Curriculum Vitae
SHAMIM ARA MOLLAH
e-mail: samollah1@gmail.com
phone: (917)370-9854

EDUCATION

2013 – 2019 **University of California, San Diego,** CA
Ph.D. Bioinformatics and Systems Biology
Emphasis in onco-proteogenomics. Thesis: Integrative approaches to
targeting chromatin remodeling in breast cancer therapy.
Thesis advisor: Dr. Shankar Subramaniam

2001- 2003 **Columbia University**, New York, NY
M.A. Biomedical Informatics
Emphasis in Natural Language Processing. Thesis: Automatic learning of the
morphology of medical language.
Thesis Advisor: Dr. Stephen B Johnson

1994-1999 **Indiana University, South Bend**, IN
B.S. Computer Science
Emphasis in Artificial Intelligence.
Research advisor: Dr. Morteza Shafii-Mousavi

1994-1999 **Indiana University, South Bend**, IN
B.A. Applied Mathematics
Emphasis in mathematical modeling. Research project: Mathematical model of
the hunting strategy of a velociraptor using reinforcement learning.
Research advisor: Dr. Morteza Shafii-Mousavi

WORK EXPERIENCE

2019 – present **Washington University School of Medicine**, St. Louis, MO
<u>Assistant Professor of Genetics</u>

2013 – present **The Rockefeller University**, New York, NY
<u>Bioinformatics scientist consultant</u>

2008 – 2013 **The Rockefeller University**, New York, NY
<u>Bioinformatics scientist</u>

Manage Bioinformatics Data Analysis Core, Center of Clinical and Translational
Science (CCTS)

Grant writing:
- T1&T2 grant (Grant #: 5UL1 RR024143/ 5KL2RR024) awarded in 2009 and
  its renewal in 2011.
- Helped design informatics research proposals for R01 grant, 2011 NIH
  CTSA grant renewal (Grant #: UL1RR024143).

Predictive and diagnostic modeling:
- Bleeding Phenotype data: Developed statistical/machine learning models for diagnostic prediction and discovering correlations between clinical phenotypes and Von Willebrand Disease patient subtypes. (**Dr. Barry Coller, Laboratory of Blood and Vascular System**)

Molecular profiling:
- Performed preprocessing and advanced analysis of Agilent, Illumina and Affymetrix gene expression and microarray data.
- Analysis of metabolic data: Analyzed the entire transcriptome of PXR mouse liver to study metabolic adjustments to enhance longevity. (**Dr Jan Breslow, Laboratory of Biochemical Genetics and Metabolism**)
- Analysis of immunological data: Analyzed gene expression of migratory dendritic cells of various tissue subsets in mouse population. (**Dr. Ralph Steinman/ Dr. Niroshana Anandasabapathy, Laboratory of Cellular Physiology and Immunology**)

Knowledge Representation/Natural Language Processing:
- Developed an ontology based comprehensive Bleeding History Phenotyping System (BHPS) to collect, curate, disseminate, facilitate standardization and analyze bleeding history data. (**Dr. Barry Coller, Laboratory of Blood and Vascular System**)
- Human Subject Database (HSDB): Developed an ontology based architecture for sharing human studies data from multiple institutions for large-scale data mining, synthesis, analysis, and reuse. (Collaboration with researchers at UCSF, Stanford, U of Washington, Johns Hopkins and Duke)

2007-2008   **Sullivan & Cromwell LLP**, New York, NY
Technical Manager, Data Analytics, Electronic Discovery & Compliances,
- Oversaw and supervised activities in the data analytic laboratory.
- Developed tools for legal text mining and image manipulation problems using natural language processing and machine learning techniques.
- Developed analytical tools in Merger and Acquisition practice areas of corporate litigation.
- Generated analytical queries to refine extracted data sets, using precision and recall tools and techniques.
- Recommended data laboratory and production equipment, software, policy & procedures.
- Provided consultation to legal teams and wall street clients on Electronic Discovery issues and concerns.

2005-2007   **Albert Einstein College of Medicine**, New York, NY
Biomedical Informatics Developer/Scientific Programmer
- Conducted exploratory and applied research in the area of computational linguistics with specific focus on algorithms applicable to medical language parsing and interpretation, information extraction and text mining. (**Dr. Christopher Cimino, Computer Based Education**)

1999-2004   **Albert Einstein College of Medicine**, New York, NY
Web Developer

- Rapid prototyping and deployment of web services.
- Maintained / supported database engines.
- Designed various IT workshops for faculty.

1998-1999   **Michael Frelas Consulting LLC**, South Bend, IN
<u>Junior Computer Systems Consultant/ Programmer</u> (Paid internship)
- Performed database conversion, mapping data elements between two systems, running automated database conversion tool, and checking converted data for accuracy.
- Developed various graphical user interfaces.

SCIENTIFIC LEADERSHIP

2017 – present   Scientific reviewer for IEEE/ACM Transactions on Computational Biology and Bioinformatics.

2016 – present   Research contributor for NIH sponsored LINCS project

2009 – present   Scientific reviewer for American Medical Informatics Association (AMIA)

2011   Session chair for "Discovering and Tracking Genomic Variants", American Medical Informatics Association (AMIA) Summit in Translational Bioinformatics

2008 – 2013   NIH Clinical and Translational Science Award (CTSA) Informatics Key Function Committee (IKFC) member

---

PUBLICATIONS

S. A. Mollah and S. Subramaniam, "Histone Signatures Predict Therapeutic Efficacy in Breast Cancer," in IEEE Open Journal of Engineering in Medicine and Biology, vol. 1, pp. 74-82, 2020. (**doi.org/10.1109/OJEMB.2020.2967105**)

Mollah SA, Subramaniam S. Global chromatin profiling fingerprints reveal therapeutic efficacy in breast cancer. 2019, CELL-REPORTS. **doi.org/10.2139/ssrn.3413902**.

Tsui B, **Mollah S**, Skola D, Dow M, Hsu C, Carter H; "Creating a scalable deep learning based named entity recognition model for biomedical textual data by repurposing biosample specimen free-text annotation", (submitted/in review GigaScience, available on BioRxiv). **doi.org/10.1101/414136.**

Mollah S, Dobrin J, Feder R, Tse S, Matos I, Cheong C, Steinman R, Anandasabapathy N; "Flt3L dependence helps define an uncharacterized subset of murine cutaneous dendritic cells", 2014, Journal of Investigative Dermatology, 134(5):1265-75. **doi.org/10.1038/jid.2013.515**.

Anandasabapathy N, Feder R, **Mollah S**, Tse S, Longhi M, Mehandru S, Matos I, Cheong C, Ruane D, Brane L, Teixeira A, Dobrin J, Mizenina O, Park C, Meredith M, Clausen B, Nussenzweig M, Steinman R; "Classical Flt3L-dependent dendritic cells control immunity to protein vaccine", 2014, Journal of Experimental Medicine 25,211(9):1875-91. **doi.org/10.1084/jem.20131397**.

Elbatarny M, **Mollah S**, Grabell J, Bae S, Deforest M, Tuttle A, Hopman W, Clark DS, Mauer AC, Bowman M, Riddel J, Christopherson PA, Montgomery RR, Zimmerman Program Investigators, Rand ML, Coller B, James PD, "Normal Range of Bleeding Scores for the ISTH-BAT: Adult and Pediatric Data from The Merging Project", 2014, Haemophilia. 20 (6), 831-835. **doi.org/10.1111/hae.12503.**

Mollah, S, James, P, Coller, B, et al, "A machine learning approach to merging and analyzing data from the condensed MCMDM1 VWD bleeding questionnaire", 2013 International Society of Thrombosis and Hemostasis proceeding.

Mollah, S, PB James, Grabell J, Barbour EM, Coller B, "Diagnostic Prediction of Von Willebrand Disease using multiple bleeding phenomics Datasets", Join Summit on Translational Bioinformatics and Clinical Research Informatics conference, March 2013. **PMID:24303262**.

Sim, I, Carini, S, **Mollah SA** et al, Ontology-based federated data access to human studies information" American Medical Informatics Association proceeding, November 2012. *Distinguished Paper Award.* **PMID:23304360**.

Mollah, SA, Mauer, AC, et al, "A practical approach to automated OWL anotation", Translational Bioinformatics proceeding, March 2011.

Mollah, SA, Sim, I, Chakrabarty, S et al, "Development of common administrative data elements to facilitate data sharing across the human studies", 56th American Medical Informatics Association proceeding, November 2010.

Sim, I, Carini, S, **Mollah SA** et al, The human studies database project: Federating human studies design data using the ontology of clinical research, AMIA Clinical Research Informatics summit 2010. *Distinguished Paper Award.* **PMID:21347149**.

Mauer, A C, Barbour, E, **Mollah, SA** et al, Initial deployment of a comprehensive, Ootology-backed, web-based bleeding history phenotyping instrument in normal individuals. J Thromb Haemost 2009,7:14. **insights.ovid.com/jthrh/200907002/00149457-200907002-00033**.

Carini, S, Pollock, B H, **Mollah, SA** et al, "Development and evaluation of a study design typology for human subjects Research ", 55th American Medical Informatics Association proceeding, November 2009. *Distinguished Paper Award.* **PMID:20351827**.

Mollah, SA, Barbour, E, Sim, I et al, "Developing an interventions ontology for federated sharing of human studies Data ", 55th American Medical Informatics Association proceeding, San Francisco, November 2009.

Mollah, SA, Cimino, C, "Indexing key concepts from biomedical text descriptions for image retrieval in medical education", 55th American Medical Informatics Association proceeding, San Francisco, November 2009.

Mauer, A C, Barbour, E, **Mollah, SA** et al, Creating an ontology-based human phenotyping system: The Rockefeller University bleeding history experience, Journal of The Society for Clinical and Translational Science, Vol2, Issue 5 2009,382:85. **PMID:20443924**.

Mollah, SA, Cimino, C, "Semi-automatic indexing of postscript files using medical

text indexer in medical education", 53rd American Medical Informatics Association proceeding, Chicago, November 2007. **PMID:18694151**.

Mollah, SA, Johnson, SB, "Automatic learning of the morphology of medical language using information compression", 49th American Medical Informatics Association proceeding, November 2003. **PMID:14728443**.

Das, A, **Mollah, SA**, Shafii-Mousavi, M, "Mathematical Model of the Hunting Strategy of a Velociraptor," Annual Indiana University Undergraduate Research Conference, November, 1998.

ORAL AND POSTER PRESENTATIONS

Mollah SA, "Determining therapeutic efficacy in breast cancer using network-based global chromatin profile fingerprinting", City of Hope cancer center, Pasadena, October 2018. (Invited talk).

Mollah SA, "Determining therapeutic efficacy in breast cancer using network-based global chromatin profile fingerprinting", Washington University, St.Louis, October 2018. (Invited talk).

Mollah SA, "Determining therapeutic efficacy in breast cancer using network-based global chromatin profile fingerprinting", Weil Cornell University, New York, March 2019. (Invited talk).

Mollah SA, Subramaniam S; "iPhDNet: Network-Based Global Chromatin Profile Fingerprinting Reveals Therapeutic Efficacy in Breast Cancer", Intelligent Systems in Molecular Biology (ISMB) conference, Chicago, June 2018. (Oral and poster presentations)

Mollah, **SA**, Grabell J, Rand ML, Coller, B, James, P, et al, "The merging project: A machine learning approach to merge and analyze data from four different bleeding questionnaires", International Society of Thrombosis and Hemostasis conference, Amsterdam, Netherland, July 2013. (Oral presentation).

Mollah, **SA**, PB James, Grabell J, Barbour EM, Coller B, "Diagnostic Prediction of Von Willebrand Disease using multiple bleeding phenomics Datasets", Join Summit on Translational Bioinformatics and Clinical Research Informatics conference, San Francisco, March 2013. (Oral presentation).

Mollah, SA, "Capturing Descriptions of Human Studies for Federated Data Sharing: Frontline Experiences from Three Institutions", Clinical Research Informatics summit, San Francisco, March 2012. (Panel speaker).

Mollah, SA, Mauer, AC, et al, "A practical approach to automated OWL annotation", Join Summit on Translational Bioinformatics and Clinical Research Informatics conference, March 2011. (Poster).

Mollah, SA, Sim, I, Chakrabarty, S et al, "Development of common administrative data elements to facilitate data sharing across the human studies", 56th American Medical Informatics Association (AMIA) conference, November 2010. (Poster).

Mollah, SA, Barbour, E, Sim, I et al, "Developing an interventions ontology for federated sharing of human studies Data ", 55th American Medical Informatics Association (AMIA), San Francisco, November 2009. (Poster).

Mollah, SA, Cimino, C, "Indexing key concepts from biomedical text descriptions for image retrieval in medical education", 55th American Medical Informatics Association (AMIA), San Francisco, November 2009. (Poster).

Mollah, SA, Cimino, C, "Semi-automatic indexing of postscript files using medical text indexer in medical education", 53rd American Medical Informatics Association (AMIA), Chicago, November 2007. (Poster).

Mollah, SA, Johnson, SB, "Automatic learning of the morphology of medical language using information compression", 49th American Medical Informatics Association conference, November 2003. (Poster).

Mollah, SA, Das, A, Shafii-Mousavi, M, "Mathematical Model of the Hunting Strategy of a Velociraptor," Annual Indiana University Undergraduate Research Conference, Indianapolis, November, 1998. (Oral presentation).

RESEARCH EXPERIENCE

2013 – 2019   **University of California, San Diego**, CA
Graduate student researcher
- Network-based global chromatin profile fingerprinting in cancer using NIH sponsored LINCS, TCGA proteomics and genomics data.
- Ontology based stratification of cancer subtypes using somatic mutation profile.
- Mapping of DNase1 footprints within the regulatory regions in breast cancer.
- Characterization of miRNA profile and its regulatory pathway in glaucoma and embryonic stem cell using high-throughput RNAseq data.

2012 – 2013   **The Rockefeller University**, New York, NY
Analysis of immunological data
Generated gene expression profile for steady state migratory dendritic cells of various tissue subsets from mouse population using this gene expression profile we analyzed their developmental and functional pathways. (Programming language /Tools: Perl, R, Agilent's GeneSpring, Ingenuity Pathway Analysis)

2012 – 2013   **The Rockefeller University**, New York
Analysis of microarray data for Metabolism and Aging
Analyzed the entire transcriptome of PXR mouse liver to study metabolic adjustments to enhance longevity by generating gene expression profile and study pathway analysis. (Programming language /Tools: Perl, R, Agilent's GeneSpring,)

2011 – 2013   **The Rockefeller University** New York, NY
Predictive and diagnostic modeling
Developed a Bayesian model for predicting Von Willebrand Disease (VWD) patient subtypes. (*Programming language /Tools: Java, Weka, R*)

2009 – 2013   **The Rockefeller University** New York, NY
Human Studies Database project
Developed an ontology based federated, computable database of the design and results of human studies from multiple institutions clinical trial management systems

to accelerate the discovery of findings from human studies, study designs and results for large-scale data mining, synthesis, analysis, and reuse (Programing language/Tool: Perl, OWL, XML, XSLT, SWRL, DQRL, MetaMap)

2008 – 2010   **The Rockefeller University** New York, NY
<u>Data Warehouse and visualization</u>
Developed a web based comprehensive Bleeding Phenotype Repository System (BPRS) to store, retrieve and visualize bleeding phenotype data from multiple institution using ontology as a framework. (*Programming language/Tool: OWL, Python, Perl, PHP, AJAX, MySQL*)

2007   **Albert Einstein College of Medicine**, *New York, NY*
<u>Protein Structure Prediction</u>
Investigated a novel method for secondary protein structure prediction using a feed-forward artificial neural network (*Programming language: C*)

2007   **Albert Einstein College of Medicine**, New York, NY
<u>Image Retrieval</u>
Developed an image retrieval system by indexing key concepts from the text descriptions of biomedical images using a natural language processing (NLP) technique. Demonstrated the potential for using NLP to automate indexing of image descriptions within a medical education curriculum (Programming language: Perl, PHP)

2005   **Albert Einstein College of Medicine**, New York, NY
<u>Key Word Generation</u>
Developed a key word searching mechanism by automating the generation of Medical Subject Heading (MeSH) terms from medline abstracts using an NLP technique (Programming language: Perl)

2003   **Columbia University**, New York, NY
<u>Language Modeling (Master's thesis)</u>
Developed an unsupervised segmentation algorithm to learn the underlying structure of medical language using a statistical modeling techniques and information theories (Programming language: Perl).

1997 – 1998   **Indiana University, South Bend**, IN
<u>Mathematical Modeling</u>
Devised a mathematical model and a computer simulation for "Hunting strategy of a velociraptor" A dynamic programming technique and a reinforcement learning technique were used (*Programming language: C++*).

TEACHING EXPERIENCE

<u>Teaching Assistant</u>
University of California, San Diego, CA
(Spring 2015) Bioinformtics applications to human diseases (Dr. Sergei Pond, graduate level course)
(Spring 2016) Algorithms in computational biology (Dr. Vineet Bafna, graduate level course)

<u>Instructor</u>
The Rockefeller University, New York, NY
(Summer 2011) Lectured on "Natural Language Processing in Medicine" to the clinical scholars
Albert Einstein College of Medicine, New York, NY
(Fall 2006) Taught an introductory course in biomedical informatics to medical students and IT personnel

<u>Research Mentor</u>
University of California, San Diego, CA
(Sept 2018 - present) Bioengineering graduate student, Ashay Patel
(Feb 2018 - present) High school student, currently Computer Science undergraduate student, Melba Nuzen, California Institute of Technology
(Jan 2018- present) Bioengineering undergraduate student, Mitchell Kong
(Jan 2016 – 2017) Bioengineering and Neuroscience undergraduate student, Marissa Eliot
(Mar 2016 – 2017) Bioengineering undergraduate student, Shoichiro, Ma
(Jan 2016 – Sept 2016) Bioengineering undergraduate student, Heye, Zhou

The Rockefeller University, New York, NY
Mobile Bleeding Phenotype Recording Instrument Project through SMART program:
(Summer 2010) Computer science undergraduate student at Boston University, David Lam

Albert Einstein College of Medicine, New York, NY
Protein structure prediction Project through SMART program:
(2005 - 2007) Biomedical Engineering undergraduate student at Columbia University, Shun Yu, investigated a novel method for secondary protein structure prediction based using a deep learning feed-forward artificial neural network (Programming language: C).
Intel Competition Project:
(2004) High school student, Shun Yu, optimized National Center for Biotechnology Information (NCBI) Basic Local Alignment and Search Tool (BLAST) using genetic algorithm technique (Programming language: C)

<u>Tutor, Department of Mathematics and Computer Science</u>
Indiana University, South Bend, IN
(1995-1998) Assisted regular, special aid, under-represented minority undergraduate students with math and computer programming course work.

AWARDS

- 2013-2014 National Library of Medicine (NLM) predoctoral fellowship
- 2012 Distinguished Paper Award at American Medical Informatics Association Informatics Summit
- 2010 Distinguished Paper Award at American Medical Informatics Association, Clinical Research Informatics Summit
- 2009 Distinguished Paper Award at American Medical Informatics Association Informatics Summit
- 1995 - 1998 Indiana University International Student Scholarship
- 1998 Indiana University SMART (Student/Mentor Academic Research Team) Summer Fellowship
- 1998 WHO'S WHO among students in American Colleges and Universities in recognition of outstanding merit and accomplishment as a student at Indiana University

---

Curriculum Vitae
SHAMIM ARA MOLLAH

AFFILIATIONS

- 2016-present   Member of Association for Woman in Science (AWIS), San Diego.
- 2010-present   International Society of Computational Biology (ISCB).
- 2002-present   American Medical Informatics Association (AMIA)
- 2002-2013   Association of Computational Linguistics (ACL)
- 1995-1998   Association for Computing Machinery (ACM)

COMPUTER SKILLS

- Programming /Scripting Languages: R, Python, C/C++, Perl, Matlab, Java, Visual Basic, ASP, Javascript, PHP, OWL, XML,Web 20, AJAX
- Tools: Weka, R studio, Agilent GeneSpring, Ingenuity Pathway Analysis, Lexus Nexus Law
- Database Management: SQL, MySQL, PostgreSQL
- Web Technology: Apache, Microsoft IIS
- Computer Platforms: UNIX (HP), Linux (Red Hat, Gentoo), MS Windows, Mac OS X