

**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: Saltz, Joel H.

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

POSITION TITLE: Distinguished Professor of Biomedical Informatics, Chair of Biomedical Informatics, Vice Chair of Pathology

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
University of Michigan, Ann Arbor Michigan	BS	12/1977	Mathematics, Physics
University of Michigan, Ann Arbor Michigan	MA	12/1978	Mathematics
Duke University, Durham North Carolina	MD/PhD	09/1985	NIH MSTP; Computer Science PhD
Johns Hopkins Medical Institutions, Baltimore MD	Residency	07/1998	Clinical Pathology

**A. Personal Statement**

I am dedicated to the development of principles, techniques and tools that can be used by cancer researchers to assemble a coherent biomedical picture by integrating information from multiple complementary microscopy, Radiology imaging and molecular data sources to better understand cancer invasion, heterogeneity, metastasis and microenvironment and to develop and coordinate cancer therapies, predict outcome and response to treatments, generate basic insights into pathophysiology, and identify new treatment targets. My research in Digital Pathology dates back twenty years and funded efforts to integrate Digital Pathology and Radiology imaging began with my leadership of a P20 NIBIB funded center P20EB000591 dedicated to integrative Pathology and Radiology image analysis. My approach consists of closely coordinated efforts in image analysis, machine learning, database design and high end computing. The proposed work extends and leverages tools and methods I developed through years of funded projects supported by a wide range of institutes and agencies including NCI, NLM, NIBIB, NSF, DARPA, AFOSR, NASA, DOD and DOE to develop innovative techniques, methodologies, algorithms and software systems to support integrative data analyses, large scale digital microscopy, high-performance computing, data management, and data federation.

I have been deeply committed to biomedical informatics research, training and mentorship. I have founded three highly successful departments of biomedical informatics each with PhD training programs. The first two were at Ohio State University and Emory University. In 2013, I came to Stony Brook as Vice President for Clinical Informatics and Founding Department Chair of Biomedical Informatics – to create a living laboratory for biomedical informatics and to create a third unique biomedical informatics department dually housed in the School of Medicine and the College of Engineering.

I have been deeply engaged with the CTSA community since 2008 when I assumed leadership of the Atlanta CTSA ACTSI Biomedical Informatics Program. In that capacity, I led the development of machine learning based predictive analytics methods to support Emory's population health initiatives. I also collaborated with Morehouse School of Medicine to foster the development of AME Church based programs to improve diabetic control. I also co-led (joint with Bill Hersh from OHSU) a set of workshops to establish recommendations for the use of operational electronic health record data in comparative effectiveness research. More recently, I have played foundational roles in the N3C, which is a consortium of the CTSA hubs supported by the National Center for Advancing Translational Sciences (NCATS) and the National

Center for Data to Health (CD2H). My group collaborated with the rest of the N3C leadership team in the development of the N3C architecture and contributed tools and analysis methods to the N3C enclave for data curation, quality control, and analysis. We also became a partner in the NYU led COVID RECOVER effort jointly with CTSA sites University of Colorado, Johns Hopkins University, and the University of North Carolina. Our group has recently successfully concluded an NCATS funded project with CTSA site U Arkansas linking COVID-19 image data in the Cancer Imaging Archive with N3C clinical data. We work closely with U Arkansas and Emory in the development of integrated methods for management, curation and visualization of Radiology and Pathology image data.

Over the past eight years at Stony Brook, I have led the development of an initial multi-modal research data warehouse, as described in the Master Facilities and Equipment section. In this application, I will provide overall direction and management for Element D, module D3 and Element E project that targets development of AI Methods for Leveraging Multi-Modal data in Clinical Data Analyses.

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

UH3 CA225021

PI: Saltz; MPIs Foran, Sharma, Durban

4/1/2018 - 3/31/2023

Methods and Tools for Integrating Pathomics Data into Cancer Registries

U24 CA180924

PI: Saltz

9/1/2014 - 8/31/2020

Tools to Analyze Morphology and Spatially Mapped Molecular Data

U24 CA215109

MPIs: Saltz, Prior, Sharma; Prior Corresponding PI

9/22/2017 - 7/31/2022

TCIA Sustainment and Scalability - Platforms for Quantitative Imaging Informatics in Precision Medicine

OT2HL161847-01

PI: Haendel

10/1/2021 – 5/23/2022

Post-Acute Sequelae of SARS-CoV-2 Infection Initiative: NYU Langone Health Clinical Science Core, Data Resource Core

7U24TR002305-05

PI: Haendel

8/15/2021-6/30/2022

A National Center for Digital Health Informatics Innovation

3U24TR002306-04S1

PI: Haendel

7/1/2020 – 6/30/2021

COVID-19 Non-Clinical Research: CD2H - National COVID Cohort Collaborative (N3C)

Citations:

1. Hou, L., Samaras, D., Kurc, T., Gao, Y., Davis, J., **Saltz, J.**, Patch-based Convolutional Neural Network for Whole Slide Tissue Image Classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016*, Las Vegas, NV. PMID: PMC5085270
2. Hou, L., Singh, K., Samaras, D., Kurc, T., Gao, Y., Seidman, R., **Saltz, J.**, Automatic histopathology image analysis with CNNs. In *Proceedings of the IEEE New York Scientific Data Summit 2016*.
3. Hou, L., Samaras, D., Kurc, T., Gao, Y., **Saltz, J.**, Neural Networks with Smooth Adaptive Activation Functions for Regression, arXiv, 2016

4. Hou, L., A. Agarwal, D. Samaras, T. M. Kurc, R. R. Gupta and J. H. Saltz (2019). Robust Histopathology Image Analysis: To Label or to Synthesize? CVPR19,2019; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition

## B. Positions, Scientific Appointments, and Honors

- 2020- Vice Chair Department of Pathology, Stony Brook University, Stony Brook Medicine  
2013- Professor of Radiology, Pathology and Computer Science, Stony Brook University, Stony Brook  
2013- Vice President for Clinical Informatics, Stony Brook Medicine, Stony Brook NY 2015-  
Program Co-Leader for Imaging, Bioinformatics & Engineering Sciences, Stony Brook Cancer  
Center
- 2013- Cherith Endowed Professor and Chair, Department of Biomedical Informatics, Stony Brook  
University, Stony Brook NY
- 2011-2013 Professor and Chair, Department of Biomedical Informatics, Emory University, Atlanta, GA  
2009- Fellow, American College of Medical Informatics  
2009-2013 Fellow - Georgia Cancer Coalition  
2008-2013 Georgia Research Alliance Eminent Scholar  
2008- 2013 Director, Center for Comprehensive Informatics, Emory University, Atlanta, GA  
2008- 2013 Professor, Department of Pathology, Mathematics and Computer Science, Biostatistics and  
Bioinformatics, School of Medicine, Emory University, Atlanta, GA
- 2004-2008 Endowed Chair, Dorothy M. Davis Cancer Fund, Arthur G. James Cancer Hospital, The Ohio  
State University
- 2002-2008 Professor and Vice Chair, Department of Pathology, The Ohio State University, Columbus, OH  
2001-2008 Professor, Dept. of Computer and Information Science, The Ohio State University, Columbus  
2001-2008 Professor and Chair, Department of Biomedical Informatics, College of Medicine and Public  
Health, The Ohio State University, Columbus, OH
- 1998-2001 Director, Informatics Division, Department of Pathology, The Johns Hopkins Medical Institutions,  
The Johns Hopkins University, Baltimore, MD
- 1998-2001 Professor, Department of Pathology, The Johns Hopkins Medical Institutions, The Johns  
Hopkins University, Baltimore, M
- 1997-2001 Professor, Department of Computer Science and Institute for Advanced Computer Studies,  
University of Maryland, College Park, MD
- 1997-1998 Associate Professor, Department of Pathology, The Johns Hopkins Medical Institutions, Johns  
Hopkins University, Baltimore, MD
- 1992-1997 Associate Professor, Department of Computer Science and Institute for Advanced Computer  
Studies, University of Maryland, College Park, MD
- 1989-1992 Lead Computer Scientist, Institute for Computer Applications in Science and Engineering  
(ICASE), NASA Langley Research Center, Hampton, VA
- 1986-1989 Assistant Professor, Department of Computer Science, Yale University, New Haven, CT  
1985-1986 Staff Scientist, Institute for Computer Applications in Science and Engineering (ICASE), NASA  
Langley Research Center, Hampton, VA

## C. Contributions to Science

1. **Pioneering Whole Slide Image Methods:** I led the team at Johns Hopkins and the University of Maryland College Park in 1997 that was the first to develop the “Virtual Microscope,” and pioneered developments in digital pathology whole slide image navigation, data management, caching strategies and computer aided classification. My team’s initial efforts included development of the first whole slide image viewer followed by several years of effort to develop efficient methods to support whole slide image visualization, query, caching, data management along with methods for efficient systems software support for whole slide data analytic pipelines. This work targeted clusters and HPC systems. We went on to develop methods capable of analysis and visualization of 3-D Pathology images generated from serial sections and management of data obtained through multi-resolution whole slide image capture.
  - a. Ferreira R, Moon B, Humphries J, Sussman A, Saltz J, Miller R, et al. The Virtual Microscope. Proc AMIA Annu Fall Symp. 1997;449–53.

- b. Catalyurek U, Beynon MD, Chang C, Kurc T, Sussman A, Saltz J. The virtual microscope [Internet]. Vol. 7, IEEE Transactions on Information Technology in Biomedicine. 2003. p. 230–48. DOI: 10.1109/titb.2004.823952
- c. Pantanowitz L, Sharma A, Carter AB, Kurc T, Sussman A, Saltz J. Twenty Years of Digital Pathology: An Overview of the Road Travelled, What is on the Horizon, and the Emergence of Vendor-Neutral Archives. *J Pathol Inform.* 2018 Nov 21;9:40.
- d. Beynon M, Chang C, Catalyurek U, Kurc T, Sussman A, Andrade H, et al. Processing large-scale multi-dimensional data in parallel and distributed environments. *Parallel Comput.* 2002 May 1;28(5):827–59.

**2. Visualization, Query, Caching, Data Management, Pipeline Execution:** In the past eight years since I came to Stony Brook, much focus has shifted to management, visualization and quality control of derived data products of such instance semantic segmentation datasets generated by AI algorithms.

- a. Saltz J, Sharma A, Iyer G, Bremer E, Wang F, Jasniewski A, et al. A Containerized Software System for Generation, Management, and Exploration of Features from Whole Slide Tissue Images. *Cancer Res.* 2017 Nov 1;77(21):e79–82.
- b. Foran DJ, Yang L, Chen W, Hu J, Goodell LA, Reiss M, et al. ImageMiner: a software system for comparative analysis of tissue microarrays using content-based image retrieval, high-performance computing, and grid technology. *J Am Med Inform Assoc.* 2011 Jul;18(4):403–15.
- c. Sharma A, Tarbox L, Kurc T, Bona J, Smith K, Kathiravelu P, et al. PRISM: A platform for imaging in Precision Medicine. *JCO Clin Cancer Inform.* 2020 Jun;4(4):491–9.
- d. Mosaliganti K, Pan T, Ridgway R, Sharp R, Cooper L, Gulacy A, et al. An imaging workflow for characterizing phenotypical change in large histological mouse model datasets. *J Biomed Inform.* 2008 Dec;41(6):863–73.

**3. Machine learning, Computer Vision and AI Based Digital Pathology Methods:** My group has developed a variety of machine learning methods to target analysis of whole slide images. Early work in this area included development of machine learning methods for Neuroblastoma and Lymphoma classification. More recently, my group has developed innovative multi-instance learning methods to classify whole slide H&E images using coarse grained case level training data. The 2016 CVPR paper describing this work has been highly influential and has been cited over 340 times (Google Scholar, January 2021). My group has also developed innovative methods that leverage generative adversarial networks to generalize nuclear segmentation across tissue types. This work was described in a 2019 CVPR paper, we then went on to use the method to create a dataset consisting of roughly 5 billion segmented nuclei; this dataset is public and documented in our 2020 Nature Scientific Data publication. We have developed a rich set of pipelines to employ a variety of convolutional network algorithms to compute biologically significant Pathology features from H&E and multiplex IHC images, including spatial maps of tumor infiltrating lymphocytes (TILs). Development of one of our early CNN methods, designed to classify TILs in multiple tumor types, was motivated by our participation in the Pan Cancer Atlas project.

- a. Hou L, Agarwal A, Samaras D, Kurc TM, Gupta RR, Saltz JH. Robust histopathology image analysis: To label or to synthesize? In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019. p. 8533–42.
- b. Hou L, Samaras D, Kurc TM, Gao Y, Davis JE, Saltz JH. Patch-based Convolutional Neural Network for Whole Slide Tissue Image Classification. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit.* 2016 Jun;2016:2424–33.
- c. Sertel O, Kong J, Shimada H, Catalyurek UV, Saltz JH, Gurcan MN. Computer-aided Prognosis of Neuroblastoma on Whole-slide Images: Classification of Stromal Development. *Pattern Recognit.* 2009 Jun;42(6):1093–103.
- d. Fassler DJ, Abousamra S, Gupta R, Chen C, Zhao M, Paredes D, et al. Deep learning-based image analysis methods for brightfield-acquired multiplex immunohistochemistry images. *Diagn Pathol.* 2020 Jul 28;15(1):100.

**4. Runtime Compilation:** In the late 1980's through the middle 1990's, I developed and refined the Inspector Executor paradigm. Irregular array accesses arise in many scientific applications including

sparse matrix solvers, unstructured mesh partial differential equation (PDE) solvers and particle methods. Traditional compilation techniques required that indices to data arrays be symbolically analyzable at compile time. A common characteristic of irregular applications is the use of indirect indexing to represent relationships among array elements. This means that data arrays are indexed through values of other arrays, called indirection arrays. Inability to characterize array access patterns symbolically can prevent compilers from generating efficient code for irregular applications. The inspector/executor strategy involves using compilers to generate code to examine and analyze data references during program execution.

- a. Saltz JH, Mirchandaney R, Crowley K. Run-time parallelization and scheduling of loops [Internet]. Vol. 40, IEEE Transactions on Computers. 1991. p. 603–12. DOI: 10.1109/12.88484
- b. Ponnusamy R, Saltz J, Choudhary A, Hwang Y-S, Fox G. Runtime support and compilation methods for user-specified irregular data distributions [Internet]. Vol. 6, IEEE Transactions on Parallel and Distributed Systems. 1995. p. 815–31. DOI: 10.1109/71.406958
- c. Hwang Y-S, Moon B, Sharma SD, Ponnusamy R, Das R, Saltz JH. Runtime and language support for compiling adaptive irregular programs on distributed-memory machines [Internet]. Vol. 25, Software: Practice and Experience. 1995. p. 597–621. DOI: 10.1002/spe.4380250603
- d. Agrawal G, Saltz J. Interprocedural Compilation of Irregular Applications for Distributed Memory Machines. In: Supercomputing '95: Proceedings of the 1995 ACM/IEEE Conference on Supercomputing. 1995. p. 48–48.

- 5. Data Science Middleware:** In the mid-1990's through roughly 2010, I developed a variety of methods to support what is now called edge computing. These methods include Active Disks, one of my several papers in this area has over 500 citations as of January 2021 and continues to be cited. Another research project along these lines was DataCutter which consisted of light weight portable processes designed to form distributed pipelines with processes that could be instantiated or moved close to data or to computational resources.

The Active Data Repository project was a spatial database project from the mid 1990's through early 2000's that in many respects presaged Hadoop and Spark; this consisted of a system optimized to support computations that made use of generalized reduction operations. This effort incorporated spatial indexing and optimized cluster/de-clustering methods to optimize I/O in clusters and parallel machines. After the development of Hadoop and Spark, in the 2013-2017 timeframe, my team went on to develop analogous software systems that employed these tools. The VLDB article on this topic from 2013 has been cited over 600 times (as of January 2021) and the Hilbert space filling curve article in IEEE TKDE from 2001 has been cited over 850 times (as of January 2021).

- a. Baig F, Vo H, Kurc T, Saltz J, Wang F. SparkGIS: Resource Aware Efficient In-Memory Spatial Query Processing. Proc ACM SIGSPATIAL Int Conf Adv Inf [Internet]. 2017 Nov;2017. PMID: 30035278, PMCID: PMC6054321
- b. Aji A, Wang F, Vo H, Lee R, Liu Q, Zhang X, et al. Hadoop-GIS: A High Performance Spatial Data Warehousing System over MapReduce. Proceedings VLDB Endowment [Internet]. 2013 Aug;6(11). PMID: 24187650, PMCID: PMC3814183
- c. Moon B, Jagadish HV, Faloutsos C, Saltz JH. Analysis of the clustering properties of the Hilbert space-filling curve. IEEE Trans Knowl Data Eng. 2001 Jan;13(1):124–41.
- d. Uysal M, Acharya A, Saltz J. Evaluation of active disks for decision support databases. In: Proceedings Sixth International Symposium on High-Performance Computer Architecture HPCA-6 (Cat NoPR00550). 2000. p. 337–48.

#### List of References:

<http://www.ncbi.nlm.nih.gov/sites/myncbi/joel.saltz.1/bibliography/40137974/public/?sort=date&direction=descending>