

1. Biographical Information

Joseph Daniel Romano, PhD

Born October 25, 1990 in Raleigh, North Carolina

Married to Sarah Prehn Romano, DDS (m. August 10, 2019)

2. Professional Information

CURRENT POSITION: Postdoctoral Researcher

INSTITUTION: University of Pennsylvania

EMAIL ADDRESS: joseph.romano[at]pennmedicine.upenn.edu

INSTITUTIONAL AFFILIATIONS:

Department of Biostatistics, Epidemiology, & Informatics;

Institute for Biomedical Informatics;

Center of Excellence in Environmental Toxicology.

ADDRESS:

D206 Richards Building

3700 Hamilton Walk

University of Pennsylvania

Philadelphia, PA 19104-6116

3. Academic History

University of Vermont (January 2010–May 2014):

B.S. in Molecular Genetics , Honors Scholar, May 2015.

Research Advisor: Neil Sarkar, PhD, MLIS, FACMI

Columbia University (September 2014–May 2019):

M.A. in Biomedical Informatics, December 2016

MPhil in Biomedical Informatics, May 2018

PhD (Biomedical Informatics), May 2019

Dissertation: “Computational Toxinology”

Doctoral Advisor: Nicholas Tatonetti, PhD, FACMI

University of Pennsylvania (June 2019–):

Postdoctoral Research Fellow

Research Mentors: Jason H. Moore, PhD, FACMI and Trevor M. Penning, PhD

4. Employment Record

Graduate Research Assistant, Columbia University, New York, New York, 2014–2019.

Postdoctoral Researcher, University of Pennsylvania, Philadelphia, Pennsylvania, 2019–.

5. Funding Awarded

Discovering clinical endpoints of toxicity via graph machine learning and semantic data analysis

K99-LM013646 (P.I.—Romano); National Library of Medicine (NLM)

Project start/end dates: 12/2021 – 11/2026

Total costs: \$934,540

6. Publications

6.1. Papers (peer-reviewed)

- P1.** Romano JD, Tharp WG, & Sarkar, IN. (2014). Exploring Complex Disease Gene Relationships Using Simultaneous Analysis. *UVM Honors College Theses*, 35(2014).
- P2.** Romano JD, Tharp WG, & Sarkar, IN. (2014). Adapting Simultaneous Analysis Phylogenomic Techniques to Study Complex Disease Gene Relationships. *Journal of Biomedical Informatics*, 54, 10–38.
- P3.** Romano JD, & Tatonetti, NP. (2015). VenomKB, a new knowledge base for facilitating the validation of putative venom therapies. *Scientific Data*, 2(1), 1-9.
- P4.** Boland MR, Jacunski A, Lorberbaum T, **Romano JD**, Moskovitch R, & Tatonetti NP. (2016). Systems biology approaches for identifying adverse drug reactions and elucidating their underlying biological mechanisms. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 8(2), 104–122.
- P5.** Romano JD, & Tatonetti NP. (2016) Using a Novel Ontology to Inform the Discovery of Therapeutic Peptides from Animal Venoms. *AMIA Summits on Translational Science Proceedings, 2016*, 209–218. Awarded second place, TBI Student Paper Competition.
- P6.** Romano JD, Bernauer M, McGrath S, Nagar SD, & Freimuth R. (2019). A Decade of Translational Bioinformatics: A Retrospective Analysis of “Year-in-Review” Presentations. *AMIA Informatics Summit Proceedings, 2019*, 335.
- P7.** Romano JD, & Tatonetti NP. (2019). Informatics and computational methods in natural product drug discovery: A review and perspectives. *Frontiers in Genetics*, 10, 368.
- P8.** Manduchi E, Fu W, **Romano JD**, Ruberto S, & Moore JH. (2020) Embedding covariate adjustments in tree-based automated machine learning for biomedical big data analyses. *BMC Bioinformatics*, 21(1), 1-13.
- P9.** Romano JD & Moore JH. (2020) Ten Simple Rules for Writing a Paper About Scientific Software. *PLOS Computational Biology*, 16(11): e1008390.
- P10.** Romano JD, Le TT, Fu W, & Moore JH. (2021) TPOT-NN: Augmenting tree-based automated machine learning with neural network estimators. *Genetic Programming and Evolvable Machines*, 22(2), 207–227.
- P11.** Romano JD, Hao Y, & Moore JH. (2021) Improving QSAR Modeling for Predictive Toxicology using Publicly Aggregated Semantic Graph Data and Graph Neural Networks. *Pacific Symposium on Biocomputing* 27, 187–198.
- P12.** Manduchi E, **Romano JD**, & Moore JH. (2021) The promise of automated machine learning for the genetic analysis of complex traits. *Human Genetics*, (2021): 1-16.
- P13.** Romano JD, Le TT, La Cava W, Gregg JT, Goldberg DJ, Chakraborty P, Ray NL, Himmelstein D, Fu W, & Moore JH. (2021) PMLB v1.0: An open source dataset collection for benchmarking machine learning methods. *Bioinformatics*, btab727.

6.2. Publications in progress (In preparation, under review, etc.)

- Q1.** Romano JD, Hao Y, Moore JH, & Penning T. ComptoxAI—A data resource for AI research in computational toxicology. (*in preparation*).
- Q2.** Romano JD, Napoletano T, Cook D, Li H, Realubit R, Douglass E, Califano A, Karan C, Ryan T, Holford M, Tatonetti NP. Discovering venom-derived drug candidates using differential gene expression. (*in preparation*).

6.3. Other Publications (preprints and other non-refereed contributions)

- R1.** Romano JD, Nwankwo V, & Tatonetti NP. VenomKB v2.0: A knowledge repository for computational toxinology. *bioRxiv* (2018), 295204.

7. Peer-Reviewed Conference Presentations

- C1.** Observational Health Data Sciences and Informatics (OHDSI). *Innovation and Application @ Columbia* (2015), New York, NY.
- C2.** Data-Driven Clinical Research Generalizability Assessment and Improvement. *8th Annual Mid-Atlantic Healthcare Informatics Symposium* (2015), Philadelphia, PA.
- C3.** Venom Peptides as Therapeutic Agents: Can we use Phylogenetics to Inform Drug Discovery? *13th Annual Rocky Mountain Bioinformatics Conference* (2016), Aspen, CO.
- C4.** Building a Centralized Resource for Computational Venom Research. *National Library of Medicine Informatics Training Conference* (2016), Columbus, OH.
- C5.** High-sensitivity Stroke Phenotyping using Recurrent Neural Networks. *AMIA Joint Summits on Translational Science* (2017), San Francisco, CA.
- C6.** Deep recurrent neural networks identify transgender patients. *AMIA Annual Symposium* (2017), Washington, DC.
- C7.** Computational Resources for Personalized Genomics: High Performance Clusters and Bioinformatics Resources for Analysis and Functional Interpretation of Next-Generation Sequencing Data. *AMIA Informatics Summit* (2018), San Francisco, CA. Tutorial session available to conference attendees.
- C8.** **VenomSeq**—A platform for drug discovery from animal venoms using differential gene expression. *AMIA Informatics Summit* (2018), San Francisco, CA.
- C9.** Engineering Innovative Informatics Solutions to Support Venom-Based Drug Discovery. *Gordon Research Conference on Venom Evolution, Function, and Biomedical Applications* (2018), West Dover, VT.
- C10.** Analyzing Whole Exome Sequencing: Unstructured Data to Variant Interpretation. *AMIA Informatics Summit* (2019), San Francisco, CA.
- C11.** ComptoxAI: A toolkit for AI research in computational toxicology. *Society of Toxicology Annual Meeting & ToxExpo* (2020), Anaheim, CA (virtual conference).
- C12.** Information retrieval and network analysis in ComptoxAI—a data infrastructure for AI research in computational toxicology. *Society of Toxicology Annual Meeting & ToxExpo* (2021), Orlando, FL (virtual conference).

8. Invited Talks

- Do's and don'ts for writing impactful papers about scientific software. *National Cancer Institute; Informatics Technology for Cancer Research initiative* (2021), Bethesda, MD (virtual talk).

9. Teaching Experience

9.1. Coursework Taught

NIHON UNIVERSITY:

Fall 2011:

English language teaching assistant.

UNIVERSITY OF VERMONT:

Spring 2012:

MMG 104—Introduction to Recombinant DNA Technology; Teaching assistant.

COLUMBIA UNIVERSITY:

Fall 2016:

BINF G4000—Acculturation to Programming and Statistics; Teaching assistant.

BINF G4006—Translational Bioinformatics; Guest lecturer—2 lectures.

Spring 2017:

BINF G4003—Symbolic Methods in Biomedical Informatics; Teaching assistant.

Spring 2018:

BINF G4003—Symbolic Methods in Biomedical Informatics; Guest lecturer.

YALE DIVINITY SCHOOL:

Summer 2018:

World Scholars SAT+ Summer Program; Guest lecturer.

UNIVERSITY OF PENNSYLVANIA:

Fall 2019:

BMIN 525—Introduction to Python Programming; Teaching assistant / guest lecturer.

Spring 2020:

BMIN 521—Foundations of AI in Health; Guest lecturer.

Fall 2020:

BMIN 525—Introduction to Python Programming; Teaching assistant / guest lecturer.

Spring 2021:

BMIN 521—Foundations of AI in Health; Guest lecturer.

Fall 2021:

BMIN 525—Introduction to Python Programming; Teaching assistant / guest lecturer.

BMIN 503/EPID 600—Data Science for Biomedical Informatics; Guest lecturer—2 lectures.

9.2. Formally Mentored Trainees

Marine Saint-Mézard (MS; ENSTA-ParisTech)—Summer research intern, Columbia University; Summer 2018.

Rachit Kumar (MD/PhD; University of Pennsylvania)—MD/PhD student, University of Pennsylvania; Fall 2020–present.

Tom Pan (BS; University of Pennsylvania)—Undergraduate researcher, University of Pennsylvania; Spring 2021–present.

Erica Suh (PhD; University of Pennsylvania)—PhD rotation student, University of Pennsylvania; Spring 2021–present.

10. Professional Society Memberships and Elected Positions

American Association for the Advancement of Science, 2014–.
American Medical Informatics Association, 2012–.
Communications director, Genomics and Translational Bioinformatics Working Group, 2015–.
Chair, Student Working Group, 2018–
American Society of Clinical Pharmacology and Therapeutics, 2017–.
International Society for Computational Biology, 2014–.
International Society on Toxinology, 2016–.
New York Academy of Sciences, 2014–2019.
Society of Toxicology, 2019–.
T_EX Users Group, 2019–.

11. Awards, Fellowships, and Honors

AP Scholar with Honors, College Board, 2009.
Eagle Scout, Boy Scouts of America, 2009.
National Merit Scholar, National Merit Scholarship Corporation, 2009.
Vermont Scholar, University of Vermont, 2010–2014.
Honors Scholar, University of Vermont, 2014.
Graduate Research Fellow in Biomedical Informatics, Columbia University, 2015–2017.
Parent grant: T15-LM007079 (PI: Hripcsak).
Funding institute: National Library of Medicine
Postdoctoral Research Fellow in Environmental Health Sciences, University of Pennsylvania, 2019–.
Parent grant: T32-ES019851 (PI: Penning).
Funding institute: National Institute of Environmental Health Sciences.

12. Review Activities

12.1. Journals

BioData Mining.
Journal of the American Medical Informatics Association (JAMIA).
Journal of Biomedical Informatics (JBI).
Journal of Open Source Software (JOSS).
Nature Communications.
Nature Scientific Reports.
PLOS Computational Biology.
PLOS ONE.
Scientific Data.

12.2. Conferences and Symposia

AMIA Annual Symposium.
AMIA Informatics Summit.
Scientific Program Committee Member, Translational Bioinformatics, 2020.
Scientific Program Committee Member, Translational Bioinformatics, 2019.
Translational Bioinformatics “Year-in-Review” Committee Member, 2016–2018.
IEEE International Conference on Healthcare Informatics.
Pacific Symposium on Biocomputing.
Symposium on Artificial Intelligence for Learning Health Systems (SAIL)
Organizing Committee Member, 2020–2021.
World Congress on Health and Biomedical Informatics (MEDINFO).
Assistant Editor, 2015–2017.