

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 UEI: NPM2J7MSCF61 Legal Name: PENNSYLVANIA STATE UNIVERSITY-UNIV PARK Department: Division: College of Medicine Street1: Office of Sponsored Programs Street2: 200 Innovation Blvd., Suite 110 City: University Park County/Parish: Centre State: PA: Pennsylvania Province: Country: USA: UNITED STATES ZIP / Postal Code: 16802-7000			
Person to be contacted on matters involving this application Prefix: First Name: Thomas Middle Name: J Last Name: Brydebell Suffix: Position/Title: Director, Grants Administration Department: Division: Street1: 500 University Drive Street2: P.O. Box 850 City: Hershey County/Parish: Dauphin State: PA: Pennsylvania Province: Country: USA: UNITED STATES ZIP / Postal Code: 17033-0850 Phone Number: 717-531-8495 Fax Number: 717-531-0040 Email: e-grants@pennstatehealth.psu.edu			
6. EMPLOYER IDENTIFICATION NUMBER (EIN) or (TIN): 1-246000376-A1			
7. TYPE OF APPLICANT: X: Other (specify) Other (Specify): State Related Institution of Higher Education 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:	
11. DESCRIPTIVE TITLE OF APPLICANT'S PROJECT: High-Resolution, Wide-Field 3D Histopathology for the Morphological Characterization of Soft-Tissue Tumor Biopsies			
12. PROPOSED PROJECT: Start Date 07/01/2024 Ending Date 06/30/2028		13. CONGRESSIONAL DISTRICT OF THE APPLICANT: PA-015	

14. PROJECT DIRECTOR/PRINCIPAL INVESTIGATOR CONTACT INFORMATION

Prefix: First Name: Andrew Middle Name: Last Name: Suffix:
 Position/TITLE: Grad Student Organization Name: PENNSYLVANIA STATE UNIV HERSHEY MED CTR
 Department:
 Division: College of Medicine
 Street1: 500 University Drive
 Street2: P.O. Box 850
 City: Hershey County/Parish: Dauphin State: PA: Pennsylvania
 Province: Country: USA: UNITED STATES ZIP / Postal Code: 17033-0850
 Phone Number: 443-683-0398 Fax Number: Email: aqs6915@psu.edu

15. ESTIMATED PROJECT FUNDING

a. Total Federal Funds Requested	\$260,385.00
b. Total Non-Federal Funds	\$0.00
c. Total Federal & Non-Federal Funds	\$260,385.00
d. Estimated Program Income	\$0.00

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:
 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: Thomas Middle Name: J Last Name: Brydebell Suffix:
 Position/TITLE: Director, Grants Administration Organization Name: PENNSYLVANIA STATE UNIV HERSHEY MED CTR
 Department: Research Affairs
 Division: College of Medicine
 Street1: 500 University Drive
 Street2: P.O. Box 850
 City: Hershey County/Parish: Dauphin State: PA: Pennsylvania
 Province: Country: USA: UNITED STATES ZIP / Postal Code: 17033-0850
 Phone Number: 717-531-8495 Fax Number: 717-531-0040 Email: e-grants@pennstatehealth.psu.edu

Signature of Authorized Representative

Date Signed

20. Pre-application File Name: Mime Type:**21. Cover Letter Attachment** File Name: cover_letter_flat1034805333.pdf Mime Type: application/pdf



December 1, 2023

To Whom it May Concern:

Enclosed is a submission of an F30 Ruth L. Kirschstein National Research Service Award (NRSA) Individual Predoctoral Fellowship award, entitled "High Resolution Wide Field 3D Histopathology for the Morphological Classification of Prostate Cancer". This application is submitted in response to funding opportunity PA-23-260.

The following referees will be providing letters:

Joshua Warrick, MD
Associate Professor, Department of Pathology and Laboratory Medicine
Associate Professor, Department of Urology
Associate Professor, Department of Surgery
Chief, Division of Anatomic Pathology
Penn State College of Medicine

Patrick J. La Riviere, PhD
Professor of Radiology
Committee on Medical Physics
University of Chicago

Blanton S. Tolbert, PhD
Jacob Gershon Cohen Professor of Biochemistry and Biophysics
Perelman School of Medicine, University of Pennsylvania

Thank you for your consideration of this application.

Sincerely,

A handwritten signature in black ink, appearing to read "Andrew Sugarman".

Andrew Sugarman
G2, MD/PhD Candidate
Penn State College of Medicine
asugarman@pennstatehealth.psu.edu

Project/Performance Site Location(s)

Project/Performance Site Primary Location

Project/Performance Site Location 1

Additional Location(s)

RESEARCH & RELATED Other Project Information

1. * Are Human Subjects Involved? <input checked="" type="radio"/> Yes <input type="radio"/> No		
1.a. If YES to Human Subjects		
Is the Project Exempt from Federal regulations? <input checked="" type="radio"/> Yes <input type="radio"/> No		
If yes, check appropriate exemption number		
Exemption Number: - 1 - 2 - 3 - <input checked="" type="checkbox"/> 4 - 5 - 6 - 7 - 8		
If no, is the IRB review Pending? <input type="radio"/> Yes <input checked="" type="radio"/> No		
IRB Approval Date:		
Human Subject Assurance Number 00004251		
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 checked="" 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 checked="" 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	abstract1034805332.pdf	Mime Type: application/pdf
8. Project Narrative	project_narrative1034805452.pdf	Mime Type: application/pdf
9. Bibliography & References Cited	bibliography1034805453.pdf	Mime Type: application/pdf
10. Facilities & Other Resources	facilities_resources1034805340.pdf	Mime Type: application/pdf
11. Equipment	equipment1034805342.pdf	Mime Type: application/pdf

PROJECT SUMMARY/ABSTRACT

The diagnosis and grading of cancer rely on the examination of abnormal tissue and the morphology of the cells within. For example, the clinical evaluation of prostate cancer relies on the assessment of glandular and cellular morphology from histopathology images. However, prostate cancer patients suffer from high rates of inter-observer variability amongst pathologists in the clinic. Additionally, recent studies have shown that the angle and depth of slide sectioning also contribute to significant variation in tumor grading, further illustrating the need for a quantitative, 3-dimensional, volumetric approach to prostate cancer whole biopsy imaging. In the proposed work, we leverage high-resolution, wide-field micro-CT to generate volumetric images of entire prostate needle core biopsies within FFPE tissue samples and without the need for contrast-enhancing stain. We term this technique 3D histopathology, we propose to apply this method to image whole-biopsy prostate needle core biopsies and quantify the variation of glandular shape as a function of position in 3D space. We also propose the development of a computational topology-based summary statistic to measure the measurement of tumor architecture in 3D space *without relying on a black-box model*. The central hypothesis of this fellowship application is that micro-CT can be adapted to generate 3D whole-biopsy images of prostate cancer and other soft-tissue tumor biopsies, revealing previously unmeasured variation in glandular structure and providing novel insight into previously unmeasured phenotypic heterogeneity. Preliminary results support the ability of our team to conduct this work, as we were able to collect proof-of-concept micro-CT images of needle core biopsy sections that demonstrate readily discernible glandular lumen and cell nuclei. This advancement of 3D histopathology and computational topology will serve public health needs by improving the diagnosis of prostate cancer and potentially other soft-tissue malignancies. Through micro-CT parameter testing and 3D atlasing (Aim 1), histological comparison and non-inferiority testing (Aim 2), and 3D topological modeling (Aim 3), this proposal will contribute to our ability to measure tumor phenotype and heterogeneity.

PROJECT NARRATIVE

Cancer diagnosis relies on the assessment of cell and tissue morphology from histology, but methods to quantify these features are limited due to the 2-dimensional nature of slide-based imaging. This project aims to address these shortcomings by developing imaging and analytical approaches to perform 3D histopathology of soft tissue biopsies, with an emphasis on prostate cancer. We propose that this will yield unbiased visualization and measurement of tumor architecture otherwise inaccessible to traditional histopathology.

- Primakov SP, Ibrahim A, van Timmeren JE, Wu G, Keek SA, Beuque M, Granzier RWY, Lavrova E, Scrivener M, Sanduleanu S, Kayan E, Halilaj I, Lenaers A, Wu J, Monshouwer R, Geets X, Gietema HA, Hendriks LEL, Morin O, Jochems A, Woodruff HC, Lambin P. Automated detection and segmentation of non-small cell lung cancer computed tomography images. *Nat Commun.* Nature Publishing Group; 2022 Jun 14;13(1):3423.
- Katiyar P, Schwenck J, Frauenfeld L, Divine MR, Agrawal V, Kohlhofer U, Gatidis S, Kontermann R, Königsrainer A, Quintanilla-Martinez L, la Fougère C, Schölkopf B, Pichler BJ, Disselhorst JA. Quantification of 10. 11. 12. 13. 14. 15. 16. 17. 18. intratumoural heterogeneity in mice and patients via machine-learning models trained on PET–MRI data. *Nat Biomed Eng.* Nature Publishing Group; 2023 Aug;7(8):1014–1027.
- Lawson P, Sholl AB, Brown JQ, Fasy BT, Wenk C. Persistent Homology for the Quantitative Evaluation of Architectural Features in Prostate Cancer Histology. *Sci Rep.* 2019 Feb 4;9:1139. PMID: PMC6361896
- Crawford L, Monod A, Chen AX, Mukherjee S, Rabadán R. Predicting Clinical Outcomes in Glioblastoma: An Application of Topological and Functional Data Analysis. *Journal of the American Statistical Association.* Taylor & Francis; 2020 Jul 2;115(531):1139–1150.
- Gordetsky J, Epstein J. Grading of prostatic adenocarcinoma: current state and prognostic implications. *Diagn Pathol.* 2016 Mar 9;11:25. PMID: PMC4784293
- Epstein JI. Prostate cancer grading: a decade after the 2005 modified system. *Mod Pathol.* Nature Publishing Group; 2018 Jan;31(1):47–63.
- Liu JTC, Glaser AK, Bera K, True LD, Reder NP, Eliceiri KW, Madabhushi A. Harnessing non-destructive 3D pathology. *Nat Biomed Eng.* Nature Publishing Group; 2021 Mar;5(3):203–218.
- Ozkan TA, Eruyar AT, Cebeci OO, Memik O, Ozcan L, Kuskonmaz I. Interobserver variability in Gleason histological grading of prostate cancer. *Scand J Urol.* 2016 Dec;50(6):420–424. PMID: 27416104
- Koyuncu C, Janowczyk A, Farre X, Pathak T, Mirtti T, Fernandez PL, Pons L, Reder NP, Serafin R, Chow SS, Viswanathan VS, Glaser AK, True LD, Liu JT, Madabhushi A. Visual assessment of 2D levels within 3D pathology datasets of prostate needle biopsies reveals substantial spatial heterogeneity. *Lab Invest.* 2023 Oct 17;100265. PMID: 37858679
- Flach RN, Willemse PPM, Suelmann BBM, Deckers IAG, Jonges TN, van Dooijeweert C, van Diest PJ, Meijer RP. Significant Inter- and Intralaboratory Variation in Gleason Grading of Prostate Cancer: A Nationwide Study of 35,258 Patients in The Netherlands. *Cancers (Basel).* 2021 Oct 27;13(21):5378. PMID: PMC8582481
- Albers J, Pacilé S, Markus MA, Wiart M, Vande Velde G, Tromba G, Dullin C. X-ray-Based 3D Virtual Histology-Adding the Next Dimension to Histological Analysis. *Mol Imaging Biol.* 2018 Oct;20(5):732–741. PMID: 29968183
- Katsamenis OL, Olding M, Warner JA, Chatelet DS, Jones MG, Sgalla G, Smit B, Larkin OJ, Haig I, Richeldi L, Sinclair I, Lackie PM, Schneider P. X-ray Micro-Computed Tomography for Nondestructive Three-Dimensional (3D) X-ray Histology. *Am J Pathol.* 2019 Aug;189(8):1608–1620. PMID: PMC6680277

13. Töpperwien M, Markus A, Alves F, Salditt T. Contrast enhancement for visualizing neuronal cytoarchitecture by propagation-based x-ray phasecontrast tomography. *Neuroimage*. 2019 Oct 1;199:70–80. PMID: 31129306
14. Pinkert-Leetsch D, Frohn J, Ströbel P, Alves F, Salditt T, Missbach-Guentner J. Three-dimensional analysis of human pancreatic cancer specimens by phase-contrast based X-ray tomography – the next dimension of diagnosis. *Cancer Imaging*. 2023 May 2;23(1):43.
15. Frost J, Schmitzer B, Töpperwien M, Eckermann M, Franz J, Stadelmann C, Salditt T. 3d Virtual Histology Reveals Pathological Alterations of Cerebellar Granule Cells in Multiple Sclerosis. *Neuroscience*. 2023 Jun 1;520:18–38.
16. Yakovlev MA, Vanselow DJ, Ngu MS, Zaino CR, Katz SR, Ding Y, Parkinson D, Wang SY, Ang KC, La Riviere P, Cheng KC. A wide-field micro-computed tomography detector: micron resolution at half-centimetre scale. *J Synchrotron Rad. International Union of Crystallography*; 2022 Mar 1;29(2):505–514.
17. Robbins & Cotran Pathologic Basis of Disease - 10th Edition [Internet]. [cited 2023 Nov 16]. Available from: shop.elsevier.com/books /robbins-and-cotran-pathologic-basis-of-disease/kumar/978-0-323-53113-9
18. Glaser AK, Bishop KW, Barner LA, Susaki EA, Kubota SI, Gao G, Serafin RB, Balaram P, Turschak E, Nicovich PR, Lai H, Lucas LAG, Yi Y, Nichols EK, Huang H, Reder NP, Wilson JJ, Sivakumar R, Shamskhou E, Stoltzfus CR, Wei X, Hempton AK, Pende M, Murawala P, Dodt HU, Imaizumi T, Shendure J, Beliveau BJ, Gerner MY, Xin L, Zhao H, True LD, Reid RC, Chandrashekhar J, Ueda HR, Svoboda K, Liu JTC. A hybrid open-top lightsheet microscope for versatile multi-scale imaging of cleared tissues. *Nat Methods*. Nature Publishing Group; 2022 May;19(5):613–619.
19. Cheng KC, Xin X, Clark DP, La Riviere P. Whole-animal imaging, gene function, and the Zebrafish Phenome Project. *Curr Opin Genet Dev*. 2011 Oct;21(5):620–629. PMCID: PMC3413372
20. Ding Y, Vanselow DJ, Yakovlev MA, Katz SR, Lin AY, Clark DP, Vargas P, Xin 34. 35. 36. X, Copper JE, Canfield VA, Ang KC, Wang Y, Xiao X, De Carlo F, van Rossum DB, La Riviere P, Cheng KC. Computational 3D histological phenotyping of whole zebrafish by X-ray histotomography. White RM, Stainier DY, Ekker SC, editors. eLife. eLife Sciences Publications, Ltd; 2019 May 7;8:e44898.
21. Töpperwien M, van der Meer F, Stadelmann C, Salditt T. Correlative x-ray phase-contrast tomography and histology of human brain tissue affected by Alzheimer's disease. *Neuroimage*. 2020 Apr 15;210:116523. PMID: 31935519
22. Katz SR, Yakovlev MA, Vanselow DJ, Ding Y, Lin AY, Parkinson DY, Wang Y, Canfield VA, Ang KC, Cheng KC. Whole-organism 3D quantitative characterization of zebrafish melanin by silver deposition micro-CT. White RM, Ekker SC, White RM, editors. eLife. eLife Sciences Publications, Ltd; 2021 Sep 16;10:e68920.
23. Busse M, Müller M, Kimm MA, Ferstl S, Allner S, Achterhold K, Herzen J, Pfeiffer F. Three-dimensional virtual histology enabled through cytoplasm-specific X-ray stain for microscopic and nanoscopic computed tomography. *Proceedings of the National Academy of Sciences*. *Proceedings of the National Academy of Sciences*; 2018 Mar 6;115(10):2293–2298.

24. Metscher BD. MicroCT for comparative morphology: simple staining methods allow high-contrast 3D imaging of diverse non-mineralized animal tissues. *BMC Physiology*. 2009 Jun 22;9(1):11.
25. Frohn J, Pinkert-Leetsch D, Missbach-Güntner J, Reichardt M, Osterhoff M, Alves F, Salditt T. 3D virtual histology of human pancreatic tissue by multiscale phase-contrast X-ray tomography. *J Synchrotron Radiat*. 2020 Nov 1;27(Pt 6):1707–1719. PMCID: PMC7642968
26. Mohan KA, Forien JB, Sridhar V, Cuadra JA, Parkinson D. Non-Linear Phase-Retrieval Algorithms for X-ray Propagation-Based Phase-Contrast Tomography [Internet]. arXiv; 2023 [cited 2023 Jul 28]. Available from: arxiv.org/abs/2305.00334
27. Polikarpov M, Vila-Comamala J, Wang Z, Pereira A, van Gogh S, Gasser C, Jefimovs K, Romano L, Varga Z, Lång K, Schmeltz M, Tessarini S, Rawlik M, Jermann E, Lewis S, Yun W, Stampanoni M. Towards virtual histology with X-ray grating interferometry. *Sci Rep*. Nature Publishing Group; 2023 Jun 3;13(1):9049.
28. Riedel M, Taphorn K, Gustschin A, Busse M, Hammel JU, Moosmann J, Beckmann F, Fischer F, Thibault P, Herzen J. Comparing x-ray phasecontrast imaging using a Talbot array illuminator to propagation-based imaging for non-homogeneous biomedical samples. *Sci Rep*. Nature Publishing Group; 2023 Apr 28;13(1):6996.
29. Norvik C, Westöö CK, Peruzzi N, Lovric G, van der Have O, Mokso R, Jeremiassen I, Brunnström H, Galambos C, Bech M, Tran-Lundmark K. Synchrotron-based phase-contrast micro-CT as a tool for understanding pulmonary vascular pathobiology and the 3-D microanatomy of alveolar capillary dysplasia. *American Journal of Physiology-Lung Cellular and Molecular Physiology*. American Physiological Society; 2020 Jan;318(1):L65–L75.
30. google/neuroglancer: [cited 2023 Nov 19]; Available from: zenodo.org/records/5573294
31. Rebello RJ, Oing C, Knudsen KE, Loeb S, Johnson DC, Reiter RE, Gillessen S, Van der Kwast T, Bristow RG. Prostate cancer. *Nat Rev Dis Primers*. Nature Publishing Group; 2021 Feb 4;7(1):1–27.
32. Wasserman L. Topological Data Analysis. *Annual Review of Statistics and Its Application*. 2018;5(1):501–532.
33. Chazal F, Michel B. An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists. *Front Artif Intell*. 2021 Sep 29;4:667963. PMCID: PMC8511823
34. Yakovlev MA, Liang K, Zaino CR, Vanselow DJ, Sugarman AL, Lin AY, Riviere PJL, Zheng Y, Silverman JD, Leichty JC, Huang SX, Cheng KC. Quantitative Geometric Modeling of Blood Cells from X-ray Histotomograms of Whole Zebrafish Larvae. *eLife* [Internet]. eLife Sciences Publications Limited; 2023 Oct 16 [cited 2023 Dec 7];12. Available from: elifesciences.org/reviewed-preprints/89432

FACILITIES AND OTHER RESOURCES

The overall research and scientific environment available to me will strongly support the proposed research through both technological resources and abundant collaboration.

Laboratory: I primarily work in Dr. Keith Cheng's laboratory (co-sponsor) located on the 7th floor of the Biomedical Research Building in Hershey Medical Center. Dr. Cheng has provided me with a personal desk and access to multiple printers and whiteboards, in addition to a shared workstation equipped with 64Gb of RAM and Quadro RTX5000 GPU dedicated for image processing and reconstruction.

The Silverman lab has also provided me with a dedicated workspace that includes a desk, monitor, keyboard, and mouse that I utilize when I commute to University Park weekly. Our lab is located within the Westgate building at the Pennsylvania State University, and is a short walk down the hallway from Dr. Silverman's office where we regularly meet in person and at the whiteboard to discuss our research. Dr. Silverman has also provided me with my primary work computer, a Lenovo Thinkpad X1 Carbon with an intel i7 processor, 32GB of ram, and 1 terabyte of storage. Apart from my personal computing resources the Silverman lab also maintains access to the Roar Computing Cluster at University Park.

Computing Resources: There three image reconstruction computers used by the Cheng Lab: i) (Recon 1) running Windows 11 Enterprise, with 256Gb RAM, 40TB storage and GTX1080i GPU; ii) (Recon 2) with dual boot for Windows 11 Enterprise and Ubuntu, 256Gb RAM, 48TB and nVIDIA RTX8000; iii) (Recon 3) with Windows 11 Enterprise, 256Gb RAM, 48Tb and two nVIDIA A6000 GPU with nvlink. n. Connectivity between the above workstations, the scanner computer and network storage is gigabit Ethernet. The lab is also equipped with one Wacom Cintiq Pro 24, 6 Wacom Intous Pro and 1 Wacom One for segmentation. The laboratory also has 4 iMacs, and one dedicated workstation for genome analysis with 64GB of RAM and 12 TB of storage. The lab is equipped with 2 local network storage units with 80Tb and 40TB respectively. There is also one Canon color laser printer and scanner and an HP all-in-one color printer and scanner. The lab also has an Aperio AT2 slide scanner with a dedicated computer.

Penn State College of Medicine hosts a High Performance Computing (HPC) system that was purchased to provide researchers the computational and storage tools needed to efficiently and effectively process data. The HPC system is dedicated for use by College of Medicine researchers for processing genomic, DNA sequencing, imaging, and other scientific analysis. To comply with requirements for managing grant-supported research data, we will utilize the HPC system for computational and storage services. The Penn State College of Medicine's Research Informatics department provides full support of both the operation and maintenance of the HPC environment. The HPC system is physically located on the Penn State Hershey campus and the data center was designed to Tier III data center standards. The system contains 3 administrative, 10 standards, and 3 high memory compute nodes. The system provides 1 Petabyte of enterprise storage that is divided into 100 Terabytes (TBs) of high-speed scratch space and 900 TBs of usable storage space. The ten compute nodes provide 240 2.5GHz Intel v3 cores (480 threads) with a total of 2560 Gigabytes (GBs) of RAM and the three high memory nodes provide 2.3 GHz 96 v3 cores (192 threads) with a total of 2304 GBs of RAM. The total compute capacity and total RAM of the system is 4864 GBs with the standard and high memory compute nodes providing 10.5 GBs of RAM and 24 GBs of RAM per core respectively. For additional computing tasks, I have access to the Roar computing system at University Park. I specifically am able to use the Silverman Lab's Roar cluster allocation through the Institute for Computational and Data Science at the College of IST. I will have access to the entire cluster in addition to priority (± 1 min wait time) access to my personal allocation which includes one high-memory node (40 cores, 1TB RAM), two standard-memory nodes (each with 40 cores and 256GB RAM), 10TB Active Group Storage, and 20TB Nearline/Archive Storage.

Institutional Environment: Pennsylvania State University (Penn State) is the land grant institution of Pennsylvania. This proposal includes work conducted at both the College of Medicine campus and the University Park campus, the latter of which is the largest of Penn State's 24 campuses. Of note, Penn State University Libraries rank among the top 10 North American research libraries based on the Association of Research Libraries Library Investment Index Rankings. The library system consists of 36 libraries at 24 locations throughout the Commonwealth of Pennsylvania. The University Libraries house a collection of nearly 6 million items, with annual additions of roughly 100,000 volumes. The libraries have access to 579 online databases and other e-resources and subscribe to nearly 118,000 online, full-text journals.

Penn State College of Medicine, where I will primarily work for the duration of this proposal, also provides an

elite setting for me to grow as a physician scientist. I both live and primarily work within a five minute walk each from the Emergency Department where I conduct my clinical exposure training, the Penn State Cancer Institute where I regularly shadow, and the Department of Pathology where I conduct research and read slides with my mentors. Across both campuses where I will work throughout this proposal, I will interface with interdisciplinary graduate students, faculty, post doctoral researchers, physicians, and other healthcare providers that will enhance my training as a physician-scientist. Penn State will allow me to not only develop expertise at the bench and within the field of oncology, but will also encourage me to become well versed across disciplines.

EQUIPMENT

Micro-CT Imaging Equipment: The Cheng lab maintains a laboratory-based micro-CT imaging system with a 10K x 7K CMOS detector with specially-designed optics: Dr. Cheng led the design, building, integration and testing of a custom microCT system with image quality, resolution and throughput sufficient for tissue phenotyping of mm- to cm-scale specimens. The system is enclosed by a custom lead and steel enclosure for the standard laboratory space. To ensure precision of motion we custom-designed a matched rotational and linear actuator system for sub-pixel, helical, and mosaic helical scans of large specimens of tens of mm in size. Because of our unique need for large field-of-view, high-resolution, and high-throughput, we designed a custom optical system which has a 5×3.5 mm² field of view, a $0.5 \mu\text{m}$ diffraction limited resolution and a process to fabricate scintillator wafers with 3-fold higher sensitivity improvement. In its basic imaging configuration of $0.5 \mu\text{m}$ pixel size, the image is in fact under-sampled for its resolution. Sub-pixel ample shifting by $0.25 \mu\text{m}$ will allow us to reconstruct images with an effective pixel size of $0.25 \mu\text{m}$ to achieve the full $0.5 \mu\text{m}$ optical resolution capability of the optics.

RESEARCH & RELATED Senior/Key Person Profile (Expanded)

PROFILE - Project Director/Principal Investigator				
Prefix:	First Name*: Andrew	Middle Name	Last Name*: Sugarman	Suffix:
Position/Title*:	Grad Student			
Organization Name*:	PENNSYLVANIA STATE UNIV HERSEY MED CTR			
Department:				
Division:	College of Medicine			
Street1*:	500 University Drive			
Street2:	P.O. Box 850			
City*:	Hershey			
County:	Dauphin			
State*:	PA: Pennsylvania			
Province:				
Country*:	USA: UNITED STATES			
Zip / Postal Code*:	17033-0850			
Phone Number*:	443-683-0398	Fax Number:	E-Mail*: aqs6915@psu.edu	
Credential, e.g., agency login: ASUGARMAN				
Project Role*:	PD/PI			
Degree Type:	BA			
Other Project Role Category:				
Degree Year: 2015				
File Name				
Attach Biographical Sketch*:	sugarman_biosketch1034805109.pdf			
Attach Current & Pending Support:				

PROFILE - Senior/Key Person				
Prefix:	First Name*: Justin	Middle Name	Last Name*: Silverman	Suffix:
Position/Title*:	Assistant Professor			
Organization Name*:	PENNSYLVANIA STATE UNIVERSITY-UNIV PARK			
Department:				
Division:				
Street1*:	E339 Westgate Building			
Street2:				
City*:	University Park			
County:	Centre			
State*:	PA: Pennsylvania			
Province:				
Country*:	USA: UNITED STATES			
Zip / Postal Code*:	16802-6823			
Phone Number*:	814-863-8304	Fax Number:	E-Mail*: jds6696@psu.edu	
Credential, e.g., agency login: JSILVE24				
Project Role*:	Other (Specify)			
Degree Type:	MD			
Other Project Role Category: Sponsor				
Degree Year: 2020				
File Name				
Attach Biographical Sketch*:	silverman_biosketch_flat1034805346.pdf			
Attach Current & Pending Support:				

PROFILE - Senior/Key Person				
Prefix:	First Name*: Keith	Middle Name C	Last Name*: Cheng	Suffix:
Position/Title*:	PROF PATH,BIOC,PHARM			

Organization Name*:	PENNSYLVANIA STATE UNIV HERSHEY MED CTR	
Department:	Pathology	
Division:	College of Medicine	
Street1*:	500 University Drive	
Street2:	P.O. Box 850, H059	
City*:	Hershey	
County:	Dauphin	
State*:	PA: Pennsylvania	
Province:		
Country*:	USA: UNITED STATES	
Zip / Postal Code*:	17033-0850	
Phone Number*:	717-531-5635	Fax Number: 717 531-5634
E-Mail*:	kcheng76@gmail.com	
Credential, e.g., agency login: kcheng		
Project Role*:	Other (Specify)	Other Project Role Category: Co-Sponsor
Degree Type:	MD/PhD	Degree Year: 1987
Attach Biographical Sketch*:	File Name drcheng_biosketch1034805107.pdf	
Attach Current & Pending Support:		

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Andrew Lee Sugarman

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

POSITION TITLE: MD/PhD Student

EDUCATION/TRAINING (*Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.*)

INSTITUTION AND LOCATION	DEGREE <i>(if applicable)</i>	Start Date MM/YYYY	Completion Date MM/YYYY	FIELD OF STUDY
Oberlin College	B.A.	08/2015	05/2019	Biochemistry with Honors
Pennsylvania State University	M.D.	06/2020	-	Medical School
Pennsylvania State University	P.h.D	06/2020	-	Bioinformatics and Genomics

A. Personal Statement

My long-term goal is to become a practicing physician-scientist whose work translates novel imaging technology and statistical methods to improve patient outcomes. Upon completion of MD/PhD training, I plan to pursue a research-track residency in internal medicine with a fellowship in hematology and oncology, where I will develop my clinical skills and progress towards independent academic research. I became determined to train as a physician-scientist after shadowing Dr. Mark Levis on the bone marrow transplant service at Johns Hopkins Hospital. I found the disease processes underlying malignant hematology to be especially interesting, and the patient population left a lasting impression on me that persisted through my first two years of medical school. During and after my BA at Oberlin, my time as a research assistant in the Tolbert lab at Case Western heavily influenced my graduate pursuits and cemented my desire to become a physician scientist. Engaging in a range of technical basic science research prior to graduate school has positioned me well to take on a project that applies developing imaging techniques and computational methods to a rapidly emerging problem in virtual histology. Diagnosis of human cancer relies on histology, a qualitative assessment with life-changing implications on treatment decisions. However, histology is limited by field of view and sample size, requiring destruction of the sample in the process. I have sought a co-mentorship between experts in 3D imaging and statistical methods to prepare myself to address these problems. 3-dimensional imaging of tumors at histological resolution will be invaluable to this, and the Cheng Lab has published on custom micro-computed tomography technology that has achieved this field of view and resolution with model organisms (Yakovlev et. al 2022, Ding, et. al 2019). Analysis of these images requires segmentation of tissues and cells of interest, often requiring the use of complex, black-box machine learning models such as neural networks (Yakovlev et. al 2023). The Silverman Lab is well versed in statistical methods that deal with uncertainty, which will be a crucial step in the deployment of these models to the quantification of cell morphology. Combining these unique technologies towards the study of human cancer will prepare me for a career in translational research, one in which I aspire to specialize in hematologic oncology and treat patients in bone marrow transplant. I look forward to how I can contribute to these projects as I train to become a well-rounded physician scientist.

B. Positions, Scientific Appointments and Honors

2020 - 2021	NIH T32 Fellowship, Penn State College of Medicine Medical Scientist Training Program
2019 - 2020	Research Assistant, Tolbert Lab, Case Western Reserve University Department of Chemistry
2019	American Chemical Society Organic Division Undergraduate Award
2019	Sigma Xi Scientific Research Honor Society
2018	Summer Internship, Tolbert Lab, Case Western Reserve University Department of Chemistry
2015 - 2019	John F. Oberlin Scholarship

Leadership Positions:

2020-Present Treasurer of the Hematology/Oncology Interest Group, Penn State College of Medicine

2020-Present Secretary of the Exercise is Medicine Group, Penn State College of Medicine

2020-Present Treasurer of the Cycling Club, Penn State College of Medicine

2020-Present Student Government IT Chair, Penn State College of Medicine

2022-Present Social Chair, Penn State MSTP PSSA

C. Contributions to Science

Undergraduate Research: I began my research career in the lab of Dr. Albert Matlin, where I performed organic synthesis experiments testing enantioselective natural products synthesis. I enjoyed the daily opportunity to solve new problems, and after taking advanced organic chemistry courses in the spring of my sophomore year I pivoted to the lab of Dr. Duy (Zoey) Hua for the entirety of my junior year. I performed molecular dynamics simulations of human spleen tyrosine kinase, and developed a strong interest in structural biology and an inclination towards computational work. This motivated me to seek departmental funding to support a summer internship in the Tolbert lab where I continued conducting research after graduation. My final year of research at Oberlin College focused on chemical modifiers for self-polymerizing polydopamine reactions, which culminated in a successful honors thesis defense.

Research Assistantship: I joined the lab of Dr. Blanton Tolbert in the Department of Chemistry at Case Western Reserve University. I primarily performed molecular dynamics simulations, integrating NMR and Small Angle X-ray Scattering (SAXS) data to solve the structures of RNA, RNA-drug, RNA-protein, and RNA-drug-protein interactions. Our work culminated in the analysis of a small molecule that selectively binds to a stem loop region of the ribosomal entry site in enterovirus-71 (1) and solution structures of HIV-1 RNA structural elements.

1. Davila-Calderon J, Patwardhan NN, Chiu LY, **Sugarman A**, Cai Z, Penutmutchu SR, Li ML, Brewer G, Hargrove AE, Tolbert BS. IRES-targeting small molecule inhibits enterovirus 71 replication via allosteric stabilization of a ternary complex. *Nat Commun.* 2020 Sep 22;11(1):4775. doi: 10.1038/s41467-020-18594-3. PMID: 32963221; PMCID: PMC7508794.
2. Chiu LY, Emery A, Jain N, **Sugarman A**, Kendrick N, Luo L, Ford W, Swanstrom R, Tolbert BS. Encoded Conformational Dynamics of the HIV Splice Site A3 Regulatory Locus: Implications for Differential Binding of hnRNP Splicing Auxiliary Factors. *J Mol Biol.* 2022 Sep 30;434(18):167728. doi: 10.1016/j.jmb.2022.167728. Epub 2022 Jul 21. PMID: 35870649; PMCID: PMC9945881.
3. Luo L, Chiu LY, **Sugarman A**, Gupta P, Rouskin S, Tolbert BS. HnRNP A1/A2 Proteins Assemble onto 7SK snRNA via Context Dependent Interactions. *J Mol Biol.* 2021 Apr 30;433(9):166885. doi: 10.1016/j.jmb.2021.166885. Epub 2021 Mar 5. PMID: 33684393; PMCID: PMC8091503.
4. Donlic A, Swanson EG, Chiu LY, Wicks SL, Juru AU, Cai Z, Kassam K, Laudeman C, Sanaba BG, **Sugarman A**, Han E, Tolbert BS, Hargrove AE. R-BIND 2.0: An Updated Database of Bioactive RNA-Targeting Small Molecules and Associated RNA Secondary Structures. *ACS Chem Biol.* 2022 Jun 17;17(6):1556-1566. doi: 10.1021/acschembio.2c00224. Epub 2022 May 20. PMID: 35594415; PMCID: PMC9343015.

Graduate Research: Under the co-mentorship of Dr. Justin Silverman and Dr. Keith Cheng I have contributed to the acquisition and analysis of high-resolution micro-computed tomography images of soft tissue samples.

For example, in recent work led by Maksim Yakovlev in Elife I performed statistical analysis on 3-dimensional

segmentation masks of Zebrafish red blood cells. This work aims to build a framework for Zebrafish researchers to quantitatively compare phenotypes within models of disease.

5. Yakovlev Maksim A., Liang Ke, Zaino Carolyn R., Vanselow Daniel J., **Sugarman Andrew L.**, Lin Alex Y., La Riviere Patrick J., Zheng Yuxi, Silverman Justin D., Leichty John C., Huang Sharon X., Cheng Keith C. (2023) Quantitative Geometric Modeling of Blood Cells from X-ray Histotomograms of Whole Zebrafish Larvae eLife 12:RP89432. doi.org/10.7554/eLife.89432.1

I aim to build upon this work and apply similar methods to the quantitative analysis of tumor biopsies. My current project focuses on the phase-contrast based imaging of tumor tissue blocks, from which our group aims to extract measurements of microanatomical features, measure 3D spatial variation in key diagnostic elements such as prostate glands, and build methods to better quantify tumor heterogeneity in multiple tissue types. My thesis work will also focus on the computational methods required to refine, analyze, and distribute the findings from these imaging experiments. I will focus on the use of methods from topological data analysis in the analysis of 3D data, which will sharpen my skills in programming and statistics, preparing me to contribute to a wide range of computational problems. I look forward to improving my knowledge of the basic and clinical sciences as I prepare for a career as an independent investigator.

D. Scholastic Performance

YEAR	COURSE TITLE	GRADE
2023	STAT 555: Statistical Genomics	A
2023	MCIBS: 554 Bioinformatics 1	A
2023	BIOL 428: Population Genetics	A
2023	BGEN 600: Thesis Research	A
2023	BGEN 590: Colloquium	A
2022	BMMB 852: Applied Bioinformatics	A
2022	BGEN 551: Genomics	A
2022	BGEN 597: Special Topics Bioinformatics and Genomics Core	A-
2022	MICRO 581: Immunology A	A-
2021	NBS 723: Neural and Behavioral Sciences	P
2021	SHS 721: Health Systems	P
2021	NBS 723: Neuroanatomy	P
2021	MEP 721: Med Ethics Prof	P
2021	HMN 723: Communications	P
2021	GI 723: GI	P
2021	FPCC 723: Foundations of Patient Centered Care - 3	P
2021	ENREP 721: Endo and Repro	P
2021	BMS 506A: Human Health and Disease A	P
2021	REN 713: Renal Med	P
2021	HMN 715: Critical Thinking	P
2021	FPCC 714: Foundations of Patient Centered Care - 2	P
2021	FORM 713: Form and Function	P
2021	BMS 506B: Human Health and Disease B	P
2020	CARES 713: CardioResp Med	P
2020	HMN 714: Mind-Body	P
2020	HDHR 711: Host Defense Host Response	P
2020	HMN 713: Medical Humanities	P
2020	FPCC 713: Foundations of Patient Centered Care - 1	P
2020	SPM 711: Scientific Principles of Medicine	P

2020	SHS 711: Health Systems	P
2020	TRANS 711: Transitions to Medical School	P
2019	CHEM 405: Topics in Organic Chemistry	B
2019	PHIL 201: Reason and Argument	A-
2019	CHEM 339: Quantum Chemistry and Kinetics	B-
2019	CHEM 526F Research - Full	P
2019	CHEM WT002: Honors Research	Full
2018	CHEM 327: Synthesis Laboratory	B
2018	CHEM 374: Biochemistry	B
2018	CHEM 525: Research - Half	P
2018	STAT 114: Introduction to Biostatistics	B+
2018	BIOL 213: Mol Biol, Cell Biol, & Biochem	B
2018	CHEM 213: Inorganic Chemistry	B
2018	CHEM 526F: Research - Full	P
2018	ENGL 140: Arthurian Fictions	B+
2018	CHEM WT002: Computational Structural Bio	Full
2017	PSYC 100: Intro to Psychological Sciences	A
2017	HIST 101: Medieval and Early Modrn Eur Hist	A
2017	CHEM 525F: Research - Full	P
2017	CHEM 211: Analytical Chemistry	B
2017	PHYS104: Elementary Physics II	B+
2017	CHEM 525H: Research - Half	P
2017	CHEM 325: Organic Mechanisms and Synthesis	A
2017	CHEM 254: Bioorganic Chemistry	A
2017	CHEM WT001: Organic Chemistry Research	Full
2016	RELG 153: Intro to Relg: Purity and Pollution	A-
2016	PHYS 110: Mechanics and Relativity	B
2016	PHIL 126: Problems of Philosophy	A-
2016	CHEM 205: Principles of Organic Chemistry	B+
2016	GEOL 122: Natural Hazards	A-
2016	ECON 101: Principles of Economics	A
2016	CHEM 102: Chemical Principles	B+
2016	BIOL 100: Organismal Biology	B+
2016	WT004: Shadowing of Trauma Surgeon	Full
2015	MATH 134: Calculus II	B
2015	PYSP 034: Values of Higher Education	B
2015	CHEM 101: Structure and Reactivity	A-
2015	ATHL 159: Ind. Baseball Skills Training	P
2015	ANTH 102: Human Origins	B

* The first two years of medical school coursework at Penn State College of Medicine are graded pass("P")/fail only

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. DO NOT EXCEED FIVE PAGES.

NAME: Silverman, Justin D

eRA COMMONS USER NAME (credential, e.g., agency login): JSILVE24

POSITION TITLE: Assistant Professor

EDUCATION/TRAINING (*Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.*)

INSTITUTION AND LOCATION	DEGREE (if applicable)	END DATE MM/YYYY	FIELD OF STUDY
Johns Hopkins University, Baltimore, Maryland	BS	05/2011	Physics and Biophysics
Duke University, Durham, North Carolina	PHD	05/2019	Computational Biology and Bioinformatics
Duke University, Durham, North Carolina	MD	05/2020	

A. Personal Statement

I am a statistician (PhD) and physician (MD). My research focuses on developing robust and efficient statistical methods for the analysis of complex biomedical data. From a methodological perspective I focus on the analysis imperfect data that lead to partially identified statistical models. This includes but is not limited to challenges stemming from unmeasured confounding, informative missingness, and measurement bias. From an applied perspective I have particular expertise in the analysis of non-standard high-dimensional data such as compositional data, functional data, and shape data. I have applied my work widely in fields ranging from genomics to neuroscience.

I have successfully mentored numerous graduate students in a variety of programs ranging from statistics to animal sciences. Within the past year, three of my mentees have been awarded graduate fellowships and/or funding through T32 funded training grants. I have a long-standing collaboration with Dr. Cheng who is co-sponsoring this proposal. Some of our recent work modeling the shape of blood cells imaged by micro-CT was recently released and is currently in revision at eLife.

Overall I have the interest, technical expertise, and mentoring experience needed to facilitate the successful completion of the proposed work and assist Andrew in achieving his long-term career goals.

1. Yakovlev,Maksim,A, Liang,Ke,, Zaino,Carolyn,R, Vanselow,Daniel,J, Sugarman,Andrew,L, Lin,Alex,Y, La Riviere,Patrick ,J, Xheng,Yuxi,, Silverman,Justin,D, Leichty,John,C, Huang,Sharon,X, Cheng,Keith,C. Quantitative Geometric Modeling of Blood Cells from X-ray Histotomograms of Whole Zebrafish Larvae. *eLife* [Preprint]. 2023 May 23. Available from: <https://elifesciences.org/reviewed-preprints/89432v1/reviews#tab-content> DOI: 10.7554/eLife.89432.1
2. Silverman JD, Roche K, Holmes ZC, David LA, Mukherjee S. Bayesian Multinomial Logistic Normal Models through Marginally Latent Matrix-T Processes. *Journal of Machine Learning Research*. 2022 February 01; 23(7):1-42. Available from: <http://jmlr.org/papers/v23/19-882.html>
3. Silverman JD, Hupert N, Washburne AD. Using influenza surveillance networks to estimate state-specific prevalence of SARS-CoV-2 in the United States. *Sci Transl Med*. 2020 Jul 29;12(554) PubMed Central PMCID: PMC7319260.
4. Silverman JD, Washburne AD, Mukherjee S, David LA. A phylogenetic transform enhances analysis of compositional microbiota data. *Elife*. 2017 Feb 15;6 PubMed Central PMCID: PMC5328592.

B. Positions, Scientific Appointments and Honors

Positions and Scientific Appointments

2020 - Assistant Professor, Penn State University, State College, PA

Honors

2022	Dean's Circle of Teaching Excellence, The Pennsylvania State University
2017	Mitchell Meritorious Research Travel Award, Duke University
2017	Best Young Presentation, Compositional Data Analysis Workshop
2011	H. Keffer Hartline Award for Outstanding Scholarship in Biophysics, Johns Hopkins University
2011	Donald E. Kerr Memorial Award for Excellence in Physics, Johns Hopkins University
2011	Phi Beta Kappa, Johns Hopkins University
2010	Barry M. Goldwater Scholarship, Barry Goldwater Foundation
2010	Provost Undergraduate Research Award, Johns Hopkins

C. Contribution to Science

1. **Bayesian count-compositional modeling of sequence count data.** While popular, direct log-ratio transformation of sequence count data are plagued by largely heuristic approaches to handling zero values and an inability to account for variation due to counting also present in these data. To address these limitations I have developed many Bayesian count-compositional models demonstrating how core concepts from the field of compositional data analysis could be applied as different choices of link functions in Bayesian hierarchical multinomial logistic-normal models. Beyond linear models, I have developed longitudinal extensions of these models based upon generalized dynamic linear models, extensions for non-linear regression based upon generalized Gaussian process regression models, and dimensionality reduction methods based upon joint modeling of multi-omic studies. I have also developed popular software for inferring these models including fido and Songbird. Underpinning fido, I developed the theory of Marginally Latent Matrix-t Processes as a means of developing scalable and accurate Posterior estimation for these models that is often 10,000-100,000 times more efficient than MCMC with provable error bounds. These methods have been adopted by numerous research groups and are being used in projects across the NIH and NSF. Beyond methodological development, I have also demonstrated how these models, combined with novel experimental designs can be used to mitigate PCR bias and have established best practices for modeling zeros in sequence count data.

- a. Silverman JD, Roche K, Holmes ZC, David LA, Mukherjee S. Bayesian Multinomial Logistic Normal Models through Marginally Latent Matrix-T Processes. *Journal of Machine Learning Research*. 2022 February 01; 23(7):1-42. Available from: <http://jmlr.org/papers/v23/19-882.html>
- b. Silverman JD, Bloom RJ, Jiang S, Durand HK, Dallow E, Mukherjee S, David LA. Measuring and mitigating PCR bias in microbiota datasets. *PLoS Comput Biol*. 2021 Jul;17(7):e1009113. PubMed Central PMCID: PMC8284789.
- c. Silverman JD, Roche K, Mukherjee S, David LA. Naught all zeros in sequence count data are the same. *Comput Struct Biotechnol J*. 2020;18:2789-2798. PubMed Central PMCID: PMC7568192.
- d. Silverman JD, Durand HK, Bloom RJ, Mukherjee S, David LA. Dynamic linear models guide design and analysis of microbiota studies within artificial human guts. *Microbiome*. 2018 Nov 12;6(1):202. PubMed Central PMCID: PMC6233358.

2. **Developments in Scale Reliant Inference** I am responsible for the conceptualization and the theoretical foundations underlying the emerging field of Scale Reliant Inference (SRI). By posing SRI as an estimation problem involving partially identified models I was able to prove the first non-axiomatic limits on the analysis of multivariate survey data where the variation in the sum of each

measurement is arbitrary and does not reflect the scale (i.e., size) of the underlying systems being surveyed. With application to the analysis of sequence count data, these results have resolved long-standing debates within the bioinformatics community regarding the extend to which modeling assumptions (e.g., normalizations) can overcome the limitations of the observed data. Through theoretical and empirical studies, I have shown that avoiding a troubling statistical phenomena called unacknowledged bias requires considering uncertainty in those modeling assumptions themselves. To facilitate such analyses I developed a specialized type of Bayesian hierarchical models called Scale Simulation Random Variables (SSRVs). I have shown that by addressing unacknowledged bias SSRVs can often reduce false positive rates while retaining statistical power in a wide range of applications from differential abundance analysis to inter-gene correlation analysis. Numerous groups have started to apply tools and theory from SRI to their studies of sequence count data and SSRVs are now integrated into popular modeling tools such as ALDEx2.

- a. McGovern,Kyle,C, Nixon,Michelle,P, Silverman,Justin,D. Addressing Erroneous Scale Assumptions in Microbe and Gene Set Enrichment Analysis. *bioRxiv* [Preprint]. 2023 March 23. Available from: <https://www.biorxiv.org/content/10.1101/2023.03.10.532120v1> DOI: 10.1101/2023.03.10.532120
- b. Nixon,Michelle,P, Letourneau,Jeffrey,, David,Lawrence,A, Lazar,Nicole,A, Mukherjee,Sayan,, Silverman,Justin,D. Scale Reliant Inference. *arXiv* [Preprint]. 2022 January 10 [revised 2023 February 10]. Available from: <https://arxiv.org/abs/2201.03616> DOI: 10.48550/arXiv.2201.03616

3. **Merging phylogenetics and compositional data analysis to enhance the analysis of microbiome data.** In addition to informing the relative abundance of taxa, sequence similarity in microbiome data can be used to estimate the evolutionary relationships between taxa. Just as dogs and wolves are more likely to be found in similar environments utilizing similar resources than dogs and dolphins, evolutionary structure in microbial communities can provide critical insights on the forces that structure these communities. Early in my career I developed the phylogenetic isometric log-ratio transform (PhILR). PhILR uses the phylogenetic relationships to overcome limitations with existing compositional data analysis methods. Under this transform, microbial compositions are projected from the compositional simplex into Real space with coordinates that take on a phylogenetic interpretation: the balance in the abundance between neighboring phylogenetic clades. While this provided new avenues for interpreting these data, this transform itself had a number mathematical advantages over the more standard additive or centered log-ratio transforms at the heart of compositional data analysis. Building on this work I participated in the development of the generalized phylofactorization model which is a compositional factor model with factors constrained by the phylogenetic relationships between microbial taxa. Together these methods have been widely used to identify forces structuring both host-associated and environmental microbial communities. The PhILR transform has also been used by a number of groups as a pre-processing step to improve the performance of black-box machine learning algorithms.

- a. Washburne AD, Silverman JD, Morton JT, Becker DJ, Crowley D, Mukherjee S, David LA, Plowright RK. Phylofactorization: a graph partitioning algorithm to identify phylogenetic scales of ecological data. *Ecological monographs*. 2019 February 19; 98(2):e01353. DOI: 10.1002/ecm.1353
- b. Silverman JD, Washburne AD, Mukherjee S, David LA. A phylogenetic transform enhances analysis of compositional microbiota data. *Elife*. 2017 Feb 15;6 PubMed Central PMCID: PMC5328592.
- c. Washburne AD, Silverman JD, Leff JW, Bennett DJ, Darcy JL, Mukherjee S, Fierer N, David LA. Phylogenetic factorization of compositional data yields lineage-level associations in microbiome datasets. *PeerJ*. 2017;5:e2969. PubMed Central PMCID: PMC5345826.

4. **Developed novel surveillance systems for quantifying and tracking COVID-19 burden.** Stating in

January 2020, my expertise in multivariate time-series analysis and clinical medicine were called upon to aid in the response to the COVID-19 pandemic. Mounting an effective, early responses to an emerging infectious disease requires an accurate picture of the prevalence of the disease and the geographic spread. Yet this type of tracking is often hindered by resource limitations that hinder the widespread use of specific molecular tests. Early in 2020, we found that the rate at which patients present to their physician complaining of influenza-like symptoms could be used to quantify and track the burden of COVID-19 at the state-level in the United States. Using a large surveillance network designed by the CDC to track influenza (ILINet) we provided some of the earliest evidence that the vast majority of cases were going undetected. Between March 8 and March 28th, we found that approximately one in every 80 cases were identified and included in nationally reported cases counts. This work made national news and led to a number of collaborations with local, state, and national governmental organizations helping to direct the response to the COVID-19 pandemic. Later, as key variants of concern were emerging out of England, South Africa, and Brazil, I developed a number of statistical tools that were used to quantify the relative fitness, transmissible, and burden of these variants. Through this work we provided some of the earliest evidence that the B.1.1.7 variant emerging in England was more transmissible and conferred higher mortality risk than previously dominant the wild type.

- a. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, Pearson CAB, Russell TW, Tully DC, Washburne AD, Wenseleers T, Gimma A, Waites W, Wong KLM, van Zandvoort K, Silverman JD, Diaz-Ordaz K, Keogh R, Eggo RM, Funk S, Jit M, Atkins KE, Edmunds WJ. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*. 2021 Apr 9;372(6538) PubMed Central PMCID: PMC8128288.
 - b. Silverman JD, Hupert N, Washburne AD. Using influenza surveillance networks to estimate state-specific prevalence of SARS-CoV-2 in the United States. *Sci Transl Med*. 2020 Jul 29;12(554) PubMed Central PMCID: PMC7319260.
5. **Open Source Scientific Software.** I have authored and currently maintain 13 open source software packages including multiple software packages for statistical analysis of sequence count data (e.g., *PhILR*, *RcppHungarian*, *RcppCoDA*, *NVC*, *driver*, and *fido*). I am committed to long-term support and maintenance of my software. I authored *PhILR* at the start of 2016 and continue to routinely collaborate with the community to enhance this software: improving documentation, reviewing pull requests, authoring new features, and fixing bugs. *PhILR* is published on Bioconductor and is downloaded approximately 3000 times per year. To date *PhILR* has been used in numerous research projects including studying the role of gut microbiota in body composition (Ang et al. *eLife*, 2021) to uncovering the interaction between host genetics and microbiota in autism spectrum disorder (Buffington et al. *Cell*, 2021), and even inferring changing ecology of human oral microbiota over the past 100,000 years (Yates et al. *PNAS*, 2021). I am also the author of *RcppHungarian*, a fast C++ header library with bindings to the R programming language for solving minimum cost bipartite matching problems. *RcppHungarian* is published on CRAN and has supported a number of studies including novel methods for large-scale analysis of spectral imaging data (Paradis, International Journal of Applied Earth OBservation and Geoinformation, 2022). One of my post popular software packages, *fido*, is a fast and flexible implementation of a wide range of Bayesian Multinomial Logistic Normal models. While still in a beta release, *fido* has already been applied in a number of studies ranging from identifying the role of short-chain fatty acid production in pediatric obesity (Holmes et al. *mBio*, 2020) to biomarker discovery in Parkinson's disease (Pereira et al. *bioRxiv*, 2021). Recently, we even used *fido* quantify the influenza-like illness surge at the start of the COVID-19 epidemic in the United States (Silverman et al. *Science Translational Medicine*, 2020). Of note, *fido* is more commonly known as *stray* yet was recently renamed due to a name collision with another package on CRAN.

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Cheng Keith C.

ERA COMMONS USER NAME (credential, e.g., agency login): kcheng

POSITION TITLE: Professor

EDUCATION/TRAINING (*Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.*)

INSTITUTION AND LOCATION	DEGREE (if applicable)	Completion Date MM/YYYY	FIELD OF STUDY
Harvard University, Cambridge, MA	B.A.	06/1976	Biochemical Sciences
New York University School of Medicine, NYC	M.D.	05/1980	Medicine
Brigham & Women's Hospital, Boston, MA	Residency	06/1981	Anatomic Pathology
University of Washington Hospitals, Seattle, WA	Residency	03/1988	Anatomic Pathology
Fred Hutchinson Cancer Research Center & University of Washington, Seattle, WA	Ph.D.	09/1987	Genetic Recombination
University of Washington, Seattle, WA	Sr. Fellow	03/1992	Mechanisms of Mutation

A. Personal Statement

My background in biochemistry, the molecular genetics of recombination and DNA repair, anatomic pathology, virology, and the genetics of cancer led to a desire to understand the molecular and cellular mechanisms that underlie cancer progression and complex traits. My work in pathology led to both an understanding of the cellular and tissue architectural characteristics of cancer and other diseases and a maddening frustration with our inability to quantitatively define the morphological features that tell us what cells are, and what physiological or disease states they are in. Over more some 20 years, we pursued the use of mutant phenotypes to inform us about all the functions of each gene, including the use of histology for forward and reverse genetic screens in zebrafish. I realized during this work that 2D histopathology would need to become 3-dimensional to make whole-organism computational histopathology possible.

Micron-scale computerized tomography (microCT) was a theoretical possibility, but all instances of commercial and synchrotron microCT utilized commercial microscope lenses whose specifications make it impossible to fulfill the necessary combination of centimeter field of view AND submicron voxel resolution. Breaking this barrier would require the design and creation of large-field, high-NA optics similar to that used in photolithography (used in the manufacture of computer chips) so that we could scan 5-10 mm wide fixed and metal-stained biological samples such as juvenile and adult zebrafish and mammalian tissue samples. The large image files required massive computer power to process, visualize, and analyze, initially using large local workstations, now shifting towards high-performance computing. Our Penn State working group decided to pilot a computational, distributional phenomics, whose goal is to bring quantitative, and statistical rigor to tissue analysis. We have used supervised machine learning and modeling to measure the geometric features of virtually every blood cell in whole zebrafish larvae. Given that cells and nuclei can be considered 3D derivatives of spheres, we are collectively working towards defining all of the finite number of cell types in all of nature in health and disease. We have called this idea, which includes *whole-organism* histopathological phenotyping, *Geometry of Life and Disease*. My present passion is to complete pilot experiments required to define the firm foundation needed to make histotomography accessible and useful to scientists world-wide. The anticipated distributional and computational phenomics will facilitate an integrated understanding of how genes, epigenetics, environment and disease define organismal phenotype.

My exploratory, collaborative and interdisciplinary bent is reflected by our: 1) Deciphering of the history of the evolution of human skin color that includes our discovery of a key contributor to the lighter skin color of Europeans (Lamason et al. 2005), 2) a decade's effort to define the key genetic determinants of East Asian and Native

American skin color, 3) Conceptualizing, planning, and creating a web-based atlas of microanatomy (bio-atlas.psu.edu), 4) Organizing and completing a new NIH/Penn State funded Zebrafish Functional Genomics Core, 5) Creating a microCT-based, pan-cellular, 3D, imaging tool for whole, optically opaque vertebrate animals that allows 3D histopathology, X-ray histotomography at Argonne National Laboratory and Lawrence Berkeley National Laboratory, and 6) establishing the Penn State Computational Organismal and Tissue Phenomics initiative, now Geometry of Life and Disease. The latter is being applied across biomedical sciences and science education. Our interdisciplinary research environment, reflected by our recent publications, is well-suited for our proposition to lay hardware and software foundations for user-accessible histotomography.

Ongoing and recently completed projects that I would like to highlight include:

DOE ALS-11657

Cheng (PI)

10/15/22 - 9/30/25 (approximate)

Exploring applications of novel, wide-field, submicron resolution lens and camera systems for microCT (towards defining the Geometry of Life)

2R24OD018559

Cheng (PI)

08/15/19 – 07/31/23

Groundwork for a Synchrotron MicroCT Imaging Resource for Biology (SMIRB)

1R01NS108407

Kim (PI)

09/1/18 – 05/31/23

Architecture of the Neurovascular Unit and Its Function in the Whole Mouse Brain

1R01MH116176

Kim (PI)

06/1/18 – 02/28/23

Brain-Wide Input and Output Wiring Diagram of Oxytocin Neurons and Its Function in Claustrum-Endopiriform Complex

B. Positions, Scientific Appointments, and Honors

Positions and Employment

2019-present	Director, Penn State Computational Phenomics and Geometry of Life and Disease (GOLD) Initiatives
2008-2018	Director, Division of Experimental Pathology, Pennsylvania State University College of Medicine
2007-present	Professor, Departments of Pathology and Biochemistry & Molecular Biology, Pennsylvania State University College of Medicine
1998-2007	Associate Professor; Professor, Departments of Pathology and Biochemistry & Molecular Biology, Pennsylvania State University College of Medicine
1992-1998	Assistant Professor; Professor, Departments of Pathology and Biochemistry & Molecular Biology, Pennsylvania State University College of Medicine

Other Experience and Professional Memberships

Honors

2021-2022 Penn State Institute for Computational and Data Science (ICDS) Faculty Scholar Award

2020-2022 Penn State Institute for Computational and Data Science (ICDS) Faculty Advisory Council

2020-2021 Co-Chair, NIH ORIP Validation of Animal Models Workshop Committee

2019-present, Member, PSU Research Computing and Cyberinfrastructure (RCCI) executive committee

2012-present, Member, Penn State Institute for Personalized Medicine, 2018-present Member, Penn State Research Computing and Cyberinfrastructure/ Group Leader for Cognitive and Immersive Technologies, renamed Emerging and Evolving Technologies

2010-2012 Member, NCRR Linking Animal Models of Human Disease (LAMHDI) Project Team
2009-present, Research Computing Advisory Group, PSCOM
2005-2015 Founding Co-Director, Penn State IBIOS Intercollege program in Bioinformatics and Genomics
2004-present, Founding Curator of the Zebrafish Atlas of Microanatomy, renamed bio-atlas in 2016
2005 FASEB Symposium Chair: Systems Morphogenetics: Biological Context for the Genome Project
2004-present Editorial Board, *Zebrafish*
2004-2005 Co-director, Cross-campus Biology of Neoplasia Course at Pennsylvania State University
2003 Penn State Symposium organizer: Genetic and Functional Genomics Approaches in Model Organisms
1998-2000 Chair, Genetics and Functional Genomics Strategic Planning Committee, PSCOM
1992-present Graduate faculty appointments in Biochemistry & Molecular Biology, Cell & Molecular Biology, and
Intercollege Genetics, Biomedical Sciences (BMS) graduate programs
1992-present Ad hoc reviewer for NSF, *Carcinogenesis*, *Cancer Research*, *Biotechniques*, *Genetics*, *Science*,
Genes & Development, *Nature*, *PLoS Genetics*, *PLoS Biology*

C. Contributions to Science

Hotspots of Homologous Recombination. As a graduate student in the laboratory of Gerald Smith (Fred Hutchinson Cancer Research Institute), I worked on Chi hotspots of recombination defined by the sequence 5'GCTGGTGG3. We were the first to establish a range of genetic hotspot activity associated with sequences similar to Chi in bacteriophage lambda, and to then show that their degree of genetic activity was determined by the degree of single stranded DNA cleavage induced by *E. coli*'s RecBCD enzyme. To test the predictions of models of recombination induced by Chi hotspots, I used a series of clear-plaque mutations to map the distribution of heteroduplex DNA in the region of Chi on phage lambda that were preserved using a mismatch-correction-deficient background.

- **Cheng KC** and Smith GR (1984) Recombinational hotspot activity of Chi-like sequences. *J. Mol. Biol.* 180:371-377. PMID: 6239928.
- **Cheng KC** and Smith GR (1987) Cutting of Chi-like sequences by RecBCD enzyme. *J. Mol. Biol.* 194:747-750. PMID: 2958631.
- **Cheng KC** and Smith GR (1989) Distribution of Chi-stimulated recombinational exchanges and heteroduplex endpoints in phage lambda. *Genetics* 123:5-17 (cover article) PMID: 2530132. PMCID: PMC1203790.

Oxidative DNA damage. As a Damon-Runyon fellow in the laboratory of Larry Loeb (U Wash), I established the mutagenic spectrum of the chemical carcinogen vinyl chloride, and of one of the most important mutagenic intermediates of oxidative DNA damage, which is a byproduct of oxidative metabolism and inflammation leading to formation of the 8-hydroxy (aka 8-oxo) derivative of guanine. We performed a comprehensive in vivo assessment of the mutagenic spectrum of this moiety both as template and substrate. We established mechanisms by which 8-hydroxyG can cause either of two types of mutation, (GC → TA or AT → CG) by mispairing with A, depending upon its presence in template vs. substrate. This work I believe constitutes the first detailed description of how a single DNA base modification can cause two types of mutation.

- **Cheng KC** et al. 1991, The vinyl chloride DNA derivative, N^{2,3}-ethenoguanine, causes G to A transitions in *E. coli*, *Proc. Natl. Acad. Sci., U.S.A.* 88:9974-9978. PMCID: PMC52849.
- **Cheng KC** et al. 1992, 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G to T and A to C substitutions, *J Biol Chem.* 267:166-172. PMID: 1730583; 2335 Google Scholar citations.

Zebrafish as a disease model. My lab's first projects included two of the first genetic screens in zebrafish for cancer phenotypes: a somatic mutator screen (Moore et al. 2006), and the first histological screen for genes involved in nuclear atypia, a key histological feature of cancer (Mohideen et al 2003). We also established tools for large-scale histological screening of zebrafish larvae (Tsao-Wu et al 1998; Moore et al 2002) and participated in the Zebrafish Encode Project (Yang et al. 2020).

- Moore JL, Rush LM, Breneman C, Mohideen MA, **Cheng KC** (2006) Zebrafish genomic instability mutants and dominant cancer susceptibility. *Genetics*. 174(2):585-600. PMID: 11848405 PMCID: PMC1602069. cover article.
- Mohideen MA, Beckwith LG, Tsao-Wu GS, Moore JL, Wong AC, Chinoy MR, **Cheng KC** (2003) Histology-based screen for zebrafish mutants with abnormal cell differentiation. *Dev Dyn.* 228(3): 414-23. PMID: 14579380. First genetic screen in a vertebrate based on histology.
- Lin AY, Ding Y, Vanselow DJ, Katz SR, Yakovlev MA, Clark DP, Mandrell D, Copper JE, van Rossum DB, **Cheng KC** (2018). Rigid Embedding of Fixed and Stained, Whole, Millimeter-Scale Specimens for

Section-free 3D Histology by Micro-Computed Tomography. *J Vis Exp* Oct 17;(140). PMID: 30394379
PMCID: PMC6235553.

- Yang H, Luan Y, Liu T, Lee HJ, Fang L, Wang Y, Wang X, Zhang B, Jin Q, Ang KC, Xing X, Wang J, Xu J, Song F, Sriranga I, Salameh T, Li D, Choudhary MNK, Topczewski J, Wang K, Gerhard GS, Hardison RC, Wang T, **Cheng KC**, Yue F (2020) A map of cis-regulatory elements and 3D genome structures in zebrafish. *Nature*, i588:337-343. PMID: 33239788. PMCID: PMC8183574 ["Zebrafish Encode Project"]

Skin color and Personalized Medicine. We discovered what appears to be the central determinant of the lighter skin phenotype of people of European ancestry: the *A111T* allele of *SLC24A5*. The lighter color of a zebrafish pigment variant was associated with a decrease in the number, size, and pigmentation of the melanosomes. That these features also characterize European skin suggested to me that a polymorphism in the variant gene in zebrafish, *slc24a5*, could play a role in human pigmentation. This suspicion led to the corresponding realization that a coding polymorphism in the orthologous human *SLC24A5* gene is homozygous in every European (CEU) individual in the then-new HapMap database of human polymorphisms. The human gene rescued the zebrafish *golden* phenotype, and African/European admixed individuals are generally lighter in the presence of the mutation (see cover article in *Science* (Lamason et al. 2005). Related work includes the first chapter of scientific considerations of race and skin color, the identification of the zebrafish *albino* gene as *slc45a2*, a known albinism gene in humans, and published in the context of a whole-animal assay for assessing phenotypes caused by human coding polymorphisms (Tsetskhadze ZR et al., 2012 in *PLoS One* 7(10): e47398 10.1371/journal.pone.0047398). Our haplotype analysis of the genomic region surrounding *SLC24A5* in global human populations supports our hypothesis that natural selection for the *A111T* mutation occurred once in human history in the middle east, potentially around the time of the last glacial maximum about 15,000 years ago (Canfield et al. 2013). We have since been probing the genetic origin of light skin in East Asians and Amerindians.

- Lamason RL et al and **Cheng KC** (2005) SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans, *Science* 310:1782-1786 PubMed PMID: 16357253. (1201 Google Scholar citations)
- Tsetskhadze ZR, et al, Kawakami K, **Cheng KC** (2012) Functional assessment of human coding mutations affecting skin pigmentation using zebrafish. *PLoS One* 7(10): e47398 10.1371/journal.pone.0047398. PMID: 23071798 PMCID: PMC3468441.
- Canfield VA, et al, Oppenheimer S, **Cheng KC** (2013) Molecular phylogeography of a human autosomal skin color locus under natural selection, *G3* 3(11): 2059-2067. doi: 10.1534/g3.113.007484. PMID: 24048645 PMCID: PMC3815065.
- Ang KC,et al., **Cheng KC** (2023) Native American genetic ancestry and pigmentation allele contributions to skin color in a Caribbean population. *eLife* 12, e77514. PMID: 37294081. PMCID: PMC10371226.

Phenomics. Our contributions to phenomics began with the idea of analyzing histological images (Canada et al. 2006, 2007, 2008a, 2008b, 2011). Histology, while a powerful tool, has significant limitations: two-dimensional assessments of 3D tissue architecture, and limited sampling. Our current dream is to automate whole-body tissue phenotyping for every cell type and tissue to study the function of every protein-encoding gene in the vertebrate genome, and to assess the phenotypic impact of >100,000 chemicals to which we are potentially exposed during the manufacture of drugs and materials. These issues are covered by a new, evolving field, phenomics (see Cheng KC et al., 2011, 2012, 2016). We have used a systems approach towards creating infrastructure for scientists to study every cell of an entire specimen in 3 dimensions, cutting by computer in any direction, allowing analysis of specific cell types, and enabling quantitative volumetric and pattern analysis. This new imaging method is based on microCT, more specifically X-ray tomographic reconstruction of fixed and stained tissues, utilizing both the monochromatic and parallel beam geometry of the x-rays at the Advanced Photon Source at Argonne National Laboratory based on optimizations (Ding et al. 2018, 2019) and Lawrence Berkeley Laboratory. We are beginning to integrate spatial -omic studies into our development of Computational Phenomics (Van Nuffel et al. 2020 PMID: 33112610).

- **Cheng KC**, Xin X, Clark DP, and La Riviere PJ (2011) Whole-animal imaging, gene function, and the Zebrafish Phenome Project. *Current Opinion in Genetics & Development* 21:620–629. PMID: 21963132 PMCID: PMC3413372.
- **Cheng KC**, Katz SR, Lin AY, Xin X, and Ding Y (2016) Chapter Four, Whole-Organism Cellular Pathology: A Systems Approach to Phenomics. *Advances in Genetics* 95:89-115. PMID 27503355 PMCID 6592046.

- Yakovlev MA, Vanselow DJ, Ngu MS, Zaino CR, Katz SR, Ding Y, Parkinson D, Wang SY, Ang KC, La Riviere P and **Cheng KC** (2022) A wide-field micro-computed tomography detector: Micron resolution at half-centimetre scale. *J. Synchrotron Rad.* 29, doi.org/10.1107/S160057752101287X.
- **Cheng, KC**, Burdine RD, Dickinson ME, Ekker SC, Lin AY, Lloyd KCK, Lutz CM, MacRae CA, Morrison JH, O'Connor DH, Postlethwait JH, Rogers CD, Sanchez S, Simpson JH, Talbot WS, Wallace DC, Weimer JM, Bellen HJ. (2022) Promoting validation and cross-phylogenetic integration in model organism research. *Dis Model Mech.* Sep 1;15(9): dmm049600. doi: 10.1242/dmm.049600. Epub 2022 Sep 20. PMID: 36125045. Special article.

Other Publications:

<http://www.ncbi.nlm.nih.gov/sites/myncbi/keith.cheng.1/bibliography/40809552/public/?sort=date&direction=ascending>

PHS Fellowship Supplemental Form

Introduction

1. Introduction

(for Resubmission applications)

Fellowship Applicant Section

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

applicantgoals_backgroundforresearch1034805356.pdf

Research Training Plan Section

3. * Specific Aims

specific_aims1034805362.pdf

4. * Research Strategy

research_strategy1034805372.pdf

5. * Respective Contributions

respective_contributions1034805375.pdf

6. * Selection of Sponsor and Institution

selection_of_sponsor_inst1034805380.pdf

7. Progress Report Publication List

(for Renewal applications)

8. * Training in the Responsible Conduct of Research

rcr1034805383.pdf

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

9. Sponsor and Co-Sponsor Statements

sponsor_co_sponsor_statement1034805387.pdf

10. Letters of Support from Collaborators, Contributors and Consultants

LOS1034805390.pdf

Institutional Environment and Commitment to Training Section

11. Description of Institutional Environment and Commitment to Training

Institutional_environment_flat1034805400.pdf

12. Description of Candidate&s Contribution to Program Goals

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

13. 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

14. Vertebrate Animals

PHS Fellowship Supplemental Form

Other Research Training Plan Information

15. Select Agent Research

16. Resource Sharing Plan

resource_sharing_plan1034805451.pdf

17. Other Plans

18. Authentication of Key Biological and/or Chemical Resources

Additional Information Section

19. 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):

20. Alternate Phone Number: 410-683-7207

21. Degree Sought During Proposed Award:

Degree:	If "other", indicate degree type:	Expected Completion Date (MM/YYYY):
OTH: Other	MD/PhD	06/2028

22. * Field of Training for Current Proposal: 158 Cancer Biology

23. * 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)
Predoctoral	Institutional	07/01/2020	06/30/2021	T32 GM118294

24. * Applications for Concurrent Support? Yes No

If yes, describe in an attached file:

25. * 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:*

26. Change of Sponsoring Institution

PHS Fellowship Supplemental Form

Budget Section

All Fellowship Applicants:

27. * Tuition and Fees:

None Requested	<input checked="" type="checkbox"/> Funds Requested
	Year 1 \$4,886.00
	Year 2 \$4,984.00
	Year 3 \$61,257.00
	Year 4 \$62,482.00
	Year 5 \$0.00
	Year 6 (when applicable) \$0.00
	Total Funds Requested: \$133,609.00

28. * Childcare Costs:

<input checked="" type="checkbox"/> None Requested	Funds Requested
	Year 1
	Year 2
	Year 3
	Year 4
	Year 5
	Year 6 (when applicable)
	Total Funds Requested:

Senior Fellowship Applicants Only:

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

30. 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

31. Appendix

APPLICANT'S BACKGROUND AND GOALS FOR FELLOWSHIP TRAINING

A. Doctoral Dissertation and Research Experience

I was first captivated by science during my time doing research in organic synthesis as a sophomore at Oberlin College under the direction of Dr. Albert Matlin. The work was difficult and we struggled to develop a reaction that was as enantioselective as we had hoped. Despite this, I was paired with another student who helped make our time in lab memorable regardless of the outcome. He joined me in a mission to learn the “reaction of the day”, and our shared value of curiosity and focus on performing better and better experiments made this research enjoyable. I spent six months in the Matlin Lab, but the ethos of learning the “reaction of the day” has stuck with me to the present, and I hope to impart this attitude upon my own students when I am an independent investigator.

Pre-Doctoral Research Experience

Undergraduate Research I began my research career in the lab of Dr. Albert Matlin midway through my sophomore year at Oberlin College. I had decided to major in Biochemistry, and I had just completed organic chemistry. The course had been challenging but I became fascinated with the subject matter, especially in mechanisms and synthesis. This prompted me to seek out a winter term project in the Matlin Lab and to take advanced organic chemistry courses in the spring semester of my sophomore year (CHEM 254 and CHEM 325). In addition to learning advanced synthesis reactions, enantioselective catalysis, and multiple chromatography approaches I was able to work in a close-knit and motivated research team from the start. Dr. Matlin and my research partner both inspired me to value curiosity and patience at the bench regardless of results. They taught me how to approach my science the right way.

CHEM 254 was taught by Dr. Duy Hua, a structural biologist. After being introduced to foundational concepts in protein structure and conformational change in her class, I pivoted to conduct computational structural biology research in her lab for the entirety of my junior year at Oberlin including winter term. My research in Dr. Hua's lab involved molecular dynamics simulations of human spleen tyrosine kinase (SYK) structures acquired from phosphorylated and dephosphorylated crystal structures. During my time in Dr. Hua's lab I was fortunate to develop my bash scripting skills and learn how to manage jobs on a supercomputing cluster, in addition to furthering my understanding of protein structural change and signaling. I became fascinated with molecular dynamics and sought opportunities to dive deeper into the mysteries of how biomolecular structure governed the biology of disease. In addition to providing me with an opportunity to discover my passion for computational research, Dr. Hua provided me with a standout memore in professional development. I struggled severely in a presentation on molecular dynamics I had been designated to give to our group, and Dr. Hua made sure I was aware of the preparation and attention to detail necessary to deliver an effective talk. Dr. Hua ultimately encouraged me to pursue molecular dynamics research and ultimately to apply to MD/PhD programs, and I think of the talk I gave in her lab as the moment I began to take ownership of my science communication in preparation for an academic career.

Throughout my junior year I reached out and applied to labs that performed similar research, and ultimately chose a position to work for Dr. Blanton Tolbert (see reference letter) in the Department of Chemistry at Case Western Reserve University. I began my work in Dr. Tolbert's lab in the summer after my junior year at Oberlin and began several projects performing molecular dynamics (MD) simulations of non-coding RNAs and RNA-protein complexes. In my first summer in the lab I was able to integrate NMR and small-angle x-ray scattering (SAXS) data into MD simulations to predict 3D solution structures of key regulatory RNAs in HIV-1. I presented my work that summer at the Meeting for Structural Biology Related to HIV/AIDS at the NIH, and I gave an oral presentation *Hacking the Viral Mechanisms of HIV with Molecular Dynamics Simulations* at Oberlin my senior year.

I completed and successfully defended an Honors thesis in Biochemistry during my senior year at Oberlin. My project focused on modulating the self-polymerization of polydopamine with a variety of small molecules in the lab of Dr. Jason Belitsky. This organic synthesis/materials chemistry project developed my skills in the wet lab and took me out of my comfort zone as a chemist. The challenging process of writing and defending an undergraduate thesis helped shaped me as an academic scientist and cemented my desire to work towards a PhD. I presented my Honors thesis research in an oral presentation at the ACS Meeting in Miniature (MiM) at John Carroll University and in two seminar talks at Oberlin College. My thesis defense was difficult, but the rewarding process of seeing the project to completion and learning from my many mistakes made me certain in my desire to pursue a career in science.

Postgraduate Research I joined the Tolbert Lab as a Research Assistant upon graduation from Oberlin and continued my research on RNA 3D structure and its role in infectious diseases. In addition to continuing my work conducting MD simulations of non-coding RNAs in HIV, I also joined a project investigating RNA-drug interactions in enterovirus-71 (EV-71), a causative agent of hand, foot, and mouth disease. We collaborated with Dr. Amanda Hargrove's Lab at Duke and together discovered that an amiloride-based small molecule could bind to a stem-loop structure within the virus and prevents transcription. I performed MD simulations that integrated NMR and SAXS data to reveal a binding site for a heterogenous nuclear ribonucleoprotein which ultimately blocked transcription of the virus. This work culminated in a paper published in Nature Communications, on which I am an author.

I also spent significant time in the Tolbert lab conducting molecular dynamics simulations of other non-coding RNAs that govern transcription in HIV-1. I translated NMR data such as residual dipolar couplings (RDCs) to constrain simulations and worked with other members of the lab to validate structures in the context of the molecular pathways they participate in. This work resulted in two other publications on which I am an author - one in the Journal of Molecular Biology and another in the Journal of Biological Chemistry.

Conducting research on RNA viruses and RNA drug interactions affirmed my desire to become a physician scientist. In the Tolbert lab, we constantly investigated dynamic structures for which we held only clues derived from experimental data. I became fascinated with the inference of conformational states and 3D structure, and it was through this work I was first introduced to challenge of statistical uncertainty. This curiosity laid the groundwork for my interest in the Silverman lab, and the familiarity of 3D structural research made the Cheng lab a great partner in my efforts to contribute to puzzle-solving in biology.

Doctoral Dissertation Research My background in computational research supplemented by synthesis work in the wet lab led me to rotate in both labs that ultimately formed my co-mentorship team. I sought out a rotation with Dr. Cheng because of the unique imaging experiments conducted in the lab as well as the novel computational problems that arise in the analysis of these images. My project during the rotation focused building a pipeline for the automated segmentation of blood cells in micro-CT scans of Zebrafish. This project would allow researchers who study zebrafish models of disease and zebrafish genetics to computationally and quantitatively phenotype whole organisms based on the shape characteristics of their blood cells. This rotation in the Cheng Lab provided a strong foundation for my interest in micro-CT and continues to compel my affinity for computational problems. This project resulted in a paper published in eLife, on which I am an author. I also presented aspects of this work in an oral presentation at the 2021 MD/PhD National Conference and in a poster at the 2023 Mid-Atlantic Regional Zebrafish meeting at the NIH.

I became familiar with Dr. Silverman's work early on in the automated segmentation project and became interested in his group's work on statistical inference. I sought out a rotation in his lab to gain exposure to a more theory-driven approach to machine learning. During my rotation, we decided to search for methods from topological data analysis and apply them to the classification of complex shapes in biomedical data. Dr. Silverman and I met a minimum of once per week and studied the methods I was learning at the whiteboard often for more than two hours. I received hands-on training in the concepts underlying the computational methods we were studying. I first applied methods from persistent homology during this rotation, where I used the Vietoris-rips filtration to classify wild-type and mutant zebrafish blood cells based on shape.

B. Training Goals and Objectives

1. Career Goals and Research Interests My long-term goal is to become a physician scientist leading a translational research lab applying novel measurement and computational methods to clinical problems. I plan to apply to a research track internal medicine residency followed by a hematology-oncology fellowship and aim to become a tenure track faculty member at a top academic institution. I will work towards this goal in the short term by continuing to hone my clinical skills throughout my graduate years and by anchoring my thesis research in translational problems. Both of my co-mentors have a background in clinical medicine and are familiar with the training and experience required for effective medical training.

2. Research Training Goals I have focused my PhD training on novel imaging technology and statistical methods, both disciplines that will facilitate the completion of the proposed work and prepare me for my future career as a physician scientist. Under the direction of Dr. Cheng, I will receive training in experimental pathology and high-resolution imaging with a focus on both synchrotron and laboratory micro-CT methods. The micro-CT experiments conducted in our lab are motivated by histology and the diagnosis of human disease. I will work closely with Dr. Warrick (see letter of reference) to both develop my understanding of diagnostic medicine and

guide our analysis of micro-CT images. In preparation for this proposal, I was able to work closely with Dr. Warrick to get an IRB approved to perform an imaging study of prostate cancer tissue blocks. Dr. Warrick's office is located in the same building as the Cheng Lab, and he maintains an open-door policy that sees us meet nearly on a weekly basis and allows us to discuss data as soon as it is generated. This close partnership with Dr. Warrick directly contributed to preliminary data collection during a recent synchrotron trip I led to LBNL, a key role of my work in the Cheng Lab. Through my first two years in the lab I have participated in 3 synchrotron trips and taken a leadership role in one. My research goal for the continuation of this work is to further my understanding of high-resolution optics and their potential applications to the diagnosis and investigation of cancer. In pursuit of this goal I will also have several opportunities to develop my leadership skills in the setting of technically demanding experiments.

To analyze imaging results and complete aims 2 and 3 of this proposal, I will study statistical methods and machine learning with a focus on topological data analysis under the direction of Dr. Silverman. I have two goals for this phase of my research training. First, I aim to develop and optimize algorithms that will contribute to the analysis of complex 3D images, specifically aimed at methods applied to cancer biopsies. Aim 2 and especially aim 3 of this proposal will provide me with a unique opportunity to do so. Second, I strive to cultivate a generalizable skillset and understanding of data science and statistical methods that I will use throughout every facet of my career as a physician-scientist.

3. Clinical Training Goals I will improve my clinical skills and specialty exposure throughout the duration of this award to prepare me both for success in medical school and my future career as a physician-scientist. The MD/PhD program at PSCOM has 3 main avenues of support for ongoing clinical training during the graduate years. First, students enroll in the BMS 802 course after they pass their comprehensive exams, during which they work with a mentor in a hospital setting a minimum of 5 half-days per semester. Second, students also participate in clinical research conference (CRC) once every two months in which a physician-scientist and a group of three students prepare a presentation and lead the program through a real clinical case. This involves interactive building of a differential diagnosis, discussion of pathophysiology, and review of a relevant research article.

In addition to clinical exposure through the BMS 802 course and shadowing, I will also continue to volunteer in monthly free general medicine clinics at the Bethesda Mission in Harrisburg through the Lioncare student group at PSCOM. This will supplement the clinical training I receive in the pediatric emergency department with additional repetitions of history taking, physical exams, and note writing in the setting of a different patient population. My goal is to become a well rounded medical student with experience treating a diverse patient base and consistent practice of the fundamentals of medicine.

4. Career Development Goals I will improve my scientific writing, communication, and interpersonal skills and experiences throughout the duration of this award. I have sought out a co-mentorship that will provide unmatched training in these areas, as both co-sponsors of this grant hold MD and PhD degrees, and will have significant roles in shaping my development into a physician scientist. Dr. Silverman and Dr. Cheng each have a unique skillset that will both challenge and support me along my journey towards becoming an independent investigator. Their areas of research are independent but also complementary: Dr. Silverman is an expert in statistical methods and the analysis of biomedical data, while Dr. Cheng pathologist by training who has an extensive background in both imaging and genetics research.

The individual expertise of each of my co-sponsors will not only play crucial roles in mentoring me as I carry out the proposed work, but will also provide me with unmatched career development opportunities that will be unique to my co-mentorship. For example, on behalf of the Cheng lab and in close collaboration with Dr. La Riviere (see letter of reference), I am able to conduct highly specialized research in x-ray physics and interact with beamline scientists who directly inform my work. With the guidance of Dr. Cheng and through collaboration with Dr. Warrick I will present my work to clinical pathologists who directly encounter the problems defined in this proposal in their daily practice, and I will integrate their feedback into my future experiments. To further pursue the input of pathologists I will apply to present this work at conferences such as the United States and Canadian Academy of Pathology (USCAP) Annual Meeting. To complement their feedback I will also present my work to an audience of computational researchers and statisticians. I will apply to present at meetings such as the SPIE medical imaging conference to improve my understanding of computational research relevant to this project. I will also apply to present my work on topological data analysis (A3) at JSM in 2026.

C. Activities Planned Under This Award

A timeline summarizing the expected completion of these activities and goals is illustrated in Figure 1.

Research Training Activities: Through the experiments planned under this proposal and the guidance of my co-sponsors and broader mentorship team, I will accomplish my research goals and build on the research experience I have acquired to date.

- **Histopathology Sample Preparation and Slide Reading:** Throughout my training I will gain experience in handling sensitive tissue specimens and become familiar with the standard workflow for histology. In close collaboration with Dr. Warrick, I will also continue to study disease processes via histopathology and use these insights to guide the design of 3D imaging experiments.
- **Synchrotron Micro-CT Imaging:** Imaging experiments in the Cheng Lab involve using micro-CT resources both on campus at Hershey Medical Center (HMC) and at Lawrence Berkeley National Laboratory (LBNL). I will gain experience with x-ray imaging in each setting. In service of aims 1 and 2 of this proposal I will lead 2 experimental imaging trips to beamline 8.3.2 of the Advanced Light Source at the LBNL.
- **Image Reconstruction:** Reconstruction and analysis of phase-contrast images requires programming experience and an understanding of the physics underlying micro-CT imaging. Throughout my training under this award I will build on the programming skills I have already acquired by continuing to perform and refine 3D image reconstructions in python. Additionally, I will work closely with Dr. Patrick La Riviere from the University of Chicago in collaboration on phase-contrast experiments. Through this collaboration I will deepen my understanding of medical physics and the mathematics that govern x-ray imaging.
- **Statistical Methods and Study Design:** I have already spent time studying inference and probability using the Statistical Inference book by Casella and Berger in my first year in the Silverman lab. Under this proposal, I will also study non-inferiority trials, ANOVA, and mixed effects models in support of aim 2. This training will improve my literacy in statistical methods that will translate beyond basic science research. Completion of these aspects of my training will result in a robust foundation in statistical theory that will mold me into a well-rounded physician-scientist capable of contributing to basic and clinical research alike.
- **Topological Data Analysis:** I first learned about topological data analysis and persistent homology during my rotation in the Silverman lab. I developed an interest in these concepts and continued discussing them with Dr. Silverman. This led to aim 3 of our proposal, the execution of which will leave me with an expertise in the applications of persistent homology and a skillset that will transfer to my career as an independent researcher.

Clinical Training Activities

- **MSTP Clinical Exposure Program (CEP):** All MSTP students at PSCOM are required to participate in CEP. The CEP requires that these students join an advisor in clinic for a minimum of 6 half days each semester. Through this program, I am working with Dr. Lilia Reyes, a Pediatric Emergency Medicine Physician at Penn State Health. She will evaluate my clinical skills and note writing in preparation for clerkship training during the third year of medical school.
- **MSTP Clinical Research Conference:** The MSTP program hosts CRC once every two months, during which students are guided by a physician scientist through a presentation of a challenging clinical case, including detailed discussion of the pathophysiology and a pertinent paper. In 2023 I presented the pathophysiology of pre B-cell ALL. In 2024, I will present the paper related to a different clinical case.
- **Objective Structured Clinical Exams (OSCEs):** OSCEs are used to adjudicate the clinical skills of medical students throughout training at PSCOM. Students have the opportunity to enroll in OSCEs during the PhD portion of training, and I will participate in and pass one OSCE per semester throughout the duration of this award.
- **Lioncare Volunteering:** To supplement formal clinical training, I will continue my participation in the student-run Lioncare free clinic treating underserved populations at the Bethesda Mission in Harrisburg a minimum of once every two months.
- **Subspecialty Shadowing:** I will join Dr. Raymond Hohl (thesis committee member) for one full day in outpatient Hematology/Oncology clinic per semester to connect my clinical training to the subspecialty I hope to practice as an independent physician-scientist.

Coursework/Seminars: I have been fortunate to develop strong computational skills throughout my time at Oberlin and especially through research in the Hua and Tolbert Labs. Specifically, I wrote several programs in bash, zsh, and python to conduct and analyze MD simulations. I built on these skills through my coursework in the

	Year 1 (G3)	Year 2 (G4)	Year 3 (M3)	Year 4 (M4)	
Research Training	Complete aim 1 and collect imaging data for aim 2, draft and submit manuscript	Finalize aim 2, complete aim 3, submit computational manuscript	Finalize revisions to thesis and submit remaining manuscripts	Clinical oncology and statistics research during medical school research blocks	
	Meet weekly with co-sponsors individually and regularly with mentorship team	Defend PhD Thesis			
Clinical Training	Work monthly with Dr. Reyes in the Pediatric ED for primary care experience Work monthly with Dr. Hohl in Heme/QDC clinic		Clerkships/Rotations	Acting internships	
			OSCE Exams	OSCE Exams	
Coursework	Linear Algebra guided self-study with Dr. Silverman		Profession of Medicine	Health systems science in the clinical setting	
	Additional reading in statistics and machine learning				
	Topology course in Math Department at PSU				
Professional Development	Present research at MD/PhD biannual retreats and Seminar Series				
	Attend and submit abstract for the MD/PhD National Conference Submit for USCAP, SPIE, JSM conferences				
	Present research at Bioinformatics and Genomics Program Retreat, Penn State Cancer Research Day, Experimental Pathology Colloquium				

Figure 1: Roadmap of training components and milestones planned under this award.

Bioinformatics and Genomics program at PSU, especially in BMMB 802, MCIBS 554, and STAT 555, all courses I was able to excel in. Through STAT 555 and my work in the Silverman Lab, I have also developed skills in R. However, despite continuous programming experience, I desire to bolster my background in statistical theory and mathematics to support the proposed work and to achieve the level of expertise I need to contribute from a data science perspective. Topological data analysis requires a greater understanding of linear algebra than I currently have. I will address this deficiency by completing guided self-study in linear algebra with Dr. Silverman. During this self study I will utilize the textbook *Matrix Algebra from a Statistician's Perspective* by David A. Harville. Upon completion of this material, I will take the Topology course in the statistics/math department at PSU. This will not only support Aim 3 of this proposal, but will also advance my skills as a data scientist and build a foundation for critical analysis of other biomedical research problems. To further supplement our learning and improve student writing skills, students in Dr. Silverman's lab also conduct a self-study reviewing *The Sense of Structure: Writing from the Reader's Perspective* by George Gopen.

I am currently completing the BMS 591 course in Biomedical Research Ethics at Penn State College of Medicine which has supplemented the training in medical ethics and humanities I received through the first two years of the medical curriculum. Penn State also emphasizes the Science of Health Systems (SHS) courses during the medical years of our training. During SHS courses, we continue to discuss ethical problems in the setting of hospital medicine and biomedical research while also delving into other topics such as quality improvement research and biostatistics. Discussion of quality improvement and systems approaches to care improvement have had influence in my research as we strive to build translational methods that will positively supplement the work of clinicians and scientists alike. This coursework will supplement my studies in data science and statistics and

promote the application of my research experience to projects that improve the delivery of care to patients.

Professional Development: I have sought out a unique co-mentorship that will also provide especially beneficial professional development opportunities. The Silverman Lab meets weekly, alternating between professional presentations given by students and open-discussion journal club. The lab has an expertise in statistical theory and methods applied to biological problems, and I will benefit not only from hands-on training from Dr. Silverman and Dr. Michelle Nixon, but also from regular lunch meetings and close working relationships with other students in the lab. In addition to group lab meetings, I meet with Dr. Silverman individually on a weekly basis and additionally when needed. The Cheng Lab also meets weekly in a different format, with each student sharing brief individual updates following Dr. Cheng's, with student-led journal clubs approximately once per month. I also meet weekly with Dr. Cheng to plan projects and get feedback. I have been fortunate to form a close relationship with both of my co-sponsors and have benefited greatly from their personal mentorship.

To support this co-mentorship efficiently, I commute to University Park on Wednesdays where I have a dedicated workspace in the Silverman Lab. The Cheng Lab meets every Wednesday morning at 9am, and then the Silverman Lab meets in person every Wednesday at 1:30pm. Dr. Silverman and I then typically meet in person at 4pm before I return to Hershey. To supplement our regular one-on-one meetings, Dr. Silverman also meets with me before and after each presentation I deliver, as he does with all of his mentees. Overall, our arrangement has not only minimized time lost through commuting but has ensured a balanced investment into each aspect of my project with consistent guidance from each mentor regardless of their primary campus. Dr. Silverman and Dr. Cheng also will continue to have regular meetings both for their ongoing collaborations and the joint supervision of my project. We are able to meet approximately monthly via zoom and additionally as needed. This has benefitted me as a student and supported both the publication of a paper on which I am an author as well as my successful completion of the comprehensive exam in the Bioinformatics and Genomics program.

Medical School (2 Years):

After successfully defending my PhD thesis, I will reenter medical school for the third and fourth years of clinical training and this award. I will focus during this period of training on my clinical skills and prioritize my development into a well-rounded physician. I will continue to develop my research, leadership, and professional skills in addition to my clinical training throughout this award. During the third year of medical school (M3 in table 1) I will complete clerkships in all major specialties and select elective rotations in hematology/oncology. Upon completion of M3 I will sit for the USMLE Step 2 medical board exam. During the fourth year of medical school (M4 in table 1) I will utilize at least 3 of my 12 4-week blocks to conduct research in malignant hematology. The expected balance of these disciplines under the award is summarized in Figure 2. As I complete the final year of this award and complete medical school I will apply to a research-track internal medicine residency program with a combined fellowship in hematology-oncology.

Year	Research	Coursework/Seminars	Prof. Development	Clinical Training
G3	80%	10%	5%	5%
G4	85%	5%	5%	5%
M3	5%	1%	1%	93%
M4	40%	5%	15%	40%

Figure 2: Allocation of time for training components across the years requested under this award.

SPECIFIC AIMS

Abnormal tissue morphology is a hallmark of many soft tissue tumors. Characterization of morphology at the macroscopic or microscopic level can provide insights critical to diagnosis, prognosis, therapeutic design, and even basic research. At the macroscopic level, a variety of three-dimensional (3D) imaging technologies (e.g., MRI, CT, and PET-CT) exist and allow clinicians to stage cancer and direct therapies based on the size and shape of lesions [1, 2]. Complementing these imaging technologies, machine learning and statistics advances allow researchers to quantify morphologic characteristics to aid cancer diagnosis and prognosis [3, 4]. Morphology is just as critical at the microscopic level. For example, in prostate cancer, glandular morphology is paramount to cancer staging and is the main component of the popular Gleason score [5, 6]. However, at the microscopic level, clinicians and researchers have been limited primarily to qualitative assessment of two-dimensional, slide-based imaging [7]. Assessing morphology in only two dimensions can be sub-optimal [8, 9, 10]. For example, the angle at which a two-dimensional tissue slice is taken can alter conclusions about glandular morphology [11] and is hypothesized to lead to the large variability in prostate cancer grading between observers [11, 12]. In short, there is a paucity of tools available for assessing 3D tissue morphology at the microscopic level.

Recent micro Computed Tomography (micro-CT) advances can facilitate high-resolution 3D imaging of biological tissue [13, 14]. In particular, synchrotron micro-CT techniques have demonstrated the ability to resolve cellular structures. However, these methods often require dedicated beamlines or are limited in their field of view [15, 16, 17]. To address this problem, Dr. Cheng's lab has shown how a novel wide-field detector can generate almost centimeter scale images at sub-micrometer resolution [18]. Recently, we developed a Propagation-Based phase-contrast CT (PBCT) imaging protocol  eliminates the need to stain tissue blocks and thereby permits direct imaging of clinical biopsies, which are often paraffin-embedded and challenging to stain. Using this technology, we have been able to produce 3D images of prostate cancer and melanoma with the highest resolution to field of view ratio to date. In what follows, We propose to expand upon this technology and demonstrate its ability as both a clinical and research tool. In Aim 1, we will optimize the PBCT protocol to ease future technology adoption. In Aims 2 and 3, we will evaluate the capabilities of this imaging technology as a clinical and research tool for prostate cancer.

Aim 1: Optimize PBCT imaging of Formalin-Fixed Paraffin-Embedded (FFPE) soft-tissue biopsies. Our preliminary results already establish that PBCT is capable of producing high-resolution images of soft-tissue biopsies even in settings where contrast-enhancing staining of tissue blocks is not possible. To optimize this protocol, we will develop a multi-tissue atlas demonstrating how key protocol parameters (e.g., x-ray energy and propagation distance) affect image characteristics including resolution and contrast in different tissue types such as mouse breast cancer, human oropharyngeal cancer, and human prostate cancer. This atlas will assist researchers in determining optimal imaging parameters for their studies.

Aim 2: Evaluate utility of PBCT imaging for the grading of prostate cancer. We will collect the first micro-CT dataset of prostate cancer. Biopsies from disease subjects and age-matched healthy controls will be used to assess and test the hypothesis that PBCT grading is non-inferior to standard slide-based imaging. Beyond this pilot study, these data will be made available publicly as a resource for the community.

Aim 3: Quantify the morphological heterogeneity in prostate glands using computational persistent homology. Computational topology provides a rich suite of tools for quantifying 3D shapes in such a way that morphology can be included as covariates in statistical and machine learning-based cancer prognostics [1, 4]. In collaboration with Dr. Silverman, we will identify which topological summary statistics can best capture the morphologic differences between cancerous and healthy prostate glands. Using these statistics, we will then provide the first quantification of how glandular morphology varies spatially within a biopsy as a function of age and disease status. This quantification will provide novel biological insights into tumor heterogeneity and will inform the design of future studies of prostate cancer.

Completing these aims will provide three critical outcomes: (1) the first micro-CT imaging protocol for the high-resolution 3D imaging of whole tissue blocks, (2) a clinical evaluation of this technology in the context of prostate cancer grading, and (3) the first quantitative study of morphological variation in prostate cancer. Finally, this research will provide me with the skills and expertise to pursue my long-term career goal of leading an independent research program in Oncology.

RESEARCH STRATEGY

Significance

Abnormal morphology is a hallmark of many cancers and is critical in diagnostic, prognostic, and treatment decisions [19]. Both macroscopic morphology (e.g., the shape of a tumor as a whole) and microscopic morphology (e.g., the shape of prostate glands) are clinically important. While clinicians and researchers have numerous tools available for assessing and quantifying morphology at the macroscopic level, comparable tools are often unavailable at the microscopic level. At the macroscopic level, clinicians assess morphology using 3D imaging technologies such as magnetic resonance imaging (MRI) or computed tomography (CT). Yet at the microscopic level, morphological assessment is almost exclusively done by traditional 2D slide-based imaging. Numerous studies have found that 2D imaging techniques can be sub-optimal for assessing cancer morphology [9, 10, 16]. For example, in staging prostate cancer, Koyuncu et al. (2023) [10] found that the slice angle used to prepare 2D slides could alter the apparent morphology of glands and alter grading decisions. It is hypothesized that such problems may be partially responsible for the large variability in grading between individuals [8]. Outside of the imaging technologies themselves, the ubiquity of 3D images of cancer have led to numerous statistical and machine learning advances that may ultimately improve clinical care. For example, Crawford et al. (2020) [4] showed that computational topology  could be used to identify the morphological biomarkers for MRI imaging of Glioblastoma that were more predictive of patient outcomes than traditional clinical metrics or molecular biomarkers. Tools such as those developed by Crawford et al. [4, 3] cannot be applied at the microscopic level due to the lack of 3D imaging data. In short, assessment of cancer morphology is limited at the microscopic level due to a lack of available imaging technologies.

Several approaches to 3D microscopic imaging have recently been proposed. These include the hybrid Open-Top Light-Sheet microscopy (OTLS) illustrated in Glaser et al. (2022) [20] and a number of related methods collectively called micro-CT [14, 15, 17, 21, 22]. OTLS is a 3D imaging method that relies on tissue clearing and fluorescent staining of tissue samples. OTLS permits high-quality volumetric imaging of whole biopsies without requiring repeated sectioning. Yet, OTLS requires complicated sample preparation procedures involving tissue clearing and staining with fluorescent markers. Micro-CT however can provide 3D imaging of biological samples at $0.5\mu\text{m}$ isotropic voxel resolution, doing so without the use of tissue clearing or the addition of any fluorescent markers [16, 18, 22, 23].

While multiple groups have shown that micro-CT is capable of producing high-resolution cellular and sub-cellular imaging, these technologies have been limited by their narrow field-of-view. For example, while Pinkert-Leetsch et al. demonstrated they could resolve nuclei within pancreatic adenocarcinoma, their entire field of view was limited to 1.6 by 1.4mm [16]. This field-of-view is much smaller than most biopsy specimens, requiring the user to acquire many scans to cover just one sample. The Cheng Lab recently addressed this issue and demonstrated that a novel wide-field detector allowed for nearly centimeter scale (whole-biopsy) imaging at sub-micrometer ($0.5\mu\text{m}$) isotropic voxel resolution [18].

Contrast-enhancing stains are typically required to obtain high-quality micro-CT images of biological tissue [22, 24, 25, 26]. Yet many clinical samples are Formalin-Fixed and Paraffin-Embedded (FFPE) and do not take up those stains without significant alterations to the sample such as deparaffinization. Recently, Frohn et al. (2020) [27], demonstrated that high-resolution stain-free micro-CT of FFPE pancreatic biopsies could be obtained using Propagation-Based phase-contrast CT and phase-retrieval algorithms developed for traditional (macroscopic) CT [28]. In brief, PBCT reconstructs high-resolution 3D images by using the fact that the phase of photons change differently depending on their incident energy and the type of material they pass through. This is in contrast to traditional CT imaging that primarily focuses on the magnitude of attenuation as photons pass through a material [24, 26]. However, two factors limit the utility of their approach. First, they had the same limited field-of-view typical of micro-CT. Second, their approach utilized a number of non-standard tools as part of the imaging apparatus including Kirkpatrick-Baez (KB) mirrors and a waveguide, or an interferometer [17, 29, 30].

Recently we have demonstrated that our wide-field detector could be combined with a modified PBCT protocol to obtain high-resolution, whole-biopsy, stain-free micro-CT images of clinical biopsies (Figure 1). Notably, our modified PBCT protocols do not require KB mirrors or waveguides and are therefore simpler than some of the prior PBCT protocols with no appreciable decrease in resolution. With this development effective 3D microscopic imaging of clinical biopsies is possible. In what follows, we propose to expand upon this prior work. In Aim 1, we propose to optimize our modified PBCT protocol and provide a multi-tissue atlas that researchers can use to determine optimal imaging parameter for different tissue types. Aims 2 and 3 then focus on the application of this

technology to imaging prostate biopsies. Aim 2 evaluates the utility of this technology in clinical settings while Aim 3 focuses on adapting quantitative tools developed for macroscopic morphometrics to the microscopic domain. Beyond demonstrating the technology, Aims 2 and 3 provide novel insights into prostate cancer. For example, Aim 3 will provide the first quantitative study of how the morphological features distinguishing healthy versus cancerous prostate glands varies spatially within a tumor, between individuals, as a function of age, and as a function of disease status. Among other uses, that analysis can directly inform the design of future studies of prostate cancer as it will directly inform estimates of effect sizes and sampling variability.

Innovation

The proposed work will provide several innovations in the context of 3D microscopic imaging and the study of prostate cancer.

- The modified PBCT protocol optimized in Aim 1 represents the highest-resolution, stain-free approach to 3D imaging of whole-biopsies to date.
- The multi-tissue atlas developed in Aim 1 will be the first resource that researchers can use to guide their own micro-CT imaging studies of human biopsies.
- Aim 2 will develop, and make publicly available, the first 3D microscopic imaging study of prostate cancer.
- Aim 2 will provide the first assessment of the utility of micro-CT in prostate cancer grading.
- Aim 3 will provide the first quantitative study of morphological characteristics of that distinguish healthy from cancerous prostate glands.
- Aim 3 will provide the first quantitative study of how morphological characteristics that distinguish cancer vary spatially within the prostate and as a function of age and disease status.

Approach

Aim 1: Optimize PBCT Imaging of Formalin-Fixed Paraffin-Embedded (FFPE) Soft-Tissue Biopsies.

We have already developed a PBCT protocol that permits high-resolution, stain-free 3D imaging of whole-biopsies (See demonstration in Figure 1). In contrast to standard contrast-based micro-CT, our approach can be applied to FFPE clinical samples without deparaffinization. While we have already optimized this technology for the identification of glandular structure in prostate biopsies, further optimization may be needed for different imaging tasks. Broadly, within the emerging field of PBCT, questions remain about optimal image acquisition parameters such as free space propagation distance (R) and beam energy (K) [16, 31]. There are presently no resources that researchers can use to determine these parameters.

The goal of this aim is to provide a publicly available multi-tissue imaging atlas which researchers can evaluate whether their questions can be answered using PBCT and to determine imaging parameters (R and K) that would best capture the tissue structures of interest. We have worked in close collaboration with Dr. Patrick La Riviere (see letter of reference) to devise and optimize our PBCT experiments and will continue to refer to his expertise. To create this atlas we will image two tumor biopsies and two healthy controls from four distinct tissue types including human prostate cancer, human oropharyngeal cancer, and mouse breast cancer (see letters of support from our collaborators Dr. Edward Gunther and Dr. Jiafen Hu who will provide samples). For each sample, we will image at a range of distances R from 30mm to 200mm and three different beam energies K (14keV, 20keV, and 26keV). This imaging will be repeated within two different locations to characterize how optimal imaging parameters generalize between laboratories. First, we will image using a polychromatic bench-top cone-beam source maintained within the Cheng Lab. Second, we will repeat this imaging at the Lawrence Berkeley National Lab (LBNL) synchrotron 8.3.2 beamline of the Advanced Light Source (ALS) where Dr. Cheng maintains dedicated time (see sponsor and co-sponsor statement). As these imaging techniques are non-destructive, we will also provide paired slide-based histology images of each sample for comparison. All slide based images, raw scan data, and 3D reconstructions will be made publicly available as an online resource which will allow researchers to determine if their research questions might be studied using PBCT and to quickly identify optimal imaging parameters. Our images will be served online through the open-source web visualization software Neuroglancer [32]. This atlas will be extensible and will serve as a centralized community resource where researcher can share images and the parameter used for imaging.

Potential Pitfalls and Alternative Approaches Overall this aim will be straightforward to complete as we have already established proof-of-concept (e.g., Figure 1). The major complication that might arise are measurement artifacts called “edge-enhancement” which can occur with propagation-based phase contrast imaging [21, 22].

Notably, these artifacts have been largely absent from our prostate biopsies. That said, if such artifacts are

encountered, we will evaluate different publicly-available phase retrieval algorithms which have been designed for the task [28].

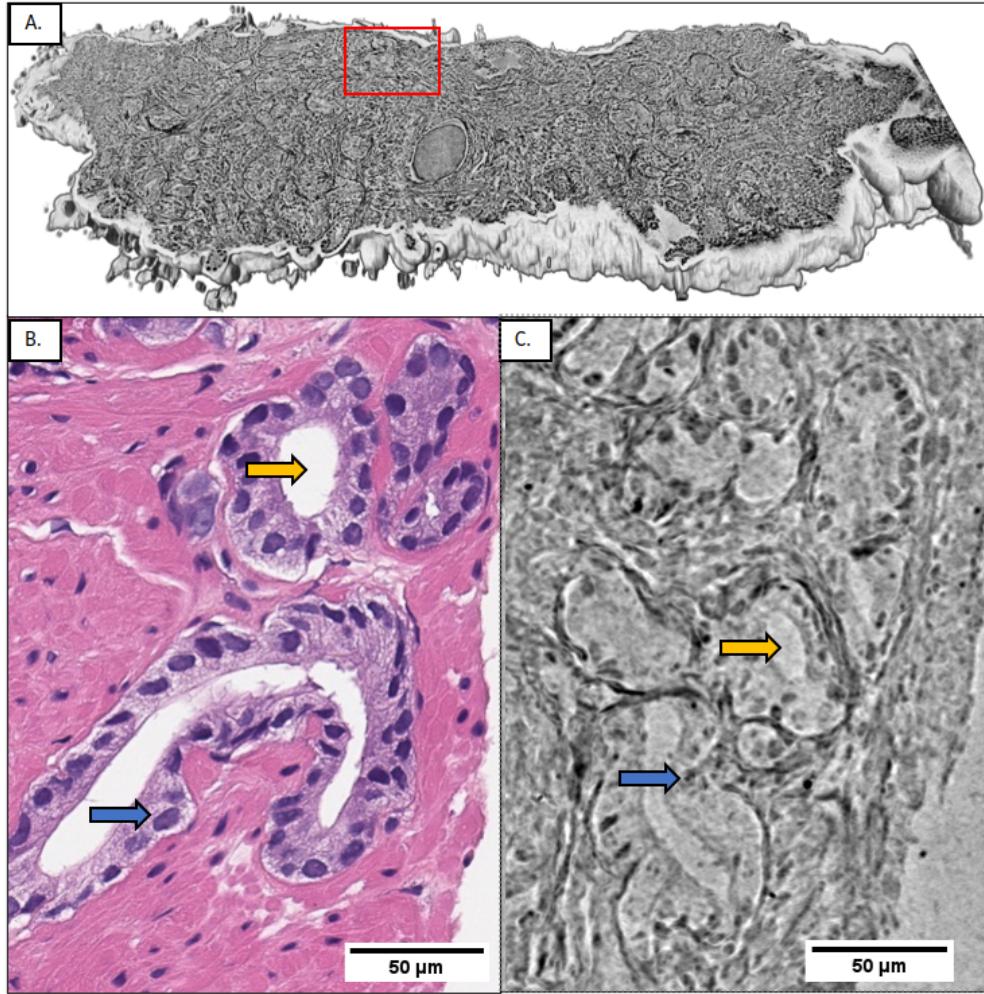


Figure 1: PBCT imaging can resolve glandular structure in prostate biopsies. (A) 3D rendering of unstained FFPE prostate cancer scanned at LBNL, with a red box around the approximate region of panel C. (B) Histopathology slide taken from a malignant biopsy core, 20X magnification cropped to highlight a section of malignant glands. Nuclei (blue arrow) and lumen (orange arrow) are marked for comparison. (C) 50 μm 2D projection of a 3D image obtained from a PBCT scan of the same sample zoomed in to roughly the same region of malignant glands as shown in Panel B. Nuclei (blue arrow) are prominent. The lumina of glands (orange arrow) along with their borders are clearly defined. We obtained this PBCT image at the LBNL synchrotron at $K = 20\text{keV}$ and $R = 80\text{mm}$ without contrast-enhancing stain. The entire biopsy was imaged, only a small subset of that scan is included for brevity.

images may be superior to scoring based on 2D images. Unfortunately, directly testing this hypothesis confronts a form of *familiarity bias* since clinicians are trained on traditional slide-based imaging. As a result, here we focus only on a proof of concept focused on testing whether 3D imaging is non-inferior to 2D imaging. To test this hypothesis we will collect, and make available, the first 3D imaging study of prostate cancer.

Develop 3D Imaging Study of Prostate Cancer. In collaboration with Dr. Warrick at PSU we will obtain whole-biopsy PBCT images of 20 age-matched benign and malignant prostate samples (see letter of reference). We will follow the same protocol used for generation of Figure 2B which already established that we can collect high-quality whole-biopsy images of similar biopsies provided by Dr. Warrick. For comparison, after imaging, each sample will be sliced and prepared for slide-based imaging using standard laboratory protocols. To make this data available as a public resource, both the slide-based images as well as the raw and 3D rendered PBCT images will be made available online using the same web-server developed in Aim 1.

Evaluate the Hypothesis That 3D Imaging Can Be Used For Scoring Prostate Cancer. In collaboration with Dr. Warrick (see letter of reference), the Chief of Clinical Pathology at Penn State College of Medicine (PSCOM), we will recruit up to 10 clinical pathologists (Fellows and Attendings) within and external to PSCOM to score 3D and 2D images of prostate biopsies. Each pathologist will be asked to provide Gleason scores for each of the prostate biopsies. For each biopsy a pathologist will be shown at random either the 3D rendering first or the 2D

PBCT experiments will take place at the ALS of LBNL, which will undergo maintenance later next year (2024). If we are unable to complete data collection prior to maintenance, Dr. Cheng maintains an active collaboration with the Advanced Photon Source (APS) of Argonne National Laboratory. We will conduct remaining imaging at the APS if necessary.

Aim 2: Evaluate Utility of PBCT Imaging For the Grading of Prostate Cancer.

Histopathology of prostate tissue biopsies is the gold standard for the diagnosis and grading of prostate cancer [6]. Grading is based on the Gleason score [6, 8, 12], which typically ranges from 3 to 5 and is primarily defined by glandular morphology within the prostate. However, prior work has shown that Gleason scores can be highly variable both between clinicians and even when the same clinician is shown different slices of the same prostate biopsy [10]. This high variability can lead to either under- or over-treatment. It is hypothesized that scoring based on 3D im-

slide-based image first. After providing an initial score, they will then be shown the other imaging type and asked if they need to update their score based on the additional information provided by the second imaging modality. To avoid potential biases, we will randomize the order of samples for each pathologist and for each biopsy randomize whether pathologists are shown the 2D or 3D images first. Ultimately this will result in a two-by-two contingency table Y tabulating the number of individuals shown the 2D or 3D images first and whether or not they found the second imaging modality added information compared to the first.

We will test our hypothesis using standard tools from categorical data analysis. We will use a binomial test to evaluate the null hypothesis that the proportion of times where 3D imaging added information to 2D imaging (p_1) is greater than the proportion of times where 2D imaging added information to 3D imaging (p_0). To account for repeated measures we will formulate this test using binomial mixed effects models. To account for familiarity bias we will formulate this as a non-inferiority test with null hypothesis $H_0 : \log p_1/p_0, \leq \delta$ where we will choose $\delta > 0$ rather than $\delta = 0$ to provide a non-inferiority margin. We will chose δ based on the results of a small pilot which will have a similar design to this one but 2D maximal intensity projections (MIPs) from 3D images will be shown rather than full 3D renderings. As in Figure 1, we will show the same region of biopsies using each imaging modality to ensure that the information content between the two imaging types is the same. Participants will be blinded to this fact and therefore we expect that we can estimate $\delta = \log p_1/p_0$ using this pilot study. Note, our sample size is designed to achieve $> 85\%$ power under an alternative $\log p_1/p_0 \geq 0.2$ with a non-inferiority margin $\delta \leq -0.2$.

Potential Pitfalls and Alternative Strategies The major limitation we foresee is the complication of *familiarity bias* wherein pathologists will be less likely to find PBCT useful solely because they are unfamiliar with the technology. If our preliminary study estimates $\delta \geq 0.4$, we will compare gray-scale slide-based imaging to PBCT imaging. While standard slide based imaging is often colored, removing this color will likely mitigate the familiarity bias between these two imaging modalities while still allowing for meaningful comparisons.

Aim 3: Quantify the Morphological Heterogeneity in Prostate Glands Using Computational Persistent Homology.

The morphologic features of prostate glands that indicate cancer are abstract and difficult to quantify using standard statistical tools. For example, Figure 2 shows different glandular morphologies associated with different stages of cancer. The differences between Gleason pattern 5 (more diseased) and pattern 1 (healthy) are complex: it is not just a difference in size or volume but a difference of more abstract qualities like “tortuousness”. It is not clear how to quantify these abstract morphologic features let alone how to incorporate some quantitative metric of this characteristic into models.

Yet, there are numerous potential benefits from quantifying these characteristics. For example, if we could quantify Gleason patterns it would provide a more quantitative approach to staging cancer and might improve upon the current Gleason scoring system. Moreover, if we could quantify these characteristics then we could use those characteristics in statistical or machine learning models to help predict patient outcomes and guide treatment decisions. In this aim we borrow computational tools developed for macroscopic morphologic analysis of cancer and adapt those tools to the microscopic study of prostate cancer. We then demonstrate those tools by performing the first analysis of morphologic variation within prostate cancer.

Overview of Computational Persistent Homology Topology is a field of mathematics that, like geometry, studies shapes (See Wasserman et. al [34] for a review on the distinction between these fields). Homology is a subfield of topology that specifically studies the voids or holes in shapes. In the context of prostate cancer, homology can be thought of as the natural field of mathematics for modeling glands which are essentially defined by their voids (i.e.,

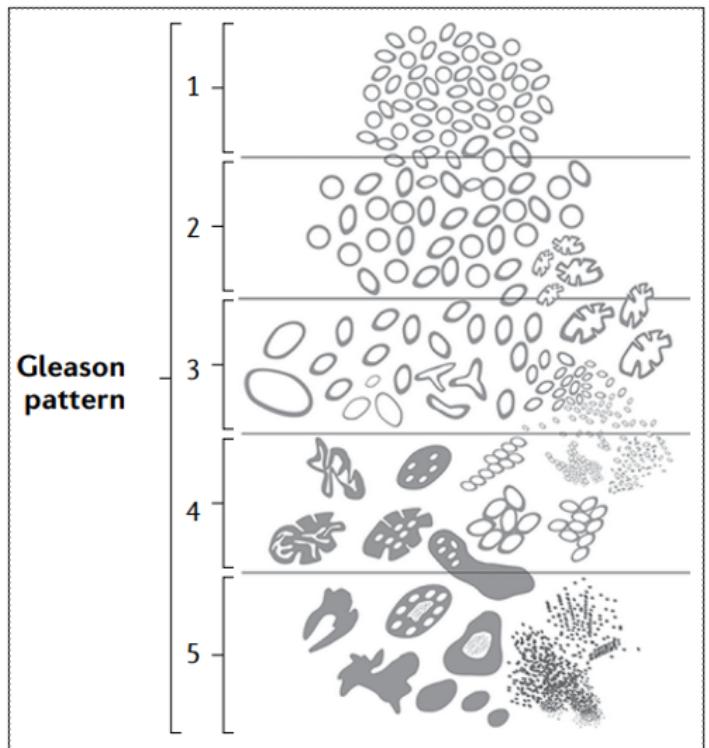


Figure 2: Diagram of different patterns in glandular morphology and the associated Gleason grade. This figure was taken from a review by Rebello et. al [33].

lumen). In recent years, homology has moved from a field of abstract math to a powerful framework for modeling shapes thanks to advances in an area of applied statistics called computational persistent homology [34, 35]. In brief, suppose that our shape is defined by a point cloud (e.g., a series 3D points each representing the nucleus of an epithelial cell of a gland). Next pretend that there is a ball of radius r centered about each point. Wherever two balls connect, create an edge between the corresponding points. This forms a graph. Rather than just forming one graph, create a series of graphs as r increases from 0 to infinity. In the language of computational persistent homology this is called a series of simplicial complexes (graphs) formed via a Rips filtration (formed by creating graphs as a function of r). The field of computational persistent homology has created a wide variety of statistics which summarize this series of simplicial complexes in such a way that the resulting statistics capture all the relevant information about homology (points, holes, voids) of the original shape. Those summary statistics can then be used in standard statistical and machine learning models.

Identify Morphologic Features Distinguishing Healthy and Cancerous Prostate Glands In collaboration with Dr. Silverman, we will use 3D renderings of PBCT scans of prostate cancer developed in Aim 2 for the identification of morphologic features that distinguish healthy from cancerous prostate glands. Using those scans we will create a labeled dataset of over 100,000 images using standard synthetic-data techniques from computer vision: each rendering will be subset into 500 contiguous, randomly selected sub-regions; each sub-region will be randomly rotated, flipped, and inverted to create no fewer than 10 synthetic copies of the original sub-region; additional noise will be added to those synthetic copies to ensure robustness of learned morphologic features to measurement noise. Contrast-based thresholding of each image will be used to extract point clouds focused on gland cell nuclei. Each point-cloud will then be transformed via the Smooth Euler Characteristic Transform (SECT; which is a statistic calculated from a filtration over the point-cloud) that was recently introduced by Crawford et al. (2020) [4] for predicting patient outcomes in Glioblastoma Multiforme from  morphologic features in MRI images. Importantly, the SECT maintains information about the homology of the point cloud but transforms that information into the space of smooth functions (into a Sobolev space). This is important as there are numerous statistical and machine learning tools available for modeling such smooth functions.

Using this function representation, we will turn the problem of identifying morphologic features into an equivalent problem of identifying a small number of simple functions that can distinguish whether an image came from a cancer or healthy biopsy. More specifically, after turning each synthetic image into a continuous function, those functions will be decomposed into a cubic-spline basis. We will then use standard variable selection techniques from machine learning to identify a small number of components splines that are best able to classify the images. Such tools include penalized regression based methods (e.g., ℓ_1 penalized logistic regression) as well as algorithmic approaches (e.g., forward selection). Standard techniques such as cross validation will be used to help ensure the generalizability of identified morphologic features. Combined with our synthetic data techniques and our use of an age-matched cohort, this approach will ensure that the morphologic features we identify are robust. Ultimately this study will provide a set of morphological features that can be used to answer these questions directly in a low-dimensional (e.g., <5 dimensional) vector, providing a concise yet maximally informative representation of the critical morphological characteristics that distinguish healthy versus cancerous prostate tissue.

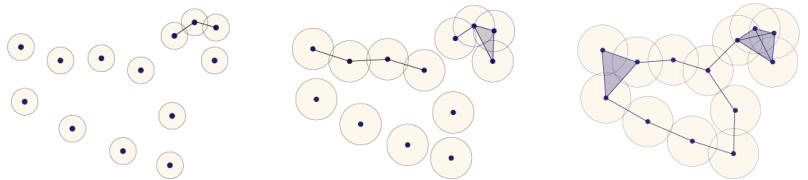


Figure 3: Illustration of a Vietoris-Rips complex being formed about a point cloud

Characterize Morphologic Variation in Prostate Glands Is unknown whether the variation in the morphologic features distinguishing cancer is greater within a single prostate biopsy, between biopsies, or separately as a function of age. Answering such questions may provide fundamental insights into the heterogeneity of prostate cancer and are important considerations when designing clinical studies. Yet it has not been possible to directly answer these questions as we have lacked quantifiable morphologic features indicative of prostate cancer. In this aim, we use image features to perform the first quantitative analysis of morphologic variation in prostate glands and thereby answer such questions.

We will use the results of the prior subaim to calculate the morphological characteristics distinguishing healthy versus cancerous glandular structure. We will model the resulting data using multivariate mixed-effects models with variance components to identify the relative variation in morphologic structure within a biopsy, between biopsies of

age- and disease-matched individuals, between biopsies taken over disease-matched biopsies taken from subjects of different ages, and between age-matched biopsies that differ in disease status. Spatial correlation between morphologic features will be accounted for using a first-order conditional-autoregressive variance component. Overall, this will result in a form of multivariate ANOVA where the relative contributions of different factors (e.g., space, age, disease status) to overall morphologic variation will be quantified.

Potential Pitfalls and Alternative Approaches Overall the modeling components of this aim are relatively straightforward as they make use of established statistical and machine learning tools that are commonly used within the Silverman lab. The major complication we might encounter is challenges in sampling glandular point-clouds from 3D rendered PBCT images. Based on our preliminary analyses of PBCT imaging of prostate biopsies (e.g., Figure 1), this does not seem likely. Still, it is possible that variation within disease phenotypes such as variation in stromal density or patterns of necrosis may complicate this task. Should we encounter this challenge, we will switch from simple contrast-based thresholding to newer deep-learning based image segmentation algorithms (e.g., [Ilastik](#), [W](#)hich we have previously used to segment specific structures in micro-CT images [36]. Once segmented, point-clouds can be generated by uniformly sampling within the segmented gland regions.

Timeline

The proposed aims are synergistic and the success of each aim helps to further subsequent aims. Still, we already have sufficient preliminary data such that the aims can be performed in parallel and the success of each aim is not dependent on previous aims. We have already optimized PBCT for imaging prostate biopsies and therefore Aim 2 can be performed without Aim 1. Similarly, we already have high-resolution full-biopsy images from a small number of healthy and cancerous prostate biopsies and therefore we can begin preliminary work on Aim 3 even before obtaining the full dataset proposed in Aim 2. Moreover, each aim will correspond to at least one first-author publication. The proposed aims will be distributed evenly across the years of this award. Preliminary data and current work (Y0) provide a foundation for the experiments of Aim 1 that will extend into year 1 of the award. Additional imaging time at LBNL in year 1 will allow data collection for Aim 2. In between imaging allocations we will recruit pathologists and perform a preliminary study. Upon completion of imaging experiments we will test the non-inferiority of our method. Aim 3 will begin with preliminary concept study and algorithm development independent of images and will result in a morphologic analysis of variation over years 1 and 2 of the award. In years 3 and 4 of the award (medical years denoted M) thesis and manuscript edits will be made during dedicated research blocks. Below we provide a timeline for completion of each major task.

Aims	Yr0	Yr1	Yr2	M
A1: Perform Imaging for PCBT Atlas				
A1: Develop Neuroglancer Web Interface for Atlas				
A2: Perform PCBT Imaging for Prostate Cancer Study				
A2: Recruit and Survey Clinical Pathologists				
A2: Perform Data Analysis to Test Non-Inferiority				
A3: Perform Morphologic Analysis of Variation				
All: Edit thesis and additional manuscripts				

Future Directions

At the completion of this study, we will have developed the first protocol for high-resolution, stain-free, 3D imaging of whole biopsies and we will have evaluated the utility of this technology in the context of prostate-cancer. This work will open up multiple promising avenues for future research. We highlight just three. First, the PBCT imaging protocol we develop and demonstrate in Aim 1 is not limited to prostate cancer. Based on this technology, my approaches to prostate cancer in Aim 2 and 3 could be easily adapted to study other soft-tissue cancers. Second, Aim 2 only serves to evaluate the utility of PBCT imaging in the context of prostate cancer. Should Aim 2 establish that PBCT imaging can be used to grade prostate cancer, future work is needed to more formally study best practices for how to integrate PBCT imaging into clinical-practice. Finally, our analysis of morphologic variation is just one potential use for the morphologic features we develop in Aim 3. Future studies will be needed to evaluate the utility of these features as biomarkers for diagnosis, prognosis, and treatment of disease.

RESPECTIVE CONTRIBUTIONS

The goals and expected outcomes of the proposed research build on studies published in the Cheng Lab, specifically the work by prior MD/PhD student Yifu Ding (Ding et. al 2019). However, micro-CT in general has only been applied to soft tissue cancer research relatively recently. In developing this proposal, we determined a gap in the field that our lab was uniquely positioned to contribute to. My long term goals are centered around contributing to cancer research, and I identified micro-CT as a technology with increasing capability to improve our ability to measure and even diagnose malignancy. This proposal builds on current literature in micro-CT and combines the technology with novel statistical methods to better understand prostate cancer and other malignancies. This work does so while prioritizing my training in statistics and programming. In writing this proposal, I read Dr. Silverman's R01 (1R01GM148972-01) and studied the structure of his writing and the strategy with which he supported his aims. I have been discussing the elements of this proposal since July of 2023. I defended my comprehensive exam in August and incorporated feedback from my committee to refine the aims of my research. Through joint discussions with Dr. Warrick (see reference letter) and Dr. Silverman in the wake of my exam, we decided to pivot to make prostate cancer a focus of this proposal. I worked diligently with Dr. Warrick (see letter of reference) to write and submit an IRB proposal ahead of a trip I led to the LBNL in early October of this year. Our IRB application was approved and Dr. Warrick was able to identify samples that I went on to prepare and image at beamline 8.3.2 (data shown in figure 2 of this proposal). I collaborated with Dr. Patrick La Riviere (see letter of reference).

The preliminary data collected under the direction of Dr. Warrick supports the feasibility of this proposal and I will continue to build on this work as I begin with aim 1. Dr. Warrick, Dr. Gunther, and Dr. Hu will contribute samples for aim 1. Under the direction and guidance of Dr. Cheng and Dr. La Riviere, I will conduct micro-CT imaging acquisition and reconstruction in aim 1, along with data acquisition, visualization, and processing for aim 2. Dr. Silverman and Dr. Warrick will mentor and assist me through the study design and statistical analysis of aim 2, and Dr. Silverman will directly oversee algorithm development and programming in aim 3. I will review results and prepare manuscripts in close discussion with Dr. Silverman, Dr. Cheng, Dr. Warrick, and Dr. La Riviere. I will review data and critique with additional support faculty members such as Dr. Jiafen Hu and Dr. Edward Gunther (see letters of support) to improve my knowledge, writing ability, and navigate the peer review process. I am thankful to be supported by a multi-disciplinary research team that is led by the close guidance of Dr. Silverman, who has set a great example from the start of how I aim to run my own lab when I am an independent investigator.

SELECTION OF SPONSOR AND INSTITUTION

I decided to pursue a career as a physician scientist after successfully defending my undergraduate honors thesis and performing full time research as a member of the Tolbert lab. I applied broadly to programs with strong computational research opportunities. I interviewed at Penn State College of Medicine (PSCOM) and discovered that this institution not only provided rigorous research opportunities, but also provided a holistic, supportive, and thorough training environment that extended to clinical medicine as well. The humanities curriculum at PSCOM is unmatched and interfaces with every aspect of our training - research and clinical initiatives are each grounded in problems that plague our patients.

I started my MD/PhD journey with a pre-matriculation rotation in the Cheng Lab that formed the early foundation for this research proposal. Dr. Cheng is an outstanding research mentor whose track record as an investigator is unmatched. He has a long history not only of high-impact publications and NIH funding, but also of leadership and collaboration in several fields. The micro-CT experiments pioneered in his lab prior to and during my rotation were compelling, but the aspect of the Cheng lab that attracted me the most was the interdisciplinary team that he leads in his initiative for computational phenomics. For example, the lab has brought in experts in the field of medical physics such as Dr. La Riviere (see letter of reference) for experimental design and guidance. Members of his lab conduct research in model organisms and human disease alike, exposing students to a wide array of expertise. Also, the lab recently pioneered the use of novel micro-CT detectors, generating 3D datasets that were entirely unique in their resolution to field-of-view ratio. Few opportunities within this lab to conduct research on the 3D imaging of human cancer biopsies. The unique technology and collaborative research environment cemented my decision to work with Dr. Cheng.

I completed my next laboratory rotation by spending a month working with Dr. Silverman in the summer between the first and second years of medical school. I had first learned of Dr. Silverman's work through his collaboration with the Cheng lab, and I decided to pursue a rotation in his lab because of my background in computational work and the unique data analysis problems I had encountered during my first rotation. Dr. Silverman immediately held me to a high standard while simultaneously investing time in walking me through detailed and difficult concepts in math and statistics. To this day he does not hesitate to work through problems on the whiteboard with myself or any of his students, and it is this investment in trainees that has guided me as a grad student and motivated me to pursue a career leading a team of my own.

I have built a strong foundation over the course of the past year and a half that will allow the co-sponsorship between Dr. Silverman and Dr. Cheng to be successful. We have a common goal in our work to conduct research that will help patients, pathologists, and the broader imaging community, and this guides our choices in the lab. Both Dr. Cheng and Dr. Silverman have encouraged my independence and curiosity while holding me accountable for and helping me understand my missteps. It is their understanding and flexibility that allowed me to perform imaging experiments that parted from the lab's status quo and generated the preliminary data for this grant proposal. They have provided me with research opportunities I will forever be grateful for - due to their collaboration my research takes me from the bench in Hershey to a synchrotron beamline in California and back to a whiteboard in Happy Valley. Their ability as mentors to both look out for my success and push me out of my comfort zone towards rigor and consistent progress will contribute greatly to my growth into an independent physician-scientist.

In addition to Dr. Silverman and Dr. Cheng, I am supported by a strong thesis committee, faculty group, and broader research team who will support, collaborate with, and train me throughout the duration of this award. Several of these faculty maintain an open-door policy and have helped me to design experiments and prepare samples (Dr. Warrick, Dr. Gunther, and Dr. Hu). Dr. Warrick is the Chief of Anatomic Pathology at the College of Medicine (see letter of reference), and he has directly overseen our sample selection and preparation in the proposed work. Dr. La Riviere (see letter of reference) is a long-time collaborator of Dr. Cheng and Dr. Silverman, and we have already initiated multiple projects focused on the improvement of phase-contrast imaging and its applications to the investigation of human cancer. I am thankful to have assembled a supportive, knowledgeable, and versatile team that will drive me to not only complete the proposed work, but also to meet and exceed my long-term career goal of becoming an independent physician-scientist.

TRAINING IN THE RESPONSIBLE CONDUCT OF RESEARCH

In addition to its emphasis on the humanities, PSCOM invests a significant amount of resources and training time in ensuring that students and faculty who conduct research within the University are well-versed in ethical conduct even in sensitive and difficult materials.

Biomedical Research Ethics (BMS 591)

All graduate students at Penn State take an ethics course within the first two years of graduate school to set the standard for their research and scholarship. BMS 591 is an in-person course offered at Hershey that employs weekly quizzes, guest lecturers, and team-based learning (TBL) to instill a rigorous understanding of ethical principles among students. The course meets weekly on Fridays and covers a new topic each week. Students are provided with assigned prereading from the textbook the *Introduction to the Responsible Conduct of Research* by Nicholas Steneck. Additionally, the course directors provide students with additional summative powerpoint presentations, news articles, and other media that they are expected to have studied ahead of taking an individual closed-book quiz at the start of the class each week. A passing grade in the class is required for progression through graduate programs at Penn State.

Collaborative Institutional Training Initiative (CITI):

PSCOM requires students and faculty to study online modules that cover standard operating procedures for sensitive materials such as research involving bloodborne pathogens and human subjects research. Each module contains a written and pre-recorded video component that instructs researchers and clinicians on how to ethically and responsibly navigate research studies and potentially hazardous situations in the laboratory. To receive certification, participants must take a quiz and receive a score of at least 80 percent. Through the CITI training portal I have completed and received certification for "Protection of Human Research Subjects", "Responsible Conduct of Research", "OSHA Bloodborne Pathogens", and more. Throughout the duration of this award I will continue to maintain these certifications and complete additional ethics and safety training as new experiments are founded.

Institutional Review Board (IRB) Study:

I have applied and received approval for an IRB study that will cover imaging experiments using human prostate cancer samples. This IRB was compiled under the direction of Dr. Joshua Warrick (see reference letter), and I received hands-on training in the responsible conduct of research with a focus on the management of protected health information (PHI). Our approved study is evidence of our productive collaboration and sets a precedent for a thorough understanding of the responsible conduct of research throughout the execution of the proposed work.

PSCOM MSTP Program Annual Responsible Conduct in Research Seminar:

All students in the MD/PhD program at PSCOM annually attend a seminar where they review responsible conduct of research in a didactic format. Each session is led by the Ethics Consultation Service at PSCOM and is required for MD/PhD students.

Silverman Lab Meetings:

The Silverman lab hosts regular discussions on responsible data analysis and ethical use of statistical models. Our weekly lab meeting regularly involves discussion of literature articles in the field of sequence count data, conformal prediction, and causal inference. Dr. Silverman and Dr. Michelle Nixon regularly provide instruction at the whiteboard to review concepts in statistics and walk the lab through discussion of logical and rigorous interpretation of data.

Cheng Lab Meetings:

The Cheng lab also covers topics in the responsible conduct of research during its meetings, especially in the context of accurate representation of data and the biological context of images. Dr. Cheng routinely reviews medically relevant concepts such as the fundamentals of histopathology with grad students and emphasizes rigor and reproducibility in image reconstruction and interpretation.

SPONSOR AND CO-SPONSOR STATEMENTS

Justin D. Silverman, M.D., Ph.D.

Research Support Available

Source	ID	Title	PD/PI	Start	End	Amount
NIH	1R01GM148972-01	Addressing Measurement Limitations for Sequence Count Data	Silverman, JD	9/2022	8/2025	\$599,945
PNNL	1R01GM149650-01	A Novel high resolution MS platform for high-throughput screening of G protein-coupled receptors	Jacobs, J & Rajagopal S	6/2023	3/2027	\$86,000
Duke University	2R01DK116187-06A1	Dietary plant diversity and the human gut microbiome	David, LA	5/2024	5/2028	\$168,134

Beyond the funding listed above and other proposal that are currently submitted or planned, the PI has over \$400,000 in startup funds from The College of Information Science and Technology (IST) at the Pennsylvania State University which can be used to support Mr. Sugarman's research and training.

Sponsor's Previous Fellows/Trainees

I have trained and mentored two doctoral students (both graduated with Ph.D. degrees) and two masters students (both graduated with M.S. degrees) and am currently training or mentoring 3 doctoral students and 1 Assistant Research Professor. Of the two doctoral students that have graduated from my lab, one currently holds a senior position in industry and the other is pursuing postdoctoral education.

Name	Position in Silverman Lab	Current Position and Institution
Kimberly Roche, Ph.D.	Predoctoral	Senior Translational Scientist at Tempus Labs
Emily Van Syoc, Ph.D.	Predoctoral	Postdoctoral Research, PSU
Farhani Momotaz, M.S.	Masters Student	Research Associate, PSU
Zhao Ma, M.S.	Masters Student	Research Assistant, Zhejiang University

Keith C. Cheng, M.D., Ph.D.

Research Support Available

My lab also has substantial research support that will provide Andrew with a surplus of the resources needed to carry out this proposal. In addition to our track record of NIH funding, our group maintains active collaboration with synchrotron beamlines such as the Lawrence Berkeley National Laboratory where Andrew has collected preliminary data and will continue to carry out experiments to support the proposed work.

Source	ID	Title	PD/PI	Start	End	Amount
NIH	1R24OD18559	Groundwork for a Synchrotron MicroCT Imaging Resource for Biology	Cheng, KC	8/2015	7/2024 (NCE)	\$2,680,046
NIH	R24OD18559	Renovation Supplement to Groundwork for a Synchrotron MicroCT Imaging Resource for Biology	Cheng, KC	8/2022	7/2024	\$231,489
DOE/LBNL	ALS-11922	X-ray histotomography applications of new wide-field, submicron resolution lens and camera systems	Cheng, KC	7/2022	6/2025	Synchrotron Imaging Time
NIH	1R24OD035407-01A1	Building a Wide-field, High-resolution Histotomography Resource for Biology	Cheng, KC	5/2024	5/2028	\$3,862,968 (score 20) Council 1/24

Co-Sponsor's Previous Fellows/Trainees

I train and mentor doctoral (including MD, Veterinary, PhD, and post-graduate) students in multiple fields, including Genetics, Biomedical Sciences, Pathology, Bioinformatics and Genomics, and Anatomy, including several MD/PhD students. As an active participant in intercollege and interdisciplinary programs including the Huck Institute for the Life Sciences and Institute for Computational and Data Sciences involving undergraduate and graduate students and faculty on multiple Penn State's campuses, and have served on graduate committees for students in Information Sciences and Technology and Computer Sciences, I interact with students at multiple levels of training and fields. I am currently training or mentoring 3 doctoral students, a postdoctoral student, a post-baccalaureate student, and 1 Assistant Research Professor. My mentees include the following:

Name	Current Position and Institution
Rebecca Lamason, Ph.D. (postbaccalaureate)	Associate Professor of Biology, MIT
Darin Clark, Ph.D. (postbaccalaureate)	Assistant Professor of Radiology, Duke Univ
Amogh Adishesha, Ph.D. (IST)	Applied Scientist, Captions (captions.ai)
William Zinnanti, M.D., Ph.D. (Biomed Sci)	Private Practice for Child and Adult Neurology
Brian Canada, Ph.D. (Bioinformatics/Genomics)	Chair, Department of Computer Science, Univ of South Carolina
Yifu Ding, M.D. Ph.D (Biomed Sci)	Resident Physician in Radiation Oncology, Emory University
Spencer Katz, M.D. Ph.D (Biomed Sci)	Resident Physician in Pediatric Medical Genetics, Cincinnati Children's Hospital
Maksim Yakovlev, Ph.D (Biomed Sci)	Postdoctoral Researcher, Argonne National Laboratory

Sponsor Statement

Training Plan, Environment, Research Facilities

Training Plan: The goal of this F30 application is to formally train Andrew in basic and translational sciences. He will continue to receive rigorous and comprehensive training from a variety of sources including coursework, one-on-one mentorship, discussion with other graduate students, and group meetings with Dr. Silverman and Dr. Cheng and Synchrotron and University Imaging Groups.

Andrew has completed and excelled in all formal coursework requirements for the Bioinformatics and Genomics PhD program and has passed his comprehensive exam. He will supplement this coursework to advance his skills in statistics and to support aims 2 and 3 of this proposal. In addition to coursework, Andrew has made strides as

a researcher over the first year and a half of his PhD. He has studied the fundamentals of probability theory in the Silverman lab, and developed expertise in both sample-preparation, micro-CT physics, and image processing in the Cheng lab. His knowledge of statistics and machine learning and experience with synchrotron microCT allowed him to contribute as an author to a recent publication in *eLife* that constitutes part of the foundation of this proposal. With regard to sample preparation and micro-CT, Andrew has absorbed knowledge from prior students in the Cheng lab and now occupies a leadership role in synchrotron trips.

As Andrew continues to conduct experiments in support of the goals of this proposal, his training will focus on two gaps in his development as a trainee. First, Andrew will receive further one-on-one training in statistical methods under the direction of Dr. Silverman. He will work through the “Matrix Algebra from a Statistician’s Perspective” textbook by David A. Harville, and in addition will take the next course in Topological Data Analysis offered at Penn State to enhance his knowledge of persistent homology and develop his expertise in support of Aim 3.

Clinical Training Plan: Even though both of us work full-time on basic science research, both Dr. Cheng and I have completed clinical training of our own and understand the importance of continuing to cultivate knowledge and patient-centered clinical skills during the graduate studies. Investing in clinical training during Andrew’s graduate years will prepare him for his transition back to the third year of medical school and for his long term goal of becoming an independent physician scientist. In Andrew’s case, we view that this will actually enhance the value of his graduate studies, given that Andrew draws substantial motivation from what he has observed in clinic and has initiated this proposal with the hope it will be able to contribute to problems that affect cancer patients.

Andrew participates in the Clinical Exposure Program (CEP) since he has completed his comprehensive exam. CEP is a tool for MD/PhD students to prepare for the clinical portions of medical training while they complete their PhDs. He has chosen to work with Dr. Lilia Reyes in the Pediatric Emergency Department to focus on becoming a well-rounded medical student and ultimately a versatile physician scientist. To supplement this work, he will also continue to shadow Dr. Raymond Hohl, an attending in hematology/oncology at Hershey and the director of the Penn State Cancer Institute, in addition to serving as one of Andrew’s thesis committee members. Committee member Dr. Joshua Warrick, Director of Anatomic Pathology at the College of Medicine, brings invaluable perspective gained from a decade of supervising residency training, and therefore work with Dr. Cheng (who also is Board-certified in Anatomic Pathology) to ensure that the projects Andrew pursues are of maximal scientific and clinical value.

Professional Development Plan: Andrew will have career development opportunities through presenting at multiple conferences, seminars, workshops, and participating in unique research focused lab trips. He has already presented his work at the MD/PhD National Student Conference and the Bioinformatics and Genomics program retreat. Presentation skills are essential components of a successful career in science and Andrew will have ample opportunities to develop strong scientific communication abilities in both of our labs.

We will be sure that Andrew is also able to attend the MD/PhD retreat twice a year, where he will observe and interact with keynote speakers and colleagues that will also contribute mentorship to his development as a physician-scientist. Wherever possible, Andrew will also engage in other presentations and seminars on campus such as the Graduate Student Research Forum, the Bioinformatics and Genomics Student Seminar, and the Experimental Pathology Colloquium.

Environment: Andrew is a graduate student in my laboratory and is part of our regular lab meetings. These meetings are devoted to the development and application of statistical methods for complex biomedical data. A full time assistant research professor (Dr. Michelle Nixon, PhD in Statistics) is employed by the lab and is available to assist graduate students in their research. Lab meetings are held for 2 hours every other week during which time one student or faculty member presents their work with ample time for questions and discussion. Presenters at lab meeting rotate and each student or faculty member presents approximately once every other month. Beyond lab meetings, I meet with Andrew individually for between 1-2 hours weekly to discuss his progress and provide training and/or career mentorship. These meetings are held more frequently during key periods such as during manuscript preparation, preparation for committee meetings, or preparation for conference presentations. We have established Andrew’s thesis committee which meets annually to discuss his progress. Beyond these annual meetings Andrew meets regularly with each committee member to discuss aspects of his research that intersect with their own expertise.

Andrew has access and regularly interacts with a wide range of extramurally funded researchers with expertise

directly applicable to the proposed work and to Andrew's larger career goals. I am an assistant professor in the College of Information Science and Technology (IST), the Department of Statistics, and the Department of Medicine. I have formal training in both statistics (PhD) and medicine (MD) which gives me a keen appreciation of both the translational context of key biomedical questions as well as the statistical challenges associated with answering those questions. I am currently the PI on an NIH R01 award which focuses on developing key theory and tools for scale reliant inference that are complementary yet non-overlapping with the proposed work (1R01GM148972-01). Other extramurally funded faculty that Andrew regularly interacts with include Dr. Francesca Chiaromonte (Statistics, Focus on Bioinformatics and Genomics), Dr. Matthew Reimherr (Statistics, Focus on Functional Methods for Biostatistics), and Dr. Vasant Hanovar (IST, Focus on Statistical Inference and Machine Learning for Biomedical Data).

Outside of the lab and his thesis committee, there are numerous venues on campus that regularly feature current work from extra and intra-mural researchers. These include the Bioinformatics and Genomics Colloquium Series, The Statistics Seminar Series, and the Microbiome Center Seminar Series.

Research Facilities: Andrew will have access to the facilities and resources necessary to undertake and complete the proposed work. Andrew has dedicated office space in my lab located in the Westgate building at the Pennsylvania State University. This building is modern (built in 2004 and recently renovated), centrally located on campus, and within a 5 minute walk from the Statistics department and the Huck Life Sciences Institute. Andrew's office space is just down the hall from my own. Andrew has access to numerous high performance computing clusters both through the College of IST as well as through the Institute for Computational and Data Science where I have a dedicated allocation for my lab on the Roar cluster. The Roar cluster contains over 36,500 computing cores and 25 PB of storage. Andrew will have access to this entire cluster as well as having priority (<1min wait time) access to my personal allocation which includes one high-memory node (40 cores, 1TB RAM), two standard-memory nodes (each with 40 cores and 256GB RAM), 10TB Active Group Storage, and 20TB Nearline/Archive Storage. To supplement these capabilities the Roar cluster also has a help-desk service (i-Ask) which assists users with all aspects of cluster computing from debugging job execution to systems administration and database maintenance. Beyond computing resources, the Penn State University Libraries rank among the top 10 North American research libraries based on the Association of Research Libraries Library Investment Index Rankings. The library system consists of 36 libraries at 24 locations throughout the Commonwealth of Pennsylvania. The University Libraries house a collection of nearly 6 million items, with annual additions of roughly 100,000 volumes. The libraries have access to 579 online databases and other e-resources and subscribe to nearly 118,000 online, full-text journals.

Number of Fellows/Trainees To Be Supervised During the Fellowship

Three, in addition to Andrew:

- Tinghua Chen, Graduate Student (Informatics)
- Kyle McGovern, Graduate Student (Bioinformatics and Genomics)
- Won Gu, Graduate Student (Statistics)

Applicant's Qualifications and Potential for a Research Career

While I have only formally advised Andrew for the past year and a half (since he started in the PhD portion of his training), it is important to note that I have been working with Andrew for almost three and a half years. In that time, I have watched Andrew combine his intellect, creativity, and grit to great effect. For example, Andrew learned the foundations of probability theory, Frequentist statistics, and topological data analysis in just over two months after he decided he wanted to learn how to rigorously model shape in 3D images. This is a remarkable feat and, in my experience, tantamount to crushing a wall that other students would simply walk away from. Andrew displays incredible intellectual curiosity has the grit and intellect to follow that curiosity. Combined with his long-term interests in cancer, I have no doubt that Andrew will ultimately lead an independent research program which will improve human health.

Co-Sponsor Statement:

It is important to point out the complexity of Andrew's project, which requires knowledge of biological questions, sample preparation, micro-CT physics using both cone-beam based local and parallel-beam based synchrotron x-ray sources, image processing, machine learning, and interacting in an multidisciplinary environment - a greatest

strength of the Cheng lab. Andrew caught on quickly in terms of all these areas, never shying away from complexity. He showed leadership skills in first learning how synchrotron microCT imaging trips work, and then being a primary manager of the most recent experimental trip to the LBNL. He led discussion of phase-based, unstained sample microCT based on readings from the literature as personally guided by Dr. Cheng. The striking preliminary data inspired and are foundational to this proposal; its extensions across tissue and sample types will contribute to multiple publications. I have mentored numerous PhD and MD/PhD students, and have significant experience and understanding of the various types of qualities that can drive success in interdisciplinary science. Andrew possesses a wonderful set of qualities, giving me every faith that he will become an excellent physician-scientist and make impactful contributions to cancer research.



Jiafen Hu, Ph.D.
Assistant Professor
Department of Pathology
and Laboratory Medicine

Penn State College of Medicine
The Milton S. Hershey Medical Center
Jake Gittlen Cancer Research Institute, H059
500 University Drive
Hershey, PA 17033-2390

Voice: (717) 531-4700
Fax: (717) 531-5634
E-mail: fjh4@psu.edu

12-02-2023

Dear Members of the Review Committee,

It is with great enthusiasm that I extend my full support to Andrew Sugarman's F30 grant application, titled 'High-Resolution Wide-Field 3D Histopathology for the Morphological Characterization of Prostate Cancer.' Andrew's research, focusing on the optimization and application of Propagation-Based phase-contrast CT (PBCT) imaging in prostate cancer, represents a crucial advancement in the field of experimental pathology.

I am an Assistant Professor in the Department of Pathology and Laboratory Medicine within the Division of Experimental Pathology at the Penn State Cancer Institute. My expertise in papillomavirus-driven infections and cancers positions me to contribute considerably to this proposal. I have closely followed Andrew's progress under the mentorship of Dr. Keith Cheng, a collaborator of mine for many years in several multidisciplinary studies, and am impressed with Andrew's understanding of micro-CT and its transformative potential in cancer diagnosis and research. I will eagerly contribute both to the proposed research and to Andrew's development as an independent scientist in this field. Specifically, I will support this project by providing tissue blocks of human and mouse oropharyngeal cancer, aiding in the completion of aim 1 of this application. I will also support projects in the other aims of the current proposal using my expertise in experimental pathology.

I look forward to evaluating the 3D imaging data that will be generated and reviewing the results with Andrew in biological context. He has assembled a diverse and complementary team that will undoubtedly produce novel and exciting data to move this new field forward. This proposal has my full support.

Sincerely,

Jiafen Hu

Institutional Environment and Commitment to Training; Additional Educational Information

Penn State is strongly committed to the training of physician-scientists throughout the 30-year history of its MSTP program. Currently under the leadership of Co-Directors Robert Levenson, Ph.D. and Leslie Parent, M.D., and Associate Directors Aron Lukacher, M.D./Ph.D., Kirsteen Browning, Ph.D., and Melissa Rolls, Ph.D., the mission of the Penn State MSTP program is to recruit motivated, diverse, experienced students and provide them with exceptional integrated clinical and research training in preparation for entering the physician scientist workforce. We have benefitted from strong institutional support which allows us to recruit 9 fully funded students per year.

Over the past several years, the Penn State College of Medicine (PSUCOM) has undergone a major expansion in facilities with the completion of the Penn State Cancer Institute and Children's Hospital along with its recent three-story addition. Construction was completed on a 46,000-square-foot, \$54 million Technology Center which provides centralized space and enhanced security for patient, research and educational data as well as the computing power for analyzing complex large data sets. This facilities expansion has been accompanied by the recruitment of >70 MD and MD/PhD physician-scientists into the junior and mid-level ranks, as well as into key positions, including Department Chairs in Medicine, Neurology, Pediatrics, Microbiology and Immunology, as well as the Cancer Center Director. These recruiting efforts aim to provide mentors and role models for MSTP students and increasing translational research at the COM. The research opportunities for MSTP students have also been enhanced by the addition of a cadre of excellent scientists at the University Park (UP) campus. We have established joint-degree programs with Engineering Sciences and Mechanics (ESM), Molecular and Cellular Integrated Biosciences (MCIBS), and the Anthropology graduate program, allowing MSTP students to perform their dissertation research in a wide variety of disciplines at UP. The success of the Penn State CTSI, a cross-campus, NIH-funded program, has played an important role in establishing effective and fruitful collaborations between investigators. Through workshops, seminars, training programs, and pilot project funding, the CTSI has strengthened bridges between MSTP training faculty members and students across the two campuses.

Andrew Sugarman completed his undergraduate training in Biochemistry at Oberlin College in 2019 and worked for a year as a Research Assistant in the lab of Dr Blanton Tolbert in the Department of Chemistry at Case Western Reserve University where he applied NMR and Small Angle X-ray Scattering technologies to investigate the structures of RNA and its interacting proteins, which resulted in 4 co-authored publications. Andrew matriculated into the MSTP program in 2020, and has completed his preclinical medical school curriculum and Step 1 of the USMLE. He entered the Bioinformatics and Genomics graduate program in Fall 2022 and successfully passed his written and oral comprehensive exam in Summer 2023. He is co-mentored by Drs Justin Silverman (Assistant Professor of Statistics, Assistant Professor of Medicine) and Keith Cheng (Distinguished Professor, Department of Pathology and Laboratory Medicine, Department of Biochemistry and Molecular Biology, Department of Pharmacology & Penn State Cancer Institute), both of whom are MD PhDs, where he is working to adapt high resolution computer tomography to create 3-dimensional structures from biopsy images. Not only would this allow greater qualitative assessment of tissue biopsies but it would also permit quantitative analyses of tumor structures "in situ" within their broader anatomical framework. Andrew has continued to refine his clinical skills by attending pediatric emergency room clinics with Dr Lilia Reyes, an Assistant Professor in the Departments of Pediatrics and Emergency Medicine and, given his interest in pursuing a residency in internal medicine and a fellowship in hematology/oncology, has also spent considerable time attending clinics with Dr Raymond Hohl, Director of the Penn State Cancer Institute.

The MSTP program currently matriculates 9 students and benefits from the active involvement of faculty members on a variety of levels, including as members of the Steering Committee, training faculty, interviewers, teachers, mentors, and dissertation advisors. The training faculty consists of 75 members, 50 in Hershey and 25 at the UP campus. Training faculty are selected based on several criteria including (i) adequate funding to support trainee stipend and tuition; (ii) proven track record in graduate training, especially MSTP students; (iii) excellence in research as judged by grant support and publication record; and (iv) excellence in mentorship. Junior faculty members may serve as research mentors if they have promising research potential, enthusiasm for training, and funds to support a student from extramural sources or departmental start-up funds.

The MSTP program integrates medical and graduate education, providing flexibility for students to design a tailor-made educational experience that prepares them for their individual careers. During the first two years of the program (M1-M2), students complete the preclinical medical education, which is organized around organ systems and consists of both lectures and case-driven problem-based learning (PBL) via small group discussions. Students also take several graduate courses during the first two years of medical school, and complete three research rotations. Students usually choose a research mentor shortly after completing the third

research rotation and a graduate program that best suits their interests. At the end of M2, all students must pass the USMLE Step1 exam before they can begin the graduate phase of training. It is expected that students in the MSTP program will complete their research and dissertation defense within a 4-year period. Beginning with the fall semester of G2, students experience the life of a physician scientist with admixed research and clinical duties through our novel course, "Translational Research in Medicine" (BMS597). This course ensures ongoing clinical experiential learning and one-on-one training with a clinician scientist preceptor, providing an opportunity for students to build upon and hone their clinical skills while beginning to formulate a translational research project. To better align their clinical and research training after returning to medical school, MSTP students participate in a second course, "Advanced Translational Medicine" (MED797), a longitudinal clerkship during M3 in which students work one-on-one with a clinician scientist preceptor. This course provides students an opportunity to integrate their clinical and research training to address a medical problem observed in clinic. Throughout the program, students are exposed to topics in clinical, basic, and translational research via the monthly MSTP seminar series, the bimonthly Clinical Research Conference (CRC), and the annual MSTP retreat. The MSTP program requires that a student have at least one first-author, peer-reviewed manuscript accepted for publication, has successfully defend their PhD dissertation, and that the dissertation is submitted and approved by the graduate school, before progression to M3. All MSTP students are required to take "Transition to Clinical Medicine", which reinforces the basic skills and knowledge a student needs to enter the clinical training years. MSTP students in the M3 and M4 years choose a Clinical Advisor who monitors the students' progress through clerkships and advises students on applying for residency positions.

During their tenure in the program, MSTP students are required to attend the monthly seminar series where students give an annual oral presentation of their research project. Students also participate in the "Translational Research in Medicine" course during the graduate years and the "Advanced Translational Medicine" clerkship in M3. MSTP students also participate in the Clinical Research Conference (CRC), a bimonthly seminar series consisting of student presentations and discussions organized around a topic in Translational Medicine using a Case Study approach. For each CRC, a physician-scientist advisor guides students through a case study with presentations focusing on differential diagnosis, pathophysiology, and a scientific paper relevant to the study case. The MSTP program holds a two-day retreat in the spring of each year featuring invited speakers from prominent MD/PhD Programs or Physician-Scientist Training Programs (PSTPs) around the country. The retreat promotes vertical interactions amongst students, provides career development information, develops professional skills, and encourages scientific exchange and network building.

The MSTP leadership believes that individual mentoring of students is essential, and meetings with students throughout the year is a high priority. Career counseling is viewed as an ongoing, continual process. Formal meetings with the Co-Directors and/or Associate Directors are semi-annual, and more often if needed. We have implemented an overarching advising, mentoring, and career development plan that begins during orientation, continues throughout the medical and graduate years, and extends through graduation and beyond. To assist in this process, we utilize an Individual Development Plan (IDP) we specifically designed for our MSTP students which is reviewed annually with the Co-Directors and their other advisors.

The Office of Medical Education has worked closely with the Co-Directors to avoid scheduling conflicts with MSTP students when making changes to the curriculum or scheduling educational sessions. There is also flexibility in the timing (January through May) that students can reenter medical school and begin clinical rotations. Finally, the Clerkship Directors have demonstrated flexibility and support for the program by allowing MSTP students flex time so they may attend program specific events such as the MSTP seminar and CRC conference. These multiple types of commitments have helped to create a vibrant and supportive academic environment in which the training of future physician-scientists represents a top priority at Penn State.

Andrew Sugarman has performed at a very high level since matriculating in the Penn State MSTP. He has is on-track to complete his PhD and return to medical school in 2026 and graduate from the MSTP program in May of 2028, and we anticipate he will continue to grow as a talented physician-scientist.

Information prepared by Kirsteen N Browning
Assistant Director, Penn State MSTP

A handwritten signature in blue ink that reads "Kirsteen Browning". The signature is fluid and cursive, with "Kirsteen" on the top line and "Browning" on the bottom line.

RESOURCE SHARING PLAN

Data Sharing Plan

The data utilized and generated throughout this research will be made publicly available within the guidelines of IRB requirements. All data that is shared publicly will be deidentified apart from the tissue of origin and diagnosis. Micro-CT scans will be hosted on either the Cheng lab reconstruction computers or within the HPC allocations of Dr. Cheng or Dr. Silverman. All 3D data that is published will be made publicly available for manipulation and analysis by other researchers and clinicians on the web-based open-source platform Neuroglancer. The Cheng lab has adapted Neuroglancer to support our data, and we will share web links in accordance with the guidelines provided by our institution, the IRB, and the publisher.



All scripts and code we generate for data analysis will also be made available on my github. This includes R code written and utilized for the non-inferiority trial in aim 2 and python code written and adapted for computational persistent homology in aim 3. Images generated from patient specimens will be labeled according to the tissue type and diagnosis and will be otherwise deidentified. In summary, all results we publish will be made publicly available.

Sharing Model Organisms

Not applicable. All model organism samples utilized in aim 1 will be specimens extracted by collaborators from animals they have sacked.

PHS Assignment Request Form

Funding Opportunity Number: PA-23-260

Funding Opportunity Title: Ruth L. Kirschstein National Research Service Award (NRSA) Individual Fellowship for Students at Institutions with NIH-Funded Institutional Predoctoral Dual-Degree Training Programs (Parent F30)

Awarding Component Assignment Request (*optional*)

If you have a suggestion for an awarding component (e.g., NIH Institute/Center) assignment, use the link below to identify the appropriate short abbreviation (e.g. "NCI" for National Cancer Institute") and enter it below in the boxes for "Suggested Awarding Components". All suggestions will be considered; however, not all assignment suggestions can be honored.

Awarding Components: https://grants.nih.gov/grants/phs_assignment_information.htm#AwardingComponents

Suggested Awarding Component: NCI

Study Section Assignment Request (*optional*)

If you have a suggestion for a study section assignment, use the link below to identify a study section(s). Enter the short abbreviation for that study section in the boxes for "Suggested Study Sections." Remove all hyphens, parentheses, and spaces. All suggestions will be considered; however, not all assignment suggestions can be honored.

For example, enter "CAMP" if you wish to suggest assignment to the NIH Cancer Molecular Pathobiology study section, or "ZRG1HDMR" if you wish to suggest assignment to the NIH Healthcare Delivery and Methodologies SBIR/STTR panel for informatics.

Study Sections: https://grants.nih.gov/grants/phs_assignment_information.htm#StudySection

Assign to Study Section:

Each entry is limited to 20 characters

PHS Assignment Request Form

Rationale for assignment suggestions (*optional*)

*Entry is limited to
1000 characters*

List individuals who should not review your application and why (*optional*)

*Entry is limited to
1000 characters*

Identify scientific areas of expertise needed to review your applications (*optional*)

Note: Do not provide names of individuals

Expertise:

*Each entry is
limited to 40
characters*

PHS Human Subjects and Clinical Trials Information

Use of Human Specimens and/or Data

* Does any of the proposed research involve human specimens and/or data?

Yes

No

Provide an explanation of why the application does not involve human subjects research

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

Other Requested Information

Human Subject Studies

Study#	Study Title	Clinical Trial?
1	3D Imaging of Prostate Biopsies	No

Delayed Onset Studies

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

Section 1 - Basic Information (Study 1)

1.1. * Study Title (each study title must be unique)

3D Imaging of Prostate Biopsies

1.2. * Is this study exempt from Federal Regulations?

Yes

No

1.3. Exemption Number

1 2 3 4 5 6 7 8

1.4. * Clinical Trial Questionnaire

1.4.a. Does the study involve human participants?

Yes

No

1.4.b. Are the participants prospectively assigned to an intervention?

Yes

No

1.4.c. Is the study designed to evaluate the effect of the intervention on the participants?

Yes

No

1.4.d. Is the effect that will be evaluated a health-related biomedical or behavioral outcome?

Yes

No

1.5. Provide the ClinicalTrials.gov Identifier (e.g. NCT87654321) for this trial, if applicable

Section 2 - Study Population Characteristics

2.1. Conditions or Focus of Study

2.2. Eligibility Criteria

2.3. Age Limits

Min Age:

Max Age:

2.3.a Inclusion of Individuals Across the Lifespan

2.4. Inclusion of Women and Minorities

2.5. Recruitment and Retention Plan

2.6. Recruitment Status

2.7. Study Timeline

2.8. Enrollment of First Participant

Inclusion Enrollment Reports

Entry#	Enrollment Location Type	Enrollment Location
The study does not have any IERs		

Section 3 - Protection and Monitoring Plans

3.1. Protection of Human Subjects

human_subjects1034805349.pdf

3.2. Is this a multi-site study that will use the same protocol to conduct non-exempt human subjects research at more than one domestic site?

Yes No N/A

If yes, describe the single IRB plan

3.3. Data and Safety Monitoring Plan

3.4. Will a Data and Safety Monitoring Board be appointed for this study?

Yes No

3.5. Overall Structure of the Study Team

Section 4 - Protocol Synopsis

4.1. Study Design

4.1.a Detailed Description

4.1.b. Primary Purpose

4.1.c. Interventions

Type	Name	Description
------	------	-------------

4.1.d. Study Phase

Is this an NIH-defined Phase III Clinical Trial?

Yes No

4.1.e. Intervention Model

4.1.f. Masking

Yes No

Participant

Care Provider

Investigator

Outcomes Assessor

4.1.g. Allocation

4.2. Outcome Measures

Type	Name	Time Frame	Brief Description
------	------	------------	-------------------

4.3. Statistical Design and Power

4.4. Subject Participation Duration

4.5. Will the study use an FDA-regulated intervention?

Yes No

4.5.a. If yes, describe the availability of Investigational Product (IP) and Investigational New Drug (IND)/Investigational Device Exemption (IDE) status

4.6 Is this an applicable clinical trial under FDAAA?

Yes No

4.7. Dissemination Plan

Section 5 - Other Clinical Trial-related Attachments

5.1. Other Clinical Trial-related Attachments

Human Subjects Rationale - Study STUDY00023526

3-dimensional virtual histology is a new and rapidly expanding field of research aimed at generating whole biopsy images for the volumetric analysis of tissue and cellular morphology[1]. The purpose of this study is to determine the 3-dimensional structure and spatial arrangement of prostate tissue architecture and glands in order to assess whether micro-CT is a viable tool for the 3-dimensional histological analysis of prostatic samples. All human tissue samples used in this study will be archival tissue already in tissue blocks. We are not collecting any new or additional biopsies from patients.

Samples will be identified by Dr. Warrick within the EHR at HMC using a natural language search in Pathnet, the pathology informatics system. All samples will be prostate needle biopsies that were taken for pathologic diagnosis and formalin-fixed and paraffin-embedded (FFPE). All samples will be from men who subsequently underwent radical prostatectomy, yielding large amounts of tissue for any required testing. Prostate biopsies will be assigned an arbitrary identifier and provided to the Cheng Lab by Dr. Warrick. Dr. Warrick will keep a key that connects the arbitrary identifier to the pathology accession number, a unique identifier assigned to the specimen in Pathnet.

Subjects cannot be injured by this protocol. There are two theoretical risks conferred by a study as we describe here. First, the patient may need his tissue to undergo molecular testing, but the study protocol has exhausted the tissue, necessitating another biopsy. Second, protected health information could be accidentally disclosed.

Our design prevents both of these risks with a high degree of certainty. First, we will only image prostate needle biopsies on men who subsequently had a radical prostatectomy. This larger specimen will contain much more information than was present on the needle biopsy, allowing for molecular testing. Second, no protected health information on any patient will be stored as part of this study. Dr. Warrick will identify patients in Pathnet, and record their pathology accession number (which is not considered PHI). The patient's name and other identifying information will never be recorded. To add an additional layer of protection, Dr. Warrick will assign each tissue block an arbitrary identifier. This identifier will be connected to the pathology accession number in a key, which only Dr. Warrick will have access to. The specimens collected in this study will be FFPE prostate core needle biopsies after they have been analyzed by a pathologist and read for diagnosis. These samples will be stored at HMC within the Cheng or Warrick Labs until they are brought on person with Andrew Sugarman for imaging at LBNL in Berkeley, CA. After imaging, they will be brought by Andrew Sugarman back to the Cheng/Warrick Labs at HMC.

Imaging data will be stored on a Cheng Lab workstation and the HMC high-performance computing cluster. Specimens will be returned to Dr. Warrick following completion of analysis.

Proposal Summary

Proposal Number: 110432 Proposal Status:
Sponsor Deadline: 12/08/2023 Submission Method:
Submission Type: Application

INVESTIGATOR DATA

PROJECT DIRECTOR / PRINCIPAL INVESTIGATOR CONTACT INFORMATION

Prefix: First Name:	<u>Andrew</u>	Middle Name:		Last Name:	<u>Sugarman</u>	Suffix:
Position>Title:	<u>Grad Student</u>	Organization:	<u>PENNSYLVANIA STATE UNIV HERSHEY MED CTR</u>			
Department:		Division:	<u>College of Medicine</u>			
Street1:	<u>500 University Drive</u>	Street2:	<u>P.O. Box 850</u>			
City:	<u>Hershey</u>	County:	<u>Dauphin</u>			
State:	<u>PA</u>	Zip Code:	<u>17033-0850</u>			
Country:	<u>USA</u>	Employee ID:				
Phone:	<u>443-683-0398</u>	Fax:				
Email:	<u>aqs6915@psu.edu</u>					

First Budget Period Effort: Calendar: Academic: Summer:

Status of PI:

Status Waiver Required?

Signed Intellectual Property Waiver Attached?

Signed Conflict of Interest Disclosure Attached?

Agency Certification Documentation Attached?

Cost Sharing Authorization Form Attached?

SPONSOR DATA

Agency:	<u>National Institutes of Health</u>
Proposal Type	
Sponsor Mechanism:	<u>Ruth L. Kirschstein National Research Service Award (NRSA) Individual Fellowship for Students at Institutions with NIH-Funded Institutional Predoctoral Dual-Degree Training Programs (Parent F30)</u>
Sponsor Type:	
Sponsor Code:	
Sponsor Name:	
SubDivision 1:	
SubDivision 2:	

PROJECT DATA

Title of Project:	<u>High-Resolution, Wide-Field 3D Histopathology for the Morphological Characterization of Soft-Tissue Tumor Biopsies</u>	
Is This a Subcontract?	No	
If Yes, who is prime?		
Type of Proposal:		
Type of Agency:	<u>Federal</u>	
Kind of Application:	<u>New</u>	
Previous Grant # or Federal Identifier:		
Change in grantee institution?		
Type of Project:		

PROJECT ADMINISTRATION

Who is responsible for this research?		
Departmental Identification Number:	Primary:	Secondary:
Departmental Name:	<u>Primary:</u>	<u>Secondary:</u>
Primary Dept. Contact Info:		
Account Classification:	Primary:	Secondary:
Other Institutional Code:		

Proposal Summary (cont'd)

NAICS Code:

COMPLIANCE DATA

Are animal subjects used? No
Is IACUC review pending?
IACUC Protocol #
IACUC Approval Date:
Are human subjects used? Yes
Is IRB review pending?
IRB Protocol #
IRB Approval Date:
Does this project involve use of any of the following? Radioactive Material(s), Radiation Producing Devices(s), Recombinant DNA, Biohazardous Chemical(s), Class IIIb or IV Lasers, Other certifications of health, safety and/or environmental compliance.

BUDGET DATA

Performance Dates	Begin Date	End Date	
First Budget Period:			
Cumulative Budget Period:			
Cost Sharing Information Committed:	Mandatory	Voluntary	
Amount:			
Source:			
Budget Period	Direct Cost	Indirect Cost	Total Cost
Total:			

AWARD DATA

Award #:	Contract #:	Date:	
Budget Period	Direct Cost	Indirect Cost	Total Cost
Total:			

EXPORT CONTROL

1. Will the project involve participation, collaboration or access to information by foreign nationals, defined as: individuals with foreign citizenship, foreign governments, foreign associations and corporations, or foreign political parties? Note: Foreign nationals granted US citizenship, or permanent residence "green card" or granted status as a "protected individual", e.g., political refugees and political asylum holders are "EXEMPT" from deemed export rule.
2. Will the project involve the shipment of equipment, technology, software, materials data or other information?
3. Will the project involve a foreign subcontract or other foreign contractual agreement?

COMMENTS AND EXPLANATIONS

PLEASE INDICATE ANY SPECIAL INSTRUCTIONS BELOW:
