

APPLICATION FOR FEDERAL ASSISTANCE

SF 424 R&R

		3. DATE RECEIVED BY STATE	State Application Identifier
1. TYPE OF SUBMISSION <input type="radio"/> Pre-application <input checked="" type="radio"/> Application <input type="radio"/> Changed/Corrected Application		4. a. Federal Identifier b. Agency Routing Identifier c. Previous Grants.gov Tracking ID	
2. DATE SUBMITTED	Applicant Identifier		
5. APPLICANT INFORMATION Legal Name: University of Utah Department: Street1: 75 South 2000 East City: Salt Lake City Province:		Organizational DUNS: [REDACTED] Division: Street2: County/Parish: Salt Lake State: UT: Utah Country: USA: UNITED STATES ZIP / Postal Code: 84112-8930	
Person to be contacted on matters involving this application Prefix: [REDACTED] First Name: [REDACTED] Middle Name: [REDACTED] Last Name: [REDACTED] Suffix: Position>Title: GRANTS & CONTRACTS OFFICER Street1: 75 S 2000 E RM 211 City: SALT LAKE CITY Province: Phone Number: [REDACTED] Fax Number: [REDACTED] Email: [REDACTED]			
6. EMPLOYER IDENTIFICATION NUMBER (EIN) or (TIN): [REDACTED]			
7. TYPE OF APPLICANT: H: Public/State Controlled Institution of Higher Education Other (Specify): Small Business Organization Type <input type="radio"/> Women Owned <input type="radio"/> Socially and Economically Disadvantaged			
8. TYPE OF APPLICATION: <input checked="" type="radio"/> New <input type="radio"/> Resubmission <input type="radio"/> Renewal <input type="radio"/> Continuation <input type="radio"/> Revision		If Revision, mark appropriate box(es). <input type="radio"/> A. Increase Award <input type="radio"/> B. Decrease Award <input type="radio"/> C. Increase Duration <input type="radio"/> D. Decrease Duration <input type="radio"/> E. Other(specify):	
Is this application being submitted to other agencies? <input type="radio"/> Yes <input checked="" type="radio"/> No What other Agencies?			
9. NAME OF FEDERAL AGENCY: National Institutes of Health		10. CATALOG OF FEDERAL DOMESTIC ASSISTANCE NUMBER: TITLE: [REDACTED]	
11. DESCRIPTIVE TITLE OF APPLICANT'S PROJECT: Contextualizing Chaotic Metabolic Networks and Their Regulation			
12. PROPOSED PROJECT: Start Date 07/01/2020 Ending Date 06/30/2026		13. CONGRESSIONAL DISTRICT OF THE APPLICANT: [REDACTED]	

14. PROJECT DIRECTOR/PRINCIPAL INVESTIGATOR CONTACT INFORMATION

Prefix: First Name: [REDACTED] Middle Name: [REDACTED] Last Name: BERG Suffix:
 JORDAN
 Position/TITLE: Graduate Research Assistant Organization Name: University of Utah
 Department: Biochemistry Division:
 Street1: 15 N Medical Dr E RM 4100 Street2:
 City: Salt Lake City County/Parish: Salt Lake State: UT: Utah
 Province: Country: USA: UNITED STATES ZIP / Postal Code:
 Phone Number: [REDACTED] Fax Number: [REDACTED] Email: [REDACTED]
 84112-8930

15. ESTIMATED PROJECT FUNDING

- a. Total Federal Funds Requested [REDACTED]
- b. Total Non-Federal Funds [REDACTED]
- c. Total Federal & Non-Federal Funds [REDACTED]
- d. Estimated Program Income [REDACTED]

16. IS APPLICATION SUBJECT TO REVIEW BY STATE EXECUTIVE ORDER 12372 PROCESS?

- a. YES THIS PREAPPLICATION/APPLICATION WAS MADE AVAILABLE TO THE STATE EXECUTIVE ORDER 12372 PROCESS FOR REVIEW ON:
DATE: [REDACTED]
- b. NO PROGRAM IS NOT COVERED BY E.O. 12372; OR
 PROGRAM HAS NOT BEEN SELECTED BY STATE FOR REVIEW

17. By signing this application, I certify (1) to the statements contained in the list of certifications* and (2) that the statements herein are true, complete and accurate to the best of my knowledge. I also provide the required assurances * and agree to comply with any resulting terms if I accept an award. I am aware that any false, fictitious, or fraudulent statements or claims may subject me to criminal, civil, or administrative penalties. (U.S. Code, Title 18, Section 1001)

I agree

The list of certifications and assurances, or an Internet site where you may obtain this list, is contained in the announcement or agency specific instructions.

18. SFLLL or other Explanatory Documentation. File Name: Mime Type:**19. Authorized Representative**

Prefix: First Name: [REDACTED] Middle Name: [REDACTED] Last Name: [REDACTED] Suffix:
 [REDACTED]
 Position/TITLE: Director Organization Name: University of Utah
 Department: OFFICE OF SPONSORED PROJECTS Division: VP FOR RESEARCH
 Street1: 75 South 2000 East Street2:
 City: Salt Lake City County/Parish: Salt Lake State: UT: Utah
 Province: Country: USA: UNITED STATES ZIP / Postal Code:
 84112-8930
 Phone Number: [REDACTED] Fax Number: [REDACTED] Email: [REDACTED]

Signature of Authorized Representative

Date Signed

20. Pre-application File Name: Mime Type:**21. Cover Letter Attachment File Name: Cover_Letter1029235290.pdf Mime Type: application/pdf**



Division of Receipt and Referral
Center for Scientific Review
National Institutes of Health
6701 Rockledge Drive
Rockledge II, MSC-7768
Bethesda, MD 20892-7768
(301) 435-1115

03 Dec 2019

Dear Selection Committee,

I am pleased to submit the enclosed fellowship proposal, "**Contextualizing Chaotic Metabolic Networks and Their Regulation**", for consideration for the NCI Predoctoral to Postdoctoral Fellow Transition Award (F99/K00). The following application contains all of the required agency approval documentation as issued in **RFA-CA-19-057**. This proposal seeks to build tools to aid in contextualizing complex and dynamic data in metabolic networks and use these tools to answer previously unanswerable questions related to the aggregate regulation of metabolism. These tools and regulatory mechanisms will be vital to deconvoluting cancer metabolism behavior. I appreciate the opportunity to be considered for this fellowship.

List of referees:

[REDACTED LIST OF REFEREES]

Sincerely,

Jordan A. Berg
Ph.D Candidate
Department of Biochemistry
University of Utah
Emma Eccles Jones Research Building, Rm 5520
15 N. Medical Drive
Salt Lake City, UT 84112-5650

Project/Performance Site Location(s)

Project/Performance Site Primary Location

Organization Name: University of Utah

* Street1: 15 N Medical Dr E RM 4100

Street2:

* City: Salt Lake City

County: Salt Lake

* State: UT: Utah

Province:

* Country: USA: UNITED
STATES

* Zip / Postal Code:
84112-8930

DUNS Number: [REDACTED]

* Project/Performance Site Congressional District: [REDACTED]

File Name

Mime Type

Additional Location(s)

RESEARCH & RELATED Other Project Information

1. * Are Human Subjects Involved? <input type="radio"/> Yes <input checked="" type="radio"/> No		
1.a. If YES to Human Subjects		
Is the Project Exempt from Federal regulations? <input type="radio"/> Yes <input type="radio"/> No		
If yes, check appropriate exemption number		
Exemption Number: — 1 — 2 — 3 — 4 — 5 — 6 — 7 — 8		
If no, is the IRB review Pending? <input type="radio"/> Yes <input type="radio"/> No		
IRB Approval Date:		
Human Subject Assurance Number		
2. * Are Vertebrate Animals Used? <input type="radio"/> Yes <input checked="" type="radio"/> No		
2.a. If YES to Vertebrate Animals		
Is the IACUC review Pending? <input type="radio"/> Yes <input type="radio"/> No		
IACUC Approval Date:		
Animal Welfare Assurance Number		
3. * Is proprietary/privileged information <input type="radio"/> Yes <input checked="" type="radio"/> No included in the application?		
4.a.* Does the Project have an Actual or Perceived Impact – positive or negative – on the environment? <input type="radio"/> Yes <input checked="" type="radio"/> No		
4.b. If yes, please explain:		
4.c. If this project has an actual or potential impact on the environment, has an exemption been authorized or an environmental assessment (EA) or environmental impact statement (EIS) been performed? <input type="radio"/> Yes <input type="radio"/> No		
4.d. If yes, please explain:		
5.a.* Is the research performance site designated, or eligible to be designated, as a historic place? <input type="radio"/> Yes <input checked="" type="radio"/> No		
5.b. If yes, please explain:		
6.a.* Does this project involve activities outside the U.S. or partnership with International Collaborators? <input type="radio"/> Yes <input checked="" type="radio"/> No		
6.b. If yes, identify countries:		
6.c. Optional Explanation:		
7. Project Summary/Abstract	Project_Abstract1029235300.pdf	Mime Type: application/pdf
8. Project Narrative	Project_Narrative1029235299.pdf	Mime Type: application/pdf
9. Bibliography & References Cited	Bibliography_References_Cited1029235269.pdf	Mime Type: application/pdf
10. Facilities & Other Resources	Facilities_Other_Resources1029235291.pdf	Mime Type: application/pdf
11. Equipment	Equipment1029204466.pdf	Mime Type: application/pdf
12. Other Attachments	Institutional_Nomination_Letter1029235230.pdf	Mime Type: application/pdf

Project Summary/Abstract

Cancer metabolism is a complex network of perturbations to essential chemical and enzymatic reactions; however, the past century has seen a largely reductionist approach to understanding this system. While previously this approach was necessary due to technological limitations, current computer age technological advances allow us to survey, model, and explore the biological details of individual cells and populations of cells. Scientific fields, such as RNA biology and metabolism, have experienced massive strides in recent decades with the advent of RNA-seq and mass spectrometry-based metabolomics, yet our ability to contextualize and extract the full extent of these enormous datasets continues to lag and often results in focusing on only a handful of entities from a dataset. This effectively causes “big data” to become “little data”. This is problematic as these experiments are often expensive and time-consuming to produce, yet we only use a fraction of the total data produced by a given experiment. For the F99 phase of my proposal, I will address these limitations by leading the development of *Metaboverse*, a multi-omic computational analysis framework built upon our previous work to contextualize -omics datasets within customizable and global metabolic network representations. This framework will lay the foundation allowing for the exploration of complex forms of metabolic regulation in cancer. For example, we will analyze the ability of metabolic networks to undergo dispersed and low-magnitude regulation, where, rather than one or two components acting as the core regulatory actors, regulation is performed by dispersed groups of genes, proteins, or metabolites. **This framework and related regulatory research will revolutionize our ability to more holistically understand temporal metabolic shifts and gene-metabolite intra-cooperativity, as well as ensure we obtain the maximum amount of information from our data.** For the K00 phase of my proposal, I will work with a postdoctoral mentor at an NCI-Designated Cancer Center or affiliated institution that will supplement my training in machine learning and network biology to develop models that improve our ability to predict metabolic state from transcriptomic state. Doing so will allow us to harness the vast transcriptomics databases in cancer biology to better understand the role of metabolism across heterogeneous tumor cell populations. **My ultimate goal is to become a tenured professor and run an independent, NIH-funded research lab that focuses on computational cancer metabolism research and that develops methods for interrogating this emerging domain of biology.**

Project Narrative

While cancer metabolism is a robust and well-developed research field, approaches to its holistic understanding are still under-developed and hinder our ability to contextualize these complex metabolic states and their consequences. During the F99 phase, I will develop tools and methods that allow researchers to explore data in a more holistic manner than previously possible, which will be essential to elucidating more complicated regulatory mechanisms within cancer metabolism. During the K00 phase, I will develop novel machine learning algorithms that will improve our ability to predict the metabolic state from the transcriptional state, allowing us to harness the rich transcription datasets found in cancer biology for therapeutic benefit.

Bibliography

1. Diether, M., Nikolaev, Y., Allain, F. H. & Sauer, U. Systematic mapping of protein-metabolite interactions in central metabolism of Escherichia coli. *Mol. Syst. Biol.* **15**, (2019). PMCID: PMC6706640
2. DeBerardinis, R. J. & Chandel, N. S. Fundamentals of cancer metabolism. *Sci. Adv.* **2**, e1600200 (2016). PMCID: PMC4928883
3. Hirschey, M. D. *et al.* Dysregulated metabolism contributes to oncogenesis. *Semin. Cancer Biol.* **35**, S129–S150 (2015). PMCID: PMC4656121
4. Rosario, S. R. *et al.* Pan-cancer analysis of transcriptional metabolic dysregulation using The Cancer Genome Atlas. *Nat. Commun.* **9**, 1–17 (2018). PMCID: PMC6294258
5. Cho, D.-Y., Kim, Y.-A. & Przytycka, T. M. Chapter 5: Network Biology Approach to Complex Diseases. *PLOS Comput. Biol.* **8**, e1002820 (2012). PMCID: PMC3531284
6. Misra, B. B., Langefeld, C., Olivier, M. & Cox, L. A. Integrated omics: tools, advances and future approaches. *J. Mol. Endocrinol.* **62**, R21–R45 (2019). PMID: 30006342
7. Li, W. Application of Volcano Plots in Analyses of mRNA Differential Expressions with Microarrays. *J. Bioinform. Comput. Biol.* **10**, 1231003 (2012). PMID: 23075208
8. Costa-Silva, J., Domingues, D. & Lopes, F. M. RNA-Seq differential expression analysis: An extended review and a software tool. *PLOS ONE* **12**, e0190152 (2017). PMCID: PMC5739479
9. Berg, J. A. *et al.* XPRESSyourself: Enhancing, Standardizing, and Automating Ribosome Profiling Computational Analyses Yields Improved Insight into Data. *bioRxiv* 704320 (2019) doi:10.1101/704320.
10. Skinner, J. E. Low-dimensional Chaos in Biological Systems. *Bio/Technology* **12**, 596–600 (1994). PMID: 7764948
11. Sager, M. *et al.* Transcriptomics in cancer diagnostics: developments in technology, clinical research and commercialization. *Expert Rev. Mol. Diagn.* **15**, 1589–1603 (2015). PMID: 26565429
12. Pinu, F. R., Goldansaz, S. A. & Jaine, J. Translational Metabolomics: Current Challenges and Future Opportunities. *Metabolites* **9**, (2019). PMCID: PMC6631405
13. Zelezniak, A. *et al.* Machine Learning Predicts the Yeast Metabolome from the Quantitative Proteome of Kinase Knockouts. *Cell Syst.* **7**, 269–283.e6 (2018). PMCID: PMC6167078
14. Pavlova, N. N. & Thompson, C. B. The Emerging Hallmarks of Cancer Metabolism. *Cell Metab.* **23**, 27–47 (2016). PMCID: PMC4715268
15. Spinelli, J. B. & Haigis, M. C. The multifaceted contributions of mitochondria to cellular metabolism. *Nat. Cell Biol.* **20**, 745–754 (2018). PMCID: PMC6541229
16. Friedman, J. R. & Nunnari, J. Mitochondrial form and function. *Nature* **505**, 335–343 (2014). PMCID: PMC4075653
17. Luengo, A., Gui, D. Y. & Vander Heiden, M. G. Targeting Metabolism for Cancer Therapy. *Cell Chem. Biol.* **24**, 1161–1180 (2017). PMCID: PMC5744685
18. Mandrup, S. Lessons Learned from Systems Approaches to Metabolism. *Trends Endocrinol. Metab.* **26**, 669–670 (2015). PMID: 26596675
19. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003). PMCID: PMC403769
20. Cottret, L. *et al.* MetExplore: collaborative edition and exploration of metabolic networks. *Nucleic Acids Res.* **46**, W495–W502 (2018). PMCID: PMC6030842
21. Cottret, L. *et al.* MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks. *Nucleic Acids Res.* **38**, W132–137 (2010). PMCID: PMC2896158
22. García-Alcalde, F., García-López, F., Dopazo, J. & Conesa, A. Paintomics: a web based tool for the joint visualization of transcriptomics and metabolomics data. *Bioinforma. Oxf. Engl.* **27**, 137–139 (2011). PMCID: PMC3008637
23. Xia, J., Psychogios, N., Young, N. & Wishart, D. S. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res.* **37**, W652–W660 (2009). PMCID: PMC2703878
24. Do, K. T. *et al.* Characterization of missing values in untargeted MS-based metabolomics data and evaluation of missing data handling strategies. *Metabolomics* **14**, 128 (2018). PMCID: PMC6153696
25. Balázsi, G., van Oudenaarden, A. & Collins, J. J. Cellular decision making and biological noise: from microbes to mammals. *Cell* **144**, 910–925 (2011). PMCID: PMC3068611
26. Stewart-Ornstein, J., Weissman, J. S. & El-Samad, H. Cellular noise regulons underlie fluctuations in *Saccharomyces cerevisiae*. *Mol. Cell* **45**, 483–493 (2012). PMCID: PMC3327736

27. Sherman, M. S., Lorenz, K., Lanier, M. H. & Cohen, B. A. Cell-to-cell variability in the propensity to transcribe explains correlated fluctuations in gene expression. *Cell Syst.* **1**, 315–325 (2015). PMCID: PMC4662655
28. Waller, T. C., Berg, J. A., Lex, A., Chapman, B. E. & Rutter, J. Compartment and hub definitions tune metabolic networks for metabolomic interpretations. *GigaScience* (In Press).
29. Bensard, C. L. *et al.* Regulation of Tumor Initiation by the Mitochondrial Pyruvate Carrier. *Cell Metab.* (2019).
30. Gerashchenko, M. V. & Gladyshev, V. N. Translation inhibitors cause abnormalities in ribosome profiling experiments. *Nucleic Acids Res.* **42**, e134 (2014). PMCID: PMC4176156
31. Santos, D. A., Shi, L., Tu, B. P. & Weissman, J. S. Cycloheximide can distort measurements of mRNA levels and translation efficiency. *Nucleic Acids Res.* **47**, 4974–4985 (2019). PMCID: PMC6547433
32. Prlić, A. & Procter, J. B. Ten Simple Rules for the Open Development of Scientific Software. *PLOS Comput. Biol.* **8**, e1002802 (2012). PMCID: PMC3516539
33. Mason, O. & Verwoerd, M. Graph theory and networks in Biology. *IET Syst. Biol.* **1**, 89–119 (2007). PMID: 17441552
34. Chen, J. & Chen, H. A Topology-Based Approach to Pattern Recognition on Graph-Structured Data. in *2018 IEEE Intl Conf on Parallel Distributed Processing with Applications, Ubiquitous Computing Communications, Big Data Cloud Computing, Social Computing Networking, Sustainable Computing Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom)* 454–461 (2018). doi:10.1109/BDCloud.2018.00075.
35. Altman, N. S. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *Am. Stat.* **46**, 175–185 (1992).
36. Joshi-Tope, G. *et al.* Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.* **33**, D428–D432 (2005). PMCID: PMC540026
37. Schomburg, I., Chang, A. & Schomburg, D. BRENDA, enzyme data and metabolic information. *Nucleic Acids Res.* **30**, 47–49 (2002). PMCID: PMC99121
38. Orsak, T. *et al.* Revealing the allosterome: systematic identification of metabolite-protein interactions. *Biochemistry* **51**, 225–232 (2012). PMID: 22122470
39. Sambamoorthy, G. & Raman, K. Understanding the evolution of functional redundancy in metabolic networks. *Bioinformatics* **34**, i981–i987 (2018). PMCID: PMC6129275
40. Liu, X., Li, Y. I. & Pritchard, J. K. Trans Effects on Gene Expression Can Drive Omnipotent Inheritance. *Cell* **177**, 1022–1034.e6 (2019). PMCID: PMC6553491 [Available on 2020-05-02]
41. Lamb, J. *et al.* The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935 (2006). PMID: 17008526
42. Pohlert, T. Non-Parametric Trend Tests and Change-Point Detection. <https://cran.r-project.org/web/packages/trend/vignettes/trend.pdf>
43. Mahmood, R., Jia, S. & Zhu, W. Analysis of climate variability, trends, and prediction in the most active parts of the Lake Chad basin, Africa. *Sci. Rep.* **9**, (2019). PMCID: PMC6474870
44. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188 (2001).
45. Korthauer, K. *et al.* A practical guide to methods controlling false discoveries in computational biology. *Genome Biol.* **20**, 118 (2019). PMCID: PMC6547503
46. Ideker, T., Ozier, O., Schwikowski, B. & Siegel, A. F. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* **18**, S233–S240 (2002). PMID: 12169552
47. Chubukov, V. *et al.* Transcriptional regulation is insufficient to explain substrate-induced flux changes in *Bacillus subtilis*. *Mol. Syst. Biol.* **9**, 709 (2013). PMCID: PMC4039378
48. Ludwig, L. S. *et al.* Transcriptional States and Chromatin Accessibility Underlying Human Erythropoiesis. *Cell Rep.* **27**, 3228–3240.e7 (2019). PMCID: PMC6579117
49. Voss, T. C. & Hager, G. L. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat. Rev. Genet.* **15**, 69–81 (2014). PMCID: PMC6322398
50. Madsen, M. J. *et al.* Reparameterization of PAM50 Expression Identifies Novel Breast Tumor Dimensions and Leads to Discovery of a Genome-Wide Significant Breast Cancer Locus at 12q15. *Cancer Epidemiol. Prev. Biomark.* **27**, 644–652 (2018). PMCID: PMC5984724
51. Dannenfelser, R. *et al.* Data-driven analysis of immune infiltrate in a large cohort of breast cancer and its association with disease progression, ER activity, and genomic complexity. *Oncotarget* **8**, 57121–57133 (2017). PMCID: PMC5593630

52. Yao, V. et al. An integrative tissue-network approach to identify and test human disease genes. *Nat. Biotechnol.* (2018) doi:10.1038/nbt.4246. PMID: 30346941
53. Yao, V., Wong, A. K. & Troyanskaya, O. G. Enabling Precision Medicine through Integrative Network Models. *J. Mol. Biol.* **430**, 2913–2923 (2018). PMID: 30003887
54. Chen, L., Ducker, G. S., Lu, W., Teng, X. & Rabinowitz, J. D. An LC-MS chemical derivatization method for the measurement of five different one-carbon states of cellular tetrahydrofolate. *Anal. Bioanal. Chem.* **409**, 5955–5964 (2017). PMCID: PMC5737010
55. García-Cañaveras, J. C., Chen, L. & Rabinowitz, J. D. The Tumor Metabolic Microenvironment: Lessons from Lactate. *Cancer Res.* **79**, 3155–3162 (2019). PMCID: PMC6606343 [Available on 2020-07-01]
56. Zhang, Y. et al. Enhancing CD8+ T Cell Fatty Acid Catabolism within a Metabolically Challenging Tumor Microenvironment Increases the Efficacy of Melanoma Immunotherapy. *Cancer Cell* **32**, 377-391.e9 (2017). PMCID: PMC5751418
57. Krishnan, A., Taroni, J. N. & Greene, C. S. Integrative Networks Illuminate Biological Factors Underlying Gene–Disease Associations. *Curr. Genet. Med. Rep.* **4**, 155–162 (2016).
58. Taroni, J. N. et al. MultiPLIER: A Transfer Learning Framework for Transcriptomics Reveals Systemic Features of Rare Disease. *Cell Syst.* **8**, 380-394.e4 (2019). PMCID: PMC6538307 [Available on 2020-05-22]
59. Way, G. P. et al. Machine Learning Detects Pan-cancer Ras Pathway Activation in The Cancer Genome Atlas. *Cell Rep.* **23**, 172-180.e3 (2018). PMCID: PMC5918694

Facilities & Key Resources

Scientific Environment: The University of Utah School of Medicine has a long history of excellence in biochemistry, genetics, oncology, and neuroscience. NIH training grants support interdepartmental research and training in these and other areas, and the University continues to attract outstanding new faculty, students, and postdocs. Basic science and clinical faculty interact regularly at interdepartmental research meetings and at seminars with intramural and extramural speakers. These include weekly Journal Club meetings, Research in Progress meetings, and weekly Metabolism Interest Group meetings. Speakers at this meeting include members of each participating lab as well as invited external speakers from major research institutions around the country, providing a rich intellectual environment that will support the proposed project. The scientific community at the University of Utah is very open and collaborative. Investigators, postdoctoral fellows, and graduate students alike are very open and willing to share ideas and collaborate. There is a wealth of interdisciplinary science occurring at the University of Utah, which was another draw for me to come to the University for graduate school. Based on my interest in leveraging computational biology and big data to be able to ask previously difficult-to-answer questions, I chose to come to the University of Utah and work in the Rutter lab where opportunities to unlock this synergy. My dissertation committee is also highly reflective of the interdisciplinary nature of the university and my research as I have experts in cancer biology, cellular quality control and signaling regulation, metabolism, RNA biochemistry and RNA-Seq, and computational biology on my committee with appointments in Biochemistry, Oncological Sciences, and Human Genetics. My committee members are all well-established in their respective fields with a diversity of experience. Speaking to this point, Dr. Jared Rutter is currently an HHMI Investigator and members of my committee, Drs. Brenda Bass and Carl Thummel, were both HHMI Investigators for over 15 years each and Dr. Brenda Bass is a member of the National Academy of Sciences. Drs. Jason Gertz and Adam Hughes are considered leaders in their respective fields and each member is essential to the success of my dissertation. This committee has been enriched by collaborators and mentors, such as Drs. Aaron Quinlan, Bei Wang Phillips, and James Cox. To further illustrate the interdisciplinary nature of the university and research environment, Dr. Jared Rutter is a Professor in the Department of Biochemistry, holds an Ida Smith Endowed Chair of Cancer Research, and is a co-leader of the Nuclear Control and Cell Growth and Differentiation program at the Huntsman Cancer Institute and organizes the Seminars in Metabolism. This proposed project will benefit from outstanding core facilities supported by the University of Utah, such as the High-Throughput Genomics and Metabolomics Cores. All core services are provided through centralized facilities at reasonable recharge rates (<http://cores.utah.edu/>).

Laboratory: The Rutter laboratory (3461 sq ft) is in the 14-year old Emma Eccles Jones Medical Research Building at the University of Utah School of Medicine. The Rutter lab has adjacent 30°, 37°, 4°, tissue culture, radioactivity and common equipment rooms. This building houses both the Department of Biochemistry as well as the Department of Pathology and is located central to the other medical research buildings under the School of Medicine as well as many of the research cores, such as the Metabolomics core, drug discovery core, a CRISPR core, and the bioinformatics core to name a few. This setup allows for easy access to collaborators and equipment to facilitate my research projects. The Rutter lab contains the needed equipment to perform yeast genetics, molecular biology and biochemistry, equipment and resources needed to prepare high-throughput sequencing libraries, as well as the necessary BSL-2 cell and tissue culture facilities to culture, transfet, and treat the cell lines to be used in this proposal. I have a large computer desk with sufficient space for my personal workstation, along with available wet bench space.

Computer: The office of Dr. Rutter is equipped with a MacBook Pro with a studio display. Each member of the Rutter laboratory has either a Macintosh or PC computer, which is networked with access to an HP laser printer and remote access to a lab server for storage and exchange of data within the lab. The PI of this proposal has a 2018 15" Macbook Pro, 2.9 GHz Intel Core i9 laptop with 32 GB of DDR4 RAM, along with two external monitors. All workstations also have the necessary software for common laboratory functions, databases, word processing, and graphics.

Office: Dr. Rutter has a 120 sq. ft. office adjacent to the laboratory. Lab members have desk space in the laboratory and access to a dedicated trainee office for writing and reading. Administrative and accounting support and support equipment (copier, fax, conference room) are in the nearby department office.

Other: The Department of Biochemistry provides full centralized support for purchasing, accounting, and human resource management. The Rutter laboratory shares two technicians with other adjacent laboratories for yeast media preparation and glassware cleaning.

Facilities: The following facilities are available through the University of Utah: confocal microscope; flow cytometry; DNA and peptide synthesis core; electron microscopy; small animal imaging; mass spectrometry core for proteomics and metabolomics; biorepository with access to >30,000 specimens collected from cancer patients undergoing surgery at the Huntsman Cancer Institute; and a high-throughput genomics core for RNA-seq, ChIP-seq, etc.

High-Throughput Genomics Core: The University of Utah High Throughput Genomics Core is housed at the Huntsman Cancer Institute, proximal to the Emma Eccles Jones Medical Research Building where our lab is located. This core provides access to state-of-the-art Illumina sequencing platforms, as well as an Aligent Technologies 2200 TapeStation for quality control of ribosome profiling sequencing libraries. The sample-to-sequence turnaround time is reasonable, data access is easy, and the sequencing quality is reliable. In fact, the High-Throughput Genomics Core was recently part of a study comparing University sequencing cores and was shown to output the most aligning reads per sequencing run of the sequencing cores analyzed (<https://www.nature.com/articles/s41467-018-03751-6>).

Metabolomics Core: The Metabolomics Core, located in the Emma Eccles Jones Research Building, where the Department of Biochemistry and the Rutter lab are located, specializes in metabolomics sample preparation, metabolome profiling, and data analysis. The core has equipment capable of handling GC-MS, LC-MS, NMR, lipidomics, and metabolic flux analysis. The Rutter lab's SCIEX X500R QTOF MS is also housed in the Metabolomics Core.

Center for High Performance Computing (CHPC): The University of Utah CHPC manages over 22,000 high performance computing cores and over 13 petabytes of RAID configured spinning disk storage. CHPC frequently holds training workshops to aid users in leveraging their resources. Users can either purchase private computing nodes and data storage space or submit quarterly allocation requests for priority access to these freecycle nodes. The Rutter lab is currently allocated a generous 50,000 wall-clock hours per quarter. The Rutter lab also has allocated freecycle computing privileges with 5TB dedicated lab storage available in a protected environment for processing sensitive data. These resources are essential and sufficient for the processing of the data associated with this proposal.

Equipment

University of Utah Center for High-Performance Computing:

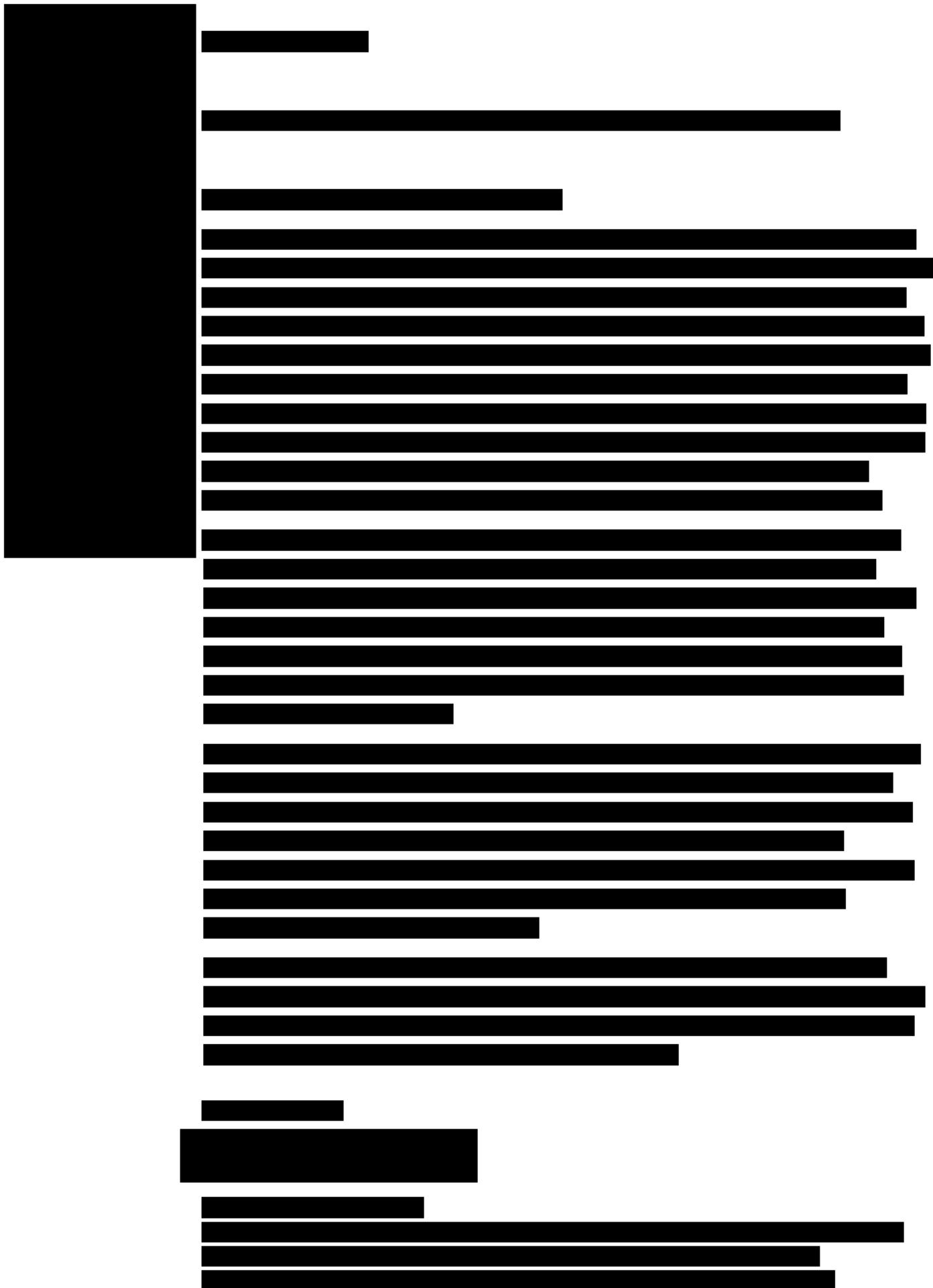
- Lab allocation of 50,000 general computing hours per quarter
- General free-cycle nodes available for all lab members with CHPC registration
- Protected computing environment free-cycle nodes and 5 TB lab project space for protected data processing and short-term storage
- Other non-allocation compute nodes available for use

Basic Lab Equipment: The Rutter lab is fully equipped to perform routine laboratory procedures involving molecular biology, yeast genetics and biochemistry, mouse husbandry and manipulations, tissue and cell culture, and standard light and fluorescence microscopy. Equipment available within the Rutter laboratory includes:

- SCIEX X500R QTOF mass spectrometer with a Shimadzu LC20 ADXR HPLC
- Zeiss LSM 880 Confocal Microscope with Airyscan
- Zeiss AxioObserver inverted microscope and ZenPro imaging system with DIC and fluorescence (including six different fluorescence filters), deconvolution, time-lapse, and tiling capabilities
- Seahorse XFe96 Analyzer for measurement of mitochondrial respiration, glycolysis, and fatty acid oxidation in live cells
- Three fridges, a deli case, a walk-in cold room
- Six -20° C freezers
- Three -80° C freezers
- Shimadzu SPD M10A VP HPLC for peptide and small molecule purification
- Roche LightCycler 480 for real-time PCR
- BioMek NX Laboratory Automation Workstation for automated microtiter plate liquid handling
- OceanOptics oxygen electrode for oxygen consumption experiments
- Eppendorf Biospectrometer kinetic spectrophotometer
- One MJ Research thermocycler and two BioRad T100 thermocyclers
- Thermo Speedvac
- Three cryostorage tanks for long term storage of cultured cells
- Five water-jacketed CO₂ incubators
- Three Thermo Sorval Legend Mirco 21 microfuges, one Eppendorf 5417C microfuge, and one Eppendorf 5415C microfuge
- Eppendorf 5424 centrifuge, Eppendorf 5417R centrifuge, Eppendorf 5810R centrifuge, a Beckman C515R centrifuge, and two Thermo Sorvall Primo centrifuges
- Balances (including analytical balance), pH meter, incubators, power supplies, protein and nucleic acid electrophoresis equipment including equipment for both semi-dry and wet transfer, heat blocks, stir plates, vortexes, orbital shakers, water baths, vacuum aspirators, Bunsen burners, light microscope, microwave oven, fume hood, and equipment for storing and handling radioactive isotopes including two handheld Geiger counters.

Biochemistry Department Equipment:

- Preparative centrifuges including ultracentrifuges
- SPEX SamplePrep 6870 Freezer/Mill
- Tetrad dissection setup
- Freeze-dryers
- Liquid scintillation counter
- Developer for Western blots and autoradiography
- Licor Odyssey IR imaging system
- NanoDrop for microvolume UV-Vis measurements
- Water bath and probe sonicators
- Dish washing/autoclaving facilities



RESEARCH & RELATED Senior/Key Person Profile (Expanded)

PROFILE - Project Director/Principal Investigator				
Prefix:	First Name*: JORDAN	Middle Name [REDACTED]	Last Name*: BERG	Suffix:
Position/Title*:	Graduate Research Assistant			
Organization Name*:	University of Utah			
Department:	Biochemistry			
Division:				
Street1*:	15 N Medical Dr E RM 4100			
Street2:				
City*:	Salt Lake City			
County:	Salt Lake			
State*:	UT: Utah			
Province:				
Country*:	USA: UNITED STATES			
Zip / Postal Code*:	84112-8930			
Phone Number*:	[REDACTED]	Fax Number:	E-Mail*:	[REDACTED]
Credential, e.g., agency login:	[REDACTED]			
Project Role*:	PD/PI			
Degree Type:	BS			
	Other Project Role Category: Degree Year: 2016			
Attach Biographical Sketch*:	File Name Biosketch_Berg1029235351.pdf			
Attach Current & Pending Support:				

PROFILE - Senior/Key Person				
Prefix:	First Name*: Jared	Middle Name P	Last Name*: Rutter	Suffix:
Position/Title*:	Professor			
Organization Name*:	University of Utah			
Department:	Biochemistry			
Division:				
Street1*:	15 N MEDICAL DR EAST RM 4100			
Street2:				
City*:	Salt Lake City			
County:	Salt Lake			
State*:	UT: Utah			
Province:				
Country*:	USA: UNITED STATES			
Zip / Postal Code*:	84112-5650			
Phone Number*:	[REDACTED]	Fax Number:	E-Mail*:	[REDACTED]
Credential, e.g., agency login:	[REDACTED]			
Project Role*:	Other (Specify)			
Degree Type:	PhD			
	Other Project Role Category: Sponsor Degree Year: 2001			
Attach Biographical Sketch*:	File Name Biosketch_Rutter1029204564.pdf			
Attach Current & Pending Support:				

PROFILE - Senior/Key Person				
Prefix:	First Name*: Bei	Middle Name	Last Name*: Wang-Phillips	Suffix:

Position/Title*:	Research Computer Scientist	
Organization Name*:	University of Utah	
Department:	SCIENT COMP & IMAG INSTIT-OPER	
Division:	SCIENTIFIC COMP & IMAGE INST	
Street1*:	72 S Central Campus Drive, RM 3750	
Street2:		
City*:	Salt Lake City	
County:	Salt Lake	
State*:	UT: Utah	
Province:		
Country*:	USA: UNITED STATES	
Zip / Postal Code*:	84112-9200	
Phone Number*:	[REDACTED]	Fax Number:
E-Mail*:	[REDACTED]	
Credential, e.g., agency login:	[REDACTED]	
Project Role*: Other (Specify)	Other Project Role Category: Co-Sponsor	
Degree Type: PhD	Degree Year: 2010	
	File Name	
Attach Biographical Sketch*:	Biosketch_Wang1029235234.pdf	
Attach Current & Pending Support:		

BIOGRAPHICAL SKETCH
DO NOT EXCEED FIVE PAGES.

NAME: Berg, Jordan A.

eRA COMMONS USER NAME (credential, e.g., agency login): JORDANBERG1

POSITION TITLE: Graduate Research Assistant, Ph.D Candidate

EDUCATION/TRAINING:

INSTITUTION AND LOCATION	DEGREE	Start Date	Completion Date	FIELD OF STUDY
Brigham Young University, Provo, UT	B.S.	06/2010	05/2016	Molecular Biology
University of Utah, Salt Lake City, UT	Ph.D.	08/2016	05/2022	Biochemistry

A. Personal Statement

While human metabolism has been elementally understood and appreciated since the time of Aristotle with the publication of “Parts of Animals”, and vigorously dissected over the past two centuries, a holistic and systematic understanding of metabolism, and particularly cancer metabolism, is incomplete. Often, cancer metabolism research takes a reductionist stance by perturbing a gene or protein in a pathway to determine its effect. But even in these scenarios, the systematic consequences are difficult to parse out. Interestingly, Drs. Ralph DeBerardinis and Navdeep Chandel (*Sci Adv*, 2016) recently published a review on cancer metabolism, where they, in their conclusions, pointed to a lack of computational methods for deconvoluting and analyzing the complex metabolic behaviors of tumor cell populations. In order to improve our understanding in this realm, my aims are two-fold. 1) Develop automated, robust tools for performing complex, system-wide interrogations of metabolism, and 2) Use these tools to address questions related to the regulation of metabolism that without such tools would be impossible to answer. With these tools, the pace of scientific discovery can modernize and accelerate, expanding the yield of our insights into the complex pathology of cancer. This information will then be used to develop novel therapeutic options that address the metabolic state of cancer.

I chose to work with Dr. Rutter due to his expertise in the mechanistic dissection of cancer metabolism and his enthusiasm to leverage computational methods to ask new and difficult questions. I chose Dr. Bei Wang Phillips as a co-mentor for her expertise in network biology, topographical data analysis, machine learning, and signal deconvolution. **With this dual-training strategy, I will be able to provide a much-needed mechanistic approach to computational dissection of cancer metabolism.** After my graduate training, I will find a position as a postdoctoral research fellow in a lab with expertise in network biology and machine learning to allow me to continue to expand this toolkit. My driving goal behind all of these career stages is to lead my own computational cancer metabolism lab where we will develop novel tools to deconvolute noisy and complex datasets and provide a more holistic view of cancer metabolism.

B. Positions and Honors

Positions and Employment

- 2014 iGEM Project Leader, 2014 iGEM Team, Brigham Young U.
2013-2016 Undergraduate Research Assistant, Julianne Grose Lab, Brigham Young U.
2014-2016 Teaching Assistant, HHMI SEAPHAGES Program, Brigham Young U.
2018, 2020 Teaching Assistant, Lit. & Problem Solving, U. of Utah
2016-present Graduate Student, Molecular Biology Program, U. of Utah
2017-present Graduate Research Assistant, Jared Rutter Lab, U. of Utah

Other Experience and Professional Memberships

- 2015 Member, American Society for Microbiology
2014-2016 Youth Mentor, Provo Youth Mentoring
2018 Lead Recruitment Host, Molecular Biology Graduate Program, U. of Utah
2018 Volunteer, Adventure Scientists

2018-present Member, SACNAS

2018-present SACNAS Webmaster/Social Media Outreach Officer

Reviewer for *Bioinformatics*, *NAR Genomics and Bioinformatics*, *Journal of Emerging Investigators*. See <https://bit.ly/2DLqcsk> for records.

Academic and Professional Honors

2014 iGEM World Jamboree, Silver Medal

2016 Outstanding Research Award, Dept. of Microbiology and Molecular Biology, Brigham Young U.

2018-2020 T32 Training Grant in Computational Approaches to Diabetes and Metabolism, U. of Utah

C. Contributions to Science

1. *Brevibacillus/Paenibacillus* Bacteriophage/Host Evolution (Undergraduate)

During my undergraduate degree, my research focus centered around the isolation, genomic characterization, and evolutionary analysis of bacteriophage infecting the bee probiotic *Brevibacillus laterosporus*, the bee pathogen *Paenibacillus larvae*, and the Rosaceae pathogen, *Erwinia amylovora*. I was the first author of two published peer-reviewed papers. I mentored 22 undergraduate research assistants and students, 16 of whom contributed as co-authors on published papers and/or primary authors on posters presented at external conferences. Additionally, I was the first author or a co-author for the annotation of 51 phage genome annotations published in NCBI's Genbank (see <https://bit.ly/2NWJ25Q> for a complete listing). I was selected to present a talk on my *Erwinia* phage research at 2015 Tri-Branche American Society for Microbiology (ASM) conference. In collaboration with the U.S. Environmental Protection Agency (EPA), I developed phage therapy cocktails for treating *Erwinia amylovora* infections. I was also a member of Brigham Young University's 2014 International Genetically Engineered Machine (iGEM) team and served as one of the project leaders.

Abstracts (selected list, 7 total):

1. **Berg JA**, Esplin IND, Brundage BM, Crockett JT, Esplin KP, Evans MR, Heaton KE, Hilton JA, Hyde JR, McBride MS, Schouten JT, Simister AR, Thurgood TL, Merrill BD, Ward AT, Breakwell DP, Burnett SH, Grose JH. (2015) Isolation of six novel *Paenibacillus larvae* bacteriophage and characterization of five *Brevibacillus laterosporus* bacteriophages to understand their evolutionary relationship to *Brevibacillus* and other bacteriophages. **Poster presentation**. 7th Annual HHMI SEA-PHAGES Symposium, Janelia Farms, Ashburn, VA, USA.
2. McBride M, Evans MR, Brundage BM, **Berg JA**, Merrill BD, Burnett SD, Breakwell DP, and Grose JH. (2015) Comparing Protein Structures of a Transcriptional Regulator Repeated in *Brevibacillus* Phages. **Poster presentation**. Tri-branch ASM meeting, Colorado State University, Fort Collins, CO, USA. Third place, **Best poster presentation**.

Oral Presentations (1 total):

1. **Berg JA**, Simister, AR, Thurgood TL, Grose JH. (2015) Characterization and analysis of six novel *Erwinia* phages reveals relationship to Enterobacteriaceae family members. **Oral presentation**. Tribranch ASM meeting, Colorado State University, Fort Collins, CO, USA.

Research Articles (selected list, 6 total):

1. **Berg JA**, Merrill BD, Breakwell DP, Grose JH, Hope S. A PCR-based method for distinguishing between two common beehive bacteria, *Paenibacillus larvae* and *Brevibacillus laterosporus*. (2018) ***Journal of Appl Environ Microbiol***. 84:e01886-18. DOI: 10.1128/AEM.01886-18.
2. Sharma R, **Berg JA**, Beatty NJ, Choi MC, Cowger AE, Duncan SG, Fajardo C, Ferguson HP, Galbraith T, Herring JA, Hoj TR, Hughes J, Hyde JR, Jensen GL, Ke K, Keele BR, Killpack S, Lawrence EEK, Nwosu I, Roark BJ, Thompson DW, Tueller JA, Ward MEH, Webb CJ, Wood ME, Wynne H, Yeates EL, Baltrus D, Breakwell DP, Hope S, Grose JH. Genome Sequences of 10 *Erwinia amylovora* Bacteriophages. (2018) ***Microbiol Resour Announc***. 7(14):e00944-18. DOI: 10.1128/MRA.00944-18.
3. Esplin IND, **Berg JA**, Sharma R, et. al. Genome sequences of 19 novel *Erwinia amylovora* bacteriophages. (2017) ***Genome Announc***. 5(46):e00931-17. DOI: 10.1128/genomeA.00931-17.
4. **Berg JA**, Merrill BD, Crockett JT, Esplin KP, Evans MR, Heaton KE, Hilton JA, Hyde JR, McBride MS, Schouten JT, Simister AR, Thurgood TL, Ward AT, Breakwell DP, Burnett SH, Grose JH.

Characterization of five novel *Brevibacillus* bacteriophages and genomic comparison of *Brevibacillus* phages. (2016) *PLoS ONE* 11(6): e0156838. DOI: 10.1371/journal.pone.0156838.

2. Computational Approaches to Deconvoluting and Understanding Metabolism (Graduate)

I have worked in a variety of labs from different fields during my rotations, but the two that were most impactful to me and my career trajectory were with Drs. Jason Gertz and Jared Rutter. During my rotation with Dr. Gertz, I developed analytical and predictive computational methods to identify patients with endometrial cancer that would respond positively to progesterone treatment. My rotation with Dr. Rutter distilled in me an amazement and excitement of metabolism and its complexity. While I still collaborate closely and often with Dr. Gertz, I like complexity and being able to make sense of complex systems, which is why I decided to do my thesis in Dr. Rutter's lab. Understanding these systems require large amounts of biological data that has been processed uniformly. To enable these scaled analyses, I developed *XPRESSyourself*, which was a collaborative project with Drs. Aaron Quinlan and Jason Gertz, as well as members of the Rutter and Quinlan labs. In this toolkit, we provide standardized and automated sequence data processing and introduced new open source tools for ribosome profiling analysis, where such tools and standardization were still missing.

With Drs. Bei Wang, James Cox, and others, we are building upon this first toolkit to create *Metaboverse*, which will standardize and enhance proteomics and metabolomics data analysis, supplemented with RNA-Seq and ribosome profiling data. We also recently completed two manuscripts where we investigated the topographical structure of the metabolic network, presented key considerations for analyzing and visualizing metabolic networks (*GigaScience, In Press*), and identified systematic changes to glucose metabolism in colorectal cancer that drives tumorigenesis (*Cell Metabolism, 2019*). These observations will serve as a foundation for visualizing and contextualizing complex metabolism and its regulation in *Metaboverse*.

Abstracts (3 total):

1. Berg JA, Belyeu JR, Morgan JT, Ouyang Y, Bott AJ, Quinlan AR, Gertz J, Rutter J. (2019) *XPRESSyourself—Enhancing, standardizing, and automating ribosome profiling computational analyses yields improved insight into data. Poster presentation.* Genome Informatics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA.
2. Berg JA, Nuebel E, Rutter JP. (2018) *Ribosome profiling reveals translation-level regulation of peroxins in response to loss of peroxisomes. Poster Presentation.* Frontiers in Metabolism, Madison, WI, USA.
3. Berg, JA, Rutter JP. (2018) *RiboPipe: A ribosome profiling pipeline optimized for easy assembly and data analysis. Poster Presentation.* RECOMB2018, Paris, FR.

Oral Presentations (3 total):

1. Berg JA, Zhou Y, Waller TC, George I, Cox J, Wang B, Rutter J. (2019) *Contextualizing chaotic metabolic networks and their regulation. Selected Talk.* EMBL/EMBO Epigenetics Meets Metabolism, Heidelberg, DE.
2. Berg JA, Belyeu JR, Morgan JT, Ouyang Y, Bott AJ, Quinlan AR, Gertz J, Rutter J. (2019) *XPRESSyourself—Enhancing, standardizing, and automating ribosome profiling computational analyses yields improved insight into data. Lightning Talk.* Genome Informatics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA.
3. Berg JA. (2019) *XPRESSyourself: Enhancing, Standardizing, and Automating Ribosome Profiling Computational Analyses Yields Improved Insights. Invited Talk.* Bioinformatics Seminar Series, Center for Computational Biology & Bioinformatics, University of California San Diego, San Diego, CA, USA.

Research Articles (*co-corresponding authors; *equal contributors; selected list, 5 total):

1. Berg JA*, Belyeu JR, Morgan JT, Ouyang Y, Bott AJ, Quinlan AR, Gertz J, Rutter J*. *XPRESSyourself: Enhancing, Standardizing, and Automating Ribosome Profiling Computational Analyses Yields Improved Insight into Data.* (2019) *bioRxiv*. 704320. DOI: 10.1101/704320. Under review.
2. Hughes CE, Coody TK, Jeong M, Berg JA, Winge DR, Hughes AL. *Amino acid toxicity drives age-related mitochondrial decline by altering iron metabolism. In Press at Cell.*
3. Bensard CL*, Wisidigama DR*, Olsen KA, Berg JA, Krah NM, Schell JC, Nowinski SM, Fogarty S, Bott AJ, Wei P, Dove KK, Tanner JM, Panic V, Cluntun A, Lettlova S, Earl CS, Namnath DF, Vázquez-Arreguin K, Villanueva CJ, Tantin D, Murtaugh LC, Evasion KJ, Ducker GS, Thummel CS, Rutter J. *Regulation of Tumor Initiation by the Mitochondrial Pyruvate Carrier.* (2019) *Cell Metabolism*.

4. Waller TC[&], Berg JA, Lex A, Chapman BE, Rutter J[&]. Compartment and Hub Definitions Tune Metabolic Networks for Metabolic Interpretations. In Press at *GigaScience*.

Complete list of publications is available on Google Scholar: <https://bit.ly/33Wu2KV>

D. Scholastic Performance

Brigham Young University

YEAR	COURSE TITLE	GRADE
2010	Cultural History of Medicinal Plants	B+
2013	Calculus 1	B
2013	Science of Biology	A
2013	Psychological Science	B+
2013	Careers in Biomedical Sciences	A
2013	Freshman Research 1	A
2013	General College Chemistry 1	A-
2013	Mentored Research	A
2014	Freshman Research 2	A
2014	General College Chemistry 2	B
2014	General Physics 1	A-
2014	General Physics Lab 1	A
2014	Mentored Research	A
2014	Synthetic Biology (2 semesters)	A
2014	General Physics 2	A-
2014	Molecular & Cellular Biology Lab	A
2014	General College Chemistry Lab	C+
2014	Molecular Biology	B+
2014	Introduction to Computer Programming	A
2014	Readings in Molecular Biology	A
2014	Genetics	B-
2015	Evolutionary Biology	B
2015	Organic Chemistry 1	C-
2015	Infection & Immunity	B-
2015	Cell Biology	C+
2015	Statistics for Engineers & Scientists	B
2015	Organic Chemistry 2	E
2015	Genomics	C+
2015	Molecular Biology Seminar	P
2016	Biochemical Engineering	B+
2016	Advanced Molecular Biology	B-
2016	Advanced Molecular Biology Lab	A
2010	Greek & Roman Mythology	A-
2010	Introduction to Music	A
2010	The New Testament	B+
2010	American Heritage	B
2010	Social/Cultural Anthropology	A
2010	First Year Swahili	A
2010	The Book of Mormon 1	A-
2013	JS & Restoration 1805-1845	A
2013	Civilization: Literature 2	A-
2013	Psychological Science	B+

YEAR	COURSE TITLE	GRADE
2013	The Book of Mormon 2	A
2013	3rd-Yr Spanish Reading, Grammar, & Culture	B
2013	Civilization: Music 1	B
2014	Survey of Eastern Religions	A
2015	Technical Communication	B+
2015	The Doctrine & Covenants	B-
2016	Beginning Mandarin	E

Cumulative Undergraduate GPA=3.23

*Undergraduate grades based on an A-E scale, E equaling a failing grade; or reported as Pass (P), Fail (F)

University of Utah

YEAR	COURSE TITLE	GRADE
2016	Faculty Research Seminar	P
2016	Protein & Nucleic Acids Biochemistry	B+
2016	Genes, Genomes, and Gene Expression	B
2016	Cell Biology	A-
2016	Research Ethics	P
2016	Research (Rotations)	P
2017	Literature & Problem Solving	P
2017	Grant Preparation	P
2017	Research (Rotations)	P
2017	Developmental Neurobiology	C+
2017	Biochemistry Journal Club	P
2017	Biochemistry Research in Progress Seminar	P
2017	Lab Research Conferences	P
2017	Thesis Research	A
2018	Biochemistry Journal Club	P
2018	Biochemistry Research in Progress Seminar	P
2018	Lab Research Conferences	P
2018	Thesis Research	A
2018	Statistics for Biomedical Informatics	A
2019	Thesis Research	A
2019	Biochemistry Research in Progress Seminar	P
2019	Lab Research Conferences	P

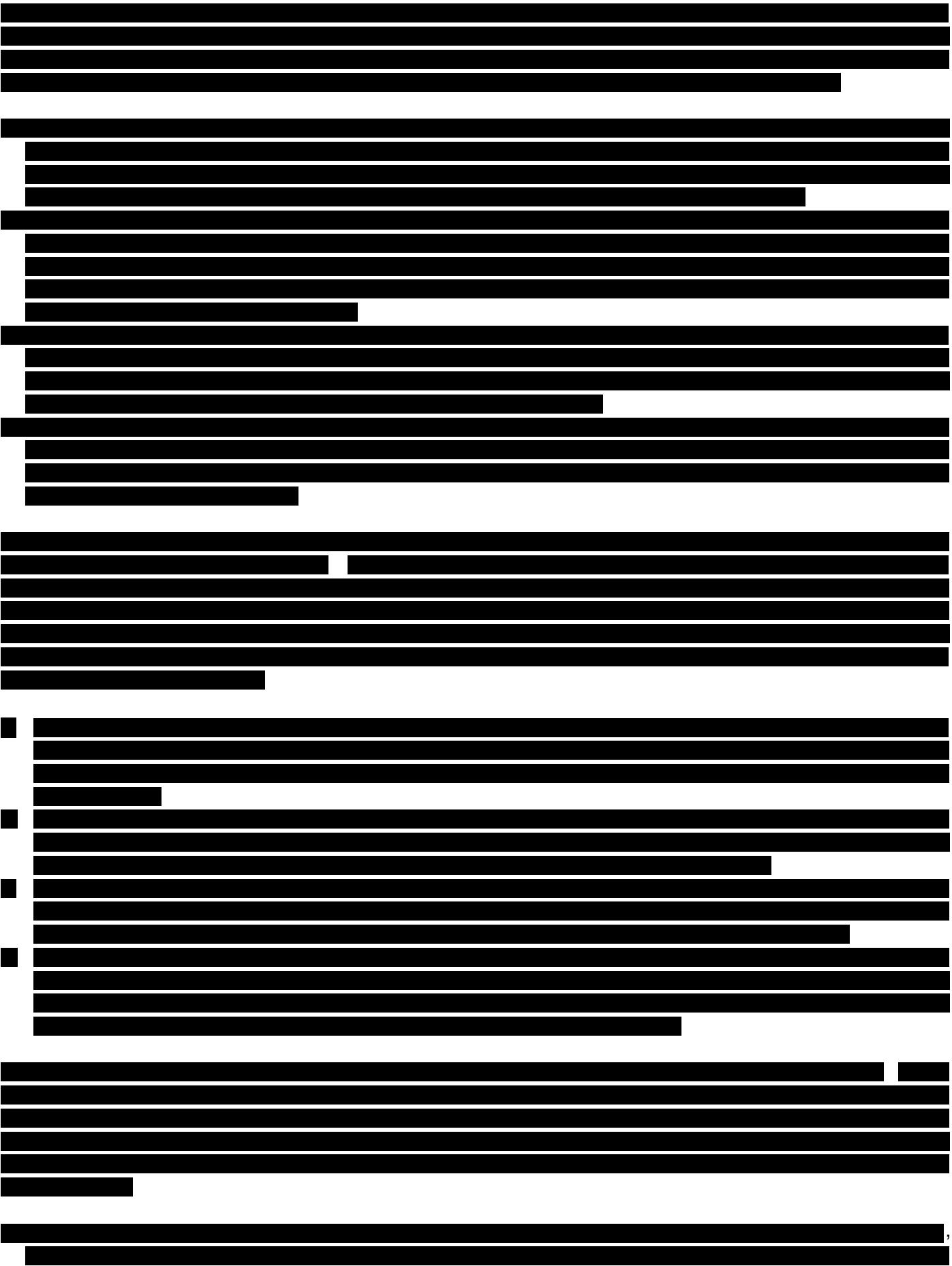
Cumulative Graduate GPA=3.807

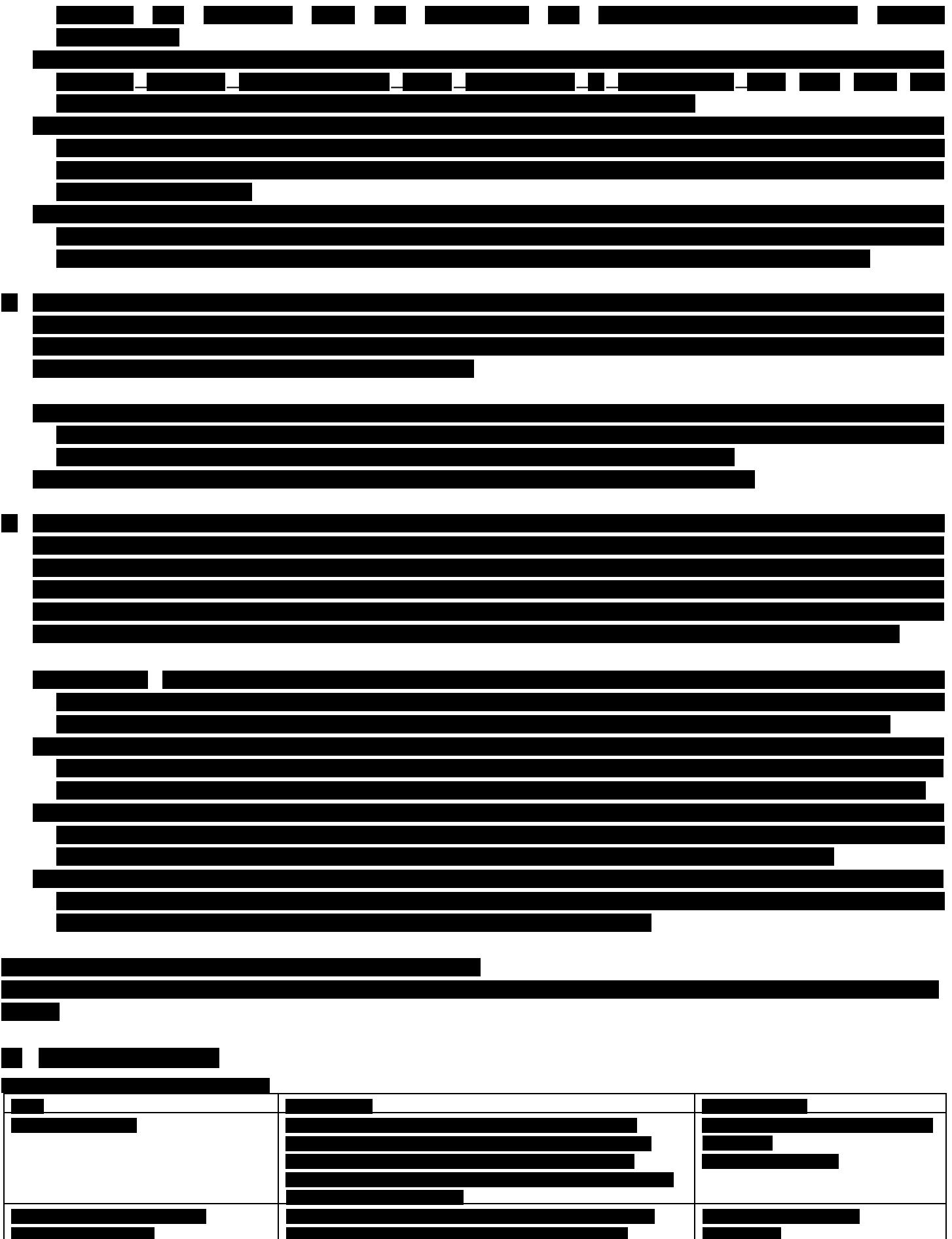
*Graduate grades based on an A-D, F scale or reported as Pass (P), Fail (F)

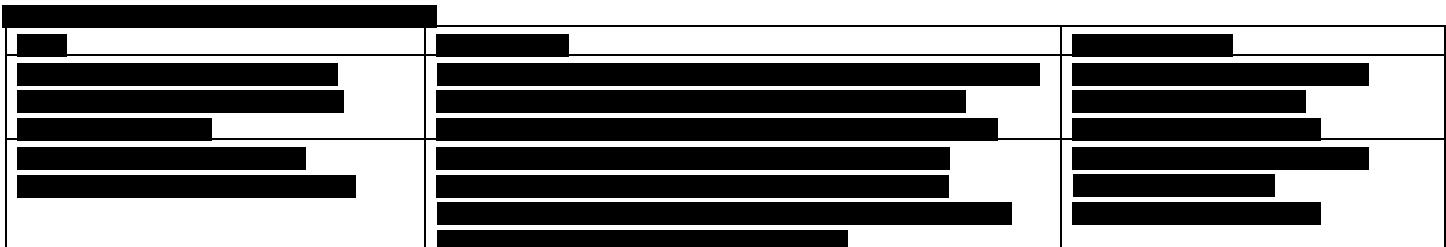
**BIOGRAPHICAL SKETCH
DO NOT EXCEED FIVE PAGES.**

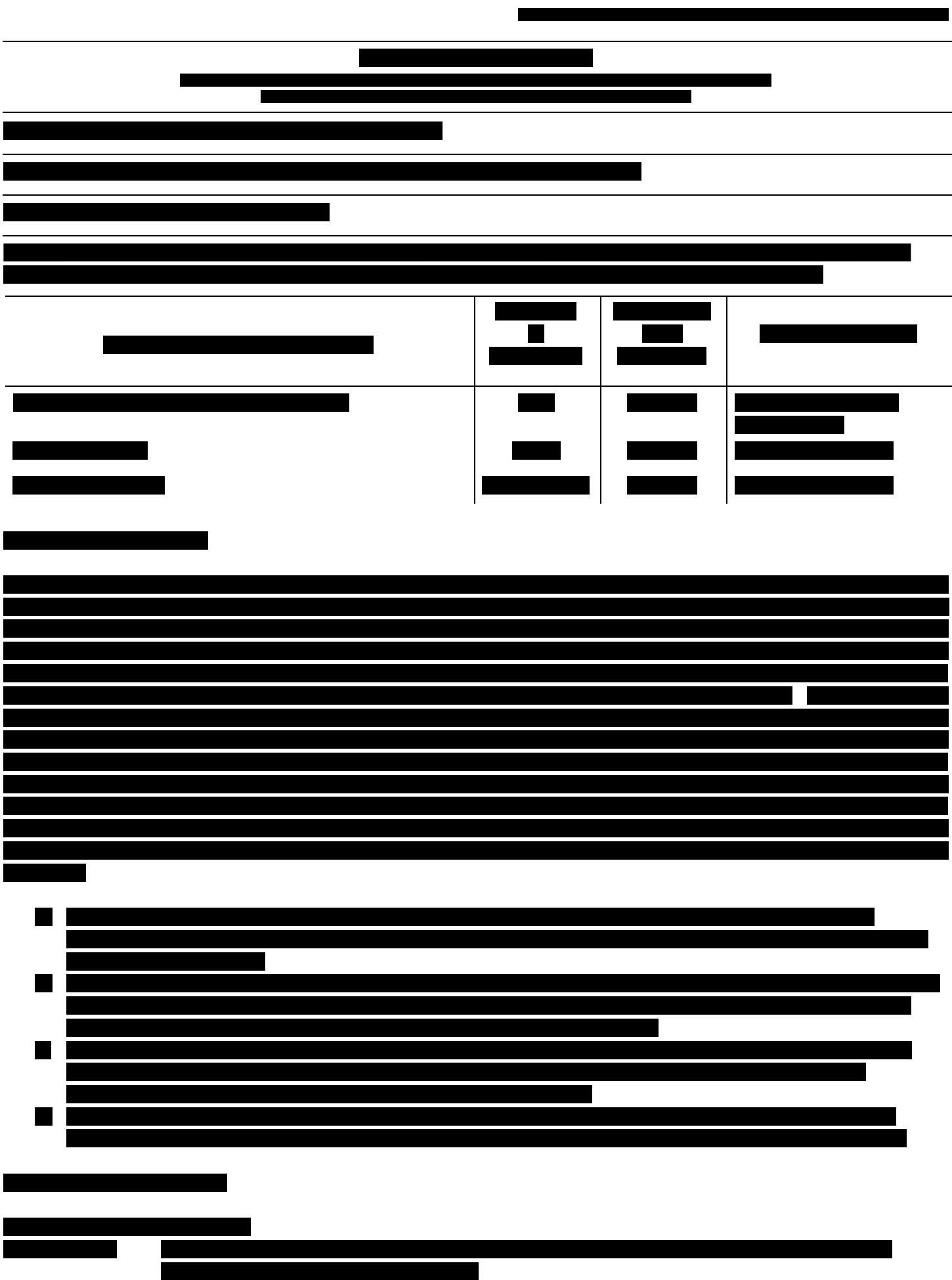
The page contains several horizontal lines of blacked-out text, indicating a redacted biographical sketch. The text is organized into approximately 15 distinct horizontal rows, with some rows being significantly longer than others. There are also a few vertical black bars and a small cluster of short horizontal bars in the upper right quadrant, likely representing redacted contact information or a signature.



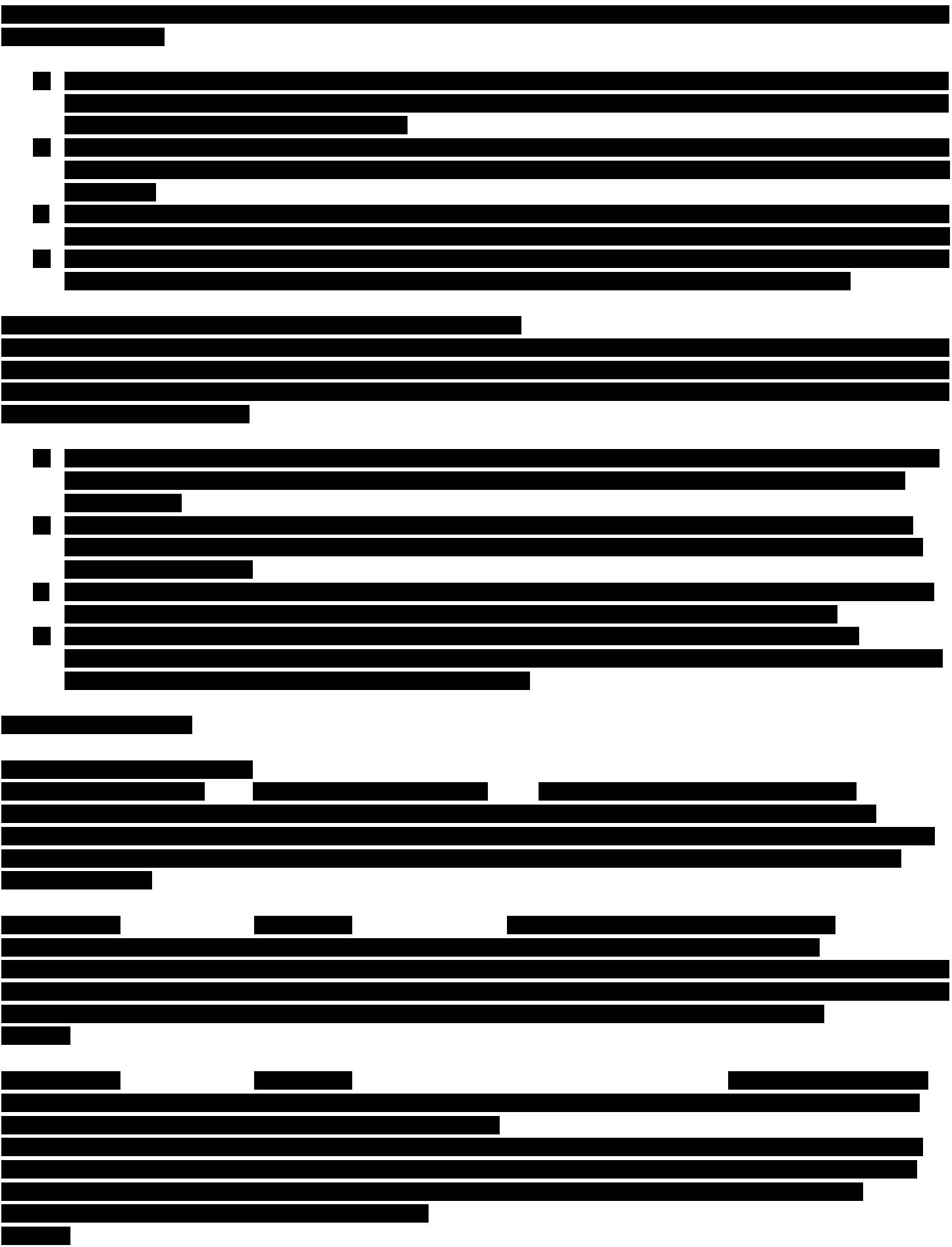


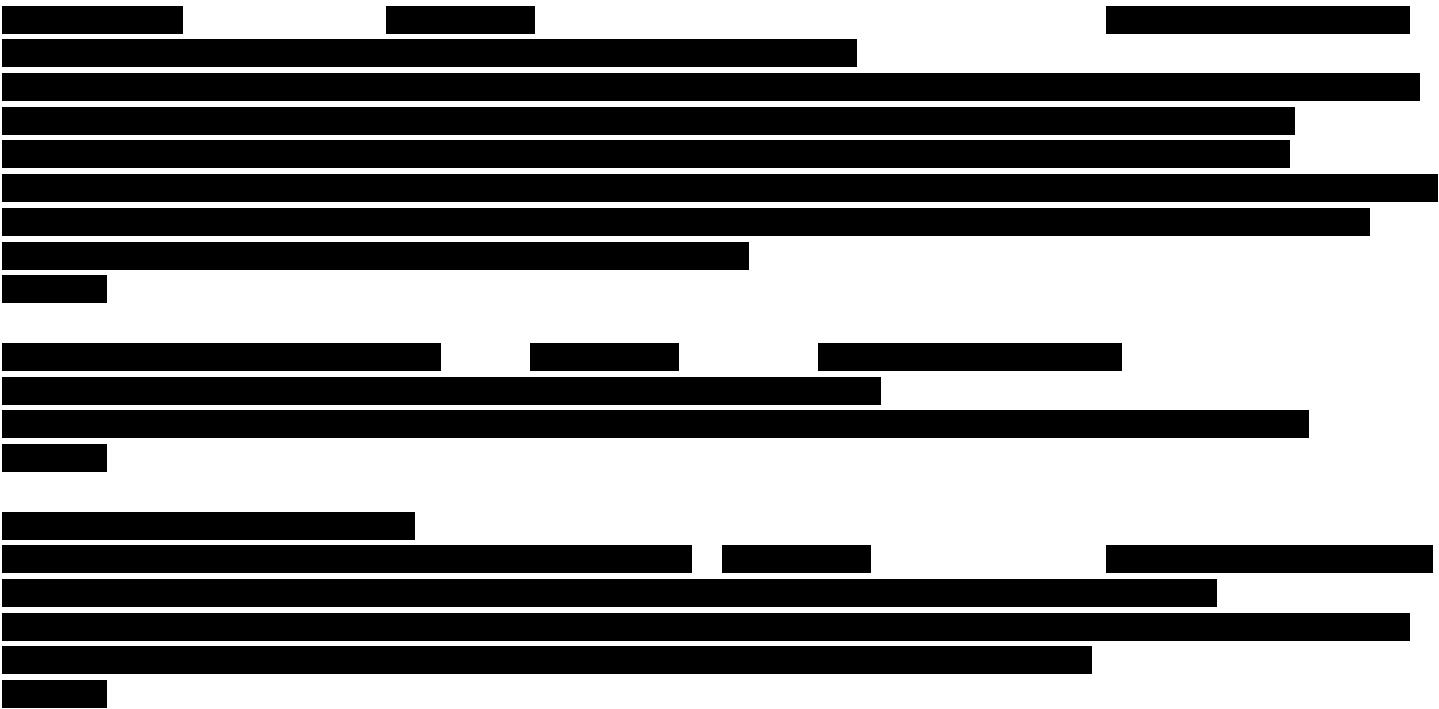












PHS Fellowship Supplemental Form

OMB Number: 0925-0001
Expiration Date: 03/31/2020

Introduction

1. Introduction

(for Resubmission applications)

Fellowship Applicant Section

2. * Applicant's Background and Goals for Fellowship Training

Applicants_Background_Goals1029235350.pdf

Research Training Plan Section

3. * Specific Aims

Specific_Aims1029235347.pdf

4. * Research Strategy

Research_Strategy1029235348.pdf

5. * Respective Contributions

Respective_Contributions1029235289.pdf

6. * Selection of Sponsor and Institution

Selection_Sponsor_Institution1029235295.pdf

7. Progress Report Publication List

(for Renewal applications)

8. * Training in the Responsible Conduct of Research

Training_Responsible_Conduct1029204555.pdf

Sponsor(s), Collaborator(s) and Consultant(s) Section

9. Sponsor and Co-Sponsor Statements

Sponsors_Statement1029235297.pdf

10. Letters of Support from Collaborators, Contributors and Consultants

Letter_Support1029235352.pdf

Institutional Environment and Commitment to Training Section

11. Description of Institutional Environment and Commitment to Training

Institutional_Environment_Training1029235287.pdf

Other Research Training Plan Section

Vertebrate Animals

The following item is taken from the Research & Related Other Project Information form and repeated here for your reference. Any change to this item must be made on the Research & Related Other Project Information form.

Are Vertebrate Animals Used? Yes No

12. Are vertebrate animals euthanized?

If "Yes" to euthanasia

Is method consistent with American Veterinary Medical Association (AVMA) guidelines?

If "No" to AVMA guidelines, describe method and provide scientific justification

13. Vertebrate Animals

PHS Fellowship Supplemental Form

OMB Number: 0925-0001
Expiration Date: 03/31/2020

Other Research Training Plan Information

14. Select Agent Research

15. Resource Sharing Plan

Resource_Sharing_Plan1029235293.pdf

16. Authentication of Key Biological and/or Chemical Resources

Additional Information Section

17. Human Embryonic Stem Cells

Does the proposed project involve human embryonic stem cells?* Yes No

If the proposed project involves human embryonic stem cells, list below the registration number of the specific cell line(s) from the following list: <http://stemcells.nih.gov/research/registry/>. Or, if a specific stem cell line cannot be referenced at this time, please check the box indicating that one from the registry will be used:

Specific stem cell line cannot be referenced at this time. One from the registry will be used.

Cell Line(s):

18. Alternate Phone Number: [REDACTED]

19. Degree Sought During Proposed Award:

Degree:

If "other", indicate degree type:

Expected Completion Date (MM/YYYY):

PHD: Doctor of Philosophy

06/2022

20. * Field of Training for Current Proposal:

104 Computational Biology

21. * Current Or Prior Kirschstein-NRSA Support?

Yes No

If yes, please identify current and prior Kirschstein-NRSA support below:

Level*	Type*	Start Date (if known)	End Date (if known)	Grant Number (if known)
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]

22. * Applications for Concurrent Support?

Yes No

If yes, describe in an attached file:

23. * Citizenship

U.S. Citizen U.S. Citizen or Non-Citizen National?

Yes No

Non-U.S. Citizen

With a Permanent U.S. Resident Visa

With a Temporary U.S. Visa

If you are a non-U.S. citizen with a temporary visa applying for an award that requires permanent residency status, and expect to be granted a permanent resident visa by the start date of the award, check here:

Name of Former Institution:*

24. Change of Sponsoring Institution

PHS Fellowship Supplemental Form

OMB Number: 0925-0001
Expiration Date: 03/31/2020

Budget Section

All Fellowship Applicants:

25. * Tuition and Fees:

None Requested Funds Requested

Year 1

[REDACTED]

Year 2

[REDACTED]

Year 3

Year 4

Year 5

Year 6 (when applicable)

Total Funds Requested:

[REDACTED]

Senior Fellowship Applicants Only:

26. Present Institutional Base Salary:	Amount	Academic Period	Number of Months
--	--------	-----------------	------------------

27. Stipends/Salary During First Year of Proposed Fellowship:

a. Federal Stipend Requested: Amount Number of Months

b. Supplementation from other sources: Amount Number of Months

Type (e.g., sabbatical leave, salary)

Source

Appendix

28. Appendix

Applicant's Background and Goals for Fellowship

A. Doctoral Dissertation and Other Research

My ultimate career goal is to become an independent investigator and a tenured professor. I envision the research program of my future lab to be a harmonious synthesis of computational metabolism and software development, enriched by collaborations with wet lab researchers and clinicians. In a recent review article by Drs. DeBerardinis and Chandel (*Sci Adv*, 2016), these leading scientists point to three primary bottlenecks to unlocking the keys of cancer metabolism for therapeutic advancement. One of the highlighted challenges was the lack of analytical and computational tools and approaches to be able to deconvolute the complex metabolic phenotypes that exist between cell populations within a tumor. My research plan is to develop the necessary tools and algorithms that would enable a more holistic understanding of the underpinnings of cancer metabolism and would eventually lead to more effective and personalized treatment options. I have already had success achieving the requisite milestones. Specifically, I have contributed significantly to collaborative projects as a co-author on manuscripts that interrogate these complex metabolic landscapes and help form the foundation of my proposed F99 phase research plan (***Mol Cell*, 2018; *Cell Metabolism*, 2019; *GigaScience*, In Press; *Cell*, In press**), and I have prepared a first-author manuscript currently under its second round of review (***bioRxiv*, 2019**) that presents a tool for the uniform and automated processing and analysis of sequence data and a validation of its performance in both cancer and neurological contexts. In this package, I also introduced vital open-source tools currently missing from the computational toolkit to handle the inherent biases in ribosome profiling data. This will be followed by my final dissertation publications wherein we will present *Metaboverse*, and where we will deconvolute and validate complex, dispersed regulatory events and signals in cancer metabolism. To date, I have presented four posters at local conferences, three posters at national and international conferences, one selected talk at a local conference, one short talk at a national conference, one selected talk at an international conference, and one invited seminar talk at the University of California San Diego. **My background and skills, augmented with the additional training outlined in this proposal, will enable me to become a leader at the intersection of cancer metabolism, computational metabolism, network biology, and machine learning.** I began preparing for these career goals early and my research experience is summarized below:

Fall 2013 - Summer 2016 (undergraduate) – *Undergraduate Research Assistant (Advisor: Julianne Grose, Ph.D)* – At the start of my second year during my undergraduate education, I began research with Dr. Julianne Grose. There I learned basic laboratory procedures and isolated novel bacteriophage that infect *Erwinia amylovora*, a fruit tree pathogen. I learned how to prepare genome sequencing libraries for this and other phages and then annotated and characterized their genomes. During my time in the Grose lab, **I published a first author paper in PLoS ONE** in collaboration with Dr. Sandra Hope describing the genomic characteristics and evolution of several *Brevibacillus* phages. Additionally, **I published another first author paper in the Journal of Applied and Environmental Microbiology** that introduces a combined computational-wet-lab methodology for distinguishing environmental microbial strains from other, closely related strains. I was also a co-author on 4 other publications related to this research. In total, I was the **first author or co-author on 51 phage genome publications** in NCBI's GenBank. While I was an undergraduate research assistant, **I mentored 22 undergraduate researchers, 16 of which were co-authors on publications or posters.** I helped prepare and present 7 posters at scientific conferences and was **selected to give a talk at the 2015 Tri-Branch American Society for Microbiology conference.**

Spring 2014 - Fall 2014 (undergraduate) – *iGEM Team Project Leader (Advisor: Julianne Grose, Ph.D)* – I was a member of the 2014 BYU International Genetically Engineered Machine (iGEM) team. For this project, we synthetically redesigned a common wastewater bacterium, *Nitrosospira multiformis*, to more efficiently break down sewage biofilm. As a part of this project, we talked with local wastewater reclamation workers to discuss key components in the design of this synthetic organism to improve the efficiency of the bioreactor and help address some of the biggest challenges in wastewater reclamation. I led the project improving the biofilm degradation ability of this organism and the design and creation of our team's website. At the end of this project, we traveled to the 2014 iGEM Giant Jamboree in Boston to present our work via an oral team presentation.

Summer 2016 (graduate) – *Rotation Project (Advisor: Daniel Leung, M.D., M.Sc.)* – Mucosal associated invariant T (MAIT) cells are a relatively newer characterized sub-class of T cells. For my rotation project, I

exposed T cells and MAIT cells to a panel of different cytokines and antibodies in an effort to induce differentiation patterns in these cells and explore immunotherapy options. I also performed single-cell PCR on patient samples to analyze cytokine profiles and classify differential signaling of MAIT cells in these patients. I used this data to develop a computational pipeline for statistical analysis of these samples, all in an effort to define the potential subtypes of this recently discovered T cell subclass. From this project, I identified several putative sub-typing factors, such as expression levels of the *TBET* and *PD-1* genes, which were differentially expressed in the various MAIT cell sub-types.

Fall 2016 (graduate) – *Rotation Project (Advisor: Jay Gertz, Ph.D)* – In endometrial cancers, some patients will respond positively to progesterone treatment alone, allowing them to avoid more invasive procedures. I analyzed endometrial cancer patient RNA-Seq data from The Cancer Genome Atlas (TCGA) to predict response to progesterone treatment based on genetic and phenotypic information. Within these groups, I identified putative genetic markers indicating susceptibility or resistance to progesterone treatment to help patients make informed decisions regarding treatment. I also developed a computational pipeline to identify these markers and showed statistical correlation between certain genes' expression levels and greater survivability in the patient cohort.

Fall 2016 (graduate) – *Rotation Project (Advisor: Jared Rutter, Ph.D)* – During my rotation with Dr. Jared Rutter, my project focused on defining the role of different proteins identified as potential actors in peroxisomal disorders. In particular, I was interested in proteins that presented a mitochondrial phenotype upon perturbation, as the role of these factors are less characterized in peroxisomal disorders. During this project, I generated several yeast strains and genetic tools that aided in our determination of these candidate proteins' role in a model of Peroxisomal Biogenesis Disorder when challenged with the deletion of the mitochondrial quality control factor, *MSP1/ATAD1*.

Spring 2017 (graduate) – *Rotation Project (Advisor: Adam Douglass, Ph.D)* – During my rotation with Dr. Adam Douglass, I worked to establish whether Pavlovian conditioning could be established in larval zebrafish. This paradigm would be consequential for neuron-mapping as larval zebrafish are transparent and one could thus map the activation of neurons during stimuli and conditioning. In particular, we were interested in using this model to map the oxytocinergic neuronal response to pain and other stimuli by fluorescent imaging of the response-triggered neurons. I was able to show that a mild conditioning response could be established in this model and began to identify modifications in the conditioning methodology that needed to be made to improve the learning of the larval zebrafish. As part of this rotation project, I also automated the image processing of these experiments to streamline the recording and analysis of these conditioning experiments.

Summer 2017 – Present (graduate) – *Ph.D Dissertation Research (Advisor: Jared Rutter, Ph.D)* – I joined Dr. Rutter's lab for a variety of reasons – his emphasis on independence and student-driven projects, the freedom he would give me to blend my dual-training in both computational and wet lab biology, and his expertise in cancer metabolism. An early collaboration (*Cell Metabolism*, 2019) made me start thinking of the challenges of contextualizing the complex cancer metabolism network. In this manuscript, we described the consequences the loss of a mitochondrial metabolite transporter, initially characterized in our lab, on tumorigenesis. In our models I began to notice interesting and complex metabolic patterns between the different stages of tumorigenesis and realized there was a void in available computational tools required to dig deeper into these questions. This project also intertwined nicely with the first phase of my dissertation research, which involved building a software package that standardized, enhanced, and automated ribosome profiling and RNA-Seq computational analysis. Early in my dissertation, I analyzed several ribosome profiling datasets and realized that, unlike RNA-seq, the standards for computational analysis of ribosome profiling data had not been established yet and a culture of analysis transparency (open source scripts) has been slow to be adopted by the community. Working with Dr. Jeffrey Morgan, an expert in ribosome profiling; Drs. Aaron Quinlan and Jay Gertz, experts in RNA-Seq and computational tool design; and other collaborators, **we built XPRESSyourself to meet these needs. The paper describing this tool is currently available as a pre-print (DOI: 10.1101/704320) and is under a second round of peer-review at PLoS Computational Biology. This paper, for which I was a corresponding author, has also helped establish me in the computational biology field and has led to reputable journals, such as Bioinformatics, reaching out to me to review papers.** These projects, and another project (*GigaScience*, In Press) will feed directly into my next project, *Metaboverse* ([github.com/Metaboverse](https://github.com/Metaboverse/metaboverse.github.com/demo); metaboverse.github.com/demo; **Aim 1.1-3**), where I will standardize and improve the methods available for the integrated analysis of sequence, proteomics, and metabolomics data.

Members of the metabolism community often struggle to make sense of their data, especially as current trends in analysis tend to focus on a list of differentially expressed entities. Understanding how these components fit into the larger network is difficult, and contextualizing these data becomes an extremely laborious and tedious task. Having assembled a team containing experts in metabolism, metabolomics, topological data analysis, network biology, and cancer biology, we are building this tool and the appropriate analytical methods to aid a user in contextualizing their data. The tool will additionally allow for the multi-omic integration of sequencing data, or any -omic for which features can be mapped to gene identifiers. Based on feedback from presentations I have given, this tool is in high demand and we are set to have a strong user base once the beta-version is released in early to mid 2020. Once this tool is finalized and published in a peer-review journal, I will focus my remaining F99 phase efforts on phase 3 of my thesis (**Aim 1.4**), in which I will use these tools developed over the past couple of years to answer new, complex, and exciting problems in metabolic regulation. It is important to note that once we have established a sufficient user base for the tools we are creating, Dr. Rutter and I will begin applying for funding support to continue to maintain and improve these tools. I will also continue to provide support for this tool and *XPRESSyourself*, regardless of funding throughout the duration of my F99 and K00 phases as they will be central in the work I will propose. As discussed above, I have presented my research at multiple venues during my graduate career, but I will highlight a selected talk on *Metaboverse* at the EMBO/EMBL Metabolism Meets Epigenetics meeting in Heidelberg, Germany, and an invited talk on *XPRESSyourself* hosted by the Center for Computational Biology & Bioinformatics at the University of California San Diego.

Aside from my research experience, several characteristics uniquely position me for excellence in computational metabolism, which I discuss in some detail below:

- **Dual Training:** I come from both computational and biology backgrounds. I completed my undergraduate degree in Molecular Biology and supplemented this training with computer science and genomics training. During my graduate career, I have focused on computational approaches, but continue to receive wet lab training and have performed my own high-throughput experiments. This background is supplemented by Dr. Rutter, who pushes me to think more mechanistically about the biological questions I pursue so I can offer unique perspectives to cancer and metabolism.
- **Leadership:** I have naturally fallen into leadership roles. During my undergraduate career, I was a project leader for our University's iGEM team and led the team wiki page design and execution. During my undergraduate training, I mentored 22 undergraduate students over three years, in addition to my service mentoring elementary school children. During my graduate career, I have continued to assume leadership roles. For example, I have assembled two separate teams with the necessary expertise for the *XPRESSyourself* and *Metaboverse* projects. I serve as a SACNAS officer and have mentored two undergraduate students.
- **Collaboration:** The teams that I have assembled have been valuable collaborators, enabling success during my research career. In addition to intra-lab collaborations, I have independently established collaborative relationships with other labs. I also attend conferences with the intention of establishing new collaborations. For example, from my interactions with others during the 2019 Genome Informatics conference, I established a connection with one of the lead PIs of the Galaxy project who requested that I work with him to integrate the *XPRESSyourself* tool into the Galaxy webserver.
- **Endurance:** I am able to persist, regardless of the barriers and difficulties I face. This was largely instilled to me during my undergraduate and graduate hobbyist running career. After running multiple marathons, each time shaving tens of minutes off of my previous personal records through discipline and training, I eventually ran a 50-mile race in the harsh southern Utah desert. While unfortunately I did not complete this race, I ran until I could hardly walk, blistered and beaten, and only stopped when I missed my final cut-off time by minutes. I have seen this same tenacity pay dividends during my research career, allowing me to publish two first-author papers on my undergraduate research, and leading the team that published my first first-author paper during my graduate career, involving weekend, 20 hours/day coding sprees to complete the project when we discovered we had a competitor. When a challenge is in front of me, I have a laser-like focus and am able to push forward. An essential part of learning endurance has been learning balance as well, where, perhaps unlike my running career where I over-trained for the 50-mile race and ran those 40 miles through the desert on a torn ligament, I have been able to find an optimum performance level in my research that allows me to drive my research forward while mitigating burn-out.

B. Training Goals and Objectives

There is great strength in the co-application of wet lab research and computational biology. While the integration of these two fields is improving, a disconnect remains and limited computational tools for handling complex datasets continues to stifle advancement. My overarching goal is to provide a bridge between these fields and further close the gap between cancer metabolism and computational biology. Completion of my goals outlined in this proposal will provide me with the training I need to provide this bridge. I have chosen a mentor, Dr. Rutter, and other collaborators, who all have expertise in metabolism, cancer biology, computational biology, network analysis, and machine learning. Both my wet lab and computational background has been a benefit to my research. I have chosen mentors with expertise across scientific domains to allow me to answer difficult and complex questions in cancer metabolism, and to approach cancer metabolism mechanistically through systems-wide approaches. The University of Utah has incredible statistical genomics and computer science communities. Members of these communities, such as my collaborators from the *XPRESSyourself* project, Dr. Aaron Quinlan, and current or continuing collaborators such as Drs. Bei Wang Phillips (see sponsor/co-sponsor statement), Jason Gertz (see letter of recommendation), and James Cox (see letter of support), have been and will be essential collaborators to ensure the research I perform is robust and high-quality. I meet with Dr. Bei Wang Phillips bi-weekly as able and with my other core collaborators at least monthly to discuss progress and generate ideas for future project directions.

I am extremely fortunate to work in Dr. Rutter's lab, where I am given a great amount of trust and independence with my project and its direction. I meet with Dr. Rutter one-on-one once a week to discuss ideas and experimental plans, and also frequently in person or via email other times throughout the week. Dr. Rutter has always provided prompt feedback that has been invaluable as I develop my projects. Dr. Rutter has truly become a mentor in every sense of the term, and someone I have learned from immeasurably. I also meet with the Rutter lab Big Data group every week to discuss the ideas proposed in this application and their progress. I am also able to provide feedback for my peers in this venue. I present my research often in lab meeting to get feedback on the direction of my projects. These meetings have been consequential to the development of my ideas as the members of the lab come from diverse backgrounds and have valuable perspectives. Additionally, I have cultivated a network of computational biologists at the University of Utah and often reach out to them with questions. The University of Utah has a strong collaborative spirit, therefore when my established network is unable to provide the necessary guidance, it is easy to reach out to someone with the necessary expertise at the University and they have always welcomed these interactions. For example, before presenting my research to a more computational audience, I meet with this network to hone my presentation for scientists from a computational background. Not only do these interactions help improve the research that I perform, but they also provide me with training to present my ideas and results in a concise and clear manner and defend those ideas in front of experienced graduate students, postdoctoral fellows, and senior scientists.

The F99 and K00 phases of my training will be vital to developing the skills I need to run a successful computational metabolism lab in the future. To ensure I obtain the proper training, I have outlined the following goals. Component steps planned to achieve these goals are detailed in Table 1 of Section C in this document.

1. Research Development: While I have years of experience in bioinformatics and computational biology, there are still several skills I am lacking in my toolkit. For instance, in order to deconvolute data within complex metabolic networks, I will need to lean on principles of graph theory. In order to develop these skills, I have partnered with Dr. Bei Wang Phillips (see co-sponsor statement) and her graduate student, Youjia Zhou, to master and implement these principles in the *Metaboverse* software package. Another principle that will be essential for the projects proposed in this application is machine learning. While some view machine learning as the cure-all in medicine, improper implementation is rampant. As during the *XPRESSyourself* project, where I worked closely with Drs. Jay Gertz and Aaron Quinlan to ensure I implemented statistical and computational principles correctly, I will work with Dr. Bei Wang Phillips to develop the necessary intellectual foundation.

Various aspects of this proposal will require wet lab validation. During my training in the Rutter lab, I have learned how to prepare high-throughput sequencing libraries. I am also fortunate to be in a lab consisting of individuals from diverse backgrounds and experiences. We currently have nine post-doctoral fellows who bring expertise in sequencing and mass spectrometry, as well as yeast genetics, mammalian cell culture, and biochemistry. I intend to continue to strengthen my wet lab skills by preparing sequencing libraries and samples for mass spectrometry (with the support of others in the Rutter lab who do this routinely) to carry out the validation studies proposed in this application. My wet lab background and training will profit immensely from these relationships and training from others in the Rutter lab.

In addition to the papers I will publish as first author, and to make full use of my unique background in computational metabolism, I plan to work on 2-4 collaboration projects per year that will lead to additional co-

author papers. While I have been able to contribute significantly to each of these projects, it has not deterred me from maintaining a high level of productivity on my personal projects.

2. Computational Training: While originally trained in C++, I primarily use Python and R for my research. These programming languages, while incredible useful, have limitations. Thus, I have begun teaching myself other programming languages, such as Julia for its speed and efficiency for data processing tasks and JavaScript for the dynamic visualization libraries available. I have also been refreshing my HTML and CSS skills to aid in design and implementation of the user front-end experience of the *Metaboverse* project. I will continue to develop these skills, as I have done with Python and R over the past several years, to ensure that I have a broad, yet firm, grasp over these languages.

3. Science Communication: In order to continue to develop my science communication skills, I will approach growth in this area through a handful of avenues. First, grant writing will be an essential skill that I must continue to work on, especially as I will run my own lab at some point in the future. While I have taken a grant writing course and successfully applied for and received a T32 training grant, Jared will also help me find opportunities to aid in grant writing for applications related to my project. The second avenue will be through publications. Dr. Rutter has already been incredibly helpful in this regard by allowing me to take the lead on drafting, submitting, and revising the manuscript for *XPRESSyourself*. I will continue to take the lead on future manuscripts for my project and write 1-2 review articles related to the developing field of computational metabolism. The third and final avenue will be in relation to presentation skills. To date, I have attended 4 external conferences while in the Rutter lab, where I have presented 3 posters and 1 talk on my research. I was also fortunate enough to be invited to present my research to the UCSD Center for Computational Biology & Bioinformatics. I will continue to attend conferences related to metabolism and complex systems and seek opportunities to give talks at these venues. I will also use these experiences to continue to build my collaborative network. I believe that clinical experiences are essential for grounding and focusing our research. As such, I will continue shadowing Dr. Stephen McKeller, a heart surgeon, and Dr. Bill McKean, an oncological physician, both of whom are associated with the Rutter lab, to ensure I develop this critical human-centered grounding to guide my research.

4. Leadership: To develop the skills I need to become a successful PI, I have taken several approaches in the development of my training plan. I am currently an officer for the SACNAS chapter at the University of Utah, where I manage the group's webpage design and upkeep and social media outreach. I also regularly review for the Journal of Emerging Investigators, a peer-reviewed journal dedicated to mentoring middle- and high-school students as they publish their own research. A central goal and task as a reviewer for this journal is to assist these young researchers in developing essential scientific communication skills. My expertise has also been solicited for peer-review at reputable journals.

I have been actively sought out opportunities for providing in-person mentorship. This past Summer I mentored an intern through the Native American Summer Research program, with whom I taught computational methods for investigating cancer metabolism regulation. I also mentor an undergraduate student who is planning on attending graduate school. They are currently assisting on the *Metaboverse* project and will soon be beginning an independent project of their own design. I will aid them as they prepare their graduate school applications and other appropriate fellowship applications. I also take the lead in meetings with collaborators and have secured computational resources through the University of Utah's Center for High Performance Computing allocation award submission project for the past three years for all projects carried out in the Rutter lab.

Dr. Jared Rutter, Dr. Bei Wang Phillips, and I are confident that this training plan will be fundamental and sufficient to prepare me for my transition from Ph.D student to postdoctoral fellow, and eventually a tenure-track professor. Upon completion of my degree, I will apply for postdoctoral positions in academic research where I can continue to develop my interdisciplinary computational/wet lab training and prepare myself for a position as a principle investigator at an NCI-Designated Cancer Center. There, I will apply systems biology approaches to complex, mechanistic questions in cancer metabolism and will continue developing methods and tools that span and bridge the computation-biology gap. To ensure I stay on track to meet these goals, and aside from our weekly meetings and day-to-day communication, Dr. Rutter and I hold a yearly progress meeting. While we discuss my goals weekly, this yearly meeting gives us the opportunity to reflect on the past year, revise training plans, and evaluate whether I am on target to achieve my goals. These meetings have been a significant in my development as a scientist.

C. Activities Planned Under this Award

The milestones and benchmarks listed in Table 1 will keep me on schedule to achieve the goals discussed in my training plan (Section B). The F99/K00 award will be consequential in providing me with the independence and confidence to drive my own cancer research program and make a lasting impact in computational metabolism and cancer biology.

Table 1. Timeline and Milestones for F99/K00 Training

Activity	F99		K00			
	Y1	Y2	Y1	Y2	Y3	Y4
<u>Research Activities:</u>	80%	80%	87%	87%	80%	80%
- Maintain XPRESyoursel software	O	O				
- Complete development of Metaboverse software (Aim 1.1-3)	O					
o Expression Pattern Search Engine (Aim 1.1)	O					
o Mechanistic and regulatory curation (Aim 1.1)	O	O				
o k-NN time-aware search algorithm (Aim 1.1)	O					
o Documentation and usability	O					
o Perform validation studies (Aim 1.2)	O					
o Prepare and submit manuscript for peer review	O	O				
- Explore and validate amplifying regulatory events in metabolic networks (Aim 1.4)	O	O				
o Develop statistical network methods for exploring complex regulatory events within Metaboverse (Aim 1.3)	O	O				
o Curate time course datasets (Aim 1.3)		O				
o Interrogate temporal and spatial networks for regulatory events (Aim 1.3)		O				
o Validate regulatory networks (Aim 1.3)		O				
o Prepare and submit manuscript for peer review		O				
- Curate massive-scale training data for metabolism prediction models (Aim 2)			O			
- Create paired time course datasets for RNA-Seq and metabolomics (Aim 2)			O	O	O	O
- Generate transcriptome-to-metabolome machine learning models (Aim 2)				O	O	O
- Validate machine learning models in vitro (Aim 2)			O	O	O	O
<u>Intellectual Training:</u>	8%	5%	3%	3%	3%	3%
- Attend weekly seminars	O	O	O	O	O	O
- Complete advanced seminar coursework	O					
- Complete 5 th year graduate journal club	O					
- Complete ethics in research course	O					
- Complete machine learning course		O				
- Attend lab meeting (weekly)	O	O	O	O	O	O
- Attend department Research in Progress (weekly)	O	O	O	O	O	O
- Clinical shadowing	O	O	O	O	O	O
<u>Professional Development:</u>	5%	5%	5%	5%	5%	5%
- Attend cancer metabolism conferences (1-2x/yr)	O	O	O	O	O	O
- Attend systems/computational biology conferences (1-2x/yr)	O	O	O	O	O	O
- Write review article(s)		O	O	O	O	O
- Volunteer as peer-reviewer	O	O	O	O	O	O
- Student mentorship	O	O	O	O	O	O
- Complete TAship requirement	O		O	O	O	O
- Meet with faculty interviewees	O	O	O	O	O	O
- Attend local metabolism and cancer conferences/symposia	O	O	O	O	O	O
<u>Milestones:</u>	2%	5%	0%	0%	5%	5%
- Prepare and submit Metaboverse and metabolic regulation manuscripts	O	O				
- Prepare and defend dissertation		O				
- Prepare K00 transition application and develop K00 research plan		O				
- Apply for NIH Pathway to Independence (K99/R00) or other R-level grants					O	O
- Identify, interview for, and obtain tenure-track faculty position					O	O
<u>Mentorship Plan:</u>	5%	5%	5%	5%	7%	7%
- Meet with sponsor (weekly)	O	O	O	O	O	O
- Meet with co-sponsor (bi-weekly)	O	O	O	O	O	O
- Attend and receive feedback at sub-group meetings (weekly)	O	O	O	O	O	O
- Present at lab meeting (3x/yr)	O	O	O	O	O	O
- Present at department Research in Progress (1x/yr)	O	O	O	O	O	O
- Present at Seminars in Metabolism or similar venue (1x/yr)	O	O	O	O	O	O
- Meet with dissertation committee (1-2x/yr)	O	O	O	O	O	O
- Progress report and planning meeting with sponsor (yearly)	O	O	O	O	O	O
- Identify K00 mentor		O				
- Prepare for faculty interviews		O			O	O

Specific Aims

Metabolism is a complex network of chemical reactions. These reactions are required to sustain life and involve a multitude of actors, namely metabolites, proteins, and the genes that encode these proteins¹. The dysregulation of metabolism is one of the hallmarks of cancer²⁻⁴. Much of the research in cancer metabolism has sought to understand this complex network using reductionist approaches by interfering with a single gene, protein, or metabolite to determine its effect². While valuable, this approach limits the overall understanding of how this perturbation and its outcomes fit into the larger network. To approach this system holistically, computational tools and more sophisticated analyses are needed to properly deconvolute these networks⁵. High-throughput methods enable us to probe these networks systematically, but it is difficult to contextualize these data on the metabolic network⁶. Additionally, researchers tend to use only a limited number of datapoints from a dataset and only focus on entities that are regulated above an arbitrary threshold^{7,8}. While this approach helps identify large magnitude changes, it neglects essential information that is encoded within lower magnitude changes within the network.

To begin to address these challenges in -omics data contextualization, we built *XPRESSyourself* to uniformly process and analyze ribosome profiling and RNA-Seq data⁹. Our collaborator and Director of the University of Utah Metabolomics Core, Dr. James Cox, has built an automated pipeline for processing metabolomics raw signal data. Together, these two software packages will form the operational foundation of *Metaboverse*, a software package for contextualizing multi-omic data on the global metabolic network. It will also allow the systematic searching of regulatory patterns within the network. For example, we are interested in identifying reactions where input abundance is high and output abundance is low, indicating some regulatory change at a reaction. We will also build more sophisticated search algorithms for identifying temporal regulatory events. Using these tools, we will be able to interrogate low magnitude changes within the network. By using the pathway neighbors of the changed entity, we can also strengthen our confidence in whether these smaller changes are a result of noise or are meaningfully changing¹⁰. Identifying and understanding these regulatory principles will allow for improved statistical confidence in metabolic events, will describe the regulatory landscape of metabolism with improved clarity, and will describe these changes within the context of tumorigenesis.

My primary goal is to become an independent investigator leading a lab that bridges the gap between mechanistic wet-lab and computational cancer metabolism research. During the F99 phase, I will lead the team developing *Metaboverse* and interrogate these more complex forms of temporal metabolic regulation. During the K00 phase

Aim 1: Dissertation Research Project (F99) – Visualizing, Contextualizing, and Analyzing Chaotic Metabolic Networks and their Regulation. (1.1) I will complete development of *Metaboverse*, implementation of novel expression pattern search algorithms, and curation of metabolic interactome databases to allow identification of novel regulatory events and infer their mechanisms. (1.2) I will demonstrate the utility of *Metaboverse* by identifying previously missed regulatory events from publicly available biological data. I will analyze flux metabolomics datasets in metabolic perturbation models generated in our lab.

Aim 2: Postdoctoral Research Direction (K00) –

With my dual training possible with my background and the unique combination of my sponsor and co-sponsor, I will be able to bring a more mechanistic perspective to computational approaches for cancer metabolism than previously possible and establish a pioneering lab in computational metabolism.

Research Strategy

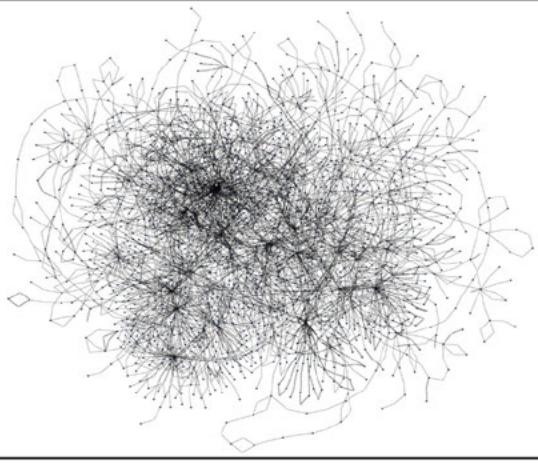
A. Significance

Cancer metabolism is complex, dysregulated, and difficult to deconvolute. Metabolism is the sum total of the chemical reactions that occur within a cell to build the necessary units for a cell to function and survive¹⁴. While commonly viewed as the “powerhouse of the cell”, mitochondria do more than just generate energy and act as a hub for cellular metabolism^{15,16}. The consequences of aberrant metabolism are far-reaching, playing a significant role in inborn errors of metabolism, diabetes, and in particular, cancer. In some cancers, altered glucose utilization leads to perturbations in metabolism and contributes to metabolic states favorable for tumorigenesis³. While a variety of pharmacological methods have been proposed to alleviate dysregulated metabolism in this context, we are far from a full understanding of the intricate dynamics of cancer metabolism systems and the consequences of their disruption^{2,3,14}.

In order to better understand cancer metabolism and develop more targeted treatment options that restore native metabolic homeostasis and discourage tumorigenesis, we must approach a tumor’s metabolism systematically to deconvolute its complex nature^{17,18}. However, a major bottleneck in metabolism research is contextualizing data within the global metabolic network¹², and analysis of these data and their place in the global metabolic system is difficult. Over the last decade, tools have been created to address this issue of data contextualization, yet they suffer from a variety of problems. All are either limited in scope of the analyses they are able to perform, unable to integrate multi-omics data, or are simply outdated, inoperable, and unusable for the general user¹⁹⁻²³. Therefore, in order to make better sense of a tumor’s metabolism system, we need to develop better computational tools that will allow for this deconvolution.

A holistic model of cancer metabolism relies on utilizing the entire dataset. Along with improvements to the way we contextualize data, we must make sense of the entire system’s ability to adapt its behavior upon perturbation. However, current analytical tools are unavailable to extract the required data, particularly in metabolomics data where missing values for metabolites are common²⁴. Additionally, analyses tend to focus on a limited list of large magnitude changes that fall outside of established, arbitrary thresholds. For example, in a metabolomics experiment, a researcher is often provided with a list of measured metabolites that fall outside a fold change greater than 2 or less than -2. However, we know that noise is inherent to biology²⁵⁻²⁷, and that ignoring entities that do not change by the arbitrary magnitude discards valuable data describing the entire state of metabolism and the genetic environment. By extracting more from our data and moving beyond reductionist approaches¹⁸, we will be able to better understand the intricate details of metabolism in cancer.

Our new computational framework currently in development, *Metaboverse*, will deconvolute and contextualize multi-omic data and shed light on complex regulatory events in cancer metabolism. This computational strategy builds on two foundational projects from the Rutter lab. The first is XPRESSyourself, a computational pipeline I developed for automating the processing and analysis of transcriptomic and translational sequence data⁹. Paired with the work of our collaborator, Dr. James Cox, the University of Utah Metabolomics Core Director, who developed a pipeline for the automation and uniform processing of metabolomics data, these software packages will be fundamental in analyzing data for this proposal. The second is a recent manuscript from our lab where we investigated the fundamental metabolic network structure and described important considerations for contextualizing data within this complex structure²⁸. In sum, these projects will allow us to visualize multi-omic metabolic data correctly.



Metaboverse will be an open-source, interactive, and scalable framework for multi-omic data visualization and analysis on the global metabolic network (**Aim 1.1-3**). With this tool, we will be able to answer questions related to more complex regulatory mechanisms in cancer metabolism.

Figure 1: The Metabolic Network. A representation of metabolism without organelle compartments or central metabolite hubs from Waller et. al., *In Press*. One can appreciate the complexity of the metabolic network and the challenge of contextualizing data within this structure. Each node (point) represents a chemical entity. Each link (line) represents a relationship between nodes.

B. Approach

Aim 1: Dissertation Research Project (F99) – Visualizing, Contextualizing, and Analyzing Chaotic Metabolic Networks and their Regulation. To date, my dissertation has been marked by several collaborative efforts, and my own effort to standardize and enhance sequence analysis, particularly for ribosome profiling data. Each of these projects, discussed below, contribute significantly to the remaining F99 work I will propose.

Dissertation progress report

a) Regulation of Tumor Initiation by the Mitochondrial Pyruvate Carrier. As part of a collaborative project, we described how tumorigenesis is drastically increased by the loss of the mitochondrial pyruvate carrier (MPC) in mouse and fly colon cancer models²⁹. I curated and analyzed the mouse model and human patient data to identify the systematic consequences of MPC-loss on pyruvate metabolism and became interested in developing methods that would allow for a more integrated analysis of cancer metabolism. During this project I developed *XPRESSplot*, a component of *XPRESSyourself*, that provides an easy-to-use package for plotting biological data⁹. Methods included in this package and used for this project included principle component analysis, differential expression analysis, systematic analysis of gene regulatory patterns, and more.

b) Compartment and Hub Definitions Tune Metabolic Networks for Metabolic Interpretations. In this work, we described key considerations for analyzing metabolic networks and data²⁸. As metabolic networks are complex and pathways are often entangled, we described the topographical structure of the network and determined that removal or simplification of “hubs” (metabolites that participated in more than 50 reactions), aided in the disentanglement of this network and made visualization and interpretation of metabolic data easier. These improvements were demonstrated by reanalyzing existing data and extracting new metabolic insights from the data using these improvements. During this project, I began developing the foundation for *Metaboverse*.

c) XPRESSyourself: Enhancing, Standardizing, and Automating Ribosome Profiling Computational Analyses Yields Improved Insights. Early into my dissertation, I noticed a lack of standardized and automated methods for processing and analyzing sequence data. This was particularly apparent for ribosome profiling, where the experimental methodology has been available for a decade, yet computational approaches are incomplete and inconsistent. I developed *XPRESSyourself* as an all-in-one toolkit for ribosome profiling, standard single-end, or paired-end RNA-Seq data processing and analysis⁹. Along with learning software development best practices, I filled voids in the ribosome profiling toolkit where publicly available options were missing, such as fast and efficient quality control modules and a curation tool that allowed for the removal of systematic biases present in ribosome profiling data^{30,31}. This project allowed me to master methods for sequence data analysis in RNA-Seq and ribosome profiling³². As part of the validation of this tool, we benchmarked *XPRESSyourself* to publicly available data. We showed that our methods could extract additional supporting biological information.

Development, Validation, and Release of Metaboverse. A challenge within the metabolomics community since its inception has been contextualizing data⁶. While it is simple to determine up- or down-regulated components of metabolism from most experiments, making sense of how these changes contribute to the larger story has traditionally been performed manually through the curation of a graphical pathway representation with the data overlaid by hand. While some tools have emerged to handle this challenge, they are severely limited in capability^{19–23}. For example, one of the most popular options, *Cytoscape*, offers robust visualization and a certain level of customizable analysis¹⁹. However, in order to perform more complex searches in a network, like those proposed in this Aim, additional infrastructure needs to be built from the ground up.

Five essential challenges remain for which a comprehensive tool is missing. **First**, how do we develop a metabolic network visualization tool that is easy to use and then how do we understand its output? **Second**, how do we integrate the full extent of available regulatory information to infer the mechanism behind the events we observe? **Third**, how do we parse out relevant information from the global network when massive amounts of data and annotated relationships are involved? **Fourth**, how do we strengthen our confidence in changes and relationships across the global metabolic network? **Fifth**, how do we integrate cross-species reaction information in a way that might inform identification of a novel reaction in our organism of interest? The development of *Metaboverse* will address each of these challenges (alpha version under development at github.com/Metaboverse). *Metaboverse* is designed to be a simple tool to help users integrate single or multi-omic datasets into the metabolic network, contextualize their data within the global network, and search the network for interesting regulatory information. It is capable of processing data for every model organism, and its network structure is customizable. To date, the basic structure of *Metaboverse* is in place and we have developed proof-of-concept tests for the majority of planned features; however, several key features are still in active

development, which will compose the sub-aims for Aim 1.1-3. A working demo demonstrating some of these planned features is available online (metaboverse.github.io/demo). Completion of these sub-aims will provide me with thorough training in metabolomics data analysis and processing, network biology, machine learning, and database curation. These sub-aims will also require tailored statistical method development. These supplementary training components will be guided by my co-sponsor, Dr. Bei Wang Phillips, with additional mentorship from Dr. Jason Gertz.

Aim 1.1: Complete analytical platform and database curation for Metaboverse.

In order to allow for the robust multi-omic integration and analysis of data in the context of the metabolic network, we will build advanced capabilities into *Metaboverse* that will allow for the systematic analysis of regulatory events in cancer metabolism and improve our ability to generate hypotheses and infer regulatory mechanisms from data. Throughout this proposal, I will generally refer to a regulatory event as a reaction where input is high and output is low, or vice versa. If an event is accompanied by data for modifiers of that reaction, such as a catalyst or inhibitor, these reactions will be prioritized for the user. However, more complex regulatory events will be analyzed by *Metaboverse*.

- *Expression and abundance pattern search engine (PaSEN):* In collaboration with Dr. Bei Wang Phillips and Youjia Zhou, we are developing *PaSEN* using principles from network topology searches for rapidly analyzing the global metabolic network for interesting regulatory events^{33,34}. In Figure 2A, the transition between L-Arginine (L-Arg) and L-Arginosuccinate (ARSUA) shows a large difference in relative abundances across the reaction (>2 fold). *PaSEN* will search for changes across a reaction using similar and more complex criteria. *PaSEN* will interrogate the entire network for potential regulatory events, and therefore avoid the bias originating from a researcher focusing on a few familiar pathways. The current metabolic model includes modifier information for reactions. In this same example in Figure 2A, we see that gene expression for ASL, whose gene products form the ASL tetramer which acts as a catalyst in this reaction, is down regulated. It could be inferred that the reduction of this catalyst is driving the change across the reaction. Results such as these will be returned to the user as a prioritized results table based on these and other considerations. This output will allow the user to quickly identify the highest confidence regulatory events to help them focus their follow-up and validation experiments on network-wide changes. The ability to track regulatory patterns across time will also be incorporated into the search engine.

- *k-Nearest Neighbors searches:* Once a regulatory event is identified, a user may want to explore how a change affects distal reactions. For example, does the decrease in glycerol (center) lead to changes in pathways other than in glucose metabolism? Using k-nearest neighbors (k-NN) search principles³⁵, users will be able to explore various reaction neighborhoods of their entity of interest (Figure 2B). This capability expanded to multiple nearest neighbors and its ability to perform trend analysis across the reaction neighborhood will be essential for monitoring regulatory waves across time and between pathways (Aim 1.3). For cases where network coverage is sparse, we will develop a reaction collapse module that will allow for summarization of reactions with missing data to bridge these gaps. Each reaction bridge will be maintained in the network representation as a separate pseudo-relationship so a user can identify these missing steps for further interrogation of those components if desired.

- *Regulatory predictions:* The current database used by *Metaboverse* is derived from Reactome³⁶ and contains reaction modifier information that can be helpful in adding context behind the patterns identified by *PaSEN*. We will

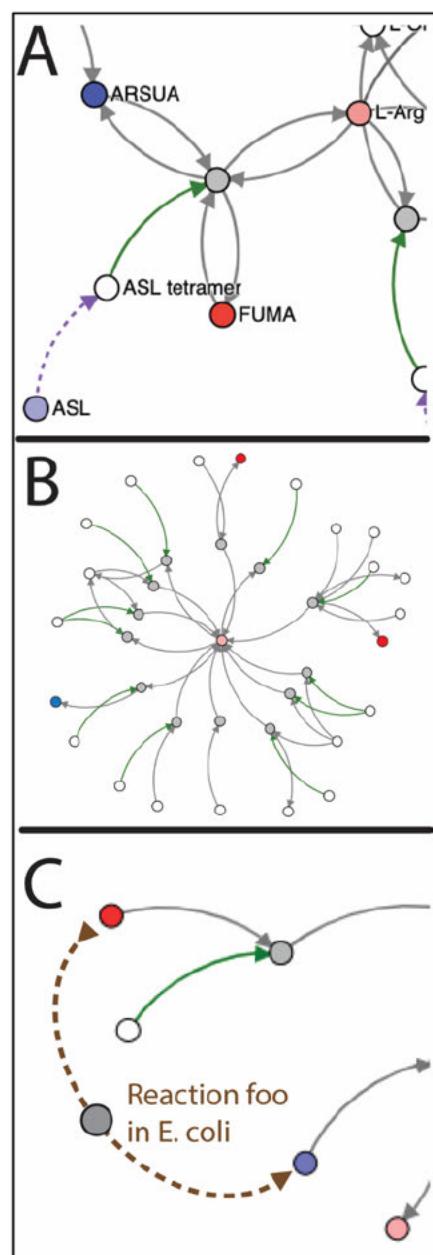


Figure 2: Features of Metaboverse.
(A) Metaboverse demo displaying test metabolomics and transcriptomics data overlaid on the Urea Cycle. Grey circles indicate reactions. White circles indicate missing data or no change. Red circles indicate up regulation, while blue circles indicate down regulation. Green curved arrows indicate a reaction catalyst. Dotted purple curved arrows indicate a gene component of a protein involved in the reaction. (B) Example of k-NN search to explore distal participation of metabolite in reactions across all pathways and between pathways. (C) Example of a regulatory event across a reaction not annotated in human but found in other organisms.

supplement this information with BRENDA, a comprehensive database for enzyme interactions across organisms³⁷. This will further improve the ability of *Metaboverse* to provide mechanistic predictions from events identified by *PaSEN*. *PaSEN* will also search putative reactions in, for example humans, that have only been identified (with high confidence) in other organisms (see Figure 2C). This may be useful in cases where a regulatory activity across an unannotated reaction in an organism exists, implying existence of an unannotated metabolic feature. These databases will be supplemented with findings from MIDAS³⁸, a platform continuing to be developed in the Rutter lab in collaboration with Calico Life Sciences, which identifies novel protein-metabolite and RNA-metabolite interactions.

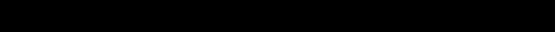
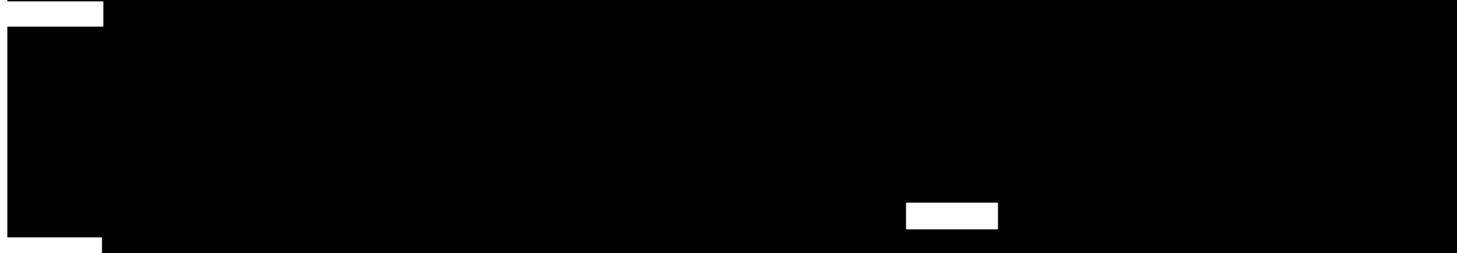
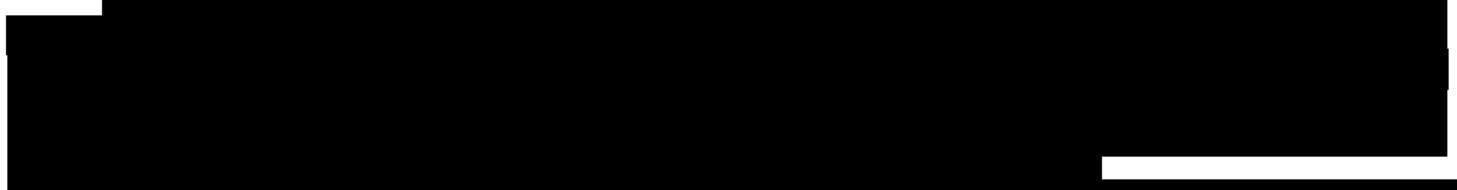
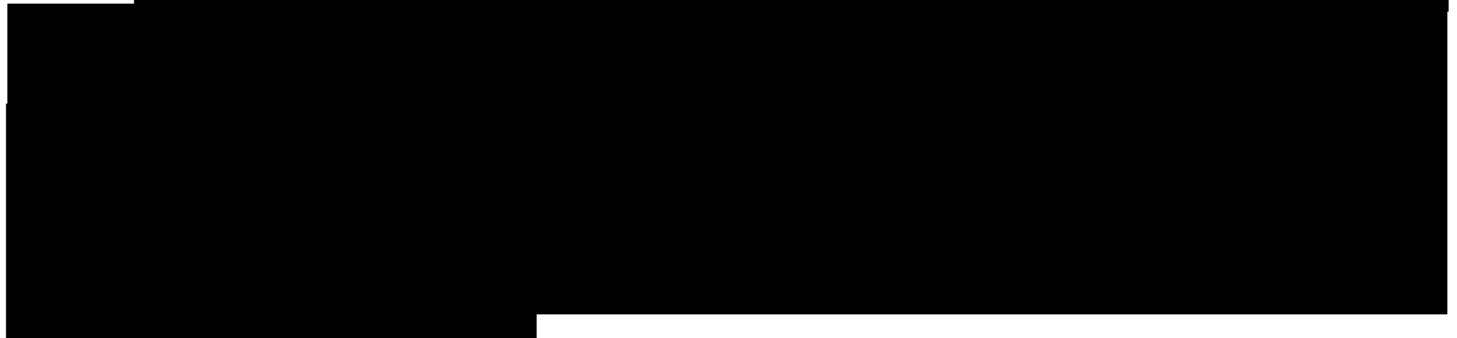
Aim 1.2: Demonstrate Metaboverse's ability to extract novel regulatory events from public and new data.

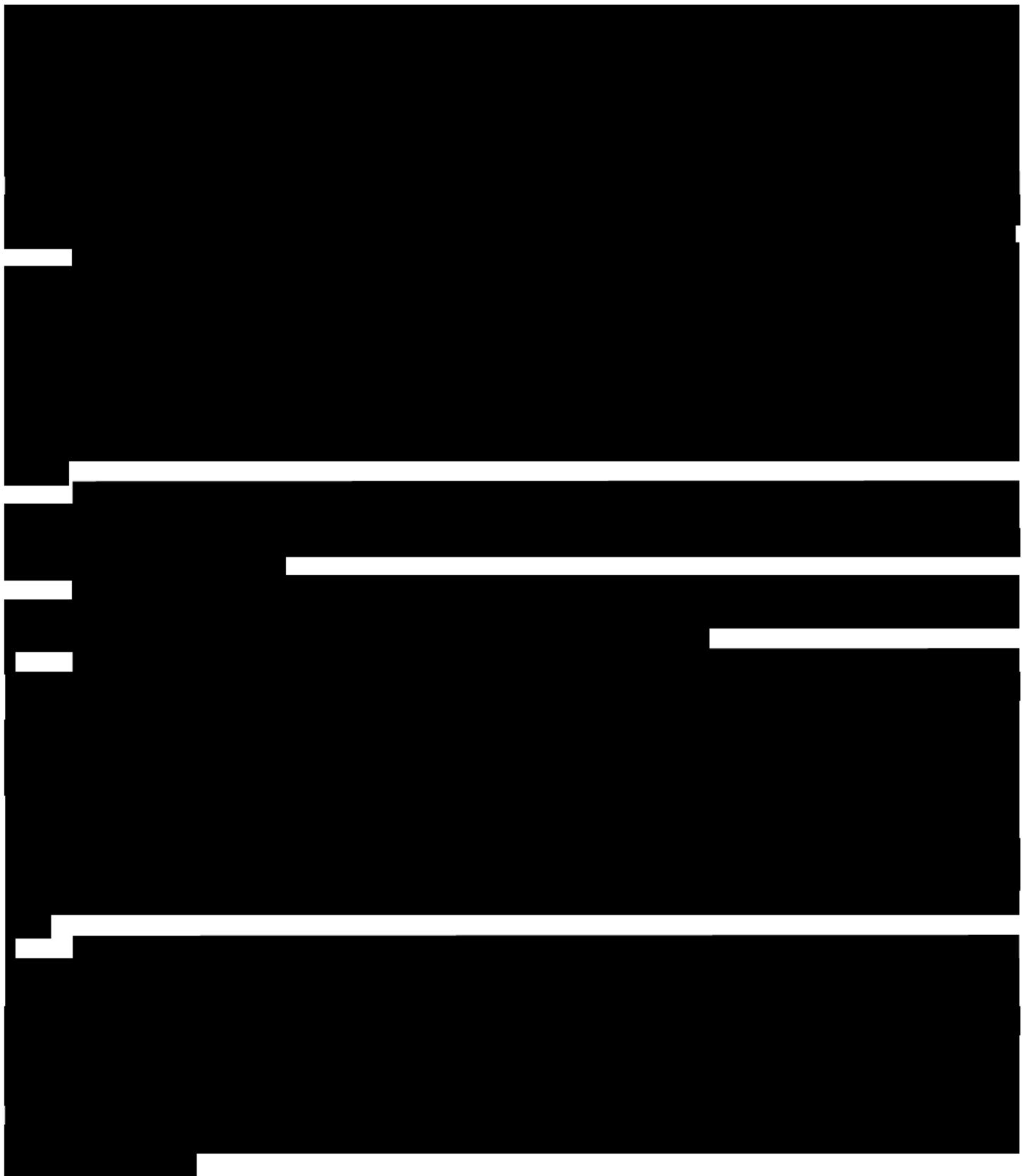
Using high-quality metabolomics datasets I identify with Dr. James Cox, I will compare the original analyses of these data with those derived from *Metaboverse* using the features described in Aim 1.1. We expect that *Metaboverse* will recapitulate the original findings and identify novel regulatory events missed in the original study using conventional analysis methods. In collaboration with Drs. Cox and Waller, we have also produced flux metabolomics datasets in yeast with varying levels of MPC expression. As described before, we have observed that MPC amplifying tumorigenesis in the colon²⁹. Using these data and the search features available in *Metaboverse*, we will interrogate the mechanistic roles of MPC loss on glucose utilization to better understand its role in tumorigenesis.

Alternative Strategies: While the amount of public metabolomics data is increasing, there is currently no NIH mandate for their publication as with gene expression data. Therefore, we may find that the number of total metabolomics datasets we can use for Aim 1.2 is limited. If this is true, we will generate paired yeast RNA-Seq and metabolomics datasets under various conditions to generate the needed data.



Aim 1.3:





Completion of the F99 phase of my research will provide a new software package for the integrated multi-omic analysis of metabolic data. Using this tool, we will explore the role of MPC and pyruvate flux on glycolysis and better understand how MPC impairment promotes tumorigenesis.

Aim 2: Postdoctoral Research Direction (K00) –

This award will be essential in my progress by allowing me the flexibility and independence needed in my research to answer questions such as those posed in this Aim and will help establish myself as an expert and pioneer in computational metabolism and cancer metabolism.

Respective Contributions

Respective contributions to the research plan

The research strategy was developed and conceived by the PI (Jordan Berg) with insight from the sponsor (Dr. Jared Rutter) and co-sponsor (Dr. Bei Wang Phillips). The proposed experiments were conceived and designed by the PI.

Respective contributions to the experiments

Yeast metabolomics data used in *Metaboverse* validation were generated by Drs. T. Cameron Waller and James Cox. *Metaboverse* is developed by the PI, with support from Dr. Bei Wang Phillips, Youjia Zhou, and Ian George. Data curation and analysis for the metabolic regulation studies proposed in this document will be performed by the PI. Any validation studies in yeast and commercial human cell lines will be performed by the PI with assistance from Dr. Alex Bott of the Rutter lab.

Selection of Sponsor and Institution

To establish myself in the field of computational metabolism, the selection of my sponsor, co-sponsor, and institution were essential and each decision came with extensive introspection and discussion. Preparing myself for my career requires several components: 1) Mentoring from an expert metabolism researcher who appreciates systematic approaches to untangling the complex features of metabolism and its role in cancer and who will train me to approach metabolism as a thoughtful biologist, 2) Mentoring from a computational expert with a background in graph theory and machine learning who will challenge and train me to think analytically about the proper computational and statistical approaches associated with this proposal, and 3) An institutional environment that will allow me to develop rich collaborative relationships and that has the resources necessary to allow me to ask important questions in cancer metabolism.

Selection of Sponsor: While I was impressed by all of the professors with whom I rotated, what ultimately drew me to choose Dr. Rutter as my mentor was his expertise in metabolism and metabolic regulation, his willingness to explore metabolism in novel, inter-disciplinary ways, and his rigor in publishing high-quality, foundational research papers. After my dissertation work, I will pursue a postdoctoral research position and a position as a tenure-track professor in academia, and I was certain Dr. Rutter would provide me with the best opportunities and mentorship to prepare myself for these career stages. This has been true manifold. Dr. Rutter's approach to thinking about biology and driving the field to new limits has been consequential in my training and will be essential to ensure I have a solid foundation in cancer biology and metabolism before my postdoctoral training in a more computationally focused lab for the K00 training phase. I also appreciate Dr. Rutter's approach to mentoring where he provides his mentees with complete independence and freedom to explore and think creatively. These combined qualities will allow me to take full advantage of my dissertation training and will help me establish the needed connections within the cancer, metabolism, and computational biology communities. Dr. Rutter has been fantastic in making himself available whenever needed for advice or administrative purposes. If he or I are traveling, he will make a point to schedule time to talk virtually while the other is in transit. In the Rutter lab, I will also be able to establish myself in the systems biology and computational metabolism fields by publishing foundational papers in the field of metabolic regulation, along with publishing papers presenting tools needed for processing and analyzing ribosome profiling, RNA-Seq, proteomics, and metabolomics datasets.

Selection of Co-Sponsor: I selected Dr. Bei Wang Phillips as my co-sponsor due to her expertise in topological data analysis, graph theory, and machine learning. It was important in the development of my training plan that I surround myself with an expert with a computational background. I was especially impressed with Dr. Phillips enthusiasm and vision when we initially discussed the *Metaboverse* project. Already, I have seen incredible dividends in my personal and intellectual growth from this relationship, along with the bolstered ability to innovate in new ways at the intersection of computer science and cancer metabolism. The training Dr. Phillips will provide will be vital to my success as a postdoctoral fellow as I will need a solid foundation in machine learning and graph theory in order to hit the ground running at the start of the K00 phase of my training.

Selection of Institution: I was drawn to the University of Utah because of its strong collaborative spirit and its multi-disciplinary approach to biomedical and basic science. I was impressed by the multiplicity of fields I could explore within the Molecular Biology/Biochemistry Graduate School Program. During my interviews at the University of Utah, I could see how excited and creative the professors in each department were, and this was encouraging to me as I saw the potential to apply my own expertise in creative and forward-thinking manners. I was also impressed with the caliber of scientists at the University of Utah. The University has been home to a Nobel Prize winner and several other prestigious scientists, and is an NCI-Designated Comprehensive Cancer Center, which speaks to the quality of science at this institution.

Another consideration in my decision to attend the University of Utah for my dissertation work was the numerous resources available. For example, in the Emma Eccles Jones Building, where the Rutter lab is located, we are a short walk away from the Metabolomics Core. This is an essential resource for the projects I have proposed and the proximity to Dr. James Cox, the core director, and his willingness to discuss ideas is consequential for this project. I have also had similar interactions with the sequencing core and the Center for High-Performance Computing. Feedback and resources from both groups have been invaluable. The close proximity to the Huntsman Cancer Institute, where my long-time collaborator, Dr. Jason Gertz, is located, is also important as there are numerous world-class cancer biologists in addition to Dr. Gertz with whom I can discuss and develop my ideas.

Training in Responsible Conduct of Research

Introduction

Training in the Responsible Conduct of Research (RCR) is an ongoing process at the University of Utah. Trainees receive informal training from faculty and mentors, and formal training through the Research Education Department's RCR courses. The Research Education RCR courses are designed to meet or exceed all NIH and NSF requirements for RCR training. All individuals who are covered by this mandate and working under this grant will either complete a 9-week Research Ethics course (PHIL 7570/1.0 credit hours) and/or complete Research Education RCR courses. Compliance is administered under the University Office of the Vice President for Research, which is responsible for administration of grants and contracts, management of intellectual property, oversight of human subject and laboratory animal research, and implementation of the university's policies on research. Mr. Jordan Berg successfully fulfilled this requirement by completing the PHIL 7570 course during the Fall 2016 semester and will continue to complete an RCR course through the University of Utah every four years, thus is scheduled to complete an RCR course Fall 2020 to fulfill this additional goal.

Instructional Components of the University of Utah Responsible Conduct of Research Course

1. **Format:** The Research Education Department offers a certificate that consists of a minimum of 10 hours of instruction including live lectures, group discussions, assigned readings, and a comprehensive online class with a proficiency examination.
2. **Subject Matter:** The program includes (but is not limited to) the following instructional areas:
 - a. Protection and Use of Human Subjects in Research
 - b. Conflict of Interest in University Research
 - c. Data Acquisition, Management, Sharing and Ownership
 - d. Research Misconduct
 - e. Publication Practices and Responsible Authorship
 - f. Mentoring/Trainee Responsibilities
 - g. Peer Review
 - h. Collaborative Research
3. Animal Care and Use in Research is also taught but is not a requirement of the Research Education Department's RCR certificate.
4. **Faculty Participation:** All live lecture and small group discussions are taught by University faculty and staff with extensive experience in the subject matter. Faculty members in charge of research training programs frequently contribute to presentations and discussions.
5. **Duration of Instruction:** Instruction consists of at least 10 hours, made up of eight (8) hours of live lecture, face-to-face instruction and supplemented by a two-hour comprehensive online course with a proficiency examination.
6. **Frequency of Instruction:** Trainees must participate in the RCR program at the earliest opportunity, at least once every four years, and at least once during their early career stage. Once the required 10 hours of RCR instruction have been successfully completed, the trainee receives a dated "Certificate of Achievement."

How Participation in RCR Instruction is Monitored

Participation in Research Education's RCR courses is monitored in several ways: signed and dated class attendance rosters, electronic transcripts, online class registration, and a dated "Certificate of Achievement". Trainees enrolled in the online RCR class are required to successfully complete review quizzes at the end of each of the 9 units and a final proficiency examination with a minimum passing grade of 85%. Upon successful completion of the online class, a Record of Class Completion is provided to the trainee. Trainees are also expected to attend all required live lecture and small group discussion sessions. Attendance rosters are the primary way to ensure that trainees have completed the appropriate training requirements. In addition to attendance, trainees are expected to be actively involved in all class discussions and activities. Because research is being done in a multitude of subject areas, having input from all trainees is essential to having a discussion that sees a situation from multiple viewpoints.

Sponsor and Co-Sponsor Statements

Sponsor: Dr. Jared Rutter, PhD

Professor of Biochemistry, Ida Smith Endowed Chair of Cancer Research, University of Utah
Investigator, Howard Hughes Medical Institute

Co-Sponsor: Dr. Bei Wang Phillips, PhD

Assistant Professor, School of Computing, Scientific Computing and Imaging (SCI) Institute, University of Utah

A. Research Support Available:

Table 1: Current Support

Table 2: Pending Support

Title	Source and PI	Amount (Annual Direct)
[REDACTED]	[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]

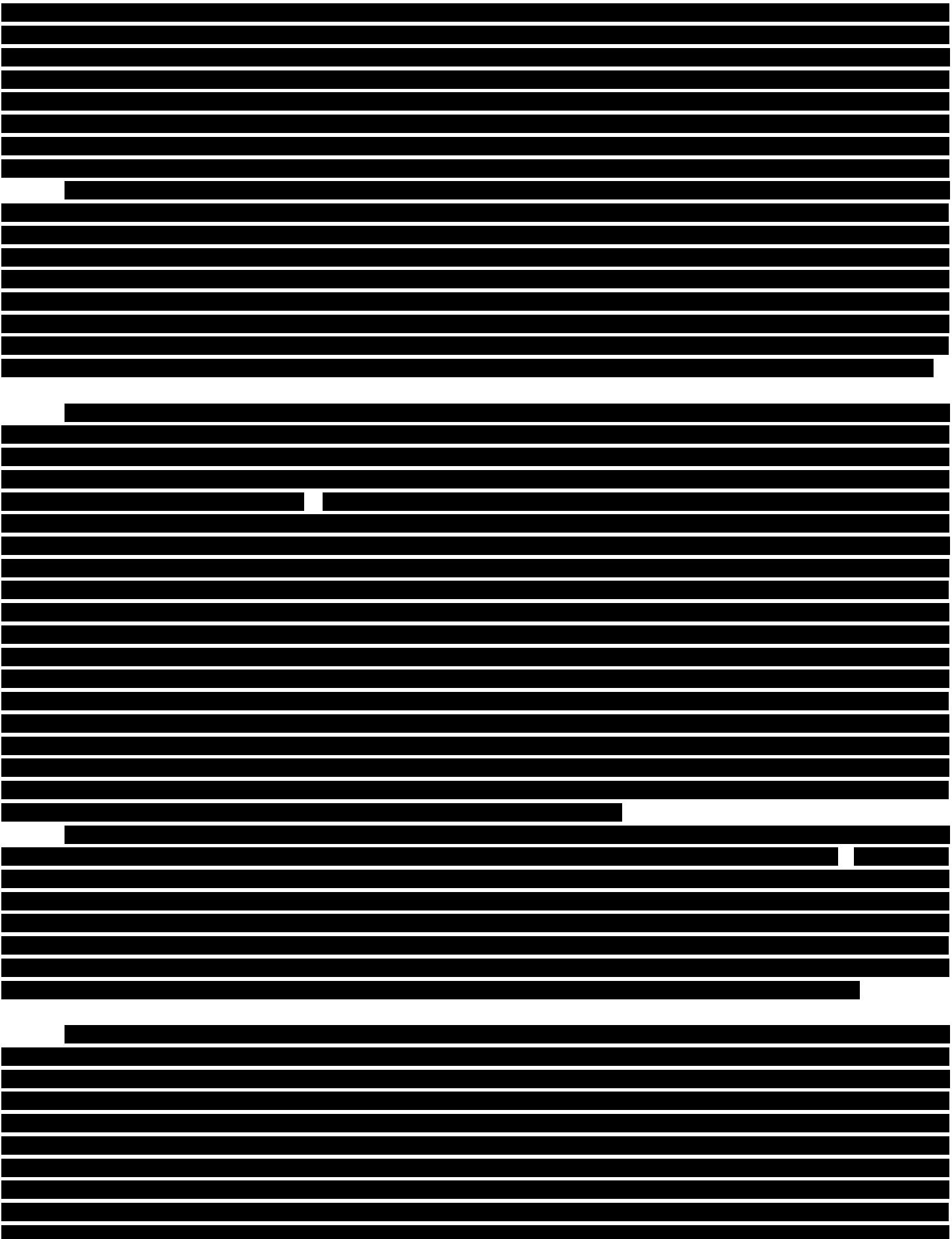
B. Sponsor's Previous Fellows/Trainees:

Table 3: Representative Trainee Information

C. Training Plan, Environment, Research Facilities:

Training Plan

A horizontal bar chart consisting of ten solid black bars of varying lengths. The bars are arranged horizontally from left to right. The first bar is the shortest, followed by a series of progressively longer bars, with the eighth bar being the longest. After the eighth bar, there is a significant gap before the final two bars, which are also black and of equal length, though shorter than the eighth bar.



The image shows a document page with several horizontal black redaction bars. There are four main groups of bars. The first group consists of a single short bar near the top left. The second group has a short bar on the left and a long bar extending across most of the page. The third group has a short bar on the left and a long bar extending across most of the page. The fourth group has a short bar on the left and a long bar extending across most of the page. The bars are solid black and vary in length, obscuring specific lines of text.

D. Number of Fellows/Trainees to be Supervised During Fellowship

Table 4: Current Trainees of Sponsor and Co-Sponsor

Table 4. Current Trainees of Sponsor and Co-Sponsor		
[REDACTED]	[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]

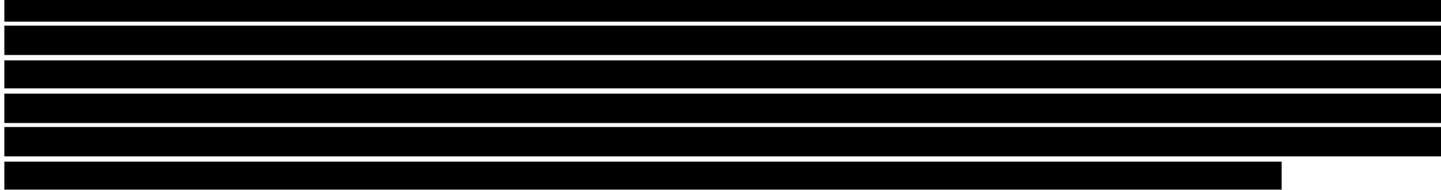
		
--	---	--

E. Applicant's Qualifications and Potential for a Research Career

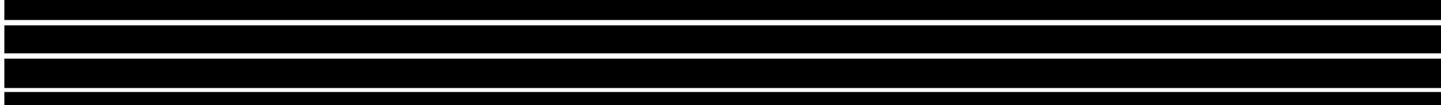
Sponsor's Statement









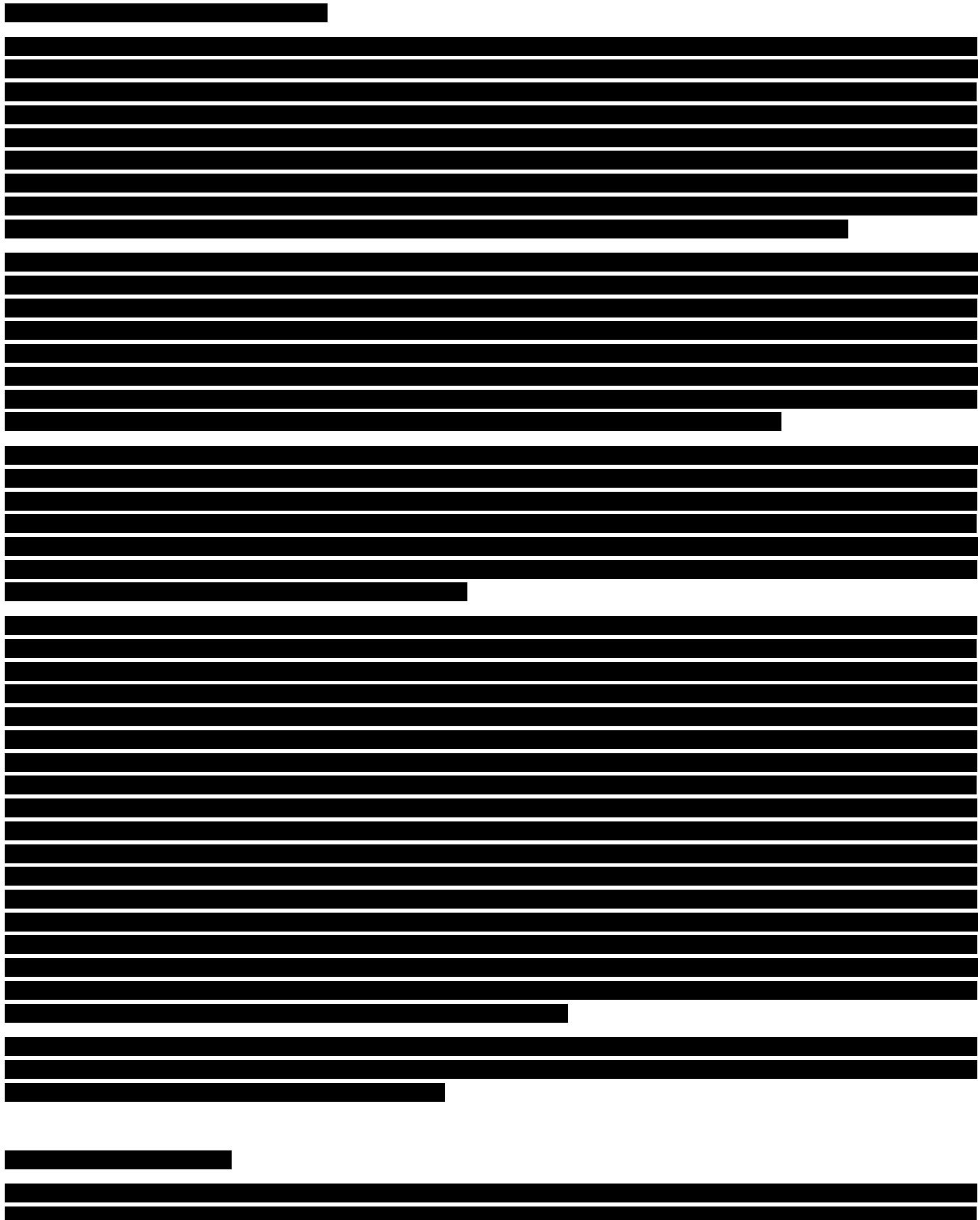


Co-Sponsor's Statement

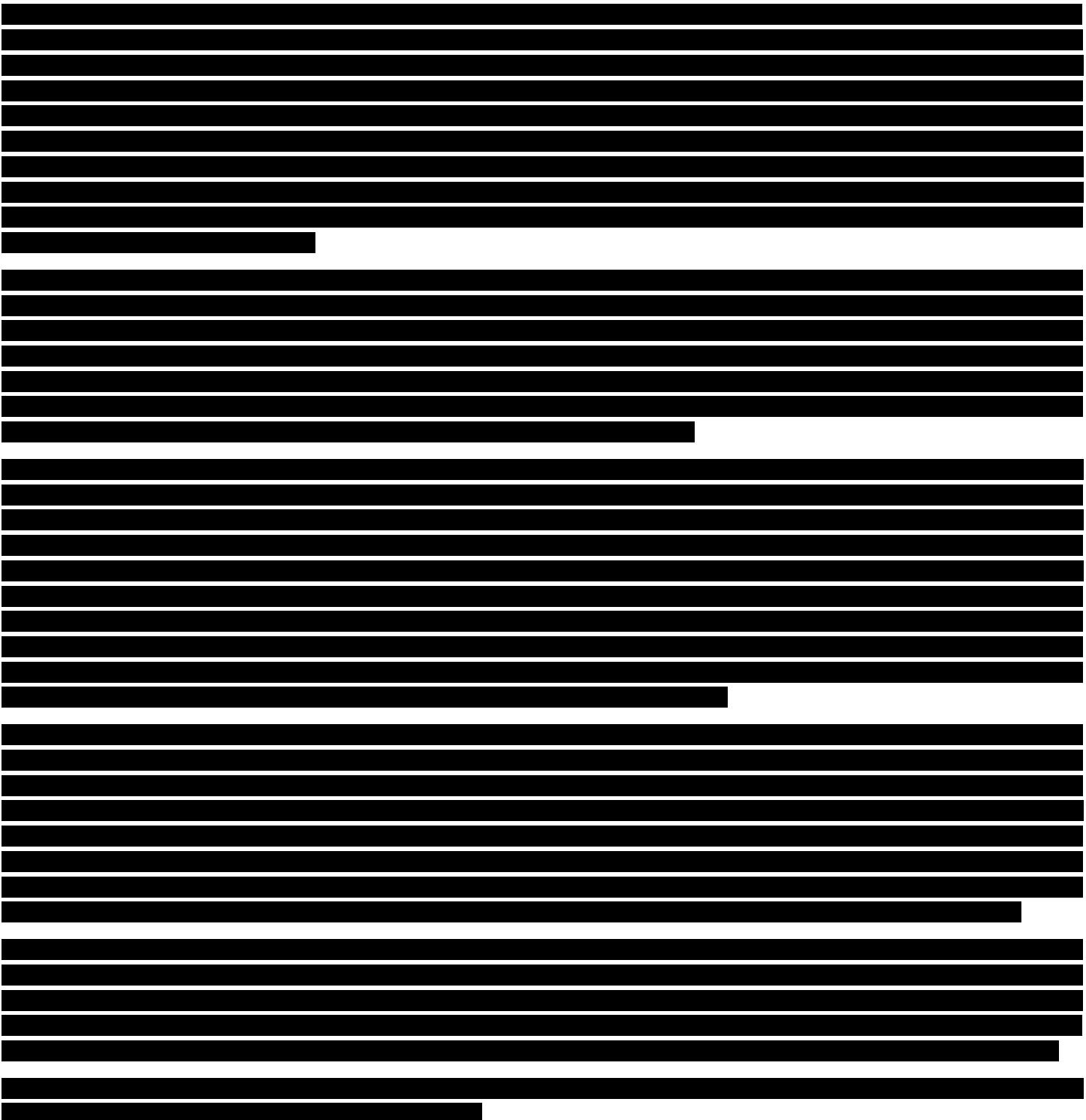


[REDACTED]

Institutional Environment and Commitment to Training



The page contains a large number of horizontal black redaction bars, indicating that the majority of the content has been removed or is confidential. The bars are evenly spaced and cover almost the entire width of the page.



Resource Sharing Plan

Software: I am fully committed to open-source software development and have a track record of adhering to this commitment (see Berg, et. al., *bioRxiv*, 2019; github.com/j-berg). Software is made available under a GNU General Public License, which allows for broad use and re-use. Whenever possible, projects in development are immediately open source during their development to ensure transparency and encourage contribution from the broader community.

Data: As with developed software, any data created and used for this proposal will be made publicly available via a repository such as the Gene Expression Omnibus or Metabolomics Workbench. All scripts developed for data analysis will be made available (see github.com/j-berg for examples). Analysis code associated with a published work will be open-access to facilitate reproducibility and will be given a digital object identifier (DOI) to act as a persistent identifier to make sure that the resource is perpetually available and free to access.

Publications: All work from this proposal will be published in peer-reviewed open-access journals under CC-BY, or similarly open, licenses. Manuscripts in review will be made immediately available via a pre-print server, such as *bioRxiv*. NIH and other resources will be acknowledged in all derivative works.

Other: Any lab reagents or cell lines generated for use as part of this proposal will be stored and managed by Dr. Jared Rutter, who will handle requests for these reagents. The Rutter lab has a rich history of making its resources available for others and has benefited from many rich collaborations as a result. To aid in dissemination of knowledge, presentations I give regarding this and related work will be made publicly available via github.com/j-berg/presentations.

PHS Human Subjects and Clinical Trials Information

Are Human Subjects Involved

Yes

No

Is the Project Exempt from Federal regulations?

Yes

No

Exemption Number

1 2 3 4 5 6 7 8

Does the proposed research involve human specimens
and/or data

Yes

No

If Yes, provide an explanation of why the applica-
tion does not involve human subjects research

Human_Subjects1029235292.pdf

Other Requested information

Delayed Onset Studies

Delayed Onset Study#	Study Title	Anticipated Clinical Trial?	Justification
The form does not have any delayed onset studies			

Human Subjects

The accompanying proposal outlined by the PI (Mr. Jordan Berg) will be completed using yeast strains generated within the Rutter lab, or commercially available mammalian cells lines. A large portion of this work will be completed by way of secondary analysis of sequencing, proteomics, or metabolomics datasets that are already publicly available by way of another published study. These datasets will be obtained through NIH-sponsored databases, such as the Gene Expression Omnibus (GEO), Metabolomics Workbench, or similar NIH-supported data repositories where proper protections for human patient data will have been taken. If protected patient data is used, it will only be accessed through the dbGAP portal where data has already been de-identified, and only after the PI and sponsor have submitted the proper request form to access the data.

In the cases outlined above where de-identified patient data is used, processing and analysis of data that could in any way be traced back to the individual will be accessed directly from the dbGAP portal, processed on the Rutter lab protected computing environment at the University of Utah Center for High-Performance Computing, and promptly destroyed after use. These data files will be accessed only through proper secure channels and will only be used if the original study followed the appropriate NIH standards for data acquisition and storage.