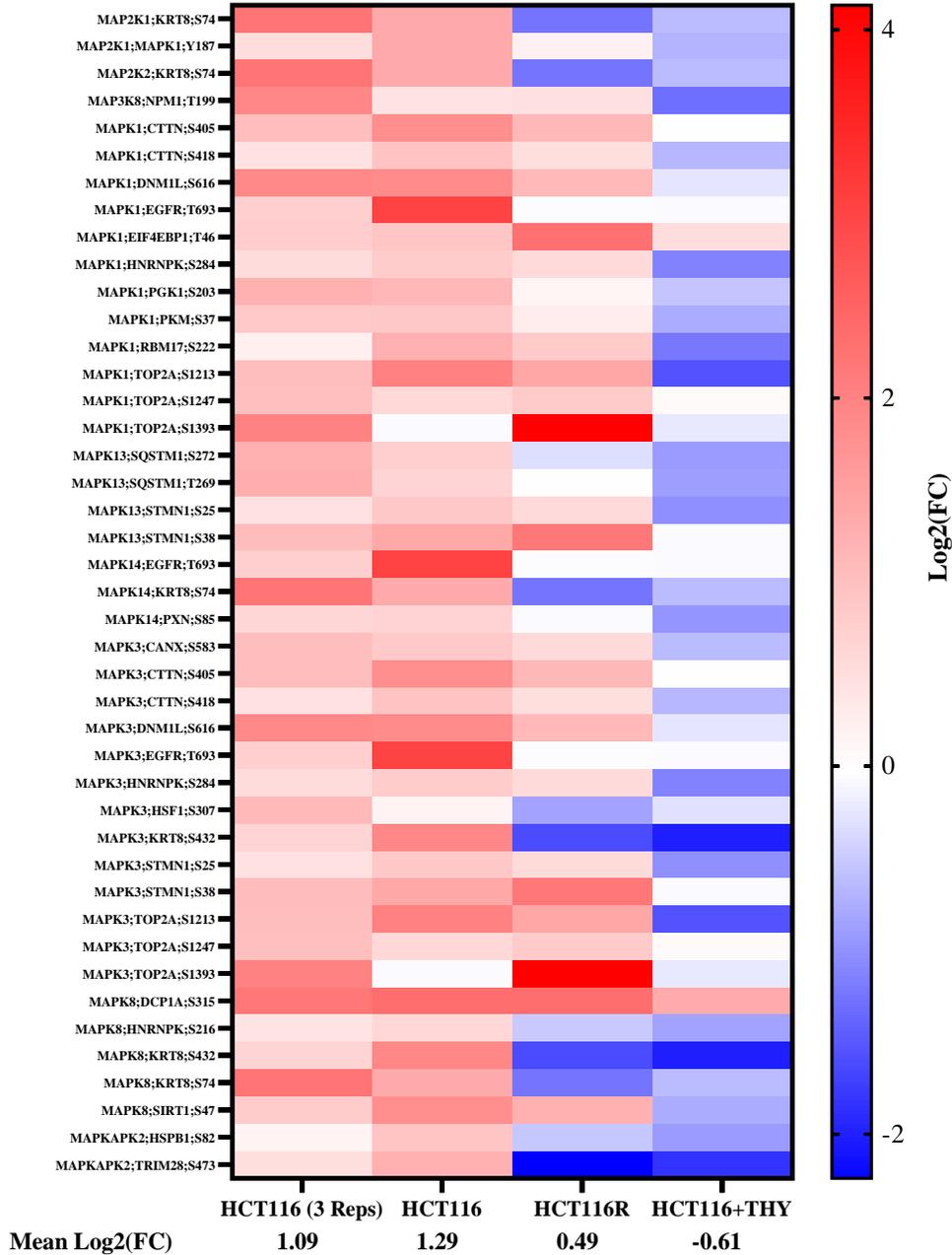


Figure 2.17. Heatmaps generated using data from 0.5-hour 5-FdUMP/LV treated HCT116, HCT116/200, and HCT116+THY samples. A) Heatmap of CDK1 substrate groups from each of the sample groups at the 0.5-hour time point. The mean log<sub>2</sub>(FC) for the substrate group is calculated for each sample group. B) Heatmap of CDK2 substrate groups from each of the sample groups at the 0.5-hour time point. The mean log<sub>2</sub>(FC) for the substrate group is calculated for each sample group. C) Heatmap of MAPK family (MAP2K1, MAP2K2, MAP3K8, MAPK1, MAPK13, MAPK14, MAPK3, MAPK8, MAPKAPK2) substrate groups from each of the sample groups at the 0.5-hour time point. The mean log<sub>2</sub>(FC) for the substrate group is calculated for each sample group.

from both datasets show similar patterns in terms of average fold change within the different substrate groups (Figure 2.17). Additionally, comparing the 0.5-hour 5-FdUMP/LV treated HCT116 cell sample heat maps to the 0.5-hour 5-FdUMP/LV treated HCT116/200 cell sample shows a slight decrease in the fold change averages across each substrate group and the 0.5-hour 5-FdUMP/LV/THY treated HCT116 cell sample heatmaps show a decrease in average fold change averages across each substrate group (Figure 2.17).

Previous reports have shown that CDK1, CDK2, and various kinases within the MAPK family are involved in response to TS inhibition, however generally these responses are shown in cells that have been subjected to the treatment for 24-72 hours rather than 0.5 hour. To our knowledge this is the first report of the early phosphorylation response to TS inhibition. As mentioned previously, phosphorylation is used within the cell as a “on” and “off” switch that signal for responses to stimuli that lead to downstream responses sometimes seen several hours after the initial phosphorylation response. The phosphorylation changes that we observed in this dataset could provide insight into how the cell uses phosphorylation to trigger downstream affects.

Furthermore, these differential responses observed between the sensitive, resistant, and rescued cells could be used to generate a model for prediction of resistance to TS inhibitor treatments. Select proteins from these substrate groups could be used to establish a quantitative assay for the specific phosphopeptides to probe TS inhibitor sensitivity (Table 2.4). This “top 20” list of phosphorylated proteins could provide clinicians with a tool to screen patients for sensitivity to TS inhibitor treatment prior to treatment to determine if TS inhibitor treatment will be effective or if alternative

treatments should be pursued. This tool could be generated by selecting the substrates within the CDK1, CDK2, and MAPK family substrate groups as biomarkers for TS sensitivity. Patient samples could then be collected, subjected to TS inhibitor treatments for 0.5 hour and then the sample could be analyzed using the workflow shown in this work.

### 2.3 CONCLUSION

TS inhibitor resistance is a major problem in the cancer treatment. To understand why innate or acquired resistance occurs, the mechanisms underlying the treatment need to be better investigated. In this study, mass spectrometry-based phosphoproteomics was used to advance our current understanding of how colorectal cancer cells respond to TS inhibitor treatments. To carry out quantitative phosphoproteomic studies we used our automated phosphoproteomic method in conjunction with SILAC-based quantitation to generate reproducible phosphoproteomic data. HCT116 cells treated for 48 hours showed an activation of several different pathways including the EIF2 signaling pathway, the NRF2 signaling pathway, and the p53 signaling pathway. Additionally, we highlighted specific phosphoproteins whose abundance ratios that were significantly changed in response to the 5-FU treatment. In our follow up study using a more targeted TS inhibitor treatment, we showed changes in the phospho-profile over several different time points. Interestingly, the earliest time point, 0.5-hour, showed significant changes in the phosphoproteome and a follow up study of that time point confirmed these changes. The 0.5-hour phosphoproteomic data show clear changes in the CDK1, CDK2, and MAPK family substrate groups that were different between TS inhibitor sensitive, resistant, and thymidine rescued cells.

Table 2.4. Top 20 phosphorylated proteins from CDK1, CDK2, or MAPK family substrate groups that show differential responses to the 5-FdUMP/LV treatment at the 0.5-hour time point.

Protein	Description	Phosphosite	Log <sub>2</sub> (Fold Change) Data			
			HCT116 (3 Reps)	HCT116	HCT116R	HCT116 +THY
DNM1L	Dynamain-1-like protein	S616	1.910	1.868	1.122	-0.236
PRKAR1 A	cAMP-dependent protein kinase type I-alpha regulatory subunit	S83	1.812	2.096	1.267	0.343
SAMHD1	SAM And HD Domain Containing Deoxynucleoside Triphosphate	T592	1.712	1.922	1.234	-0.249
EFHD2	EF-hand domain-containing protein D2	S74	1.334	1.125	0.568	-1.204
PGK1	Phosphoglycerate kinase 1	S203	1.256	1.159	0.148	-0.527
MAP4	Microtubule-associated protein 4	S787	1.176	1.405	0.737	-0.949
EIF4G1	Eukaryotic translation initiation factor 4 gamma 1	S1231	1.063	1.267	-0.146	-0.241
LMNA	Prelamin-A/C	S392	1.009	1.018	0.324	-0.869
KAT7	Histone acetyltransferase KAT7	T88	0.976	1.051	1.642	-0.295
HNRNPH 1	Heterogeneous nuclear ribonucleoprotein H	S104	0.883	1.044	-0.160	-0.468
PKM	Pyruvate kinase PKM	S37	0.869	0.888	0.305	-0.730
RANBP2	E3 SUMO-protein ligase RanBP2	S2280	0.833	0.809	0.200	-1.228
SRSF9	Serine/arginine-rich splicing factor 9	S216	0.816	1.038	-0.108	-0.667
EIF4EBP1	Eukaryotic translation initiation factor 4E- binding protein 1	T46	0.807	0.923	2.314	0.543
RPL12	60S ribosomal protein L12	S38	0.691	0.759	0.088	-0.390
PXN	Paxillin	S85	0.660	0.707	-0.037	-0.935
CDK7	Cyclin-dependent kinase 7	S164	0.596	0.509	-0.480	-0.969
NPM1	Nucleophosmin	S70	0.581	0.447	-0.063	-0.911
MCM2	DNA replication licensing factor MCM2	S41	0.530	0.377	1.092	-1.510
HNRNPK	Heterogeneous nuclear ribonucleoprotein K	S216	0.425	0.642	-0.476	-0.824

Altogether the data generated in this project, especially from the CLUE and KSEA analysis, provide helpful data in future explorations into TS inhibitor sensitivity and resistance. This data justifies that it is necessary to further explore how these specific phosphorylation changes drive sensitivity and resistance within cancer cells. Further research is needed to determine if this workflow could screen colorectal tumor samples for TS inhibitor sensitivity. Finally, we need to explore if this workflow can be applied for other cancer types.

## 2.4 MATERIALS & METHODS

### *2.4.1 Cell culture and treatment*

HCT116 cells were grown in Gibco™ DMEM/F-12, containing 15 mM of HEPES and supplemented with 10% FBS. Cells were passaged by trypsin/EDTA treatment at 80-90% confluence. The cells were passaged at least 5 times prior to treatment with 10  $\mu$ M 5-fluoro-2'-deoxyuridine-5'-monophosphate and 10  $\mu$ M folinic acid for the specified amount of time. After treatment, the cells were washed 3 times with cold 1x PBS. The cells were collected by scraping from the dish, and pelleted by centrifuging for 5 min at 500  $\times$  g. The supernatant was removed, and the cell pellets were stored at -80 °C until further processing.

### *2.4.2 Cell viability assay*

Cell viability of HCT116 and HCT116/200 cells was determined via a CellTiter-Blue (CTB) assay (Promega). Two thousand cells were seeded in 96-well plates and allowed to attach overnight. The cells were incubated with the specified drugs for 24 hours. After 24 hours, diluted CTB reagent (20  $\mu$ L of CTB to 100  $\mu$ L of media) was added and

allowed to incubate for 1 hour. The wells were measured for the florescent product (560(20)<sub>Ex</sub>/590(10)<sub>Em</sub>) using a SpectraMax M2.

### ***2.4.3 Cell lysis***

For HCT116 cell lysis,  $1 \times 10^8$  cells were lysed with 1 mL of RIPA buffer (Thermo Scientific Pierce, Rockford, IL) containing 1 $\times$  of Halt Protease and Phosphatase Inhibitor Cocktail (Thermo Scientific Pierce, Rockford, IL). Protein concentration was determined via Bicinchoninic acid assay. Based on the calculated protein concentration “heavy” and “light” samples were mixed 1:1. Proteins were reduced with addition of 10 mM tris(2-carboxyethyl)phosphine (TCEP) at 56 °C for 30 minutes, then alkylated with addition of 25 mM iodoacetamide for 30 minutes in dark. Next, 1.5 mL of ice-cold acetone was added to the sample to precipitate the protein. The mixture was placed at -80 °C for 30 minutes followed by -20 °C for 30 minutes. The sample was then centrifuged for 5 minutes at  $2000 \times g$  and the supernatant was discarded. The protein pellet was washed using 1 mL of ice-cold acetone and centrifuged for 5 minutes at  $2000 \times g$ . The supernatant was then discarded. The resulting pellet was suspended in 50 mM ammonium bicarbonate to a final concentration of 5 mg/mL followed by overnight trypsin digestion (1:50 enzyme to protein by mass ratio) at 37 °C using MS Grade Trypsin Protease (Thermo Scientific Pierce, Rockford, IL). The sample was then acidified to 1% TFA and then centrifuged for 5 minutes at  $5000 \times g$ . The supernatant was then aliquoted and stored at -80 °C until required for phosphopeptide enrichment.

### ***2.4.4 Phosphopeptide enrichment of HCT116 cell digests***

Cell digest was thawed on ice for 30 minutes. Aliquots of 20  $\mu$ L of cell digest was then diluted with 1% TFA to final volume of 200  $\mu$ L and transferred to a 1 mL 96-well plate.

The peptide sample was then desalted using 5 mg RP 300  $\mu$ L IMCStips (IMCS, Irmo, SC) and then enriched for phosphopeptides using 10  $\mu$ L PolyTi 300  $\mu$ L IMCStips (IMCS, Irmo, SC). The automated peptide desalting and phosphopeptide enrichment was performed on a Hamilton STAR system following the previously published workflow<sup>25</sup>. Once the sample was eluted, the eluted solutions were lyophilized. The samples were then stored at -80 °C until ready for analysis.

#### ***2.4.5 Orbitrap Fusion analysis for global phosphopeptide analysis***

Thermo Scientific Orbitrap Fusion run under conditions established by the staff at the Nevada Proteomics Center was used for sample analysis. Liquid chromatography mass spectrometry (LC-MS) was performed on an UltiMate 3000 RSLCnano system (Thermo Scientific, San Jose, CA) on a self-packed UChrom C18 column (100  $\mu$ m x 35 cm). The gradient used consisted of Solvent B from 2-90% (Solvent A: 0.1% formic Acid, Solvent B: acetonitrile, 0.1% formic Acid) over 175 mins at 50 °C using a digital Pico View nanospray source (New Objectives, Woburn, MA) that was modified with a custom-built column heater and an ABIRD background suppressor (ESI Source Solutions, Woburn, MA). The self-packed column tapered tip was pulled with a laser micropipette puller P-2000 (Sutter Instrument Co, Novato, CA) to an approximate id of 10  $\mu$ m. The column was then packed with 1-2 cm of 5  $\mu$ m Sepax GP-C18 (120 Å) (Sepax Technologies, Newark, DE) followed by 40 cm of 1.8  $\mu$ m Sepax GP-C18 (120Å) at 9000 psi using a nano LC column packing kit (nanoLCMS, Gold River, CA). The MS1 precursor selection range is from 400-1500 m/z at a resolution of 120K and an automatic gain control (AGC) target of  $2.0 \times 10^5$  with a maximum injection time of 100 ms. Quadrupole isolation window at 0.7 Th for MS2 analysis using CID fragmentation in the linear ion trap with a

collision energy of 35%. The AGC was set to  $4.0 \times 10^3$  with a maximum injection time of 150 ms. The instrument was operated in a top speed data-dependent mode with a most intense precursor priority with dynamic exclusion set to an exclusion duration of 60 s with a 10 ppm tolerance.

#### ***2.4.6 Database search and analysis***

The raw data files obtained from the Orbitrap Fusion were searched against a human database using SEQUEST on Proteome Discoverer (Version 2.2, Thermo Fisher). The peptide precursor mass tolerance was set to 10 ppm, and the fragment mass tolerance was set to 0.6 Da. Search criteria included a static modification of cysteine residues of +57.0214 Da and a variable modification of +15.995 Da to include potential oxidation of methionine, a modification of +79.966 Da on serine, threonine, or tyrosine for the identification of phosphorylation (three modifications allowed per peptide). To detect heavy labeled peptides, variable modifications of +6.020 Da on lysine and +10.008 Da on arginine. Searches were performed with full tryptic digestion and allowed a maximum of two missed cleavages on the peptides analyzed from the sequence database. False discovery rates (FDR) were set to less than 1% for each analysis at the PSM level. Phosphorylation site localization from CID spectra was determined by PhosphoRS on Proteome Discoverer 2.2. Confidence view was assigned a score of 0.4, indicating medium confidence. For quantification, the Minora Feature detector was used with the quantification method set to SILAC 2plex (Arg10, Lys6). GraphPad Prism 8 was used for graph generation, statistical analysis, and heat map generation. Ingenuity Pathway Analysis (IPA) was used for performing pathway analysis, generating the tox list, and visualization of signaling pathways. RStudio version 1.3 was used for implementing R-

based programs. CLUE was implemented as an R package. The source code and documentation are freely available from CRAN (<http://cran.r-project.org/web/packages/ClueR/index.html>).<sup>67</sup> KSEAapp was implemented as an R package. The source code and documentation are freely available from GitHub (<https://github.com/casecpb/KSEAapp>).<sup>68</sup>

CHAPTER 3  
PHOSPHOPROTEOMIC ANALYSIS OF TOMATO CELLS TREATED WITH  
GREEN LEAF VOLATILES.

3.1 INTRODUCTION

*3.1.1 Overview of green leaf volatiles*

Green leaf volatiles (GLVs) are emitted by plants and are known to play an important role in plant defense.<sup>69-74</sup> These GLVs are responsible for the smell associated with freshly cut grass as they are emitted following damage to plants.<sup>75</sup> GLVs were first isolated in early 20<sup>th</sup> century where *E-2-hexenal* was extracted from European hornbeam leaves and many others have since been isolated.<sup>70, 72</sup> Since their discovery, significant effort has been made to understand their role in plant-to-plant communication as well as intraplant (or within plant) communication. These 6 carbon compounds are derived from linolenic or linoleic acid and are diverse in structure and include alcohols, aldehydes, and esters. Biosynthesis of GLVs has been linked to the oxylipin pathway where enzymes convert linolenic and linoleic acids into each of the different GLVs (Figure 3.1).<sup>76</sup>

*3.1.2 Roles of green leaf volatiles in plant communication*

GLVs are known to be released by plants under both biotic and abiotic stress situations including herbivory, mechanical damage, fungal or bacterial infection, drought, and heat (Figure 3.2).<sup>69-71, 73, 77-80</sup> GLVs play multiple roles in plant defense in response to these stressors. For example, upon herbivory, the resulting GLVs can attract predators of feeding herbivores and thus they act as an indirect defense mechanism.<sup>73, 74</sup>

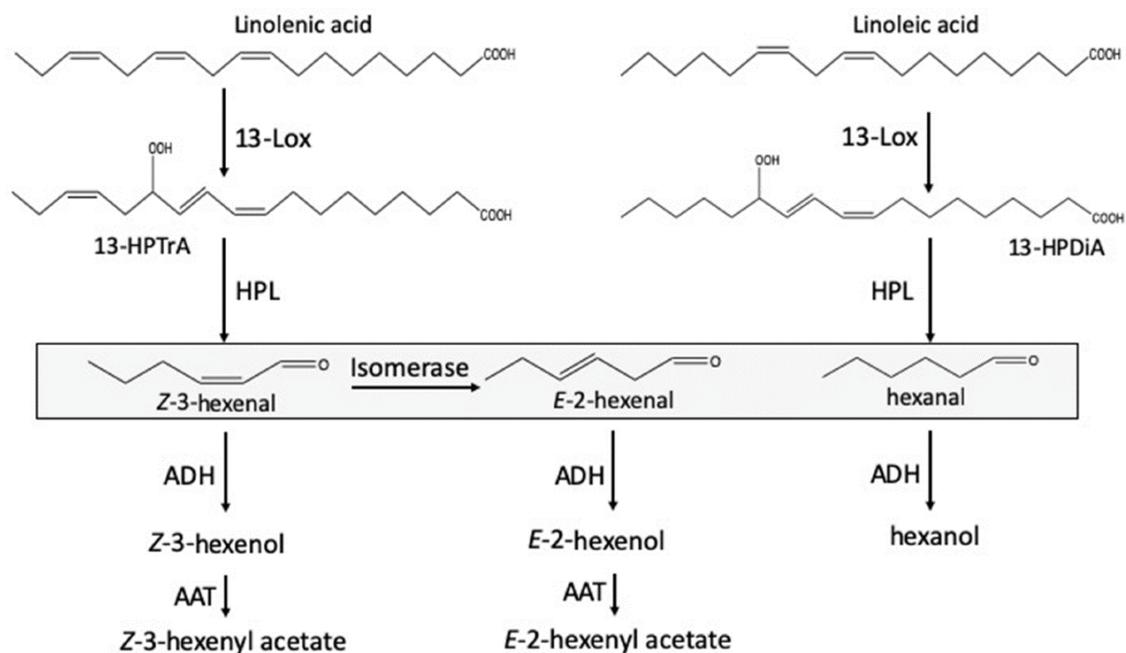


Figure 3.1. Biosynthetic pathway of GLVs. For the formation of GLVs, lipoxygenase (13-Lox) catalyzes the oxygenation of linolenic acid or linoleic acid at the 13<sup>th</sup> position to form 13-hydroperoxy octadecatrienoic and 13-hydroperoxy octadecadienoic acid (13-HPTTrA and 13-HPDiA). Fatty acid 13-hydroperoxide lyase (HPL) then metabolizes 13-HPTTrA and 13-HPDiA further to form *Z*-3-hexenal and hexanal and 12-oxo-(*Z*)-9-dodecenoic acid. *z*-3-hexenal can readily isomerize to form *E*-2-hexenal. The aldehyde groups in the newly formed GLVs can be reduced by alcohol dehydrogenases to form *Z*-3-hexenol, *E*-2-hexenol, and hexanol. Both *Z*-3-hexenol and *E*-2-hexenol can be further modified by acyltransferases to form *Z*-3-hexenyl acetate and *E*-2-hexenyl acetate. The boxed compounds are GLVs highly produced by damaged plant tissue, while the other GLVs are synthesized in cells adjacent to damaged cells. The original figure is from reference 76 and the legend has been modified from the original.<sup>76</sup>

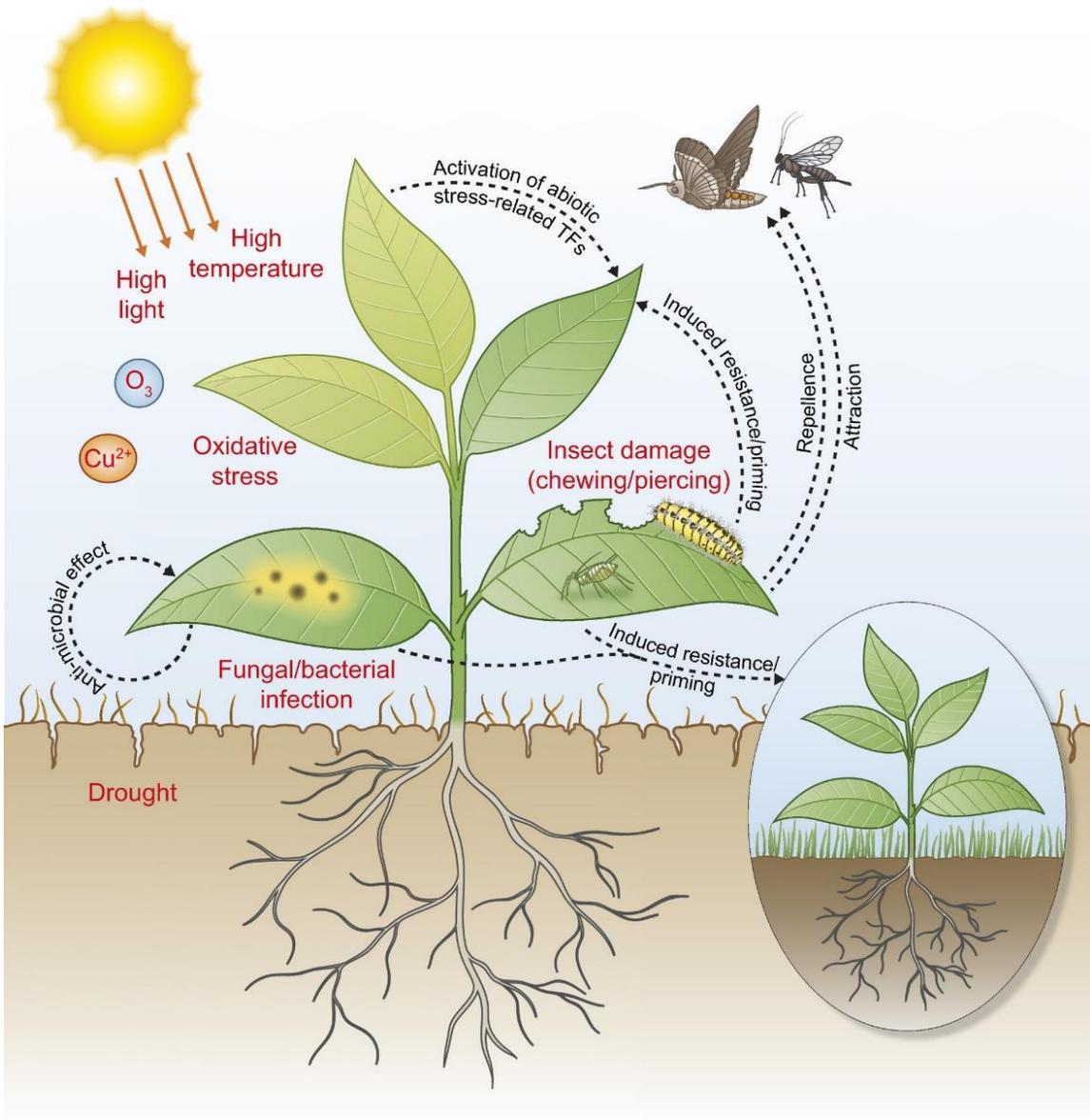


Figure 3.2. Schematic of different stressors that induce GLVs (red) and their associated roles they play in plant defense (black). The original figure is from reference 70 and the legend has been modified from the original.<sup>70</sup>

Another well documented effect of GLVs are priming effects. When GLVs are released by damaged plants, neighboring plants can perceive these volatiles and arm themselves for the impending attack.<sup>70, 74, 81-83</sup> In 2004, Engelberth *et al.* showed that, when infested with caterpillars, maize seedlings released Z-3-hexenal, Z-3-hexenol, and Z-3-hexenyl acetate. When neighboring seedlings were subsequently exposed to caterpillar regurgitate to mimic herbivory, they produced significantly higher amounts of the major defense related hormone jasmonic acid, as compared to untreated plants.<sup>74</sup> In 2006, van Hulst *et al.* reported that *Arabidopsis* treated with a low level (10 mg/L) of  $\beta$ -Aminobutyric acid (BABA), a chemical known to cause resistance to biotrophic and necrotrophic pathogens, did not directly induce defense related genes or lead to significant reductions in relative growth rate (RGR) while *Arabidopsis* treated with higher levels (60 or 200 mg/L) of BABA showed direct induction of defense related genes and a significantly decreased RGR. However, *Arabidopsis* that were primed with 10 mg/L BABA and then exposed to *H. parasitica*, showed enhanced activation of defense related genes compared to non-primed *Arabidopsis* plants.<sup>83</sup> In 2015, Ameye *et al.* used the GLV Z-3-hexenyl acetate to determine if priming effects were seen in wheat from attacks by fungus. In this work, they showed that not only did Z-3-hexenyl acetate pretreated wheat plants show greater resistance to *F. graminearum* in terms of number of necrotic lesions and jasmonic acid production, but so did leaves from seedlings derived from seeds of the primed wheat plants.<sup>81</sup> This highlights that not only do primed plants show higher resistance to fungus, but so do the offspring of the plants. Priming therefore is an efficient defense strategy for plants as it confers the ability to quickly respond to attacks without the costly benefits of a direct defense response (Figure 3.2). While this priming affect by GLVs as well as

many other chemicals has been reported for over a decade, the pathways that cause these priming effects are not well understood. Priming could have meaningful application in sustainable agriculture to arm plants against common biotic stressors in place of current pesticides, but the underlying mechanisms and pathways that lead to priming need to be elucidated for more targeted studies.

### ***3.1.3 Damage-associated molecular patterns and GLVs***

While GLVs are known to have many different functions including priming and antimicrobial effects, over the last couple of years, it has been speculated that GLVs primarily function as DAMPs or damage-associated molecular patterns.<sup>69, 71</sup> DAMPs are host biomolecules that signal danger and trigger damage associated signaling responses such as an increase in reactive oxygen species (ROS), membrane depolarization, influx of  $\text{Ca}^{2+}$ , and activation of mitogen-activated protein kinases (MAPKs).<sup>84-87</sup> These early responses take place within minutes or seconds of exposure, and they trigger production of jasmonic acid through the octadecanoid pathway that lead to the transcriptional upregulation of proteinase inhibitors and amino acid-catabolizing enzymes that have negative effects on the attacking species.<sup>70, 86, 88</sup> Extracellular ATP (eATP) is a classical DAMP. ATP has a relatively high concentration within the cell and when cells are damaged, ATP is released into the extracellular space where it is absent in undamaged tissue. Plant cells have a plasma membrane protein called Does not Respond to Nucleotides1 (DORN1) that binds to eATP and is required for ATP-induced  $\text{Ca}^{2+}$  influx and for increased MAPK signaling demonstrated in *Arabidopsis*.<sup>89, 90</sup> As expected, DORN1 mutant plants show significantly reduced response to ATP treatment and wounding, while plants overexpressing DORN1 show an elevated response to both

treatments.<sup>89</sup> The signaling peptide systemin is another well studied DAMP.<sup>71, 86, 87, 91</sup> Systemin belongs to a class of signaling peptides that are derived from plants and confers a wound response. Systemin contains 18 amino acids and is derived from the C-terminal region of prosystemin, a wound-induced prohormone.<sup>71</sup> Plants lacking prosystemin showed low levels of defense related protein upon wounding, and the overexpression of prosystemin caused continuous expression of defense related genes. The receptor of systemin has been identified as SYR1, a leucine-rich repeat receptor kinase. Once systemin is bound to SYR1, a signaling pathway is induced leading to the activation of MAPKs, ion fluxes, calcium-dependent proteins, ROS, and production of jasmonic acid.<sup>86, 87, 92, 93</sup> Stratmann *et al.* in 2007 used virus-induced gene silencing to co-silence *MPK1* and *MPK2* in tomato plants. This resulted in reduced MPK1/2 kinase activity as well as systemin-induced jasmonic acid production. Silencing of *MPK1* and *MPK2* significantly reduced the prosystemin-mediated resistance to *Manduca sexta*.<sup>86</sup>

As mentioned above, it has been hypothesized that GLVs operate as DAMPs. The early responses to GLVs such as membrane depolarization and an increase in  $[Ca^{2+}]_{cyt}$  have been reported.<sup>84, 85</sup> A complex relationship between GLVs and jasmonic acid has also been reported. Plants treated with GLVs or jasmonic acid show increases in the biosynthesis pathways of both these molecules indicating a positive feedback loop for alerting plants to attack. However, plants that were deficient in jasmonic acid production still showed levels of GLVs after damage by the caterpillar *S. exigua*, highlighting that the jasmonic acid pathway is not singly responsible for GLV production.<sup>70, 94</sup> To better understand the signaling pathways involved in response to GLVs, the signal perception and signal transduction mechanisms need to be further studied. Over the past few

decades, much of the research into GLVs has been on plant-to-insect communication rather than plant-to-plant communication. Therefore, the perception mechanism for GLVs in insects has been characterized but is not understood in plants. The primary processing of plant volatiles in insects is via odorant receptor proteins in the antennae that primarily trigger olfactory signals.<sup>70, 95</sup> These receptor proteins are highly specific and therefore induce specific responses to the volatiles. Similar mechanisms have not been elucidated in plants and it remains to be determined if such specificity of perception for GLVs, for example between Z-3-hexenal and Z-3-hexenyl acetate, is available in plant cells. Furthermore, no genes with homology to odorant receptors are present in plants.<sup>96</sup> In addition to GLV perception, understanding GLV signal transduction is also vital to understanding how the cell triggers the physiological changes seen upon treatment of cells with GLVs. As mentioned previously, these damage signals, namely membrane depolarization and cytosolic influx in  $\text{Ca}^{2+}$  occur within minutes or seconds of treatment. This implies that these changes are not driven by transcriptional changes. It has been hypothesized that phosphorylation and dephosphorylation play a major role in these damage related signals. A recent study by Ahmad *et al.* used mass spectrometry-based phosphoproteomics to study the early phosphorylation changes in response to systemin in tomato (*Solanum peruvianum*) cells.<sup>91</sup> In that work, phosphorylation changes in MAPKs, first reported by Stratmann *et al.* were confirmed.<sup>86, 91</sup> MPK2 was found to be transiently phosphorylated after 15 minutes in response to the systemin treatment. Rapid dephosphorylation of  $\text{H}^+$ -ATPase and phosphorylation of NADPH-oxidase and  $\text{Ca}^{2+}$ -ATPase were observed at the 2-minute time point in response to the systemin treatment. They proposed that dephosphorylation after 2 minutes at the T955 position of the  $\text{H}^+$ -

ATPase is responsible for the systemin-induced alkalization of the growth medium and reported 44 early systemin-responsive kinases that were transiently phosphorylated within their study.<sup>91</sup> This study provided the first phosphoproteomic evidence for the early alarm signals typically seen in response to systemin treatment.

In view of these studies, the aim of this chapter is to contribute to the understanding of GLV perception and signal transduction. In collaboration with the Stratmann group, we show via western blot an early transient phosphorylation of MAPK1 in response to *cis*-3-hexenol, a well-studied GLV. We also observed pH changes in response to *cis*-3-hexenol, which are similar to the pH changes in response to systemin as shown by Ahmad *et al.*<sup>91</sup> Based on this, we used our previously described phosphoproteomic workflow to study the signal transduction in response to *cis*-3-hexenol in tomato (*Solanum peruvianum*) cells.<sup>25</sup> These data showed similar phosphorylation changes for MAPKs, receptor like kinases (RLKs), and calcium/calmodulin signal proteins to previous studies further suggesting the role of GLVs as DAMPs.

## 3.2 RESULTS & DISCUSSION

### ***3.2.1 Activation of MAPK1 in response to GLV and systemin treatments of tomato cells***

It has been hypothesized that GLVs act as DAMPs due to recent reports that they cause depolarization of the plasma membrane and an increase in  $[Ca^{2+}]_{cyt}$ . In 2018, Dombrowski *et al.* reported early activation (3-20 mins) of MAPKs in model grass (*Lolium temulentum*) in response to various GLVs including *cis*-3-hexenol.<sup>97</sup> To confirm this response in tomato cells, in collaboration with Dr. Johannes Stratmann's group, tomato cells were treated with 28 mM *cis*-3-hexenol and cells were collected at different

time points of the treatment. Because *cis*-3-hexenol was dissolved in ethanol, ethanol alone was used as the control sample. Additionally, tomato cells were treated with 10 nM systemin for 5 minutes using the untreated sample as the control sample. These samples were collected and probed for phosphorylated MAPK1 via western blot using a previously described protocol (Figure 3.3).<sup>98</sup>

Based on the western blot, there is clear activation of phosphorylated MAPK1 in response to *cis*-3-hexenol at the 5- and 10-minute time points and in response to systemin at the 5-minute time point, while there is no difference in phosphorylated MAPK1 for any of the control samples. This early (5-10 minute) activation of phosphorylation MAPK1 is consistent with 2018 report by Domborowski *et al.* in which they showed this same activation at these same time points in response to GLVs in model grass.<sup>97</sup> Based on this data, samples treated with either *cis*-3-hexenol or systemin for 5 minutes were used for our phosphoproteomic studies to better understand how tomato cells respond using phosphorylation in response to these treatments.

### ***3.2.2 Initial testing of phosphopeptide enrichment workflow for tomato cells***

The phosphopeptide enrichment in this project was previously applied to human cell lysate. To determine if the current phosphopeptide enrichment workflow could be applied to tomato cells an initial experiment was conducted to compare the resulting phosphoprofile from tomato samples when enriched using the IMCStip-based desalting and phosphopeptide enrichment method. Three different sample types were generated for the initial experiment: tomato cells treated with 28 mM *cis*-3-hexenol for 5 minutes, tomato cells treated with ethanol for 5 minutes, and untreated tomato cells. For each sample,

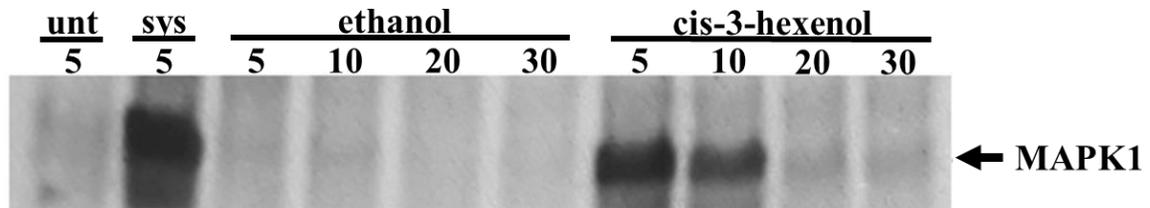


Figure 3.3. Western blot showing changes in phosphorylated MAPK1 in tomato cell lysate in response to different treatments at different time points. The lane under “unt” is untreated tomato cell lysate. The lane under “sys” is tomato cell lysate from cells treated for 5 minutes with 10 nM systemin. Lanes under “ethanol” are tomato cell lysate from cells treated for 5, 10, 20, and 30 minutes with 15  $\mu$ L of ethanol. Lanes under “*cis*-3-hexenol” is tomato cell lysate from cells treated for 5, 10, 20, and 30 minutes with 28 mM *cis*-3-hexenol. The primary antibody used for this protocol was anti-pERK MAPK (Cell Signaling Technology, Danvers, MA).

200  $\mu\text{g}$  of cell lysate was used for the desalting and phosphopeptide enrichment. Nine samples were generated ( $n = 3$  per treatment type) and a sample from each of the treatment types were analyzed by different MS systems, an LTQ Orbitrap Velos, a Q Exactive HF-X, and an Orbitrap Fusion (Table 3.1 & Figure 3.4).

The data from the initial experiment showed that different MS systems resulted in different phosphopeptide specificities from the same sample types (Figure 3.4). The LTQ Orbitrap Velos showed the highest phosphopeptide specificity for the phosphopeptide samples, with an average of 98%, while identifying the lowest number of phosphopeptides, with average of  $514 \pm 36$  phosphopeptides per sample (Figure 3.4). The Q Exactive HF-X showed the second highest phosphopeptide specificity for the phosphopeptide samples, with an average of 92%, while identifying the second highest number of phosphopeptides, with average of  $1804 \pm 154$  phosphopeptides per sample (Figure 3.4). The Orbitrap Fusion showed the lowest phosphopeptide specificity for the samples with an average of 59% but showed the highest average number of phosphopeptide identifications with  $3045 \pm 387$  (Figure 3.4). These differences in phosphopeptide specificity and phosphopeptide identifications could stem from several differences between the MS systems. However, one difference between the sample analysis is that samples analyzed by the Orbitrap Fusion system, were first separated on an LC column that was 35 cm in length, while samples analyzed by the LTQ Orbitrap Velos and the Q Exactive HF-X which used columns that were 10 cm and 15 cm in length, respectively. This column length difference could account some of the difference in phosphopeptide identifications between the Q Exactive HF-X and the Orbitrap Fusion as a longer LC column leads to better separation of the peptides prior to ionization.<sup>31</sup>

Table 3.1. Summary data from the tomato cells with different treatments and analyzed by different MS systems.

Sample	Instrument	Treatment	Protein Ids	Peptide Ids	Phosphopeptide Ids	Enrichment Specificity, %
1	LTQ Orbitrap Velos	Untreated	318	468	464	99%
2	LTQ Orbitrap Velos	Ethanol	347	551	547	99%
3	LTQ Orbitrap Velos	<i>Cis</i> -3- hexenol	359	548	531	97%
4	Q Exactive HF-X	Untreated	859	1647	1597	97%
5	Q Exactive HF-X	Ethanol	1101	2107	1968	93%
6	Q Exactive HF-X	<i>Cis</i> -3- hexenol	1142	2139	1849	86%
7	Orbitrap Fusion	Untreated	2546	5509	3003	55%
8	Orbitrap Fusion	Ethanol	2306	5118	2593	51%
9	Orbitrap Fusion	<i>Cis</i> -3- hexenol	2286	4949	3540	72%

**Phosphopeptide Enrichment Data:  
LTQ Velos v QE-HF-X v Fusion**

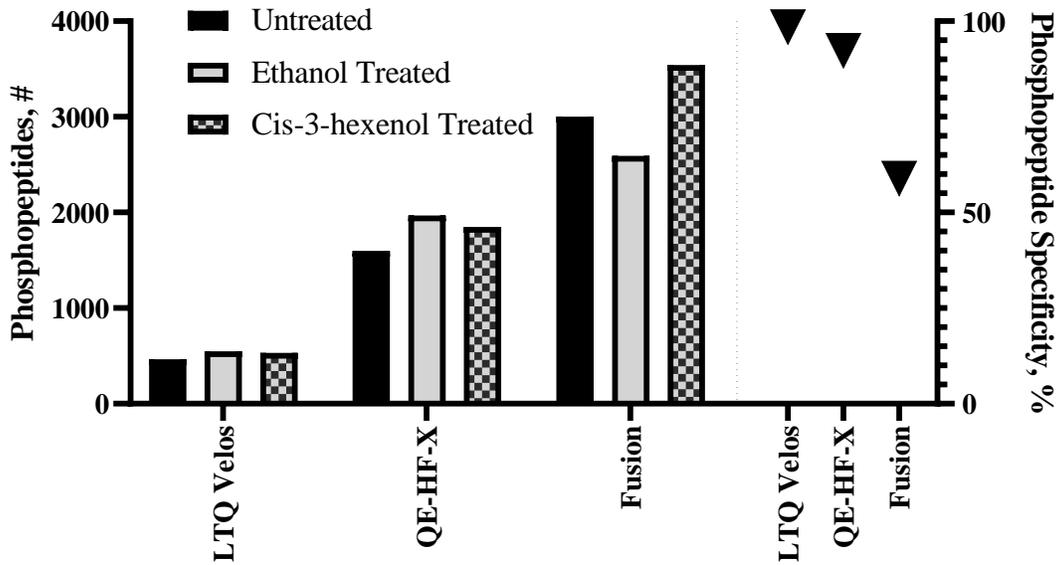


Figure 3.4. Summary data from the treated tomato cells. Three samples per treatment type were collected. A sample from each treatment type was analyzed by the Orbitrap LTQ Velos, the QE-HFX, and the Orbitrap Fusion system.

### ***3.2.3 Label-free quantitative phosphoproteomic analysis of cis-3-hexenol treated tomato cells and systemin treated tomato cells***

Based on the initial study, it was determined that our current phosphopeptide enrichment process could be applied to tomato cell lysis samples. We then applied this workflow to better understand phosphorylation changes that occur in tomato cells in response to GLV treatments. In this experiment, tomato cells were treated for 5 minutes with 28 mM *cis*-3-hexenol, in triplicate, with ethanol treated tomato cells as a control sample. For comparison, tomato cells were also treated with 10 nM systemin for 5 minutes, in triplicate, with untreated cells as the control sample. These samples were prepared in triplicate using the automated desalting and phosphopeptide enrichment workflow and was analyzed by MS. The data was processed in Proteome Discoverer 2.2. The raw data including average protein and phosphopeptide identifications were calculated as well as the phosphopeptide specificity for each treatment type (Figure 3.5).

From the raw data analysis, a total of 1575 phosphoproteins, 6097 phosphopeptides, 7866 phosphosites, and 6849 class I phosphosites were identified from all analyzed samples. On average,  $1321 \pm 51$  phosphoproteins and  $2604 \pm 179$  phosphopeptides were identified from each sample (Figure 3.5B). Additionally, average phosphopeptide specificity was determined to be  $85\% \pm 6\%$  (Figure 3.5B). This data highlights the reproducibility of the sample preparation process as the identifications across all samples are similar and each sample had good phosphopeptide specificity. To gain a better understanding of the phosphopeptides enriched during the phosphopeptide enrichment, the number of phospho-serines, phospho-threonines, and phospho-tyrosines that were detected in the samples was calculated. Of the phosphosites identified, 88.7%

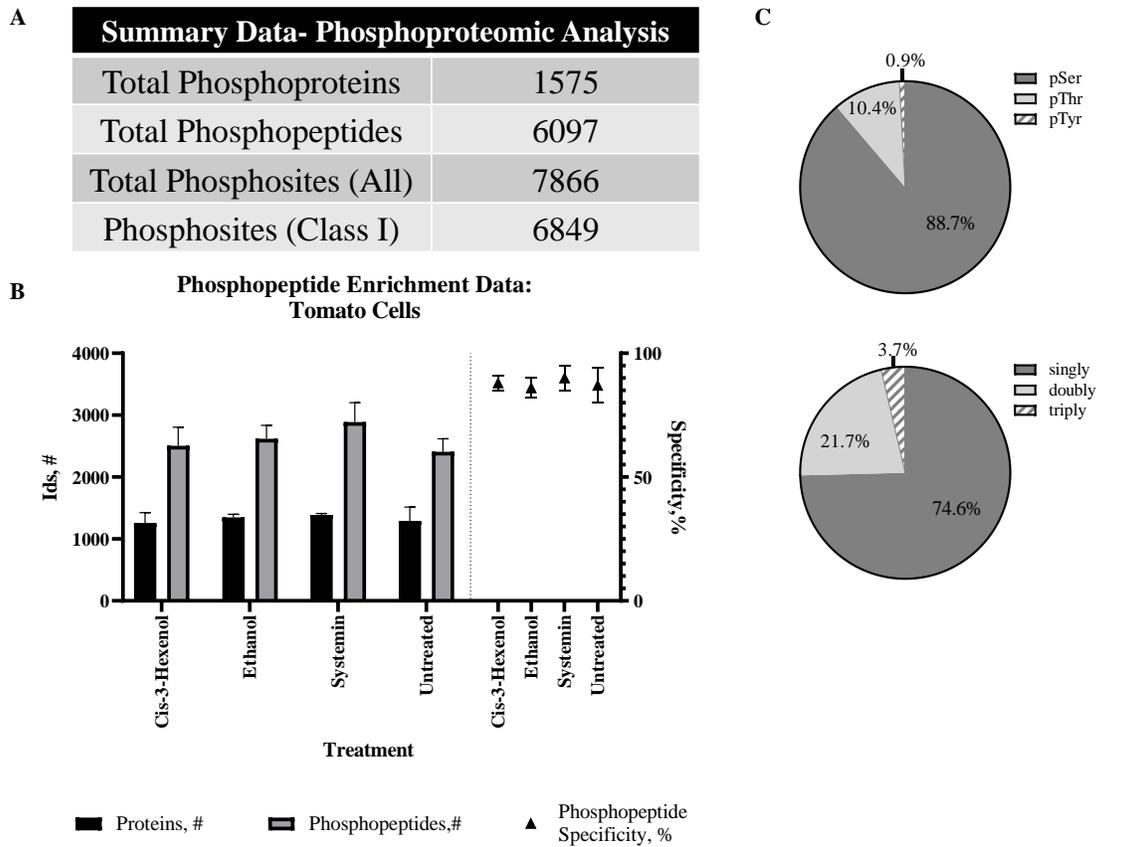


Figure 3.5. Summary of the data generated from phosphopeptide enriched samples from tomato cell lysis ( $n = 3$  per treatment, data are mean  $\pm$  s.d.). A) Summary of number of phosphoproteins, phosphopeptides, phosphosites, and phosphosites (Class 1) from the MS analysis. B) The number of proteins identified, the number of phosphopeptides identified, and the phosphopeptide specificity for each treatment type. C) Summary of the phosphopeptide profile including the percentage of phospho-serines (pSer), phospho-threonines (pThr), and phospho-tyrosines (pTyr) and the percentage of singly, doubly, and triply phosphorylated peptides from all samples.

were phospho-serines, 10.4% were phospho-threonines, and 0.9% were phospho-tyrosines (Figure 3.5C). Additionally, the number of singly, doubly, and triply phosphorylated peptides was calculated. Of the phosphopeptides identified, 74.6% were singly phosphorylated, 21.7% were doubly phosphorylated, and 3.7% were triply phosphorylated (Figure 3.5C). This is in line with our previously published phosphoproteomic data for sample enriched with PolyTi resin and in line with our SILAC-based phosphoproteomic data shown in chapter 2.<sup>25</sup>

Using this MS data, label-free quantitation was performed using the Minora node in Proteome Discoverer. The resulting quantitative data calculates the fold change in comparison to the control sample. In our experiments, all calculated fold changes are based on identifications from at least two replicates. Using Proteome Discoverer, the fold change between the sample and control proteins could be calculated for both the *cis*-3-hexenol treated samples and the systemin treated samples. The cutoff for significance for the calculated fold change was set at greater than or equal to 2 and less than or equal to 0.5 which is the standard practice for quantitative proteomics. For each treatment type, the number of proteins that have a fold change greater than or equal to 2 or less than or equal to 0.5 was calculated (Figure 3.6). For the *cis*-3-hexenol samples, 229 proteins had a fold change of greater than 2 and 7 proteins with a fold change less than 0.5. For the systemin samples, 472 proteins had a fold change of greater than 2 and 14 proteins with a fold change less than 0.5.

**Label Free Quantitation:  
Cis-3-hexenol & Systemin Treatments**

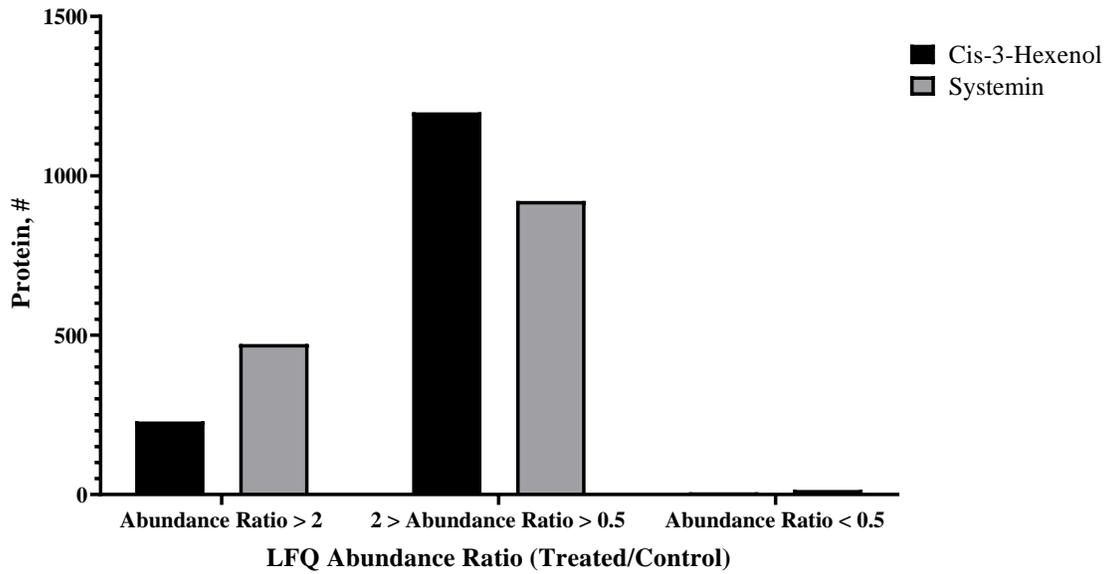


Figure 3.6. Bar graph of the number of proteins with abundance ratios greater than 2, less than 2 and greater than 0.5, and less than 0.5 for both *cis*-3-hexenol treated and systemin treated samples (n = 3 per treatment type).

### ***3.2.4 Comparison of phosphoproteome changes in response to cis-3-hexenol and systemin***

One of the aims of this work is to understand if plant cells respond to GLVs similarly to a well-established DAMP, namely, systemin. To this end, the overlap in phosphoproteins that had fold changes that were significantly increased or decreased in response to both *cis*-3-hexenol and systemin was determined. The number of phosphoproteins whose abundance ratios were significantly changed in both the *cis*-3-hexenol-treated samples and the systemin-treated samples was determined to be 84 (Table 3.2). Consistent with previous data, cells treated with *cis*-3-hexenol and systemin show increased phosphorylation of phosphosite Y223 of MAPK1 (Solyc12g019460.1.1).<sup>86, 91</sup> A representative mass spectrum of the phosphorylated peptide from the *cis*-3-hexenol sample is shown Figure 3.7. The spectrum in Figure 3.7 is annotated with the associated b/y ions from the peptide with the sequence VTSETDFMTEyVVTR. Based on the b/y ions, the phosphorylation site can be localized to the residue Y11, corresponding to Y223 of MAPK1. For both *cis*-3-hexenol and systemin, phosphorylation of phosphosite Y223 of MAPK1 showed the highest fold change of any class I phosphosite within the dataset. Two other MAPKs, MAPK 16 (Solyc01g080240.2.1) and MAPK (Solyc04g007710.2.1), had fold changes greater than 2 in both treatment types. This highlights that both treatment types activate the MAPK signaling pathway which is characteristic of DAMPs.<sup>69, 91</sup> Further, Receptor-like kinase (Solyc02g081040.2.1) and Universal stress protein (Solyc10g006760.2.1) both show increased phosphorylation for both treatments consistent with previous reports and are associated with the DAMP response.<sup>69, 99-101</sup>

Table 3.2. List of phosphoproteins whose abundance ratios were significantly increased in both the *cis*-3-hexenol-treated samples and the systemin-treated samples.

Accession	Protein Description	Average Fold Change: <i>cis</i> -3-hexenol	Average Fold Change: systemin	Modifications
Solyc12g019460.1.1	Mitogen-activated protein kinase 1	8.74	37.24	Y223
Solyc02g081040.2.1	Receptor-like kinase	7.29	5.85	S314
Solyc11g043130.1.1	Phosphatidylinositol 4-kinase	6.45	10.03	S535
Solyc06g069890.2.1	Nucleosome assembly protein family	6.00	6.94	S192
Solyc01g103780.2.1	TCP family transcription factor	5.72	2.09	S327
Solyc10g006760.2.1	Universal stress protein	5.42	4.08	T39
Solyc02g065150.1.1	LOB domain family protein	5.35	10.52	T272
Solyc03g008010.2.1	PPPDE peptidase domain-containing protein 1	5.24	15.83	S187
Solyc10g083360.1.1	Calmodulin binding protein	5.00	17.94	S381; S386
Solyc09g009690.2.1	Ribonucleoside hydrolase 1	4.73	3.18	S112
Solyc07g061930.1.1	Unknown Protein	4.69	2.87	S16
Solyc02g090410.2.1	LOB domain family protein	4.20	9.89	T295
Solyc06g074650.2.1	AP-2 complex subunit alpha	4.18	7.55	S656
Solyc07g065640.2.1	RPM1 interacting protein 4 transcript 2	4.00	2.33	S44
Solyc01g008130.2.1	DEP domain-containing protein 1B	3.78	8.76	S83; S103
Solyc05g051290.2.1	High mobility group family	3.73	15.27	S108
Solyc02g082280.2.1	Genomic DNA chromosome 5 TAC clone K2A18	3.60	14.26	S803
Solyc09g082190.1.1	Unknown Protein	3.54	12.16	S90
Solyc04g072560.2.1	IST1 homolog	3.52	3.70	S351; S376; S517
Solyc03g117420.2.1	Leucine-rich repeat family protein	3.47	9.25	S17
Solyc03g117020.2.1	Stromal cell-derived factor 2	3.40	4.42	S111
Solyc12g055830.1.1	Inorganic pyrophosphatase	3.36	3.10	S25
Solyc08g016070.1.1	Dedicator of cytokinesis 11	3.14	4.96	S388
Solyc03g116780.2.1	CUE domain containing protein	3.09	14.11	T21

Solyc08g065150.1.1	Genomic DNA chromosome 5 P1 clone MHM17	3.06	2.79	S79; S251
Solyc08g005450.1.1	CAX-interacting protein 4	3.02	5.20	S160; S219; S221; S294
Solyc04g071860.2.1	Palmitoyltransferase AKR1	2.96	3.53	S372
Solyc02g068200.1.1	TCP family transcription factor	2.94	3.26	S262
Solyc01g107880.2.1	Membrane-associated ring finger 6	2.93	2.96	S33
Solyc02g069660.2.1	Vacuolar protein sorting-associated protein 54	2.92	9.80	S30
Solyc02g090570.2.1	Transducin family protein	2.87	2.02	S1154; S1155
Solyc01g080240.2.1	Mitogen-activated protein kinase 16	2.77	2.25	T187; Y189
Solyc08g006030.2.1	Unknown Protein	2.74	2.90	S429
Solyc12g098420.1.1	Ubiquitin carboxyl-terminal hydrolase	2.72	10.63	S280
Solyc03g007740.2.1	Reticulon family protein	2.68	17.10	S33
Solyc01g086690.2.1	Double-strand-break repair protein rad21	2.64	2.60	S201; S414
Solyc07g022760.2.1	Pre-mRNA-processing factor 40 homolog A	2.63	2.47	S362; S921; S922; S957; S984
Solyc06g084440.2.1	Nuclear protein localization 4	2.62	12.18	S102
Solyc01g005430.2.1	UBX domain protein 6	2.61	10.09	S16
Solyc09g014500.2.1	Yippee zinc-binding-like protein	2.57	3.02	T126
Solyc01g091340.2.1	Peptidyl-prolyl cis-trans isomerase	2.57	2.31	S379
Solyc02g087760.2.1	Calmodulin binding protein	2.56	2.17	S49; S386; S509
Solyc04g082820.2.1	ARID/BRIGHT DNA-binding domain- containing protein	2.52	7.03	S203; T384
Solyc04g056270.2.1	Calmodulin-binding transcription activator 3	2.52	2.55	S452; S961
Solyc04g007370.2.1	Chaperone protein dnaJ 15	2.51	3.48	S364
Solyc04g007710.2.1	Mitogen-activated protein kinase	2.50	2.01	T180; Y182
Solyc03g095840.2.1	Eukaryotic translation initiation factor 2 alpha subunit family protein expressed	2.50	2.71	S322
Solyc06g071800.2.1	Serine/threonine protein kinase	2.47	2.61	S302; S305; S531
Solyc09g008940.2.1	AT-hook motif nuclear localized protein 1	2.47	3.33	S22

Solyc10g051340.1.1	Adenylyl cyclase-associated protein	2.46	4.22	S303
Solyc05g015920.2.1	GYF domain-containing protein	2.46	2.70	S348
Solyc01g095730.2.1	Os12g0114200 protein	2.44	2.53	S1336
Solyc03g034440.2.1	Eukaryotic translation initiation factor 5	2.41	5.16	S182; S408; S410
Solyc02g082270.2.1	RNA-binding protein 68390-68829	2.40	2.39	S256
Solyc06g067990.2.1	Lys-63-specific deubiquitinase BRCC36	2.39	3.89	S386
Solyc11g027880.1.1	Phosphatidylinositol transfer protein SFH5	2.36	4.02	S102
Solyc02g065630.2.1	Myosin-like protein	2.35	3.25	S289
Solyc04g074230.2.1	14-3-3 protein sigma gamma zeta beta/alpha	2.33	6.61	S211
Solyc04g081820.2.1	La domain containing protein expressed	2.32	2.12	S38; S65; T198
Solyc04g007120.2.1	UV excision repair protein RAD23	2.31	8.88	S77; S231
Solyc07g049480.2.1	Cleavage and polyadenylation specificity factor subunit 3	2.31	2.17	S51
Solyc10g081710.1.1	RING finger protein 43	2.28	10.80	S299
Solyc01g106440.2.1	Ankyrin repeat-containing protein-like	2.28	2.13	S249
Solyc04g082070.2.1	tRNA pseudouridine synthase family protein	2.27	4.40	S191
Solyc09g074940.1.1	Translocase of chloroplast 90, chloroplastic	2.24	2.46	S14; T523; S594
Solyc06g035450.2.1	ATP-dependent RNA helicase	2.22	4.36	S50; S52
Solyc05g014370.2.1	Mitotic spindle checkpoint protein MAD2	2.21	3.62	S185
Solyc05g015500.2.1	Zinc finger Ran-binding domain-containing protein 2	2.20	2.69	S273
Solyc11g020560.1.1	Pol polyprotein	2.20	4.34	T62
Solyc09g007160.2.1	TBC1 domain family member 9B	2.18	3.52	S744
Solyc12g005770.1.1	Zinc finger CCCH domain-containing protein 13	2.18	2.35	S61; S85; S314
Solyc11g072210.1.1	Zinc finger protein-like	2.18	2.12	S457
Solyc02g080470.2.1	Heat shock protein 4	2.18	4.01	S590
Solyc09g097900.2.1	Ras-related protein Rab-25	2.17	3.08	S4
Solyc03g117280.2.1	Unknown Protein	2.16	2.87	S529

Solyc05g054970.2.1	Ankyrin repeat domain-containing protein 13C-A	2.12	2.50	T566
Solyc10g083640.1.1	Splicing factor U2af subunit	2.10	2.52	S259; S265
Solyc11g069600.1.1	Inter-alpha-trypsin inhibitor heavy chain H4	2.09	2.52	S50
Solyc11g008340.1.1	Exocyst complex component 4	2.09	7.46	S331
Solyc10g006770.2.1	FACT complex subunit SPT16	2.08	3.31	S810
Solyc04g007230.2.1	IST1 homolog	2.07	2.58	S231; S243
Solyc07g062940.2.1	Serine/threonine kinase 25	2.06	3.53	S294; S477; T480
Solyc02g069650.1.1	Vacuolar sorting-associated protein	2.03	3.01	S424
Solyc04g078950.1.1	Genomic DNA chromosome 5 P1 clone MBD2	2.01	2.93	S257; S271; S272

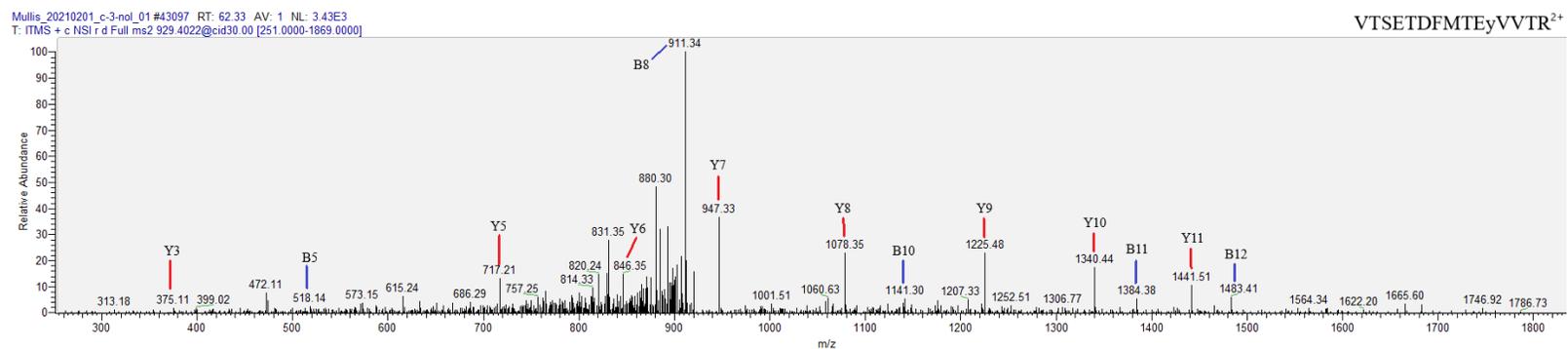


Figure 3.7. Mass spectrum showing the fragmentation pattern from an MS2 scan of a phosphopeptide with the sequence VTSETDFMTEyVVTR with a charge state of 2+. The spectrum is from the *cis*-3-hexenol treated peptide sample MS analysis. This phosphopeptide is part of MAPK1 (Solyc12g019460.1.1) and is phosphorylated at residue Y11 (Y223 for the protein). The B/Y ions were manually annotated.

As mentioned previously, it has been hypothesized that GLVs act as DAMPs due to recent research that showed activation of MAPKs and activation of calcium/calmodulin signaling. To highlight proteins associated with processes or pathways that are involved in the DAMP response, we manually curated a list of proteins from the *cis*-3-hexenol triplicate dataset that could be linked to DAMP associated processes or pathways (Table 3.3 & Figure 3.8).

From the *cis*-3-hexenol dataset, a total of 4 MAPKs were identified and their fold changes were quantified. As mentioned previously, MAPK1 was shown to be increased in response to the *cis*-3-hexenol treatment in line with the data shown in Figure 3.3 and previous reports using *Lolium temulentum*. MAPK1 is a TEK MAPK and is well characterized as being responsible for driving activation of the jasmonic acid pathway.<sup>86</sup> Two other MAPKs (Solyc01g080240.2.1 and Solyc04g007710.2.1) had more modest, but still significant increase in fold change in response the *cis*-3-hexenol. One main difference between these two MAPKs compared to MAPK1 is that they are TDY MAPKs as opposed to TEK.<sup>102, 103</sup> Currently, there has been no functional data for TDY MAPKs and are reported here for the first time to be activated in response to the *cis*-3-hexenol treatment.<sup>102, 103</sup>

As mentioned above, Ahmad *et al.* showed that the plasma membrane H<sup>+</sup>-ATPase (AHA) was rapidly dephosphorylate at the T955 position at 2 minutes in response to systemin treatments and showed a decreased activity of H<sup>+</sup>-ATPase at 5 minutes presumably due to the rapid dephosphorylation at the 2-minute time point. Ahmad *et al.* showed that there is a rapid alkalinization of the media and hypothesized that the

Table 3.3. List of phosphoproteins from the *cis*-3-hexenol triplicate dataset that are linked to DAMP associated processes or pathways.

	Accession	Protein Description	Average Fold Change: <i>cis</i> -3-hexenol	Modifications
<b>MAPKs</b>	Solyc12g019460.1.1	Mitogen-activated protein kinase 1	8.737	Y223
	Solyc01g080240.2.1	Mitogen-activated protein kinase 16	2.769	T187; Y189
	Solyc04g007710.2.1	Mitogen-activated protein kinase	2.499	T180; Y182
	Solyc06g068990.2.1	Mitogen-activated protein kinase	1.022	T270; Y272; S529
<b>H<sup>+</sup>-ATPase Plasma Membrane</b>	Solyc03g113400.2.1	H <sup>+</sup> -ATPase	1.802	T955
	Solyc07g017780.2.1	H <sup>+</sup> -ATPase	0.965	T897; S915; T964
<b>Receptor Like Kinases</b>	Solyc02g081040.2.1	Receptor-like kinase	9.8815	S314
	Solyc01g109650.2.1	Receptor like kinase, RLK	3.1575	S813; S832
	Solyc11g011020.1.1	Receptor like kinase, RLK	2.22125	S356
	Solyc11g006040.1.1	Receptor like kinase, RLK	1.847	S529
	Solyc01g080880.2.1	Receptor-like kinase	1.683	S13; Y212
	Solyc04g054200.2.1	Receptor like kinase, RLK	1.638	S732; S737
	Solyc01g108840.2.1	Receptor like kinase, RLK	1.657	S532; S534
	Solyc02g077850.2.1	LRR receptor-like serine/threonine-protein kinase	1.483	S584
	Solyc09g083210.2.1	Receptor-like protein kinase	1.479	S622
	Solyc03g111670.2.1	Receptor like kinase, RLK	1.370	S327; S354
	Solyc07g056270.2.1	Receptor-like protein kinase	1.345	S134
	Solyc10g085000.1.1	Receptor-like kinase	1.210	S25
	Solyc03g111670.2.1	Receptor like kinase, RLK	1.16975	S327; S354
	Solyc02g086270.2.1	Receptor-like kinase	1.122	S766; S767
	Solyc09g064270.2.1	Receptor-like kinase	1.0335	S550
	Solyc12g044840.1.1	Receptor-like kinase	1.014	S178
Solyc11g013880.1.1	Receptor-like protein kinase	0.939	S785	

	Solyc09g064270.2.1	Receptor-like kinase	0.889	S550
<b>Calcium/ Calmodulin Signaling</b>	Solyc01g112250.2.1	Calcium-dependent protein kinase 3	2.139	S526
	Solyc04g009800.2.1	Calcium-dependent protein kinase 2	1.571	T65; S572
	Solyc11g018610.1.1	Calcium-dependent protein kinase 4	1.225	S371; T489
	Solyc11g065660.1.1	Calcium-dependent protein kinase 8	0.637	S524
	Solyc04g049160.2.1	Calcium dependent protein kinase	1.125	T491
	Solyc01g096830.2.1	Calcium-transporting ATPase 1	1.692	S859
	Solyc02g064680.2.1	Calcium-transporting ATPase 1	1.498	S46
	Solyc10g083360.1.1	Calmodulin binding protein	5.001	S381; S386
	Solyc06g073830.1.1	Calmodulin-like protein	3.351	S92
	Solyc02g087760.2.1	Calmodulin binding protein	2.565	S49; S386; S509
	Solyc01g008950.2.1	Calmodulin 5/6/7/8-like protein	2.295	S82
	Solyc08g080470.2.1	Calmodulin binding protein	1.715	S344; S354; S492
	Solyc04g081210.2.1	Calmodulin binding protein IQ	1.538	S53; S168; S589; S686; T696; S841
	Solyc10g074740.1.1	Calmodulin-like protein	1.439	S28; S35
	Solyc01g005800.2.1	Calmodulin binding protein	1.035	S411
	Solyc04g056270.2.1	Calmodulin-binding transcription activator 3	2.521	S452; S961
	Solyc12g099340.1.1	Calmodulin-binding transcription activator 5	1.812	S334
Solyc10g006010.2.1	Calcium-activated outward-rectifying potassium channel 1	1.151	S23; S33	

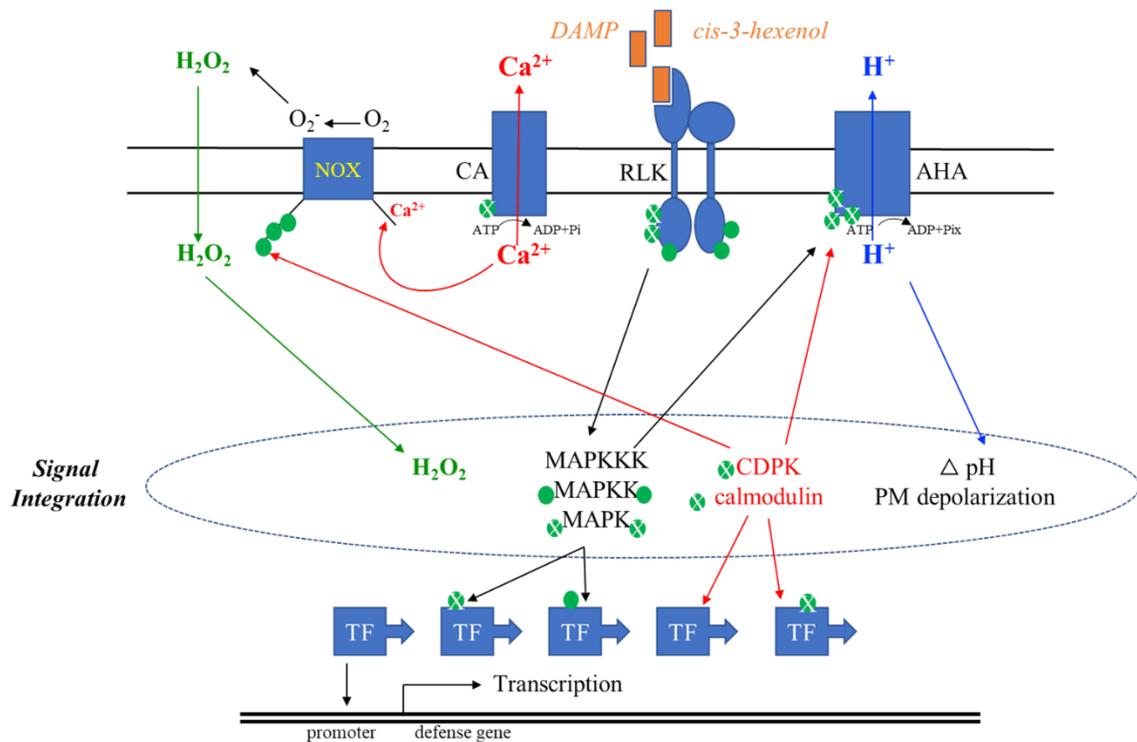


Figure 3.8. DAMP/Hexenol signal perception and signal integration pathway based on phosphoproteomic dataset. Green dots are previously reported phosphorylation sites. Green dots with an “X” are previously reported phosphorylation sites observed in our phosphoproteomic dataset. The pathway shows potential binding of *cis*-3-hexenol to RLKs (receptor like kinases) that lead to the activation of MAPK activation. Calcium dependent protein kinases (CDPK) and calmodulin proteins are activated that drive the production of reactive oxygen species (O<sub>2</sub><sup>-</sup>, H<sub>2</sub>O<sub>2</sub>, etc.) through phosphorylation of NADPH oxidase (NOX). Activation of MAPKs and CAPK/calmodulin proteins also drive activation of transcription factors leading to the expression of defense related genes. Phosphorylation of Ca<sup>2+</sup> ATPases (CA) and plasma membrane H<sup>+</sup>-ATPase (AHA) play key roles in regulating Ca<sup>2+</sup> and H<sup>+</sup> levels which are key for the DAMP response.

decreased activity of the H<sup>+</sup>-ATPases is responsible for the alkalization of the media. In their dataset, at 5 minutes the phosphorylation of both H<sup>+</sup>-ATPases were not significantly different than the starting point level and there is a subsequent increase in phosphorylation at the T955 position, more pronounced in Solyc03g113400.2.1, after 15 minutes. Similarly, Stratmann *et al.* observed alkalization of the cell media in response to different GLVs. Our dataset is in line with Ahmad *et al.* in that at the 5-minute time point there is no significant change in the phosphorylation of either H<sup>+</sup>-ATPases. In future studies, similar time point data could be generated to show these transient phosphorylation levels at the 2-minute and 15-minute time points to further confirm this data.

From Table 3.3 three receptor like kinases (Solyc01g109650.2.1, Solyc11g011020.1.1, Solyc01g108840.2.1) were found to have increased fold changes in response to the *cis*-3-hexenol treatment. Solyc02g081040.2.1 is the LysM receptor-like kinase SILYK11 and is a part of a clade of RLKs (SILYK1, SILYK12, SILYK13). An ortholog in *Arabidopsis* of SILYK11 called CERK1 is the receptor for chitin-oligomers and it has been shown that plants with silenced *SILYK11*, *SILYK1*, *SILYK12*, *SILYK13* are more susceptible to *Pseudomonas syringae*.<sup>104</sup> Overall RLKs are known to play a role in plant immunity and the activation of these RLKs could point to a general binding of *cis*-3-hexenol to RLKs for activation of downstream signaling.<sup>101</sup>

Calcium and calmodulin signaling has previously been shown in response to GLVs.<sup>84, 85</sup> In our dataset, many different calcium/calmodulin associated signaling proteins were identified as well as Ca<sup>2+</sup> ATPases (CA). Specifically, calcium-dependent protein kinase 3 (CDPK3) (Solyc01g112250.2.1) was found to be increased in response

to the *cis*-3-hexenol treatment as well as 4 calmodulin-like or calmodulin-binding proteins and 1 calmodulin-binding transcription factor. CDPKs are early responders to changes in  $[Ca^{2+}]_{cyt}$  levels and have been shown to play a key role in stress signaling and lead to synthesis of reactive oxygen species (ROS) through phosphorylation of NADPH oxidase (NOX), increase in phytohormone synthesis, and changes in gene expression through activation of transcription factors (TF).<sup>105-107</sup>

Altogether this data provides novel insight into how the phosphoproteome is changed in response to *cis*-3-hexenol and further suggest that GLVs act as DAMPs. To our knowledge this is the first phosphoproteomic analysis of plant cells treated with a GLV and thus many of these phosphoproteins found in Table 3.2 and Table 3.3 have not been reported as being involved in the cellular response to GLVs. These lists could provide insight into phosphoproteins that drive the cellular response leading to traditional DAMP signals. Further work probing these specific proteins will be needed to further understand their relevance in the cellular response to GLVs.

### 3.3 CONCLUSION

Over the past decade, researchers have found that GLVs play a key role in interplant and intraplant signaling. GLVs have been shown to lead to priming affects in neighboring plants that causes increase resistance to future attacks without the cost of an active defense attack. Additionally, GLVs are involved in intraplant communication that lead to upregulation of key defense proteins in response to active abiotic and biotic stressors. Over the past couple of years, it has been hypothesized that GLVs act as DAMPs, which show characteristic stress signaling such as an increase in MAPK signaling and increase in  $[Ca^{2+}]_{cyt}$ . However, in-depth studies of how the

phosphoproteome changes in response to GLVs have not been conducted. Here, we use mass spectrometry-based phosphoproteomics to show changes in tomato cells in response to *cis*-3-hexenol, a known GLV, and compared the phosphoproteome changes to systemin treated tomato cells. Our generated data shows phosphoproteome changes consistent with previous data, specifically increased phosphorylation of MAPKs, RLKs, and calcium/calmodulin associated signaling phosphoproteins. In future studies, our analysis could be expanded to other types of GLVs to determine if the same changes in the phosphoproteome are seen and to further our understanding of how plant cells signal for damage responses in response to GLVs. In addition, plasma membrane enrichments should be performed to gain a more comprehensive profile of plasma membrane related proteins as these proteins are key for GLV perception.

### 3.4 MATERIALS AND METHODS

#### *3.4.1 Cell suspension cultures & cell lysis*

Tomato cells were provided by Dr. Johannes Stratmann. For GLV treatments, 15  $\mu$ L of *cis*-3-hexenol diluted in ethanol was added to cells for a final concentration of 28.5 mM and 15  $\mu$ l of ethanol was added to cells as a control. For systemin treatments, systemin was added to cells for a final concentration of 10 nM. For tomato cell lysis, cells were lysed with 1 mL of RIPA buffer (Thermo Scientific Pierce, Rockford, IL) containing 1x of Halt Protease and Phosphatase Inhibitor Cocktail (Thermo Scientific Pierce, Rockford, IL). Protein concentration was determined via Bicinchoninic acid assay. Proteins were reduced with addition of 10 mM TCEP at 56 °C for 30 minutes, then alkylated with addition of 25 mM iodoacetamide for 30 minutes in dark. Next, 1.5 mL of ice-cold acetone was added to the sample to precipitate the protein. The mixture was placed at -80

°C for 30 minutes followed by -20 °C for 30 minutes. The sample was then centrifuged for 5 minutes at 2000 × g and the supernatant was discarded. The protein pellet was washed using 1 mL of ice-cold acetone and centrifuged for 5 minutes at 2000 × g. The supernatant was then discarded. The resulting pellet was suspended in 50 mM ammonium bicarbonate to a final concentration of 5 mg/mL followed by overnight trypsin digestion (1:50 enzyme to protein by mass ratio) at 37 °C using MS Grade Trypsin Protease (Thermo Scientific Pierce, Rockford, IL). The sample was then acidified to 1% TFA and then centrifuged for 5 minutes at 5000 × g. The supernatant was then aliquoted and stored at -80 °C until required for phosphopeptide enrichment.

#### ***3.4.2 Phosphopeptide enrichment of tomato cell digests***

Cell digest was thawed on ice for 30 minutes. Aliquots of 20 µL of cell digest was then diluted with 1% TFA to final volume of 200 µL and transferred to a 1 mL 96-well plate. The peptide sample was then desalted using 5 mg RP 300 µL IMCStips (IMCS, Irmo, SC) and then enriched for phosphopeptides using 10 µL PolyTi 300 µL IMCStips (IMCS, Irmo, SC). The automated peptide desalting and phosphopeptide enrichment was performed on a Hamilton STAR system following the previously published workflow<sup>25</sup>. Once the sample was eluted, the eluted solutions were lyophilized. The samples were then stored at -80 °C until ready for analysis.

#### ***3.4.3 LTQ-Orbitrap Velos Pro analysis for global phosphopeptide analysis***

Dried phosphopeptide samples were dissolved in 4.8 µL of 0.25% formic acid with 3% (vol/vol) acetonitrile and 4 µL of each sample was injected into Easy-nLC 1000 (Thermo Fisher Scientific). Peptides were separated using a 105-mm column (75 µm ID) containing Reproosil-PUR C18 resin (New Objective). The mobile phase buffer consisted

of 0.1% formic acid in ultrapure water (buffer A) with an eluting buffer of 0.1% formic acid in 80% (vol/vol) acetonitrile (buffer B) run with a linear 60-min gradient of 6–30% buffer B at flow rate of 300 nL/min. Easy-nLC 1000 was coupled online with a hybrid high-resolution LTQ-Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific). The mass spectrometer was operated in the data-dependent mode, in which a full-scan MS (from m/z 300 to 1,500 with the resolution of 30,000 at m/z 400) was followed by MS/MS of the 10 most intense ions [normalized collision energy - 30%; automatic gain control (AGC) -  $3.0 \times 10^4$ ; maximum injection time (MS1) - 100 ms; 90 s exclusion; maximum injection time (MS2) – 150 ms].

#### ***3.4.4 Q Exactive HF-X analysis for global phosphopeptide analysis***

Dried phosphopeptide samples were dissolved in 4.8  $\mu$ L of 0.25% formic acid with 3% (vol/vol) acetonitrile. Liquid chromatography was performed on a Dionex Ultimate 3000 nano-flow liquid chromatograph. Four  $\mu$ L of sample was injected onto a cartridge trap column (Thermo Fisher Scientific) and then transferred to a capillary analytical column for separation (PepMap RSLC C18, EasySpray format, 15 cm by 75  $\mu$ m packed with 3  $\mu$ m 100Å particles). The separation solvents were 2% acetonitrile (solvent A) and 80% acetonitrile (solvent B) both with 0.1 % formic acid. The solvent program started at 5% solvent B for 3 min then ramped to 55% B over 40 minutes and 95% B over 3 minutes where it remained for 9 minutes. The EasySpray nano-spray source was interfaced to a Thermo Q-Exactive HF-X quadrupole-orbitrap mass spectrometer which was operated using a standard Top-6 data dependent acquisition (DDA) protocol. MS1 spectra (scanned from 400 to 1500 m/z at 60,000 resolution) were followed by MS2 spectra of the 6 most intense multiply charged parent ions (normalized collision energy 28 ev,

resolution 30,000). Dynamic exclusion was used to prevent reanalyzing the same parent ion over the next 5 seconds.

#### ***3.4.5 Orbitrap Fusion analysis for global phosphopeptide analysis***

Thermo Scientific Orbitrap Fusion run under conditions established by the staff at the Nevada Proteomics Center was used for sample analysis. Liquid chromatography mass spectrometry (LC-MS) was performed on an UltiMate 3000 RSLCnano system (Thermo Scientific, San Jose, CA) on a self-packed UChrom C18 column (100  $\mu\text{m}$  x 35 cm). The gradient used consisted of Solvent B from 2-90% (Solvent A: 0.1% formic Acid, Solvent B: acetonitrile, 0.1% formic Acid) over 175 mins at 50 °C using a digital Pico View nanospray source (New Objectives, Woburn, MA) that was modified with a custom-built column heater and an ABIRD background suppressor (ESI Source Solutions, Woburn, MA). The self-packed column tapered tip was pulled with a laser micropipette puller P-2000 (Sutter Instrument Co, Novato, CA) to an approximate id of 10  $\mu\text{m}$ . The column was then packed with 1-2 cm of 5  $\mu\text{m}$  Sepax GP-C18 (120 Å) (Sepax Technologies, Newark, DE) followed by 40 cm of 1.8  $\mu\text{m}$  Sepax GP-C18 (120Å) at 9000 psi using a nano LC column packing kit (nanoLCMS, Gold River, CA). The MS1 precursor selection range is from 400-1500 m/z at a resolution of 120K and an automatic gain control (AGC) target of  $2.0 \times 10^5$  with a maximum injection time of 100 ms. Quadrupole isolation window at 0.7 Th for MS2 analysis using CID fragmentation in the linear ion trap with a collision energy of 35%. The AGC was set to  $4.0 \times 10^3$  with a maximum injection time of 150 ms. The instrument was operated in a top speed data-dependent mode with a most intense precursor priority with dynamic exclusion set to an exclusion duration of 60 s with a 10 ppm tolerance.

### ***3.4.6 Database search and analysis***

The raw data files obtained from the Orbitrap Fusion, Orbitrap Velos Pro, and Q Exactive HFX were searched against a *Solanum peruvianum* protein database (ITAG 2.3) using SEQUEST on Proteome Discoverer (Version 2.2, Thermo Fisher). The peptide precursor mass tolerance was set to 10 ppm, and the fragment mass tolerance was set to 0.6 Da. Search criteria included a static modification of cysteine residues of +57.0214 Da and a variable modification of +15.995 Da to include potential oxidation of methionine, a modification of +79.966 Da on serine, threonine, or tyrosine for the identification of phosphorylation (three modifications allowed per peptide). To detect heavy labeled peptides, variable modifications of +6.020 Da on lysine and +10.008 Da on arginine. Searches were performed with full tryptic digestion and allowed a maximum of two missed cleavages on the peptides analyzed from the sequence database. False discovery rates (FDR) were set to less than 1% for each analysis at the PSM level. Phosphorylation site localization from CID spectra was determined by PhosphoRS on Proteome Discoverer 2.2. Confidence view was assigned a score of 0.4, indicating medium confidence.

## CHAPTER 4

### GENERATING AN AUTOMATED EXTRACTION OF EXTRACELLULAR VESICLES FROM URINE SAMPLES.

#### 4.1 INTRODUCTION

##### *4.1.1 Overview of extracellular vesicles*

EVs are small membrane-enclosed vesicles that are released by many different cell types.<sup>108</sup> There are numerous subtypes of EVs that have been reported in the literature including exosomes, shed microvesicles (MVs), migrasomes, oncosomes, and apoptotic bodies (Figure 4.1).<sup>109</sup> The EV subtypes, including exosomes and MVs, are generally distinguished based on size, buoyant density, and protein content (Table 4.1).<sup>109</sup> Many reports have focused on exosomes and MVs, but here the term EVs will be used to refer to all secreted vesicles.

EVs are known to carry a vast array of cargo including DNA, mRNA, and proteins.<sup>108-110</sup> Over the past decade, EVs have attracted attention as it has been reported that they can be used as an intercellular communication tool, allowing both distant and local cell communication.<sup>108, 110</sup> While EVs are important for normal biological functions, they also play a role in the development of many diseases, including cancer.<sup>108-110</sup> In particular, tumor-derived EVs have been shown to carry biological material that can aid in the formation of the premetastatic niche and increase angiogenesis to facilitate cancer growth.<sup>108, 110</sup> EVs are of interest in the field of precision medicine because they are known to be secreted in many different biological fluids including serum and urine.

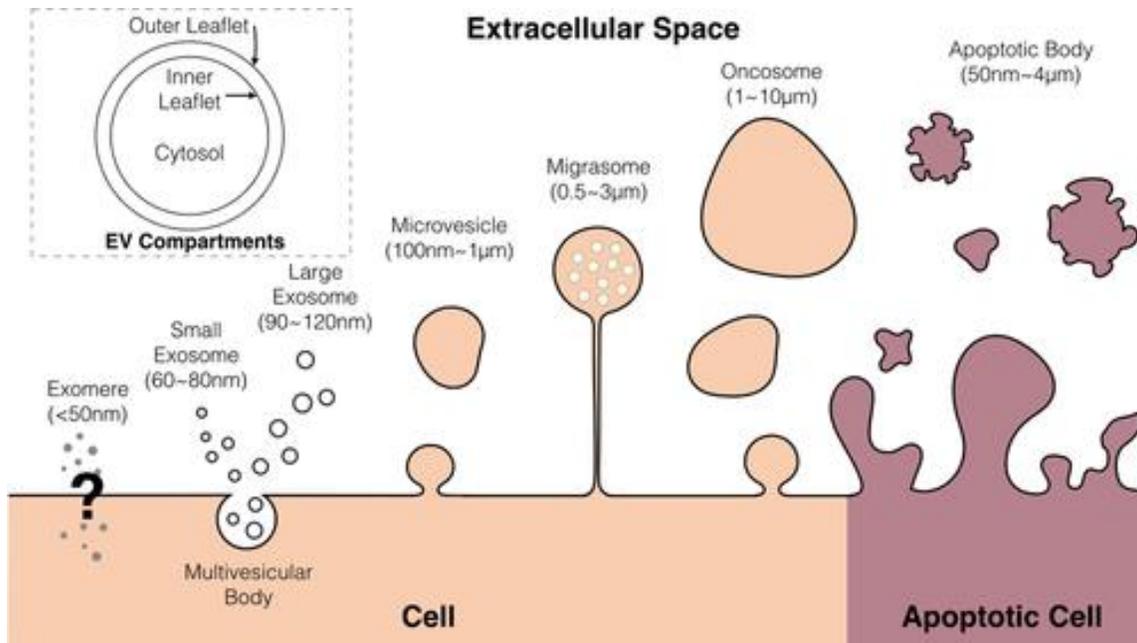


Figure 4.1. Diagram of EV subpopulations. Inlet shows three different compartments associated with EVs. Image is from reference 109.<sup>109</sup>

Table 4.1. Characteristics of exosomes and MVs that can be used to differentiate between these EV subtypes.<sup>109</sup>

	<b>Exosome</b>	<b>MVs</b>
<b>Size (diameter)</b>	40–120 nm	100–1000 nm
<b>Buoyant Density</b>	1.08–1.20 g/mL	1.13–1.19 g/mL
<b>Marker</b>	Alix, CD9, CD63, CD81, FLOT1, FLOT2, TSG101, CM	CD9, CD63, actinin-4, mitofillin
<b>References</b>	111-113	111, 112, 114, 115

This allows researchers to isolate EVs from these fluids to monitor biomarkers that can detect the onset of many different diseases, most notably, cancer.<sup>108-110</sup> It has been reported that tumor-derived EVs can be detected in biological fluids before physical detection of a tumor is possible making EVs an extremely promising biological resource for early detection of cancer.<sup>112, 116, 117</sup>

#### ***4.1.2 EVs for clinical diagnostics of colorectal cancer***

Colorectal cancer (CRC) is the third most diagnosed cancer in the United States and is expected to cause over 50,000 deaths in 2019 alone.<sup>11</sup> While better treatment of CRC is needed, early detection is key for better treatment of the disease. Currently, carcinoembryonic antigen (CEA) is a biomarker found in the blood and has been used to detect CRC. However, CEA is controversial in the oncology world mostly because of its low detection accuracy. While it has been shown to be useful under certain conditions and with the proper constraints, it has been shown ineffective for detecting early-stage CRC and only effective for 30% of stage two patients.<sup>118, 119</sup> Researchers are working to discover better biomarkers, including within EVs from patient samples, which could eventually be used to develop clinical assays to diagnose and monitor diseases. Mass-spectrometry based proteomics is one of the tools that is currently used to realize these types of clinical assays. For example, in 2017 Tomonaga *et al.* identified 725 potential candidate biomarkers for CRC based on their own generated data and a global search of the published literature.<sup>120</sup> Of these candidate biomarkers, they were able to identify 356 candidate proteins in serum or CRC cell culture EV samples and quantitate putative colorectal cancer biomarker candidates in serum extracellular vesicles. In this study, they

identified 22 proteins that showed significant differences between healthy and CRC patient samples.<sup>120</sup>

In a recent report, S. Wang *et al.* used urinary EVs to show biomarkers for neurologic disease.<sup>121</sup> In this work, they showed that, as expected, urine derived EVs were enriched in proteins associated with vesicle biogenesis and transport, however they were not significantly enriched in kidney derived proteins (Figure 4.2A). In fact, using quantitative MS-based proteomics, network analysis of their protein list showed no over-representation of proteins of any particular tissue type or biofluid (Figure 4.2B).<sup>121</sup>

In this work they identified two different proteins that could be used in as a diagnostic immunoblot assay for Parkinson's disease with better than 70% accuracy in a replication cohort. This was the first report of using urine derived EVs to be applied as a diagnostic tool for a disease unrelated to kidney function. This is significant because urine is a highly desirable diagnostic biofluid as it is noninvasive, readily available, and, in comparison to serum samples, has a relatively low protein concentration which simplifies the EV purification process. However, there are many issues that plague EV research inhibiting its translation into the clinical setting.

#### ***4.1.3 Isolation techniques for EVs purification***

One major issue hindering EV research is that current protocols for isolating and purifying EVs are time consuming, suffer from low recovery, require specialized equipment, and are low-throughput.<sup>122</sup> The most common method of isolation is differential ultracentrifugation (DUC) where samples are subjected to various centrifugation speeds to separate EVs from intact cells and cell debris. This method is time consuming (>3 hours) and is subject to inter-lab irreproducibility. Other popular

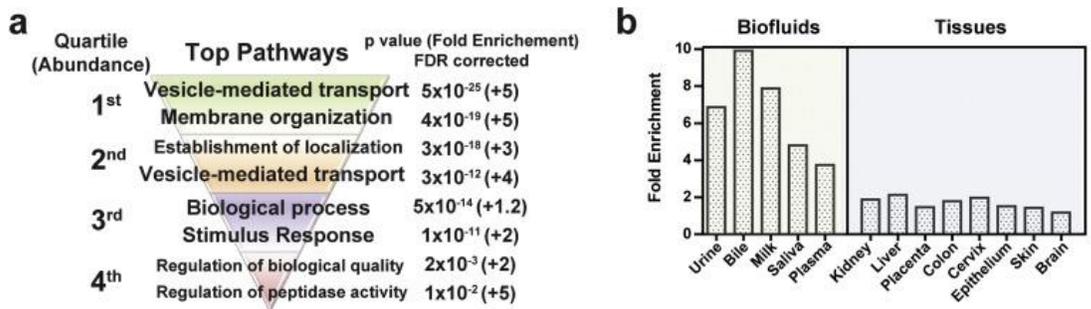


Figure 4.2. Network analysis of proteomic data generated by S. Wang et al.<sup>121</sup> The original figure is from reference 121 and the legend has been modified from the original. A) Pathway analysis of proteins generated from urinary EV samples. B) Tissue enrichment analysis of proteins generated from urinary EV samples.

methods of purifying EVs are size exclusion-based, precipitation-based, or affinity-based. Over the past 5-years significant work has been done to improve upon these isolation techniques several of which outperform the conventional DUC method.<sup>122-124</sup> In 2018, Tao *et al.* reported the use an affinity-based method called EVTRAP for efficient purification of EVs from urine samples.<sup>122</sup> In this study they showed that EVTRAP outperformed the conventional ultracentrifugation method in terms of overall recovery of both exosomes and MVs as well as overall purity of the final sample. Also, in 2018, Xing *et al.* presented an EV isolation method using anion exchange (AE) magnetic beads.<sup>123</sup> EVs have a net negative surface charge and therefore can be bound to a positively charged AE magnetic beads and subsequently eluted with a high ionic strength elution buffer. This method showed improved recovery of EVs in comparison to DUC from both cell culture supernatant and plasma samples. These two different methods are promising isolation methods, and both could be used in place of ultracentrifugation in a clinical setting.

#### ***4.1.4 EVs purification using IMCStips***

Due to the promise of these two resins, we aimed to establish a clinically relevant extraction method for EV purification using these resins in combination with the IMCStips as it has been previously established that the IMCStips allow for convenient automation of the extraction process. Toward this end, we tested a silica-based EVTRAP resin and a commercially available AE resin in the IMCStip for efficient purification of EVs from urine samples. To confirm that EVs were able to be isolated from biological samples several different techniques have been used. The most widely used method for EV detection is western blot for proteins known to be abundant in EVs. For example,

proteins in the tetraspanin family, most commonly CD9 and CD81, are known to be found on the membrane of most EVs and are therefore used for detection of EVs. Electron microscopy (EM) is also used for confirming the presence of EVs. Typically, EVs are 40-1000 nm in size and therefore EM is able to image EVs within biological samples. In the last few years, mass spectrometry has also become a useful tool in confirming the presence of EVs. This is done by performing a trypsin digestion on the purified EV sample and then injecting the digested protein sample onto an LC-MS system for detection of peptides. Then using proteomics software, the peptides are linked back to their associated proteins and the presence of EV-related proteins can be detected. Each of these methods have their drawbacks when being used for confirming the presence of EVs. For example, using western blot to confirm the presence of EVs using primary antibodies for CD9 and other EV-related proteins is highly dependent on the efficiency of the primary antibody's binding to the target protein. CD9 has primarily been used for confirming the presence of EVs because the primary antibody for this protein has been well developed but can lead to bias toward EV samples with high amounts of CD9 in the membrane and thus underreport the presence of EVs when CD9 is not as abundant. For imaging EVs using EM, the EV sample needs to be extremely pure as the presence of protein within the sample can hinder the light scattering for proper imaging of the EVs. Additionally, prior to imaging the sample is first fixed and dehydrated onto a grid. This process can deform the EVs, giving them a "cupped" shape and their native state is not able to be imaged thus, leading to error in determining the size dispersity of the EVs. For detecting EVs using MS, similar to EM, it is important that contaminating protein is removed from the EV sample prior to analyses as the contaminating protein

will be detected by the MS analysis and can swamp out the EV-related peptides thus limiting the number of EV-related proteins detected by the analysis. With these challenges in mind, it is important to confirm the presence of EVs using each of the three techniques to ensure the purity of the EV sample resulting for the extraction.

## 4.2 RESULTS & DISCUSSION

### ***4.2.1 Resin selection and initial IMCStip functionality testing***

To begin establishing an EV purification process with the IMCStips, resins first needed to be tested for basic functionality in the IMCStip. For the EVTRAP resin to be applied in the IMCStips, a silica-based EVTRAP resin was needed as previous EVTRAP extractions were done using magnetic beads. The IMCStip has a bottom frit that can retain resin with a particle size of 50 microns or higher. For initial testing, a silica resin from Sorbtech was used with a particle size of around 70 microns. After basic functionality testing to ensure there was proper dispersion of the resin inside of the IMCStip and no significant clogging issues, an extraction method was established for each of the different resin types based on previous literature and resin chemistry. For the EVTRAP resin, based on previous work by Iliuk *et al.* the EVTRAP resin needs to be first conditioned with PBS and then can incubate directly with the sample for EV extraction. The EVTRAP resin then needs to be washed with PBS twice and then the EVs are eluted using 100 mM triethylamine (TEA). The initial extraction method that was used is showed in Table 4.2.<sup>122</sup> A similar workflow was established for a weak anion exchange resin, Macro-Prep DEAE, from Biorad based on the previous work by Heath *et al* (Table 4.3).<sup>123</sup>

Table 4.2. Extraction workflow for EVTRAP resin in the 1 mL Integra IMCStip.

<b>Process</b>	<b>Solvent</b>	<b>Aspirate, <math>\mu\text{L}</math></b>	<b>Cycle Number</b>	<b>Sample, <math>\mu\text{L}</math></b>	<b>Time, mins</b>
Equilibration	1xPBS	750	3	800	3
Sample	1xPBS	1000	30	800	30
Wash 1	1xPBS	750	5	800	5
Wash 2	1xPBS	750	5	800	5
Elution	100 mM TEA	200	5	200	5

Table 4.3. Extraction workflow for DEAE resin in the 1 mL Integra IMCStip.

<b>Process</b>	<b>Solvent</b>	<b>Aspirate, μL</b>	<b>Cycle Number</b>	<b>Sample, μL</b>	<b>Time, mins</b>
Equilibratio n	25 mM NaCl, 50 mM HEPES, pH 8.2	750	3	800	3
Sample	25 mM NaCl, 50 mM HEPES	1000	30	800	30
Wash 1	25 mM NaCl, 50 mM HEPES, pH 8.2	750	5	800	5
Wash 2	25 mM NaCl, 50 mM HEPES, pH 8.2	750	5	800	5
Elution	500 M NaCl, 50 mM HEPES, pH 8.2	200	5	200	5

This IMCStip-based workflow was programmed on the Integra Assist system. For initial testing, a 1 ml urine sample was used in combination with 25 mg of EVTRAP resin or 25 mg of DEAE resin (n = 3 per resin type). Once the samples were processed using either the EVTRAP resin or the DEAE resin the sample were immediately neutralized with 1 M acetic acid to pH 7. The samples were then ran on an SDS page gel and transferred to a PVDF western blot membrane. The blot was incubated with a 1:1000 diluted CD9 primary antibody and subsequently incubated with the proper secondary antibody and ECL western blotting substrate. The resulting blot was imaged using chemiluminescence to detect the presence of CD9 in the samples (Figure 4.3).

The blot in Figure 4.3, showed that were able to detect a visible band on the blot corresponding to CD9. Additionally, the EVTRAP purified samples, based on band intensity, contained more CD9 protein than the DEAE purified samples. Based on these initial results, only EVTRAP resin was used for the remainder for the study. However, based on the band intensity, there was significant variation among the EVTRAP purified samples and further follow up experiments were needed to confirm reproducibility and to confirm the presence of EVs using different primary antibodies.

#### ***4.2.2 Troubleshooting and development of the EVTRAP IMCStip extraction workflow for EV purification from urine***

A follow up study was ran using the same parameters with the EVTRAP resin in the IMCStip. For this experiment, the flow through urine sample was collected to determine if CD9 could be detected after the EVTRAP resin binding step. For this experiment, a purified EV sample from cell culture from Cell Guidance Systems (CGS) was included in

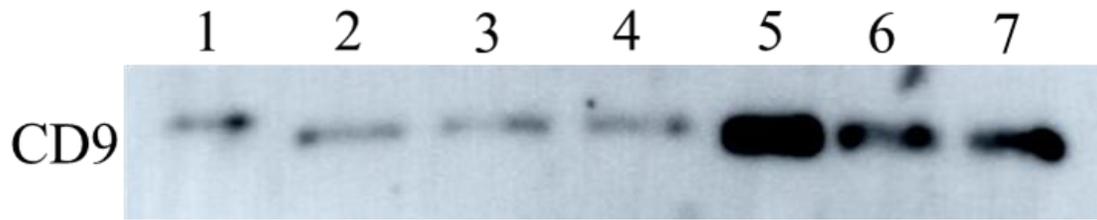


Figure 4.3. Western blot of EV sample from 1 mL of urine. Lane 1 is the undiluted urine sample. Lanes 2-4 are the DEAE purified EV samples which have a final volume of 200  $\mu$ L. Lanes 5-7 are the EVTRAP purified EV samples which have a final volume of 200  $\mu$ L.

the western blot as a control. For the western blot, a 1:500 diluted CD9 primary antibody was used as well as a 1:500 diluted CD81 primary antibody (Figure 4.4).

Based on the blots in Figure 4.4, the EVTRAP purified samples gave a clear band corresponding to the CD9 protein however using the CD81 antibody, the bands were inconsistent and needed a significantly longer exposure time compared to CD9. This could be due to several different issues including the absence of the CD81 protein in the EV sample or inefficient binding of the primary antibody to CD81 in the EV sample. Additionally, Lane 2 of Figure 4.4B shows a clear band showing that a significant amount of the EV sample is still in the flow through of the sample. During the EVTRAP extraction, significant clogging was seen leading to different amounts of sample being aspirated during the sample binding portion of the extraction. This was unexpected as this type of clogging was not seen in previous experiments. This clogging could lead to inefficient binding of the EVs to the resin as the clogging limits the exposure of the resin to the sample.

To determine the presence of EVs in the EVTRAP purified sample, transmission electron microscopy was used to image EVs in the sample. As a control, the purified EV sample from CGS was used for TEM imaging. For the TEM imaging, EVTRAP purified samples were pooled and dried. The dried sample was resuspended in 100  $\mu$ L of ultrapure water. The samples were then applied TEM grids and affixed for imaging. The resulting images are shown in Figure 4.5. Based on the TEM images in figure 4.5, the purified CGS EV sample contained a significant amount of contaminants including salts and proteins negatively affecting the TEM images. Additionally, there seems to be a high abundance of EVs within the sample although this was unable to be confirmed because

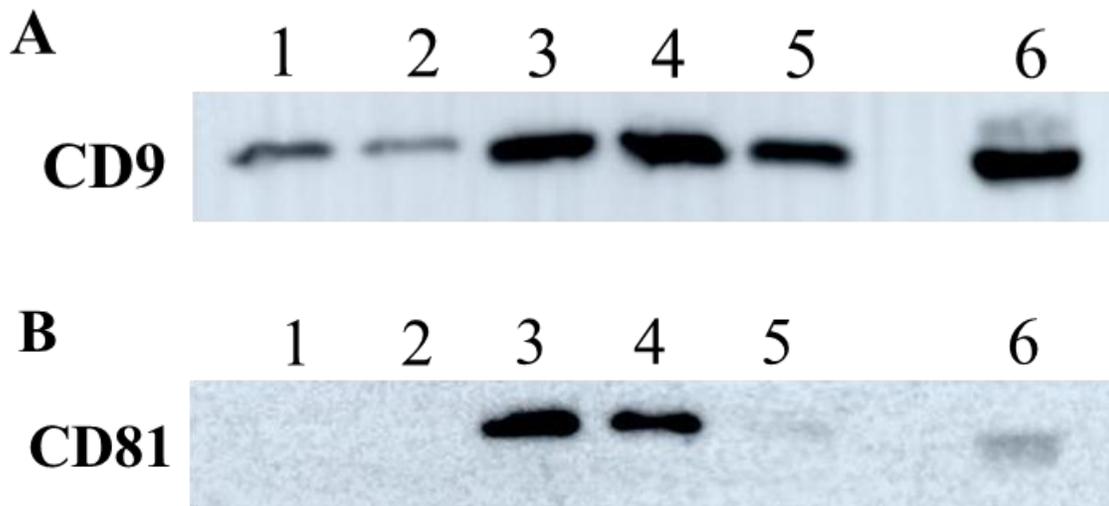


Figure 4.4. Western blot of EVTRAP purified samples from 1 mL of urine. Lane 1 is the undiluted urine sample. Lane 2 is the undiluted flowthrough urine sample. Lanes 3-5 are the EVTRAP purified EV samples which have a final volume of 200  $\mu$ L. Lane 6 is the purified EV sample from CGS. A) Western blot of EV sample from 1 mL of urine using 1:500 diluted CD9 primary antibody (exposed for 1 min). B) Western blot of EV sample from 1 mL of urine using 1:500 diluted CD81 primary antibody (exposed for 20 mins).

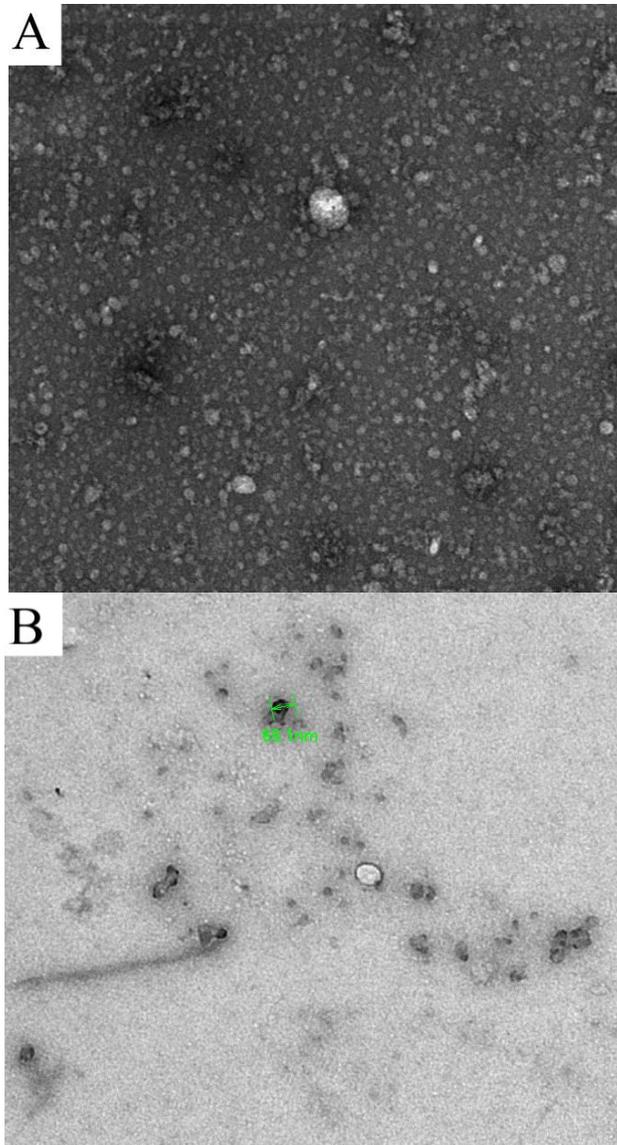


Figure 4.5. TEM images of purified EV samples. TEM images of representative of at least 3 TEM images for the denoted sample. A) CGS EV sample purified from cell culture supernatant. B) EVTRAP purified EV samples from 1 mL urine using IMCStips.

the sample was too concentrated and contained contaminants (Figure 4.5A). The EVTRAP purified EV sample contained significantly less salt and protein contamination, however the sample contain few particles that were within the 40-1000 nm range (Figure 4.5B).

To further understand the EVTRAP extraction, we collaborated with Dr. Anton Iliuk at Tymora Analytical to do further experiments with the aim of addressing the low binding efficiency of the EVTRAP and to ensure proper handling of EV samples during downstream analysis and characterization. For these experiments, a new batch of EVTRAP resin was generated to ensure that deterioration of the EVTRAP beads was not the cause of the low binding efficiency of the resin. Additionally, an update to the extraction was made to include a 10x loading buffer addition to the urine sample prior to the EVTRAP extraction. The loading buffer is a proprietary solution from Tymora Analytical that is a combination of detergent to facilitate more efficient EV binding. This loading buffer was included in the sample for all following experiments. After EVTRAP elution, the eluted sample was completely dried and resuspended in 10  $\mu$ L of 1x SDS loading buffer. For western blot analysis, the previously published western blot protocol used by Iliuk *et al.* was used which uses an Odyssey near-infrared scanner from Licor for detection of an Alexa Fluor 680 nm secondary antibody that is bound to the primary antibody. This detection protocol allows for better sensitivity and better quantification of resulting bands. In first experiment with Dr. Iliuk, two different EVTRAP resin amounts were used, 10 mg and 25 mg, to purify EVs from 500  $\mu$ L of urine. For comparison, 10 mg of EVTRAP resin was used in a manual spin tube format (Figure 4.6). Based on the blot from Figure 4.6, the urine sample purified using 10 mg of EVTRAP resin in the spin

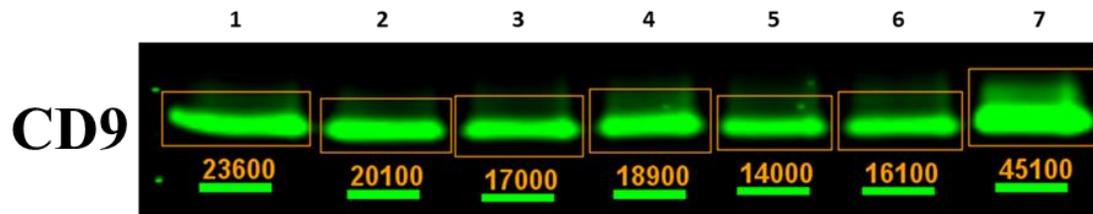


Figure 4.6. Western blot of EVTRAP purified samples from urine. The primary antibody was 1:3000 diluted CD9. Lanes 1-3 are samples purified using 25 mg of EVTRAP resin in 1 mL IMCStips. Lanes 4-6 are samples purified using 10 mg of EVTRAP resin in 1 mL IMCStips. Lane 7 is the EV sample purified using 10 mg of EVTRAP resin in a manual spin tube format.

tube format yielded a more intense CD9 band in comparison to either the 25 mg or 10 mg of EVTRAP resin in the IMCStip where the band intensity is 56% and 64% lower than the manual format band intensity, respectively. It was observed that for the IMCStip extraction the clogging during the sample bind was reduced in comparison to previous experiments however there is still inefficient binding of EVs in comparison to the manual format. For the IMCStip-based format, based on the band intensities, samples purified with 25 mg of EVTRAP resin yielded higher band intensity, an average of  $20,233 \pm 2,696$  RFU, than samples purified with 10 mg of EVTRAP resin with an average band intensity of  $16,333 \pm 2,007$  RFU.

To determine if the inefficient binding could be remedied by increasing the sample binding time, a follow up experiment was performed for which the sample binding time was set to 60 minutes rather than 30 minutes. All other aspects of the extraction were kept the same. For the experiment, 25 mg of EVTRAP resin and 10 mg of EVTRAP resin in the IMCStip were compared ( $n = 4$  per resin amount) as well as 25 mg of EVTRAP resin and 10 mg of EVTRAP resin in a manual spin tube format were compared to the IMCStip-based extraction (Figure 4.7). From the blot in Figure 4.7, the manual format yielded higher band intensity for both resin amounts, 24,700 RFU for the 25 mg EVTRAP sample and 27,600 RFU for the 10 mg EVTRAP sample, than the IMCStip-based format with an average band intensity of  $20,450 \pm 861$  RFU for IMCStips containing 25 mg of EVTRAP resin and  $12,675 \pm 3,100$  RFU for IMCStips containing 10 mg of EVTRAP resin. While the manual format still outperforms the IMCStip format, the average band intensity for the 25 mg EVTRAP resin in the IMCStip is only 26% lower

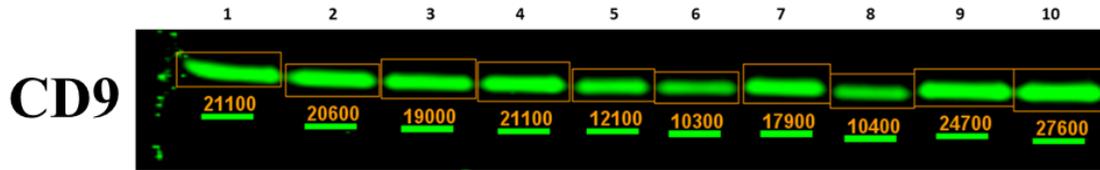


Figure 4.7. Western blot of EVTRAP purified samples from urine. The primary antibody was 1:3000 diluted CD9. For this experiment, all sample binding times were set to 60 minutes. Lanes 1-4 are samples purified using 25 mg of EVTRAP resin in the IMCStip. Lanes 5-8 are samples purified using 10 mg of EVTRAP resin in the IMCStip. Lane 9 is a sample purified using 25 mg of EVTRAP resin in a manual spin tube format. Lane 10 is a sample purified using 10 mg of EVTRAP resin in a manual spin tube format.

than the 10 mg EVTRAP manual sample compared to 56% lower in the previous experiment. In all following experiments, the sixty-minute sample binding time was used.

To further optimize the IMCStip-based extraction, the elution time was increased from 5 minutes to 10 minutes and the sample binding time was kept at 60 minutes.

Additionally, to determine the reproducibility of IMCStip-based extraction the sample number was increased to 8. For comparison, the same extraction was performed using 25 mg of EVTRAP resin in a manual spin tube format (Figure 4.8). The blot in Figure 4.8 shows significant variation in the band intensity of the 8 IMCStip-based extractions with an average band intensity of  $21,312 \pm 4857$  and is 21% lower in comparison to the manual format. This is a slight improvement from the previous experiment showing that the increased elution time marginally improved the extraction efficiency of the IMCStip-based workflow. In all following experiments, the ten-minute elution time was used.

Based on these experiments, the clogging observed during the sample binding step with the IMCStip seemed to significantly affect the efficiency of the EV binding. This is especially highlighted by the manual process outperforming the IMCStip-based extraction in terms CD9 band intensity from the western blot. Based on conversations with IMCS, the clogging of the IMCStip could be due to the resin used and suggested that a higher particle size could prevent the clogging. To this end, a silica resin with a particle size of approximately 100 microns was used for the EVTRAP conjugation. Once the new EVTRAP resin was received, a basic functionally test was performed to ensure there was no clogging using 25 mg of the resin in the 1 mL IMCStip. Once this initial testing was performed, 25 mg of the EVTRAP resin in the 1 mL IMCStip was used to extract EVs from 0.5 mL of urine ( $n = 3$ ). Following the updated EV processing method,



Figure 4.8. Western blot of EVTRAP purified samples from urine. The primary antibody was 1:3000 diluted CD9. For this experiment, all sample binding times were set to 60 minutes and the elution times were set to 10 minutes. Lanes 1-8 are samples purified using 25 mg of EVTRAP resin in the IMCStip. Lane 9 is a sample purified using 25 mg of EVTRAP resin in a manual spin tube format.

the samples were completely dried and resuspended in SDS loading buffer. For comparison, the starting sample and the flow through samples were concentrated using 30 kDa filters, dried completely, and resuspended in SDS loading buffer. The samples were then transferred to PVDF, incubated with the 1:1000 diluted CD9 primary antibody, and then incubated with the proper secondary antibody and ECL solution. The blot was then imaged under chemiluminescence using a GE imager (Figure 4.9). The blot in Figure 4.9 shows that bands in lanes 2-4, corresponding to the eluted EVTRAP samples, are wavy rather than one clear straight band in conflict with all previous blots using the EVTRAP resin. Based on conversations with Dr. Iliuk, it was suggested that the new resin had a higher binding capacity than the previous resin and therefore less resin should be used. The “wavy” bands from the western blot could stem from excessive resin binding to lipids in the urine sample, leading to improper migration of the sample during the gel electrophoresis portion of the western blot protocol. To test this hypothesis, a range of EVTRAP resin amounts were tested, 5 mg, 10 mg, and 15 mg for extracting EVs from 0.25 ml of urine. The updated western blot protocol was followed for these experiments (Figure 4.10). Based on the blot from Figure 4.10, the 15 mg EVTRAP eluted samples yield “wavy” bands, while samples using the 10 mg EVTRAP IMCStip and the 5 mg EVTRAP IMCStip yield normal bands on the blot. Additionally, the flowthrough samples show no bands corresponding to CD9 showing that all three resin amounts adequately bind all EVs from the sample, although based on the starting material band, some of the EVs may be lost during the size-based filtration process.

In a follow up experiment, we examined if the amount of urine affects the appearance of the “wavy” bands on the blot. For this experiment 0.5 mL of urine was

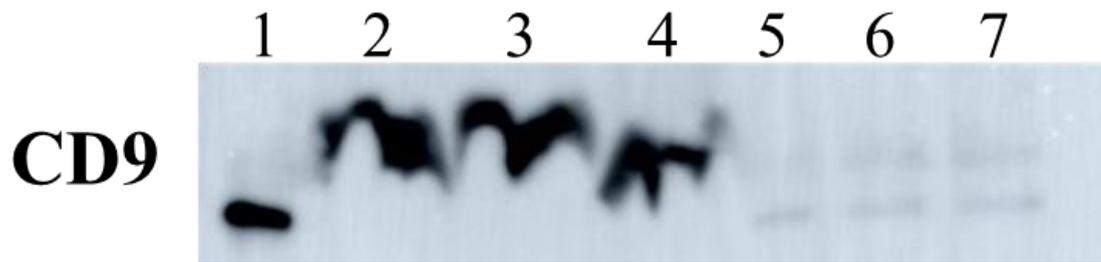


Figure 4.9. Western blot of EVTRAP purified samples from 0.5 ml of urine. The primary antibody was 1:1000 diluted CD9. Lane 1 is the starting urine sample filtered using a 30 kDa amicon filter. Lanes 2-4 are EVTRAP (~ 200-micron silica) purified samples. Lanes 5-7 are the flowthrough samples from the EVTRAP extraction process after sample binding.

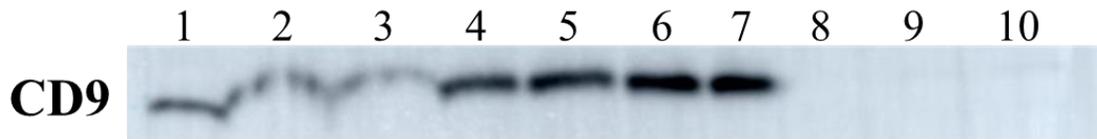


Figure 4.10. Western blot of EVTRAP purified samples from 0.25 mL of urine. Lane 1 is the starting urine sample filtered using a 30 kDa amicon filter. Lanes 2-3 are samples purified using 15 mg of EVTRAP resin in the IMCStip. Lanes 4-5 are samples purified using 10 mg of EVTRAP resin in the IMCStip. Lanes 6-7 are samples purified using 5 mg of EVTRAP resin in the IMCStip. Lanes 8-10 are the flowthrough samples from the 5 mg, 10 mg, and 15 mg EVTRAP IMCStip extraction process after sample binding.

used as the starting material and 5 mg and 10 mg of EVTRAP resin in the IMCStip was compared. The samples were processed and probed for the presence of CD9 via western blot. Based on the blot in Figure 4.11, the 10 mg EVTRAP IMCStip samples led to “wavy” bands whereas the 5 mg EVTRAP IMCStip samples had normal bands. Additionally, there was a prominent band show good EV recovery from the extraction process. Although the bands for the 5 mg EVTRAP IMCStip sample showed good EV recovery, contaminating species could be coeluting, but not enough to affect the protein migration during the gel electrophoresis process.

One of the main factors that needs to be established for the extraction is the reproducibility of the EVTRAP extraction. To test the reproducibility of the extraction, the extraction process was repeated for 8 0.5 mL urine samples using 5 mg EVTRAP IMCStips. The samples were processed and then probed for the presence of CD9 via western blot. Based on the band intensities from the blot in Figure 4.12, detailed in Table 4.4, substantial irreproducibility remains, with a relative standard deviation of 28%. This irreproducibility is unexpected and is not in line with previous experiments performed by Iliuk *et al.* or other extraction types performed with the IMCStips. Generally, due to the ability to automate the extraction process using the IMCStips, an overall increase in sample-to-sample reproducibility is seen, however there are several factors that could lead to irreproducibility including resin clumping, inhibiting the dispersion of the resin throughout the sample or frit clogging due to either the resin or particles from the sample lodging inside of the frit inhibiting proper flow of the sample through the frit. Neither of these issues were visually seen using the new EVTRAP resin with the larger particle size, however these issues could still be occurring leading to poor reproducibility.

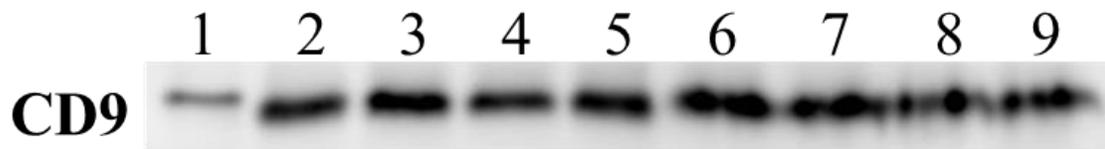


Figure 4.11. Western blot of EVTRAP purified samples from 0.5 mL of urine. Lane 1 is the starting urine sample filtered using a 30 kDa amicon filter. Lanes 2-5 are samples purified using 5 mg of EVTRAP resin in the IMCStip. Lanes 6-9 are samples purified using 10 mg of EVTRAP resin in the IMCStip.

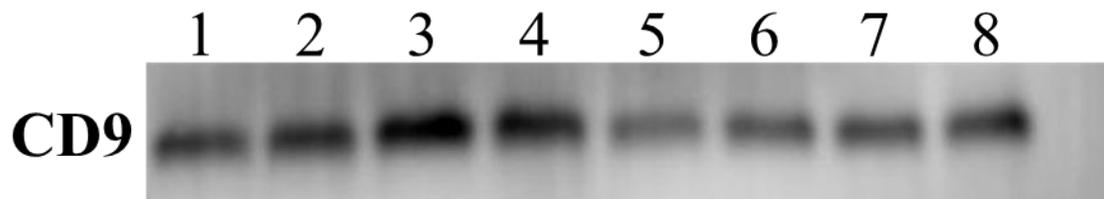


Figure 4.12. Western blot of EVTRAP purified samples from 0.5 mL of urine. Lanes 1-8 are samples purified using 5 mg of EVTRAP resin in the IMCStip.

Table 4.4. Band intensities from samples from experiments in Figure 4.12.

<b>Sample</b>	<b>Band Intensity</b>
1	10422916.04
2	11652908.37
3	14046686.64
4	10587487.05
5	5506766.83
6	6912797.15
7	7436011.28
8	8963066.26
<b>Avg.</b>	<b>9441079.95</b>
<b>RSD, %</b>	<b>28%</b>

### ***4.2.3 Mass spectrometry analysis of EV peptides samples purified using EVTRAP IMCStip extraction***

An aim of this project was to determine if EV-associated proteins could be identified through MS analysis of the purified samples. To this end, the EVTRAP extraction was used to purify EVs from 1 mL of 10x concentrated urine, concentrated using a 30 kDa filter. The purified samples were lysed and trypsin digested using our previously described protocol as well as a protocol detailed by Iliuk et. al (n = 2 per digestion protocol).<sup>122</sup> Once the samples were digested, the resulting peptides were desalted and dried completely. The samples were then resuspended and half of the sample was injected on the LC-MS/MS and the raw data was processed using Proteome Discoverer to determine proteins detected by the LC-MS/MS analysis. Based on the total ion chromatograms from the LC-MS/MS, a clear difference in the amount of protein digested can be seen. In Figures 4.13A-B, there are clear peaks for which MS2 scans were performed whereas in Figures 4.13C-D the material eluted after 57 minutes, characteristic of undigested protein, dwarfs the peaks detected between 20-57 minutes of the analysis. These differences are highlighted in the number of PSMs recorded during the LC-MS/MS analysis and the number of peptides detected within the sample (Table 4.5).

Using our current digestion protocol, we can achieve higher peptide and protein ids than using the protocol described by Iliuk. These differences can be directly attributed to the lack of digestion of protein material shown as eluting after 57 minutes in the LC-MS/MS analysis. However, the number of protein identifications we achieved is far less than previous reports by Iliuk et. al from 1 mL of 10x concentrated urine using the EVTRAP resin.<sup>122</sup> Specifically, Iliuk *et al.* achieved over 2000 protein ids and 16,000

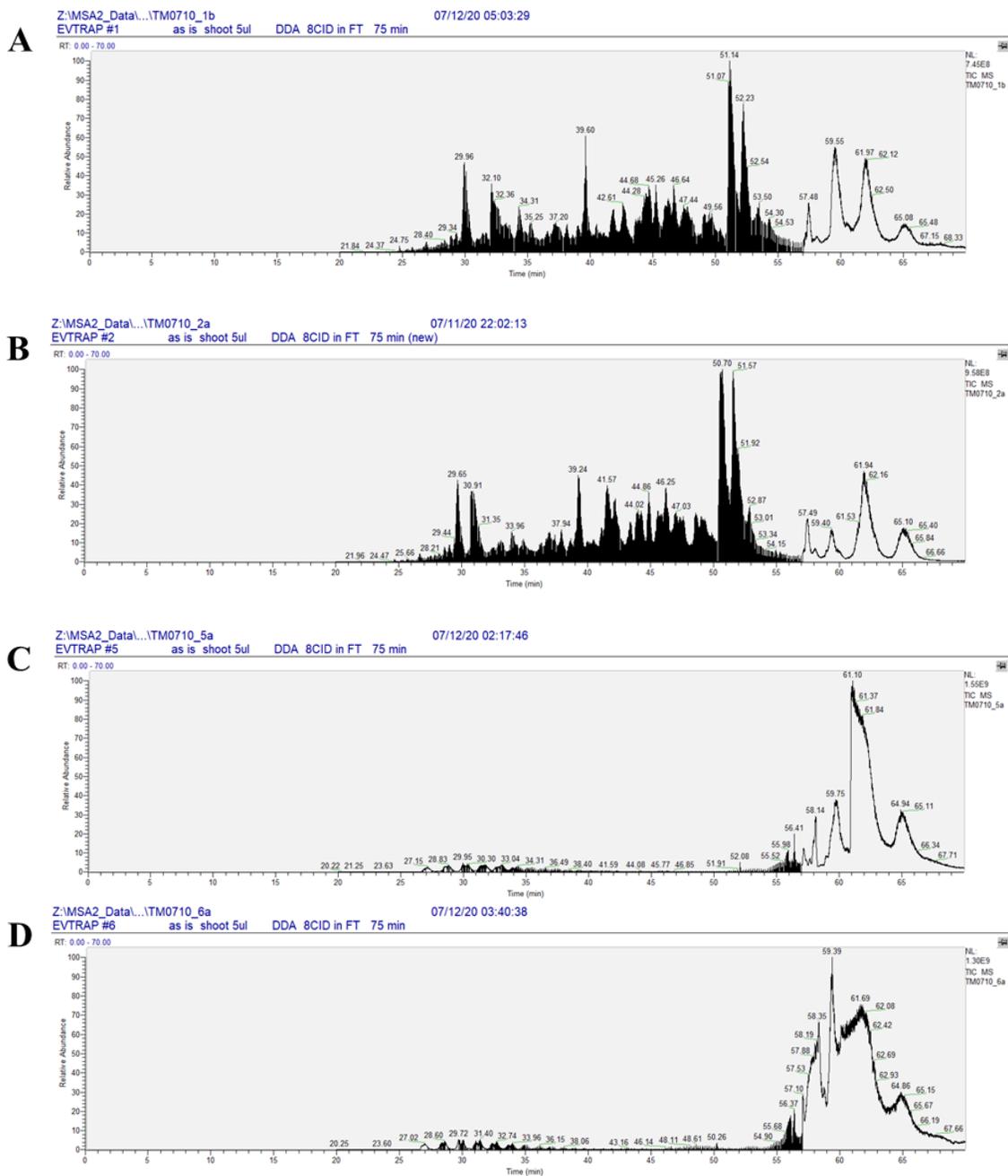


Figure 4.13. Total ion chromatograms (TICs) from LC-MS/MS analysis of EVTRAP purified peptide samples. A) TIC from EVTRAP purified sample digested using our in-house digestion protocol (sample 1). B) TIC from EVTRAP purified sample digested using our in-house digestion protocol (sample 2). C) TIC from EVTRAP purified sample digested following the digestion protocol described by Iliuk *et al.* (sample 3). D) TIC from EVTRAP purified sample digested following the digestion protocol described by Iliuk *et al.* (sample 4).

Table 4.5. Summary of protein, peptide, and peptide spectral matches from EVTRAP purified peptide samples. Sample 1 and 2 were digested following our in-house digestion protocol. Samples 3 and 4 were digested following the digestion protocol described by Iliuk *et al.*

<b>Sample</b>	<b>Protein Ids, #</b>	<b>Peptide Ids, #</b>	<b>Peptide Spectral Matches, #</b>
1	52	239	630
2	68	344	899
3	28	78	131
4	25	63	110

unique peptide ids. These differences could be due to a difference in the efficiency of the digestion protocol and the sensitivity of the MS instrument used for the analysis, however these differences also highlight that less EVs are able to be purified using the IMCStip-based extraction protocol using the new EVTRAP resin. To determine if the proteins that were identified by the MS analysis were EV associated, the list of identified proteins were compared to a top 100 EV-associated protein list published by Vesiclepedia ([http://microvesicles.org/extracellular\\_vesicle\\_markers](http://microvesicles.org/extracellular_vesicle_markers)). These proteins are ranked 1 to 100 based on the number of reports that have listed the protein as EV-associated.

From the list of proteins identified from the MS analysis from samples 1 and 2, 7 of the 68 proteins were found on the Vesiclepedia top 100 EV-associated protein list. In particular, ACTB and CD63 are well studied EV-associated proteins that are trusted markers for EVs. This enrichment of EV-associated proteins within the protein list in Table 4.6 shows that the IMCStip-based EVTRAP extraction can purify EVs from urine, however there are still contaminating proteins in the resulting EV sample and a lower amount of EVs are purified compared to previous reports with EVTRAP resin.

#### 4.3 CONCLUSION

Extracellular vesicles are a promising tool for biomarker discovery for early detection of disease including cancer. One of the main hinderances in the development of EV research for biomarker discovery are the extraction techniques currently used for EV purification are time consuming and are not easily transferable to a clinical setting. In this work we aimed to address these issues by establishing an automated IMCStip-based extraction using resins previously shown to purify EVs from urine. In this study, we showed that the established IMCStip-based extraction using EVTRAP resin can

Table 4.6. List of proteins identified by samples 1 and 2 from Table 5.

Accession	Gene Symbol	Protein Description	# Peptides	EV Associated
P13645	KRT10	Keratin, type I cytoskeletal 10	26	✓
P07911	UMOD	Uromodulin	30	
P35908	KRT2	Keratin, type II cytoskeletal 2 epidermal	27	
P35527	KRT9	Keratin, type I cytoskeletal 9	21	
P04264	KRT1	Keratin, type II cytoskeletal 1	28	✓
Q04695	KRT17	Keratin, type I cytoskeletal 17	14	
P01133	EGF	Pro-epidermal growth factor	18	
P13647	KRT5	Keratin, type II cytoskeletal 5	20	
P04259	KRT6B	Keratin, type II cytoskeletal 6B	20	
P02538	KRT6A	Keratin, type II cytoskeletal 6A	20	
P05090	APOD	Apolipoprotein D	13	
Q6EMK4	VASN	Vasorin	9	
P10909	CLU	Clusterin	11	
P02533	KRT14	Keratin, type I cytoskeletal 14	11	
P05154	SERPINA5	Plasma serine protease inhibitor	8	
P07602	PSAP	Prosaposin	6	
P02768	ALB	Serum albumin	2	✓
P08473	MME	Nepilysin	6	
P16444	DPEP1	Dipeptidase 1	5	
P01042	KNG1	Kininogen-1	6	
P12109	COL6A1	Collagen alpha-1(VI) chain	2	
P05155	SERPING1	Plasma protease C1 inhibitor	6	
P15144	ANPEP	Aminopeptidase N	5	
P19440	GGT1	Gamma-glutamyltranspeptidase 1	5	
Q86Y46	KRT73	Keratin, type II cytoskeletal 73	5	
P06702	S100A9	Protein S100-A9	3	
P05452	CLEC3B	Tetranectin	3	
P68871	HBB	Hemoglobin subunit beta	5	

Q9H8L6	MMRN2	Multimerin-2	4	
O14773	TPP1	Tripeptidyl-peptidase 1	3	
P62979	RPS27A	Ubiquitin-40S ribosomal protein S27a	2	
P01009	SERPINA1	Alpha-1-antitrypsin	2	
O96009	NAPSA	Napsin-A	2	
P15309	ACP3	Prostatic acid phosphatase	3	
Q9NZH0	GPRC5B	G-protein coupled receptor family C group 5 member B	2	
Q9BRK3	MXRA8	Matrix-remodeling-associated protein 8	1	
P02765	AHSG	Alpha-2-HS-glycoprotein	1	
P10451	SPP1	Osteopontin	1	
Q16769	QPCT	Glutaminy-peptide cyclotransferase	1	
P02760	AMBP	Protein AMBP	2	
P27487	DPP4	Dipeptidyl peptidase 4	1	
O14498	ISLR	Immunoglobulin superfamily containing leucine-rich repeat protein	2	
P04406	GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	2	✓
P01034	CST3	Cystatin-C	1	
P69905	HBA1	Hemoglobin subunit alpha	2	
P60709	ACTB	Actin, cytoplasmic 1	1	✓
Q9BZG9	LYNX1	Ly-6/neurotoxin-like protein 1	1	
P02649	APOE	Apolipoprotein E	1	
Q86YZ3	HRNR	Hornerin	2	
P31151	S100A7	Protein S100-A7	1	
P15941	MUC1	Mucin-1	1	
O43895	XPNPEP2	Xaa-Pro aminopeptidase 2	1	
O43653	PSCA	Prostate stem cell antigen	1	
P15311	EZR	Ezrin	1	✓
O95865	DDAH2	N(G),N(G)-dimethylarginine dimethylaminohydrolase 2	1	
P08962	CD63	CD63 antigen	1	✓
P05109	S100A8	Protein S100-A8	1	
O00187	MASP2	Mannan-binding lectin serine protease 2	1	

Q16651	PRSS8	Prostasin	1	
Q8WVJ2	NUDCD2	NudC domain-containing protein 2	1	
P20472	PVALB	Parvalbumin alpha	1	
P11532	DMD	Dystrophin	1	
P00749	PLAU	Urokinase-type plasminogen activator	1	
Q9UMS6	SYNPO2	Synaptopodin-2	1	
Q9UN37	VPS4A	Vacuolar protein sorting-associated protein 4A	1	
Q8IY63	AMOTL1	Angiomotin-like protein 1	1	
Q8WUT4	LRRN4	Leucine-rich repeat neuronal protein 4	1	

purify EVs from urine samples, however significant issues remain to be addressed including lack of reproducibility and low EV extraction efficiency. These issues could stem from contaminants binding to the frit inside of the IMCStip leading to clogging or could be due to clumping of the resin inside of the IMCStip. In future work, other resins including the DEAE resin could be more thoroughly tested to determine if better reproducibility and EV extraction efficiency could be achieved.

#### 4.4 MATERIALS & METHODS

##### *4.4.1 Western blot analysis*

The purified EV samples were dried using a centrivap and then reconstituted in laemmli buffer. All urine control were first filtered and dried in a vacuum centrifuge and reconstituted in laemmli buffer. Aliquots of each sample were loaded at the equivalent volumes and were separated on an 4–20% Mini-PROTEAN® TGX™ Precast Protein gel (BIO-RAD). The proteins were transferred onto a low-fluorescence PVDF membrane (MilliporeSigma), and the membrane blocked with 1% BSA in TBST for 1 hour. The membranes were then incubated with rabbit anti-CD9 (D3H4P; Cell Signaling Technology) at the specified ratio overnight in 1% BSA in TBST. The secondary antibody visualizing the binding of primary antibodies was Goat anti-Rabbit IgG-HRP (Santa Cruz Biotechnology) or Goat anti-Rabbit Alexa Fluor 680 nm (Thermo Fisher Scientific), incubated for 1 hour in 1% BSA in TBST. Using the Goat anti-Rabbit IgG-HRP, the membrane was then stained with ECL western blotting substrate (Thermo Fisher Scientific) and was scanned by a GE imager using chemiluminescence for signal detection and quantitation. Using the Goat anti-Rabbit Alexa Fluor 680 nm, the

membrane was scanned by an Odyssey near-infrared scanner (Licor) for signal detection and quantitation.

#### ***4.4.2 Transmission electron microscopy***

Transmission electron microscopy (TEM) analysis was performed by depositing 15  $\mu\text{L}$  of the purified EV sample onto 300-mesh carbon coated copper grid. The grids were stained with 2% uranyl acetate and viewed with a Hitachi HT7800 transmission electron microscope.

#### ***4.4.3 LTQ-Orbitrap Velos Pro analysis for global phosphopeptide analysis***

Dried phosphopeptide samples were dissolved in 10  $\mu\text{L}$  of 0.25% formic acid with 3% (vol/vol) acetonitrile and 4  $\mu\text{L}$  of each sample was injected into Easy-nLC 1000 (Thermo Fisher Scientific). Peptides were separated on a 105-mm column (75  $\mu\text{m}$  ID) containing Reprosil-PUR C18 resin (New Objective). The mobile phase buffer consisted of 0.1% formic acid in ultrapure water (buffer A) with an eluting buffer of 0.1% formic acid in 80% (vol/vol) acetonitrile (buffer B) run with a linear 60-min gradient of 6–30% buffer B at flow rate of 300 nL/min. Easy-nLC 1000 was coupled online with a hybrid high-resolution LTQ-Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific). The mass spectrometer was operated in the data-dependent mode, in which a full-scan MS (from  $m/z$  300 to 1,500 with the resolution of 30,000 at  $m/z$  400) was followed by MS/MS of the 10 most intense ions [normalized collision energy - 30%; automatic gain control (AGC) -  $3.0 \times 10^4$ ; maximum injection time (MS1) - 100 ms; 90 s exclusion; maximum injection time (MS2) – 150 ms].

#### ***4.4.4 Database Search and Analysis***

The raw data files obtained from the LTQ-Orbitrap Velos Pro were searched against a human database using SEQUEST on Proteome Discoverer (Version 2.2, Thermo Fisher). The peptide precursor mass tolerance was set to 10 ppm, and the fragment mass tolerance was set to 0.6 Da. Search criteria included a static modification of cysteine residues of +57.0214 Da and a variable modification of +15.995 Da to include potential oxidation of methionine, a modification of +79.966 Da on serine, threonine, or tyrosine for the identification of phosphorylation (three modifications allowed per peptide). Searches were performed with full tryptic digestion and allowed a maximum of two missed cleavages on the peptides analyzed from the sequence database. False discovery rates (FDR) were set to less than 1% for each analysis at the PSM level. Phosphorylation site localization from CID spectra was determined by PhosphoRS on Proteome Discoverer 2.2. Confidence view was assigned a score of 0.4, indicating medium confidence. Proteins were manually search against the top 100 EV-associated protein list published by Vesiclepedia ([http://microvesicles.org/extracellular\\_vesicle\\_markers](http://microvesicles.org/extracellular_vesicle_markers)).

## REFERENCES

1. Hunter, T., Why nature chose phosphate to modify proteins. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2012**, *367* (1602), 2513-6.
2. Ardito, F.; Giuliani, M.; Perrone, D.; Troiano, G.; Lo Muzio, L., The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *Int. J. Mol. Med.* **2017**, *40* (2), 271-280.
3. Pawson, T.; Scott, J. D., Protein phosphorylation in signaling--50 years and counting. *Trends Biochem. Sci.* **2005**, *30* (6), 286-90.
4. Hunter, T., Tyrosine phosphorylation: thirty years and counting. *Curr. Opin. Cell Biol.* **2009**, *21* (2), 140-6.
5. Hunter, T., Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling. *Cell* **1995**, *80* (2), 225-36.
6. Cohen, P., The role of protein phosphorylation in human health and disease. The Sir Hans Krebs Medal Lecture. *Eur. J. Biochem.* **2001**, *268* (19), 5001-10.
7. Schwartz, P. A.; Murray, B. W., Protein kinase biochemistry and drug discovery. *Bioorg. Chem.* **2011**, *39* (5-6), 192-210.
8. Gross, S.; Rahal, R.; Stransky, N.; Lengauer, C.; Hoeflich, K. P., Targeting cancer with kinase inhibitors. *J. Clin. Invest.* **2015**, *125* (5), 1780-9.
9. Imamura, H.; Wakabayashi, M.; Ishihama, Y., Analytical strategies for shotgun phosphoproteomics: status and prospects. *Semin. Cell. Dev. Biol.* **2012**, *23* (8), 836-42.
10. Riley, N. M.; Coon, J. J., Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling. *Anal. Chem.* **2016**, *88* (1), 74-94.
11. Altelaar, A. F.; Munoz, J.; Heck, A. J., Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* **2013**, *14* (1), 35-48.
12. Rauniyar, N.; Yates, J. R., 3rd, Isobaric labeling-based relative quantification in shotgun proteomics. *J. Proteome Res.* **2014**, *13* (12), 5293-309.
13. Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002**, *1* (5), 376-86.

14. Chan, C. Y.; Gritsenko, M. A.; Smith, R. D.; Qian, W. J., The current state of the art of quantitative phosphoproteomics and its applications to diabetes research. *Expert Rev. Proteomics* **2016**, *13* (4), 421-33.
15. Högberg, A.; von Stechow, L.; Bekker-Jensen, D. B.; Weinert, B. T.; Kelstrup, C. D.; Olsen, J. V., Benchmarking common quantification strategies for large-scale phosphoproteomics. *Nat. Commun.* **2018**, *9* (1), 1045.
16. Mann, M.; Ong, S. E.; Gronborg, M.; Steen, H.; Jensen, O. N.; Pandey, A., Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends Biotechnol.* **2002**, (20), p261–268.
17. Fila, J.; Honys, D., Enrichment techniques employed in phosphoproteomics. *Amino Acids* **2012**, *43* (3), 1025-47.
18. Negroni, L.; Claverol, S.; Rosenbaum, J.; Chevet, E.; Bonneau, M.; Schmitter, J. M., Comparison of IMAC and MOAC for phosphopeptide enrichment by column chromatography. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2012**, *891-892*, 109-12.
19. Matheron, L.; van den Toorn, H.; Heck, A. J.; Mohammed, S., Characterization of biases in phosphopeptide enrichment by Ti(4+)-immobilized metal affinity chromatography and TiO<sub>2</sub> using a massive synthetic library and human cell digests. *Anal. Chem.* **2014**, (86), p8312–8320.
20. Whiteaker, J. R.; Zhao, L.; Yan, P.; Ivey, R. G.; Voytovich, U. J.; Moore, H. D.; Lin, C.; Paulovich, A. G., Peptide Immunoaffinity Enrichment and Targeted Mass Spectrometry Enables Multiplex, Quantitative Pharmacodynamic Studies of Phospho-Signaling. *Mol. Cell. Proteomics* **2015**, *14* (8), 2261-73.
21. Giansanti, P.; Stokes, M. P.; Silva, J. C.; Scholten, A.; Heck, A. J., Interrogating cAMP-dependent kinase signaling in Jurkat T cells via a protein kinase A targeted immune-precipitation phosphoproteomics approach. *Mol. Cell. Proteomics* **2013**, *12* (11), 3350-9.
22. Murillo, J. R.; Kuras, M.; Rezeli, M.; Milliotis, T.; Betancourt, L.; Marko-Varga, G., Automated phosphopeptide enrichment from minute quantities of frozen malignant melanoma tissue. *PLoS One* **2018**, *13* (12), e0208562.
23. Bordin, D. C.; Alves, M. N.; de Campos, E. G.; De Martinis, B. S., Disposable pipette tips extraction: Fundamentals, applications and state of the art. *J. Sep. Sci.* **2016**, *39* (6), 1168-72.
24. Savary, B. J.; Vasu, P., Routine identity confirmation of recombinant proteins by MALDI-TOF mass spectrometry. *Methods Mol. Biol.* **2012**, *824*, 37-50.

25. Mullis, B. T.; Hwang, S.; Lee, L. A.; Iliuk, A.; Woolsey, R.; Quilici, D.; Wang, Q., Automating Complex, Multistep Processes on a Single Robotic Platform to Generate Reproducible Phosphoproteomic Data. *SLAS Discov.* **2020**, *25* (3), 277-286.
26. Schley, C.; Swart, R.; Huber, C. G., Capillary scale monolithic trap column for desalting and preconcentration of peptides and proteins in one- and two-dimensional separations. *J. Chromatogr. A* **2006**, *1136* (2), 210-20.
27. Kates, P. A.; Tomashek, J. J.; Miles, D. A.; Lee, L. A., Parallel sample processing using dispersive INtip micro-purification on programmable multichannel pipettes. *Biotechniques* **2020**, *68* (3), 148-154.
28. Searleman, A. C.; Iliuk, A. B.; Collier, T. S.; Chodosh, L. A.; Tao, W. A.; Bose, R., Tissue phosphoproteomics with PolyMAC identifies potential therapeutic targets in a transgenic mouse model of HER2 positive breast cancer. *Electrophoresis* **2014**, *35* (24), 3463-9.
29. Iliuk, A. B.; Martin, V. A.; Alicie, B. M.; Geahlen, R. L.; Tao, W. A., In-depth analyses of kinase-dependent tyrosine phosphoproteomes based on metal ion-functionalized soluble nanopolymers. *Mol. Cell. Proteomics* **2010**, *9* (10), 2162-72.
30. Thingholm, T.; Larsen, M., The Use of Titanium Dioxide for Selective Enrichment of Phosphorylated Peptides. In *von Stechow L. (eds) Phospho-Proteomics. Methods in Molecular Biology*, Springer, New York, NY: 2016; Vol. 1355, pp 135-146.
31. Hsieh, E. J.; Bereman, M. S.; Durand, S.; Valaskovic, G. A.; MacCoss, M. J., Effects of column and gradient lengths on peak capacity and peptide identification in nanoflow LC-MS/MS of complex proteomic samples. *J. Am. Soc. Mass Spectrom.* **2013**, *24* (1), 148-53.
32. Yates, J. R., 3rd, Recent technical advances in proteomics. *F1000Research* **2019**, *8*.
33. Zhou, H.; Ye, M.; Dong, J.; Corradini, E.; Cristobal, A.; Heck, A. J.; Zou, H.; Mohammed, S., Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. *Nat. Protoc.* **2013**, *8* (3), 461-80.
34. Zhou, H.; Low, T. Y.; Hennrich, M. L.; van der Toorn, H.; Schwend, T.; Zou, H.; Mohammed, S.; Heck, A. J., Enhancing the identification of phosphopeptides from putative basophilic kinase substrates using Ti (IV) based IMAC enrichment. *Mol. Cell. Proteomics* **2011**, *10* (10), M110 006452.
35. Chen, D.; Ludwig, K. R.; Krokhin, O. V.; Spicer, V.; Yang, Z.; Shen, X.; Hummon, A. B.; Sun, L., Capillary Zone Electrophoresis-Tandem Mass Spectrometry for Large-Scale Phosphoproteomics with the Production of over 11,000 Phosphopeptides from the Colon Carcinoma HCT116 Cell Line. *Anal. Chem.* **2019**, *91* (3), 2201-2208.

36. Siegel, R. L.; Fedewa, S. A.; Anderson, W. F.; Miller, K. D.; Ma, J.; Rosenberg, P. S.; Jemal, A., Colorectal Cancer Incidence Patterns in the United States, 1974-2013. *J. Natl. Cancer Inst.* **2017**, *109* (8).
37. Kweon, H. K.; Hakansson, K., Selective zirconium dioxide-based enrichment of phosphorylated peptides for mass spectrometric analysis. *Anal. Chem.* **2006**, *78* (6), 1743-9.
38. Yue, X.; Schunter, A.; Hummon, A. B., Comparing multistep immobilized metal affinity chromatography and multistep TiO<sub>2</sub> methods for phosphopeptide enrichment. *Anal. Chem.* **2015**, *87* (17), 8837-44.
39. <https://www.cancer.org/cancer/colon-rectal-cancer/about/key-statistics.html>.
40. Negrei, C.; Hudita, A.; Gingham, O.; Galateanu, B.; Voicu, S. N.; Stan, M.; Costache, M.; Fenga, C.; Drakoulis, N.; Tsatsakis, A. M., Colon Cancer Cells Gene Expression Signature As Response to 5-Fluorouracil, Oxaliplatin, and Folinic Acid Treatment. *Front. Pharmacol.* **2016**, *7*, 172.
41. Tsukahara, H.; Tsunekuni, K.; Takechi, T., Folic Acid-Metabolizing Enzymes Regulate the Antitumor Effect of 5-Fluoro-2'-Deoxyuridine in Colorectal Cancer Cell Lines. *PLoS One* **2016**, *11* (9), e0163961.
42. Ozer, U.; Barbour, K. W.; Clinton, S. A.; Berger, F. G., Oxidative Stress and Response to Thymidylate Synthase-Targeted Antimetabolites. *Mol. Pharmacol.* **2015**, *88* (6), 970-81.
43. Longley, D. B.; Johnston, P. G., Molecular mechanisms of drug resistance. *J. Pathol.* **2005**, *205* (2), 275-92.
44. Ishida, K.; Ito, C.; Ohmori, Y.; Kume, K.; Sato, K. A.; Koizumi, Y.; Konta, A.; Iwaya, T.; Nukatsuka, M.; Kobunai, T.; Takechi, T.; Nishizuka, S. S., Inhibition of PI3K suppresses propagation of drug-tolerant cancer cell subpopulations enriched by 5-fluorouracil. *Scientific Reports* **2017**, *7* (1), 2262.
45. Bracht, K.; Nicholls, A. M.; Liu, Y.; Bodmer, W. F., 5-Fluorouracil response in a large panel of colorectal cancer cell lines is associated with mismatch repair deficiency. *Br. J. Cancer* **2010**, *103* (3), 340-6.
46. He, L.; Zhu, H.; Zhou, S.; Wu, T.; Wu, H.; Yang, H.; Mao, H.; SekharKathera, C.; Janardhan, A.; Edick, A. M.; Zhang, A.; Hu, Z.; Pan, F.; Guo, Z., Wnt pathway is involved in 5-FU drug resistance of colorectal cancer cells. *Exp. Mol. Med.* **2018**, *50* (8), 101.
47. Blondy, S.; David, V.; Verdier, M.; Mathonnet, M.; Perraud, A.; Christou, N., 5-Fluorouracil resistance mechanisms in colorectal cancer: From classical pathways to promising processes. *Cancer Sci.* **2020**, *111* (9), 3142-3154.

48. Guo, X.; Goessl, E.; Jin, G.; Collie-Duguid, E. S.; Cassidy, J.; Wang, W.; O'Brien, V., Cell cycle perturbation and acquired 5-fluorouracil chemoresistance. *Anticancer Res.* **2008**, *28* (1A), 9-14.
49. Touil, Y.; Igoudjil, W.; Corvaisier, M.; Dessein, A. F.; Vandomme, J.; Monte, D.; Stechly, L.; Skrypek, N.; Langlois, C.; Grard, G.; Millet, G.; Leteurtre, E.; Dumont, P.; Truant, S.; Pruvot, F. R.; Hebbar, M.; Fan, F.; Ellis, L. M.; Formstecher, P.; Van Seuning, I.; Gespach, C.; Polakowska, R.; Huet, G., Colon cancer cells escape 5FU chemotherapy-induced cell death by entering stemness and quiescence associated with the c-Yes/YAP axis. *Clin. Cancer Res.* **2014**, *20* (4), 837-46.
50. Hwang, I. T.; Chung, Y. M.; Kim, J. J.; Chung, J. S.; Kim, B. S.; Kim, H. J.; Kim, J. S.; Yoo, Y. D., Drug resistance to 5-FU linked to reactive oxygen species modulator 1. *Biochem. Biophys. Res. Commun.* **2007**, *359* (2), 304-10.
51. Patel, B. B.; Gupta, D.; Elliott, A. A.; Sengupta, V.; Yu, Y.; Majumdar, A. P., Curcumin targets FOLFOX-surviving colon cancer cells via inhibition of EGFRs and IGF-1R. *Anticancer Res.* **2010**, *30* (2), 319-25.
52. Zhang, N.; Yin, Y.; Xu, S.-J.; Chen, W.-S., 5-Fluorouracil: Mechanisms of Resistance and Reversal Strategies. *Molecules* **2008**, *13* (8), 1551-1569.
53. Kim, M.; Yan, Y.; Kortum, R. L.; Stoeger, S. M.; Sgagias, M. K.; Lee, K.; Lewis, R. E.; Cowan, K. H., Expression of kinase suppressor of Ras1 enhances cisplatin-induced extracellular signal-regulated kinase activation and cisplatin sensitivity. *Cancer Res.* **2005**, *65* (10), 3986-92.
54. Wang, X.; Martindale, J. L.; Holbrook, N. J., Requirement for ERK activation in cisplatin-induced apoptosis. *J. Biol. Chem.* **2000**, *275* (50), 39435-43.
55. Yang, S. Y.; Miah, A.; Sales, K. M.; Fuller, B.; Seifalian, A. M.; Winslet, M., Inhibition of the p38 MAPK pathway sensitises human colon cancer cells to 5-fluorouracil treatment. *Int. J. Oncol.* **2011**, *38* (6), 1695-702.
56. Case, N.; Thomas, J.; Sen, B.; Styner, M.; Xie, Z.; Galior, K.; Rubin, J., Mechanical regulation of glycogen synthase kinase 3beta (GSK3beta) in mesenchymal stem cells is dependent on Akt protein serine 473 phosphorylation via mTORC2 protein. *J. Biol. Chem.* **2011**, *286* (45), 39450-6.
57. Koksai, A. S.; Beck, K.; Cronin, D. R.; McKenna, A.; Camp, N. D.; Srivastava, S.; MacGilvray, M. E.; Bodik, R.; Wolf-Yadlin, A.; Fraenkel, E.; Fisher, J.; Gitter, A., Synthesizing Signaling Pathways from Temporal Phosphoproteomic Data. *Cell Rep.* **2018**, *24* (13), 3607-3618.
58. De Angelis, P. M.; Svendsrud, D. H.; Kravik, K. L.; Stokke, T., Cellular response to 5-fluorouracil (5-FU) in 5-FU-resistant colon cancer cell lines during treatment and recovery. *Mol. Cancer* **2006**, *5*, 20.

59. Zhou, D.; Palam, L. R.; Jiang, L.; Narasimhan, J.; Staschke, K. A.; Wek, R. C., Phosphorylation of eIF2 directs ATF5 translational control in response to diverse stress conditions. *J. Biol. Chem.* **2008**, *283* (11), 7064-73.
60. Hamanaka, R. B.; Bennett, B. S.; Cullinan, S. B.; Diehl, J. A., PERK and GCN2 contribute to eIF2 $\alpha$  phosphorylation and cell cycle arrest after activation of the unfolded protein response pathway. *Mol. Biol. Cell* **2005**, *16* (12), 5493-501.
61. Kaehler, C.; Isensee, J.; Hucho, T.; Lehrach, H.; Krobitsch, S., 5-Fluorouracil affects assembly of stress granules based on RNA incorporation. *Nucleic Acids Res.* **2014**, *42* (10), 6436-47.
62. Giudice, A.; Arra, C.; Turco, M. C., Review of molecular mechanisms involved in the activation of the Nrf2-ARE signaling pathway by chemopreventive agents. *Methods Mol. Biol.* **2010**, *647*, 37-74.
63. Izdebska, M.; Gagat, M.; Grzanka, A., Overexpression of lamin B1 induces mitotic catastrophe in colon cancer LoVo cells and is associated with worse clinical outcomes. *Int. J. Oncol.* **2018**, *52* (1), 89-102.
64. McAllister, F.; Pineda, D. M.; Jimbo, M.; Lal, S.; Burkhart, R. A.; Moughan, J.; Winter, K. A.; Abdelmohsen, K.; Gorospe, M.; Acosta Ade, J.; Lankapalli, R. H.; Winter, J. M.; Yeo, C. J.; Witkiewicz, A. K.; Iacobuzio-Donahue, C. A.; Laheru, D.; Brody, J. R., dCK expression correlates with 5-fluorouracil efficacy and HuR cytoplasmic expression in pancreatic cancer: a dual-institutional follow-up with the RTOG 9704 trial. *Cancer Biol. Ther.* **2014**, *15* (6), 688-98.
65. Riley, N. M.; Coon, J. J., Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling. *Analytical Chemistry* **2016**, *88* (1), 74-94.
66. Du, Z.; Lovly, C. M., Mechanisms of receptor tyrosine kinase activation in cancer. *Mol. Cancer* **2018**, *17* (1), 58.
67. Yang, P.; Zheng, X.; Jayaswal, V.; Hu, G.; Yang, J. Y.; Jothi, R., Knowledge-Based Analysis for Detecting Key Signaling Events from Time-Series Phosphoproteomics Data. *PLoS Comput. Biol.* **2015**, *11* (8), e1004403.
68. Casado, P.; Rodriguez-Prados, J. C.; Cosulich, S. C.; Guichard, S.; Vanhaesebroeck, B.; Joel, S.; Cutillas, P. R., Kinase-substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells. *Sci. Signal* **2013**, *6* (268), rs6.
69. Hou, S.; Liu, Z.; Shen, H.; Wu, D., Damage-Associated Molecular Pattern-Triggered Immunity in Plants. *Front. Plant Sci.* **2019**, *10*, 646.
70. Ameye, M.; Allmann, S.; Verwaeren, J.; Smagghe, G.; Haesaert, G.; Schuurink, R. C.; Audenaert, K., Green leaf volatile production by plants: a meta-analysis. *New Phytol.* **2018**, *220* (3), 666-683.

71. Choi, H. W.; Klessig, D. F., DAMPs, MAMPs, and NAMPs in plant innate immunity. *BMC Plant Biol.* **2016**, *16* (1), 232.
72. Scala, A.; Allmann, S.; Mirabella, R.; Haring, M. A.; Schuurink, R. C., Green leaf volatiles: a plant's multifunctional weapon against herbivores and pathogens. *Int. J. Mol. Sci.* **2013**, *14* (9), 17781-811.
73. Furstenberg-Hagg, J.; Zagrobelny, M.; Bak, S., Plant defense against insect herbivores. *Int. J. Mol. Sci.* **2013**, *14* (5), 10242-97.
74. Engelberth, J.; Alborn, H. T.; Schmelz, E. A.; Tumlinson, J. H., Airborne signals prime plants against insect herbivore attack. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (6), 1781-5.
75. Dombrowski, J. E.; Kronmiller, B. A.; Hollenbeck, V. G.; Rhodes, A. C.; Henning, J. A.; Martin, R. C., Transcriptome analysis of the model grass *Lolium temulentum* exposed to green leaf volatiles. *BMC Plant Biol.* **2019**, *19* (1), 222.
76. Engelberth, J.; Engelberth, M., Variability in the Capacity to Produce Damage-Induced Aldehyde Green Leaf Volatiles among Different Plant Species Provides Novel Insights into Biosynthetic Diversity. *Plants (Basel)* **2020**, *9* (2).
77. Hsu, C. C.; Zhu, Y.; Arrington, J. V.; Paez, J. S.; Wang, P.; Zhu, P.; Chen, I. H.; Zhu, J. K.; Tao, W. A., Universal Plant Phosphoproteomics Workflow and Its Application to Tomato Signaling in Response to Cold Stress. *Mol. Cell. Proteomics* **2018**, *17* (10), 2068-2080.
78. Loreto, F.; Barta, C.; Brilli, F.; Nogues, I., On the induction of volatile organic compound emissions by plants as consequence of wounding or fluctuations of light and temperature. *Plant Cell Environ.* **2006**, *29* (9), 1820-8.
79. Gouinguene, S. P.; Turlings, T. C., The effects of abiotic factors on induced volatile emissions in corn plants. *Plant Physiol.* **2002**, *129* (3), 1296-307.
80. Croft, K.; Juttner, F.; Slusarenko, A. J., Volatile Products of the Lipoxygenase Pathway Evolved from *Phaseolus vulgaris* (L.) Leaves Inoculated with *Pseudomonas syringae* pv *phaseolicola*. *Plant Physiol.* **1993**, *101* (1), 13-24.
81. Ameye, M.; Audenaert, K.; De Zutter, N.; Steppe, K.; Van Meulebroek, L.; Vanhaecke, L.; De Vleeschauwer, D.; Haesaert, G.; Smagghe, G., Priming of wheat with the green leaf volatile Z-3-hexenyl acetate enhances defense against *Fusarium graminearum* but boosts deoxynivalenol production. *Plant Physiol.* **2015**, *167* (4), 1671-84.
82. Dicke, M., Behavioural and community ecology of plants that cry for help. *Plant Cell Environ.* **2009**, *32* (6), 654-65.

83. van Hulst, M.; Pelser, M.; van Loon, L. C.; Pieterse, C. M.; Ton, J., Costs and benefits of priming for defense in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (14), 5602-7.
84. Zebelo, S. A.; Matsui, K.; Ozawa, R.; Maffei, M. E., Plasma membrane potential depolarization and cytosolic calcium flux are early events involved in tomato (*Solanum lycopersicon*) plant-to-plant communication. *Plant Sci.* **2012**, *196*, 93-100.
85. Asai, N.; Nishioka, T.; Takabayashi, J.; Furuichi, T., Plant volatiles regulate the activities of Ca<sup>2+</sup>-permeable channels and promote cytoplasmic calcium transients in Arabidopsis leaf cells. *Plant Signal. Behav.* **2009**, *4* (4), 294-300.
86. Kandoth, P. K.; Ranf, S.; Pancholi, S. S.; Jayanty, S.; Walla, M. D.; Miller, W.; Howe, G. A.; Lincoln, D. E.; Stratmann, J. W., Tomato MAPKs LeMPK1, LeMPK2, and LeMPK3 function in the systemin-mediated defense response against herbivorous insects. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (29), 12205-10.
87. Holley, S. R.; Yalamanchili, R. D.; Moura, D. S.; Ryan, C. A.; Stratmann, J. W., Convergence of signaling pathways induced by systemin, oligosaccharide elicitors, and ultraviolet-B radiation at the level of mitogen-activated protein kinases in *Lycopersicon peruvianum* suspension-cultured cells. *Plant Physiol.* **2003**, *132* (4), 1728-38.
88. Yamauchi, Y.; Matsuda, A.; Matsuura, N.; Mizutani, M.; Sugimoto, Y., Transcriptome analysis of Arabidopsis thaliana treated with green leaf volatiles: possible role of green leaf volatiles as self-made damage-associated molecular patterns. *J. Pestic. Sci.* **2018**, *43* (3), 207-213.
89. Wang, L.; Wilkins, K. A.; Davies, J. M., Arabidopsis DORN1 extracellular ATP receptor; activation of plasma membrane K<sup>(+)</sup>-and Ca<sup>(2+)</sup>-permeable conductances. *New Phytol.* **2018**, *218* (4), 1301-1304.
90. Tanaka, K.; Choi, J.; Cao, Y.; Stacey, G., Extracellular ATP acts as a damage-associated molecular pattern (DAMP) signal in plants. *Front. Plant Sci.* **2014**, *5*, 446.
91. Haj Ahmad, F.; Wu, X. N.; Stintzi, A.; Schaller, A.; Schulze, W. X., The Systemin Signaling Cascade As Derived from Time Course Analyses of the Systemin-responsive Phosphoproteome. *Mol. Cell. Proteomics* **2019**, *18* (8), 1526-1542.
92. Wang, L.; Einig, E.; Almeida-Trapp, M.; Albert, M.; Fliegmann, J.; Mithofer, A.; Kalbacher, H.; Felix, G., The systemin receptor SYR1 enhances resistance of tomato against herbivorous insects. *Nat. Plants* **2018**, *4* (3), 152-156.
93. Scheer, J. M.; Ryan, C. A., Jr., The systemin receptor SR160 from *Lycopersicon peruvianum* is a member of the LRR receptor kinase family. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (14), 9585-90.

94. Wei, J.; Wang, L.; Zhao, J.; Li, C.; Ge, F.; Kang, L., Ecological trade-offs between jasmonic acid-dependent direct and indirect plant defences in tritrophic interactions. *New Phytol.* **2011**, *189* (2), 557-67.
95. Leal, W. S., Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. *Annu. Rev. Entomol.* **2013**, *58*, 373-91.
96. Moriyama, E. N.; Strope, P. K.; Opiyo, S. O.; Chen, Z.; Jones, A. M., Mining the *Arabidopsis thaliana* genome for highly-divergent seven transmembrane receptors. *Genome Biol.* **2006**, *7* (10), R96.
97. Dombrowski, J. E.; Martin, R. C., Activation of MAP kinases by green leaf volatiles in grasses. *BMC Res. Notes* **2018**, *11* (1), 79.
98. Bequette, C. J.; Hind, S. R.; Pulliam, S.; Higgins, R.; Stratmann, J. W., MAP kinases associate with high molecular weight multiprotein complexes. *J. Exp. Bot.* **2018**, *69* (3), 643-654.
99. Chi, Y. H.; Koo, S. S.; Oh, H. T.; Lee, E. S.; Park, J. H.; Phan, K. A. T.; Wi, S. D.; Bae, S. B.; Paeng, S. K.; Chae, H. B.; Kang, C. H.; Kim, M. G.; Kim, W. Y.; Yun, D. J.; Lee, S. Y., The Physiological Functions of Universal Stress Proteins and Their Molecular Mechanism to Protect Plants From Environmental Stresses. *Front. Plant Sci.* **2019**, *10*, 750.
100. Jung, Y. J.; Melencion, S. M.; Lee, E. S.; Park, J. H.; Alinapon, C. V.; Oh, H. T.; Yun, D. J.; Chi, Y. H.; Lee, S. Y., Universal Stress Protein Exhibits a Redox-Dependent Chaperone Function in *Arabidopsis* and Enhances Plant Tolerance to Heat Shock and Oxidative Stress. *Front. Plant Sci.* **2015**, *6*, 1141.
101. Greeff, C.; Roux, M.; Mundy, J.; Petersen, M., Receptor-like kinase complexes in plant innate immunity. *Front. Plant Sci.* **2012**, *3*, 209.
102. He, X.; Wang, C.; Wang, H.; Li, L.; Wang, C., The Function of MAPK Cascades in Response to Various Stresses in Horticultural Plants. *Front. Plant Sci.* **2020**, *11*, 952.
103. Kong, F.; Wang, J.; Cheng, L.; Liu, S.; Wu, J.; Peng, Z.; Lu, G., Genome-wide analysis of the mitogen-activated protein kinase gene family in *Solanum lycopersicum*. *Gene* **2012**, *499* (1), 108-20.
104. Liao, D.; Sun, X.; Wang, N.; Song, F.; Liang, Y., Tomato LysM Receptor-Like Kinase SILYK12 Is Involved in Arbuscular Mycorrhizal Symbiosis. *Front. Plant Sci.* **2018**, *9*, 1004.
105. Schulz, P.; Herde, M.; Romeis, T., Calcium-dependent protein kinases: hubs in plant stress signaling and development. *Plant Physiol.* **2013**, *163* (2), 523-30.

106. Tena, G.; Boudsocq, M.; Sheen, J., Protein kinase signaling networks in plant innate immunity. *Curr. Opin. Plant Biol.* **2011**, *14* (5), 519-29.
107. Boudsocq, M.; Willmann, M. R.; McCormack, M.; Lee, H.; Shan, L.; He, P.; Bush, J.; Cheng, S. H.; Sheen, J., Differential innate immune signalling via Ca(2+) sensor protein kinases. *Nature* **2010**, *464* (7287), 418-22.
108. Xu, R.; Rai, A.; Chen, M.; Suwakulsiri, W.; Greening, D. W.; Simpson, R. J., Extracellular vesicles in cancer - implications for future improvements in cancer care. *Nat. Rev. Clin. Oncol.* **2018**, *15* (10), 617-638.
109. Wu, A. Y.; Ueda, K.; Lai, C. P., Proteomic Analysis of Extracellular Vesicles for Cancer Diagnostics. *Proteomics* **2019**, *19* (1-2), e1800162.
110. Becker, A.; Thakur, B. K.; Weiss, J. M.; Kim, H. S.; Peinado, H.; Lyden, D., Extracellular Vesicles in Cancer: Cell-to-Cell Mediators of Metastasis. *Cancer Cell* **2016**, *30* (6), 836-848.
111. Haraszti, R. A.; Didiot, M. C.; Sapp, E.; Leszyk, J.; Shaffer, S. A.; Rockwell, H. E.; Gao, F.; Narain, N. R.; DiFiglia, M.; Kiebish, M. A.; Aronin, N.; Khvorova, A., High-resolution proteomic and lipidomic analysis of exosomes and microvesicles from different cell sources. *J. Extracell. Vesicles* **2016**, *5*, 32570.
112. Kowal, J.; Arras, G.; Colombo, M.; Jouve, M.; Morath, J. P.; Primdal-Bengtson, B.; Dingli, F.; Loew, D.; Tkach, M.; Thery, C., Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113* (8), E968-77.
113. Zhang, H.; Freitas, D.; Kim, H. S.; Fabijanic, K.; Li, Z.; Chen, H.; Mark, M. T.; Molina, H.; Martin, A. B.; Bojmar, L.; Fang, J.; Rampersaud, S.; Hoshino, A.; Matei, I.; Kenific, C. M.; Nakajima, M.; Mutvei, A. P.; Sansone, P.; Buehring, W.; Wang, H.; Jimenez, J. P.; Cohen-Gould, L.; Paknejad, N.; Brendel, M.; Manova-Todorova, K.; Magalhães, A.; Ferreira, J. A.; Osório, H.; Silva, A. M.; Massey, A.; Cubillos-Ruiz, J. R.; Galletti, G.; Giannakakou, P.; Cuervo, A. M.; Blenis, J.; Schwartz, R.; Brady, M. S.; Peinado, H.; Bromberg, J.; Matsui, H.; Reis, C. A.; Lyden, D., Identification of distinct nanoparticles and subsets of extracellular vesicles by asymmetric flow field-flow fractionation. *Nature Cell Biology* **2018**, *20* (3), 332-343.
114. Choi, D. S.; Lee, J. M.; Park, G. W.; Lim, H. W.; Bang, J. Y.; Kim, Y. K.; Kwon, K. H.; Kwon, H. J.; Kim, K. P.; Gho, Y. S., Proteomic analysis of microvesicles derived from human colorectal cancer cells. *J. Proteome Res.* **2007**, *6* (12), 4646-55.
115. Xu, R.; Greening, D. W.; Rai, A.; Ji, H.; Simpson, R. J., Highly-purified exosomes and shed microvesicles isolated from the human colon cancer cell line LIM1863 by sequential centrifugal ultrafiltration are biochemically and functionally distinct. *Methods* **2015**, *87*, 11-25.

116. Boukouris, S.; Mathivanan, S., Exosomes in bodily fluids are a highly stable resource of disease biomarkers. *Proteomics Clin. Appl.* **2015**, *9* (3-4), 358-67.
117. Chen, I. H.; Xue, L.; Hsu, C. C.; Paez, J. S.; Pan, L.; Andaluz, H.; Wendt, M. K.; Iliuk, A. B.; Zhu, J. K.; Tao, W. A., Phosphoproteins in extracellular vesicles as candidate markers for breast cancer. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114* (12), 3175-3180.
118. Fakih, M.; Padmanabhan, A., CEA monitoring in colorectal cancer. What you should know. *Oncology (Williston Park)* **2006**, *20* (6), 579-587; discussion 588, 594, 596 passim.
119. Saito, G.; Sadahiro, S.; Kamata, H.; Miyakita, H.; Okada, K.; Tanaka, A.; Suzuki, T., Monitoring of Serum Carcinoembryonic Antigen Levels after Curative Resection of Colon Cancer: Cutoff Values Determined according to Preoperative Levels Enhance the Diagnostic Accuracy for Recurrence. *Oncology* **2017**, *92* (5), 276-282.
120. Shiromizu, T.; Kume, H.; Ishida, M.; Adachi, J.; Kano, M.; Matsubara, H.; Tomonaga, T., Quantitation of putative colorectal cancer biomarker candidates in serum extracellular vesicles by targeted proteomics. *Scientific Reports* **2017**, *7* (1), 12782.
121. Wang, S.; Kojima, K.; Mobley, J. A.; West, A. B., Proteomic analysis of urinary extracellular vesicles reveal biomarkers for neurologic disease. *EBioMedicine* **2019**, *45*, 351-361.
122. Wu, X.; Li, L.; Iliuk, A.; Tao, W. A., Highly Efficient Phosphoproteome Capture and Analysis from Urinary Extracellular Vesicles. *J. Proteome Res.* **2018**, *17* (9), 3308-3316.
123. Heath, N.; Grant, L.; De Oliveira, T. M.; Rowlinson, R.; Osteikoetxea, X.; Dekker, N.; Overman, R., Rapid isolation and enrichment of extracellular vesicle preparations using anion exchange chromatography. *Scientific Reports* **2018**, *8* (1), 5730.
124. Wiklander, O. P. B.; Brennan, M. A.; Lotvall, J.; Breakefield, X. O.; El Andaloussi, S., Advances in therapeutic applications of extracellular vesicles. *Sci. Transl. Med.* **2019**, *11* (492).

## APPENDIX A

### REPRINT PERMISSION FOR CHAPTER 1

## SAGE's Author Archiving and Re-Use Guidelines

These guidelines should be followed by authors of Contributions published in a SAGE subscription journal, including authors whose Contributions were published under a previous version of the author guidelines. For a list of exceptions to these guidelines, please see below.

Three versions of the Contribution are referenced in these guidelines:

- **Original Submission:** the version submitted by the author before peer review
- **Accepted Manuscript:** version updated to include the author's revisions after peer review, prior to any typesetting for the journal. This is often the version accepted by the editor
- **Final Published PDF:** copy-edited and typeset Publisher's PDF, the same version published on the journal's website

### Green Open Access: SAGE's Archiving and Sharing Policy

You may share the **Original Submission** or **Accepted Manuscript** at any time after your paper is accepted and in any format. Your sharing of the **Original Submission** or **Accepted Manuscript** may include posting a downloadable copy on any website, saving a copy in any repository or network, sharing a copy through any social media channel, and distributing print or electronic copies. Please note some journals will not consider papers that have been posted as preprints prior to submission and you may check a journal's policy regarding considering previously-posted papers by referring to the journal's submission guidelines.

For information on use of Institutional Repository (IR) copies by authors and IR users, see [Posting to an Institutional Repository - Green Open Access](#).

You may use the **Final Published PDF** (or **Original Submission** or **Accepted Manuscript**, if preferred) in the following ways:

- in relation to your own teaching, provided that any electronic distribution maintains restricted access
- to share on an individual basis with research colleagues, provided that such sharing is not for commercial purposes
- in your dissertation or thesis, including where the dissertation or thesis will be posted in any electronic Institutional Repository or database
- in a book authored or edited by you, at any time after the Contribution's publication in the journal.

## APPENDIX B

### REPRINT PERMISSION FOR FIGURE 2.1

Folic Acid-Metabolizing Enzymes Regulate the Antitumor Effect of 5-Fluoro-2'-Deoxyuridine in Colorectal Cancer ...  
Hiroshi Tsukihara, Kenta Tsunekuni, Teiji Takechi

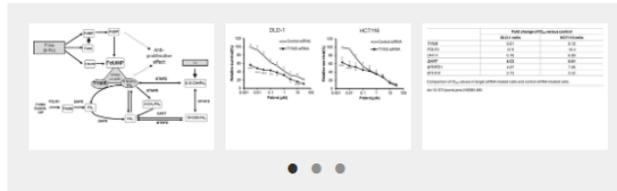


#### Abstract

Introduction  
Materials and Methods  
Results  
Discussion  
Supporting Information  
Acknowledgments  
Author Contributions  
References

Reader Comments (0)  
Figures

#### Figures



Enzyme metabolism



Statistical data



Malignant tumors



**Citation:** Tsukihara H, Tsunekuni K, Takechi T (2016) Folic Acid-Metabolizing Enzymes Regulate the Antitumor Effect of 5-Fluoro-2'-Deoxyuridine in Colorectal Cancer Cell Lines. PLoS ONE 11(9): e0163961. <https://doi.org/10.1371/journal.pone.0163961>

**Editor:** Bing Xu, Brandeis University, UNITED STATES

**Received:** May 2, 2016; **Accepted:** September 16, 2016; **Published:** September 29, 2016

**Copyright:** © 2016 Tsukihara et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability:** All relevant data are within the paper and its Supporting Information files.

**Funding:** The authors received no specific funding for this work. Taiho Pharmaceutical Co., Ltd., <http://www.taiho.co.jp/english/>, provided support in the form of salaries for authors HT, KT and TT and acquisition of research materials, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section below.

## APPENDIX C

### REPRINT PERMISSION FOR FIGURE 3.1

## MDPI Open Access Information and Policy

All articles published by MDPI are made immediately available worldwide under an open access license. This means:

- everyone has free and unlimited access to the full-text of *all* articles published in MDPI journals;
- everyone is free to re-use the published material if proper accreditation/citation of the original publication is given;
- open access publication is supported by the authors' institutes or research funding agencies by payment of a comparatively low **Article Processing Charge (APC)** for accepted articles.

### Permissions

No special permission is required to reuse all or part of article published by MDPI, including figures and tables. For articles published under an open access Creative Common CC BY license, any part of the article may be reused without permission provided that the original article is clearly cited. Reuse of an article does not imply endorsement by the authors or MDPI.

## APPENDIX D

### REPRINT PERMISSION FIGURE 3.2



#### Thank you for your order!

Dear Mr. Brian Mullis,

Thank you for placing your order through **Copyright** Clearance Center's RightsLink<sup>®</sup> service.

#### Order Summary

Licensee:	Mr. Brian Mullis
Order Date:	Mar 29, 2021
Order Number:	5038360875291
Publication:	New Phytologist
Title:	Green leaf volatile production by plants: a meta-analysis
Type of Use:	Dissertation/Thesis
Order Total:	0.00 USD

View or print complete [details](#) of your order and the publisher's terms and conditions.

Sincerely,

**Copyright** Clearance Center

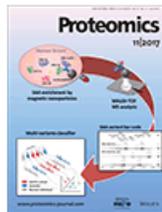
Tel: +1-855-239-3415 / +1-978-646-2777  
[customercare@copyright.com](mailto:customercare@copyright.com)  
<https://myaccount.copyright.com>



RightsLink<sup>®</sup>

## APPENDIX E

### REPRINT PERMISSION FIGURE 4.1



### Thank you for your order!

Dear Mr. Brian Mullis,

Thank you for placing your order through **Copyright** Clearance Center's RightsLink® service.

#### Order Summary

Licensee: Mr. Brian Mullis  
Order Date: Mar 29, 2021  
Order Number: 5038361201167  
Publication: Proteomics  
Title: Proteomic Analysis of Extracellular Vesicles for Cancer Diagnostics  
Type of Use: Dissertation/Thesis  
Order Total: 0.00 USD

View or print complete [details](#) of your order and the publisher's terms and conditions.

Sincerely,

**Copyright** Clearance Center

Tel: +1-855-239-3415 / +1-978-646-2777  
[customer@copyright.com](mailto:customer@copyright.com)  
<https://myaccount.copyright.com>



RightsLink®

## APPENDIX F

### REPRINT PERMISSION FIGURE 4.2

EBioMedicine 45 (2019) 351–361



Contents lists available at ScienceDirect

EBioMedicine

journal homepage: [www.ebiomedicine.com](http://www.ebiomedicine.com)

**EBioMedicine**  
Published by THE LANCET

#### Proteomic analysis of urinary extracellular vesicles reveal biomarkers for neurologic disease



Shijie Wang<sup>a</sup>, Kyoko Kojima<sup>b</sup>, James A. Mobley<sup>b</sup>, Andrew B. West<sup>a,\*</sup>

<sup>a</sup> Center for Neurodegeneration and Neurotherapeutics, Duke University, Durham, NC, USA

<sup>b</sup> Department of Surgery, University of Alabama at Birmingham, Birmingham, AL, USA

#### ARTICLE INFO

##### Article history:

Received 13 February 2019  
Received in revised form 13 June 2019  
Available online 20 June 2019

##### Keywords:

Biomarkers  
Neurodegeneration  
Extracellular vesicles  
Parkinson's disease  
Alzheimer's disease

#### ABSTRACT

**Background:** Extracellular vesicles (EVs) harbor thousands of proteins that hold promise for biomarker development. Usually difficult to purify, EVs in urine are relatively easily obtained and have demonstrated efficacy for kidney disease prediction. Herein, we further characterize the proteome of urinary EVs to explore the potential for biomarkers unrelated to kidney dysfunction, focusing on Parkinson's disease (PD).

**Methods:** Using a quantitative mass spectrometry approach, we measured urinary EV proteins from a discovery cohort of 50 subjects. EVs in urine were classified into subgroups and EV proteins were ranked by abundance and variability over time. Enriched pathways and ontologies in stable EV proteins were identified and proteins that predict PD were further measured in a cohort of 108 subjects.

**Findings:** Hundreds of commonly expressed urinary EV proteins with stable expression over time were distinguished from proteins with high variability. Bioinformatic analyses reveal a striking enrichment of endolysosomal proteins linked to Parkinson's, Alzheimer's, and Huntington's disease. Tissue and biofluid enrichment analyses show broad representation of EVs from across the body without bias towards kidney or urine proteins. Among the proteins linked to neurological diseases, SNAP23 and calbindin were the most elevated in PD cases with 86% prediction success for disease diagnosis in the discovery cohort and 76% prediction success in the replication cohort.

**Interpretation:** Urinary EVs are an underutilized but highly accessible resource for biomarker discovery with particular promise for neurological diseases like PD.

© 2019 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

#### 1. Introduction

Extracellular vesicles (EVs) are a subset of small vesicles (40–200 nm) derived from the endolysosomal system and released into biofluids by almost all cell types. EVs play a role in cell–cell communication and are involved in several diseases [1–3]. Recent studies demonstrate EVs mediating some types of immunological responses [4–6], cancer progression and metastasis [7–11], blood coagulation [12,13], and neurodegenerative diseases [1,3,14,15]. Further, emerging studies demonstrate that EVs, especially intraluminal vesicles from multi-vesicular bodies (i.e., exosomes), can readily cross the blood brain barrier and access tissues throughout the body [16–18]. The specific sources and relative contributions of different types of cells through the body to the EV pool in different biofluids remains almost completely unknown.

Compared to proteins and nucleic acids not enveloped in lipid bilayers in the extracellular space, EVs shelter biomolecules from external

nucleases, proteases, and other degradative enzymes. EVs in biofluids can be enriched with little bias using ultracentrifugation or size exclusion approaches [19,20]. Thousands of proteins not usually secreted through canonical pathways have been observed in proteomic studies in purified EV fractions [21–24]. Because of the stability of EV proteins in biofluids and diversity in proteomic composition, EVs are under evaluation for biomarker development for application in diagnostic and prognostic approaches in numerous diseases [19,24,25].

While EVs are ubiquitous in biofluids, they are also in very low relative concentration compared to more abundant secreted proteins such as immunoglobins, albumins, and lipoproteins. Higher volume collections of biofluids required for reliable measurement of EV proteins in cerebral spinal fluid, saliva, and plasma present technical challenges in routine clinical settings for biomarker discovery. In contrast, urine can be routinely collected in the hundreds of milliliters in a non-invasive manner without risk. Urinary EV proteins have shown strong promise for kidney diseases including Gitelman syndrome, familial renal hypomagnesemia, urogenital cancers, and acute kidney injury [20,26–28].

Given the polydispersity and recently recognized ability of some EVs to spread throughout the body, herein, we systematically evaluated the

\* Corresponding author at: 3 Genome Court, Durham, NC 27710, USA.  
E-mail address: [Andrew.West@Duke.edu](mailto:Andrew.West@Duke.edu) (A.B. West).

<https://doi.org/10.1016/j.ebiom.2019.06.021>

2352-3964/© 2019 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).