Ruth Johnson

+1(559)802-2442 • ruth_johnson@hms.harvard.edu
scholar.google.com/citations?user=IpaMUH0AAAAJhl=en

I am currently a Postdoctoral Fellow at Harvard Medical School in the Department of Biomedical Informatics. My research develops machine learning methods that integrate large-scale genomic biobanks and electronic health records to build translational tools. I am also interested in multi-modal foundation models and graph-based machine learning algorithms with knowledge graphs.

Education

- **Harvard Medical School** — **Boston, MA**
  Berkowitz Postdoctoral Fellow — *April 2023-present*
  Advisor: Marinka Zitnik
- **University of California, Los Angeles** — **Los Angeles, CA**
  Ph.D. Computer Science — *Sept 2017-March 2023*
  Thesis: "Leveraging genetic and electronic health record data to understand complex traits and rare diseases"
  Advisors: Bogdan Pasaniuc & Sriram Sankararaman
- **University of California, Los Angeles** — **Los Angeles, CA**
  B.S. Computational Mathematics, Minor in Bioinformatics — *Sept 2013-June 2017*

Honors and Fellowships

- **Charles J. Epstein Trainee Award for Excellence in Human Genetics Research semi-finalist**, Top 60 abstract selection out of all pre-doctoral submissions (2022)
- **Electrical Engineering and Computer Science Rising Stars**, Top 150 internationally selected predoctoral/postdoctoral scholars in EECS based on academic excellence (2020)
- **National Science Foundation Research Traineeship (NRT) Modeling and Understanding Human Behavior Fellowship** (2018)
- **Ford Fellowship Predoctoral Competition - Honorable Mention**, National Academies fellowship awarding superior academic achievement (2018)
- **National Science Foundation Graduate Research Fellowships Program (GRFP) - Honorable Mention**, Supports outstanding graduate students in STEM disciplines within the United States (2017)
- **Eugene V. Cota-Robles Fellowship** Four-year award funded by the University of California Office of the President to 85 doctoral students selected across all departments in UCLA Graduate Division (2017)

Representative Research Publications

Google scholar: `https://scholar.google.com/citations?user=IpaMUH0AAAAJ&hl=en`

1. **ClinVec: Unified Embeddings of Clinical Codes Enable Knowledge-Grounded AI in Medicine**
   Ruth Johnson, Uri Gottlieb, Galit Shaham, Lihi Eisen, Jacob Waxman, Stav Devons-Sberro, Curtis R. Ginder, Peter Hong, Raheel Sayeed, Xiaorui Su, Ben Y. Reis, Ran D. Balicer, Noa Dagan, Marinka Zitnik; *in-press npj Digitial Medicine* .
2. **Electronic health record signatures identify undiagnosed patients with Common Variable Immunodeficiency Disease**
   Ruth Johnson, Alexis V Stephens, Sergey Knyazev, Lisa A Kohn, Malika K Freund, Leroy Bondhus, Brian L Hill, Tommer Schwarz, Noah Zaitlen, Valerie Arboleda, Manish J Butte*, Bogdan Pasaniuc*; *Science Translational Medicine* 2024.

3. **Leveraging genomic diversity for discovery in an EHR-linked biobank: the UCLA ATLAS Community Health Initiative.**
Ruth Johnson, Yi Ding, Vidhya Venkateswaran, Arjun Bhattacharya, Alec Chiu, Tommer Schwarz, Malika Freund, Lingyu Zhan, Kathryn S. Burch, Christa Caggiano, Brian Hill, Nadav Rakocz, Brunilda Balliu, Jae Hoon Sul, Noah Zaitlen, Valerie A. Arboleda, Eran Halperin, Sriram Sankararaman, Manish J. Butte, UCLA Precision Health Data Discovery Repository Working Group, UCLA Precision Health ATLAS Working Group, Clara Lajonchere, Daniel H. Geschwind, Bogdan Pasaniuc; *Genome Medicine* 2022.
4. **The UCLA ATLAS Community Health Initiative: promoting precision health research in a diverse biobank**
Ruth Johnson, Yi Ding, Arjun Bhattacharya, Alec Chiu, Clara Lajonchere, Daniel H Geschwind, Bogdan Pasaniuc; *Cell Genomics* 2023.

Peer-reviewed Journal Publications

* - denotes joint authors

1. **Graph AI in Medicine**
Ruth Johnson, Michelle M. Li*, Ayush Noori*, Owen Queen*, and Marinka Zitnik; *Annual Review of Biomedical Data Science* 2024.
2. **Polygenic scores for tobacco use provide insights into systemic health risks in a diverse EHR-linked biobank in Los Angeles**
Vidhya Venkateswaran, Kristin Boulier, Yi Ding, Ruth Johnson, Arjun Bhattacharya, Bogdan Pasaniuc; *Translational Psychiatry* 2024.
3. **Global Biobank Meta-analysis Initiative: powering genetic discovery across human diseases**
Global Biobank Meta-analysis Initiative; *Cell Genomics* 2022.
4. **Exome-wide association study to identify rare variants influencing COVID-19 outcomes: Results from the Host Genetics Initiative**
Guillaume Butler-Laporte et al; *PLOS Genetics* 2022.
5. **Virtual meetings promise to eliminate the geographical and administrative barriers and increase accessibility, diversity, and inclusivity**
Juncheng Wu, Anushka Rajesh, Yu-Ning Huang, Karishma Chhugani, Rajesh Acharya, Kerui Peng, Ruth Johnson, Andrada Fiscutean, Carla Daniela Robles-Espinoza, Francisco M. De La Vega, Riyue Bao, Serghei Mangul; *Nature Biotechnology* 2022.
6. **EH3k27ac-HiChIP in prostate cell lines identifies risk genes for prostate cancer susceptibility**
Claudia Giambartolomei, Ji-Heui Seo, Tommer Schwarz, Malika Kumar Freund, Ruth Johnson, Sandor Spisak, Sylvan C. Baca, Alexander Gusev, Nicholas Mancuso, Bogdan Pasaniuc, Matthew L. Freedman; *American Journal of Human Genetics 2021* .
7. **Mapping the human genetic architecture of COVID-19**
COVID-19 Host Genetics Initiative; *Nature* 2021.
8. **Integrative analyses identify susceptibility genes underlying COVID-19 hospitalization**
Gita Pathak, Kritika Singh, Tyne Miller-Fleming, Frank Wendt, Nava Ehsan, Kangcheng Hou, Ruth Johnson, Zeyun Lu, Shyamalika Gopalan, Loic Yengo, Pejman Mohammadi, Bogdan Pasaniuc, Renato Polimanti, Lea Davis, Nicholas Mancuso; accepted at *Nature Communications* 2021.
9. **Pre-existing conditions in Hispanics/Latinxs that are COVID-19 risk factors**
Timothy S Chang, Yi Ding, Malika K Freund, Ruth Johnson, Tommer Schwarz, Julie M Yabu, Chad Hazlett, Jeffrey N Chiang, David A Wulf, UCLA Precision Health Data Discovery Repository Working Group, Daniel H Geschwind, Manish J Butte, Bogdan Pasaniuc; *iScience* 2021.
10. **Estimation of regional polygenicity from GWAS provides insights into the genetic architecture of complex traits**
Ruth Johnson, Kathryn S. Burch, Kangcheng Hou, Mario Paciuc, Bogdan Pasaniuc, Sriram Sankararaman;

PLOS Computational Biology 2021.

11. **Localizing components of shared transethnic genetic architecture of complex traits from GWAS summary data.**
    Huwenbo Shi*, Kathryn S Burch*, Ruth Johnson, Malika K Freund, Gleb Kichaev, Nicholas Mancuso, Astrid M Manuel, Natalie Dong, Bogdan Pasaniuc; *American Journal of Human Genetics* 2020.
12. **A unifying framework for joint trait analysis under a non-infinitesimal model**
    Ruth Johnson, Huwenbo Shi, Bogdan Pasaniuc*, Sriram Sankararaman*; *Bioinformatics 2019*.
13. **An automated machine learning-based model predicts postoperative mortality using readily-extractable preoperative electronic health record data**
    Brian Hill, Robert Brown, Eilon Gabel, Christine Lee, Maxime Cannesson, Loes Olde Loohuis, Ruth Johnson, Brandon Jew, Uri Maoz, Aman Mahajan, Sriram Sankararaman, Ira Hofer, Eran Halperin; *British Journal of Anaesthesia 2019*.
14. **Probabilistic fine-mapping of transcriptome-wide association studies**
    Nicholas Mancuso, Malika K. Freund, Ruth Johnson , Huwenbo Shi, Gleb Kichaev, Alexander Gusev, and Bogdan Pasaniuc; *Nature Genetics 2019*.
15. **Improved methods for multi-trait fine mapping of pleiotropic risk loci**
    Gleb Kichaev*, Megan Roytman*, Ruth Johnson, Eleazar Eskin, Sara Lindström, Peter Kraft, Bogdan Pasaniuc; *Bioinformatics 2017.*

Peer-reviewed Conference Publications

1. **Multimodal Medical Code Tokenizer**
   Xiaorui Su, Shvat Messica, Yepeng Huang, Ruth Johnson, Lukas Fesser, Shanghua Gao, Faryad Sahneh, Marinka Zitnik ; *ICML 2025.*
2. **A scalable method for estimating the regional polygenicity of complex traits**
   Ruth Johnson, Kathryn S. Burch, Kangcheng Hou, Mario Paciuc, Bogdan Pasaniuc, Sriram Sankararaman; *RECOMB 2020.*
3. **A unifying framework for joint trait analysis under a non-infinitesimal model**
   Ruth Johnson, Huwenbo Shi, Bogdan Pasaniuc*, Sriram Sankararaman*; *ISMB 2018.* (simultaneously published in *Bioinformatics*)

Pre-prints

1. **Multimodal AI predicts clinical outcomes of drug combinations from preclinical data**
   Yepeng Huang, Xiaorui Su, Varun Ullanat, Intae Moon, Ivy Liang, Lindsay Clegg, Damilola Olabode, Ruthie Johnson, Nicholas Ho, Megan Gibbs, Megan Gibbs, Alexander Gusev, Bino John, Marinka Zitnik; *arXiv 2025* (under review at *Nature Communications*).
2. **Implications of self-identified race, ethnicity, and genetic ancestry on genetic association studies in biobanks within health systems**
   Ruth Johnson & Bogdan Pasaniuc; *arXiv 2023*

Internships

- **Microsoft** — **Redmond, WA**
  Data Science PhD intern — *June 2022 – August 2022*
  Project: Explainable AI in large-scale recommendation systems.
- **Illumina** — **Foster City, CA**
  Deep Learning Engineering Intern — *June 2018-Sept 2018*
  Project: Deep-learning image segmentation for basecalling in genomic sequencing.

- **Sandia National Laboratories** — **Albuquerque, NM**
  R&D Engineering Intern — *June 2016 – August 2016*
  Project: Outlier detection for space object hyper-spectral bidirectional reflectance distribution function imaging telescope.

Oral Presentations

1. **ClinVec: Unified Embeddings of Clinical Codes Enable Knowledge-Grounded AI in Medicine**
   Program in Quantitative Genomics Annual Conference, October 2024. Boston, MA.
2. **Leveraging genomic diversity for discovery in an EHR-linked biobank: the UCLA ATLAS Community Health Initiative**
   American Society of Human Genetics Annual Meeting, October 2022. Los Angeles, CA.
3. **Electronic health record signatures identify undiagnosed patients with CVID**
   California Center for Rare Diseases Genomic Rounds, November 2021. Virtual meeting.
4. **A scalable method for estimating the regional polygenicity of complex traits**
   RECOMB, July 2020. Virtual meeting.
5. **Leveraging electronic health record signatures identify undiagnosed patients with Common Variable Immunodeficiency Disease**
   Undiagnosed Diseases Network - Steering Committee Meeting, March 2020. Los Angeles, CA, USA. (cancelled due to COVID-19)
6. **Leveraging electronic health record signatures identify undiagnosed patients with Common Variable Immunodeficiency Disease**
   Institute for Quantitative and Computational Biosciences - Research Seminar, February 2020. Los Angeles, CA, USA.
7. **Electronic health record signatures identify undiagnosed patients with of CVID**
   Medical and Population Genetics seminar - Computational Genomics and Health, November 2019. Los Angeles, CA, USA.
8. **Dissecting the genetic architecture of complex traits through local polygenicity using summary statistics from genome-wide association studies**
   Biology of Genomes 2019, May 2019. Long Island, NY, USA.
9. **Dissecting the genetic architecture of complex traits through local polygenicity using summary statistics from genome-wide association studies**
   Medical and Population Genetics, May 2019. Los Angeles, CA, USA.
10. **A scalable Bayesian model for estimating the genetic architecture of complex traits using summary statistics from GWAS**
    Probabalistic Modeling in Genomics Meeting 2018, November 2018. Long Island, NY, USA.
11. **A unifying framework for joint trait analysis under a non-infinitesimal model**
    ISMB 2018, July 2018. Chicago, IL, USA.
12. **Combining genetic correlation and colocalization into a unifying model**
    Medical and Population Genetics, May 2019. Los Angeles, CA, USA.
13. **CANVIS: Correlation Annotation VISualization**
    RECOMB Genetics Satellite Meeting, July 2017. Los Angeles, CA, USA.

Mentoring

- Jessie Chen. Undergraduate student, Bruins in Genomics Summer Program.
  Project title: "ATLAS-hub: an R Shiny App for PheWAS results on the ATLAS BioBank"
  ∗Research Excellence award and Top Presentation award

- Mario Paciuc. Undergraduate student, Rice University.
  Project title: "Genetic correlation of complex traits under a non-infinitesimal model"
  * Award for Distinction in Research and Creative Works from the Department of Statistics
- Gary Hu. Undergraduate student, Bruins in Genomics Summer Program.
  Project title: "Trans-ethnic genetic overlap in complex traits."
- Hugo Mainguy. Undergraduate student, Bruins in Genomics Summer Program.
  Project title: "Assessing the overlap of complex traits through the shared proportion of causal SNPs and genetic correlation"
- Engineering Undergraduate Research Program - Graduate Student Mentor
  Taught weekly workshops about scientific presentations and guided 10-15 students through creating abstracts, posters, and presentations about their research projects.

Teaching

- **University of California, Los Angeles** - CS146: Introduction to Machine Learning (2020); Teaching assistant
- **University of California, Los Angeles** - Math88S: Mathematics & Movies (2016); Instructor

Service

- **Conference Reviewing**: Machine Learning in Computational Biology (Program Committee), RECOMB (Program Committee), ISMB (Program Committee).
- **Journal Reviewing** Nature Genetics, Nature Communications, Cell Genomics, Genome Medicine, PLOS Genetics, Bioinformatics, Frontiers in Genetics.

Software

- **BEAVR** *Estimating regional polygenicity*

  Software that estimates the proportion of causal variants (*i.e.* polygenicity) within a given region from GWAS summary statistics and in-sample LD.

  `https://github.com/bogdanlab/BEAVR`

- **UNITY** *Quantifying genetic overlap of complex traits*

  Software that uses a fully Bayesian framework to calculate the proportion of shared causal variants between two complex traits through GWAS summary statistics. The method also explicitly models the genetic correlation present between both traits.

  `https://github.com/bogdanlab/UNITY`

- **CANVIS** *Fine-mapping visualization*

  A fine-mapping tool that visually summarizes an integrative fine-mapping experiment. The tool provides visual representation of the local correlation structure (LD), the functional annotations used, as well as association statistics and posterior probabilities for each SNP.

  `https://github.com/bogdanlab/PAINTOR/tree/master/CANVIS`