Ruth Johnson

ruthjohnson at ucla dot edu • website: ruthjohnson95.github.io

Education

- **University of California, Los Angeles** — **Los Angeles, CA**
  PhD Computer Science — *Sept 2017-present (expected Spring 2023)*
  Thesis: Scalable machine learning methods for biomedical data
- **University of California, Los Angeles** — **Los Angeles, CA**
  B.S. Mathematics, Minor in Bioinformatics — *Sept 2013-June 2017*

Academic Honors

- Charles J. Epstein Trainee Award for Excellence in Human Genetics Research semi-finalist (2022)
- Stellar Abstract Award at the Program in Quantitative Genomics annual conference Honorable mention (2020)
- EECS Rising Stars (2020)
- RECOMB Travel Fellowship (2020)
- NRT-Modeling and Understanding Human Behavior Fellowship (2018)
- Ford Fellowship Predoctoral Competition - Honorable Mention (2018)
- NSF Graduate Research Fellowships Program - Honorable Mention (2017)
- Eugene V. Cota-Robles Fellowship, UCLA (2017)

Research Experiences

- **Microsoft** — **Redmond, WA**
  Data Science PhD intern — *June 2022 – current*
  - Created Explainable AI visualization tool for ranking algorithms utilized in Whole Page Optimization production.
  - Utilized SHAP, LIME, and counterfactual methods to create both model-level and example-level explanation summaries.
  - Selected to present project to multiple levels of leadership.
  - Under supervision of Peter Xiong.
- **Machine Learning in Biomedical Lab at UCLA** — **Los Angeles, CA**
  Graduate Student Researcher — *July 2017 – current*
  - Leveraged electronic health records (EHR) to predict common variable immune deficiency and other rare disorders through phenotype risk scores
  - Research focused on combining EHR information with genetic information to study the role of genetic ancestry in the development of disease
  - Developed scalable Bayesian frameworks to model genetic architecture of complex traits
  - Emphasis on probabilistic modeling, Markov Chain Monte Carlo methods, variational inference, and graphical models
  - Under supervision of Prof. Sriram Sankararaman and Prof. Bogdan Pasaniuc
- **Illumina** — **Foster City, CA**
  Deep Learning Engineering Intern — *June 2018-Sept 2018*
  - Developed end-to-end parallelizable training data generation pipeline for deep learning base calling models
  - Performed image segmentation at subpixel level to infer regions for base-calling
  - Utilized Keras, convolutional neural networks (U-Net architecture), Python
  - Under supervision of Amirali Kia

- **Bioinformatics Lab at UCLA** — **Los Angeles, CA**
  Undergraduate Student Researcher — *March 2016 – June 2017*
  - Constructed software pipeline for fine-mapping analyses with the integration of functional annotation data
  - Created Python based data visualization tool that produces publication-ready figures of integrated fine-mapping experiments
  - Under supervision of Prof. Bogdan Pasaniuc

- **Sandia National Laboratory** — **Albuquerque, NM**
  R&D Engineering Intern — *June 2016 – August 2016*
  - Researched satellite anomaly detection methods through supervised classification algorithms such as DBSCAN
  - Designed components of an object-oriented, dynamic web UI for satellite control systems using ReactJS and ExtJS

Journal Publications

* - denotes joint authors

1. **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* in-press.
2. **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* in-press.
3. **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*.
4. **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*
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, erghei Mangul; *bioRxiv* 2021; in press at *Nature Biotechnology*.
6. **Mapping the human genetic architecture of COVID-19**
   COVID-19 Host Genetics Initiative; *Nature* 2021.
7. **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.
8. **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.

9. **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.
10. **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*.
11. **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*.
12. **A unifying framework for joint trait analysis under a non-infinitesimal model**
    Ruth Johnson, Huwenbo Shi, Bogdan Pasaniuc*, Sriram Sankararaman*; *Bioinformatics 2019*.
13. **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*.

Conference Publications

1. **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*.
2. **A unifying framework for joint trait analysis under a non-infinitesimal model**
   Ruth Johnson, Huwenbo Shi, Bogdan Pasaniuc*, Sriram Sankararaman*; *ISMB 2018*.

Preprints

1. **Electronic health record signatures identify undiagnosed patients with Common Variable Immunodeficiency Disease**
   Ruth Dolly 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*; *medRxiv* 2022.
2. **The UCLA ATLAS Community Health Initiative: promoting precision health research in a diverse biobank**
   Ruth D Johnson, Yi Ding, Arjun Bhattacharya, Alec Chiu, Clara Lajonchere, Daniel H Geschwind, Bogdan Pasaniuc; *medRxiv* 2022. (under revision at *Cell Genomics*)
3. **Global Biobank Meta-analysis Initiative: powering genetic discovery across human diseases**
   Global Biobank Meta-analysis Initiative; *medRxiv* 2021. (under revision at *Cell Genomics*)

Oral Presentations

1. **Electronic health record signatures identify undiagnosed patients with CVID**
   California Center for Rare Diseases Genomic Rounds, November 2021. Virtual meeting.
2. **A scalable method for estimating the regional polygenicity of complex traits**
   RECOMB, July 2020. Virtual meeting.
3. **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)

4. **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.
5. **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.
6. **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.
7. **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.
8. **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.
9. **A unifying framework for joint trait analysis under a non-infinitesimal model**
   ISMB 2018, July 2018. Chicago, IL, USA.
10. **Combining genetic correlation and colocalization into a unifying model**
    Medical and Population Genetics, May 2019. Los Angeles, CA, USA.
11. **CANVIS: Correlation Annotation VISualization**
    RECOMB Genetics Satellite Meeting, July 2017. Los Angeles, CA, USA.

Poster Presentations

1. **An analysis of the genetic overlap of 20 complex traits under a non-infinitesimal model**
   Ruth Johnson, Huwenbo Shi, Kathryn Burch, Bogdan Pasaniuc, Sriram Sankararaman; Annual meeting of the American Society of Human Genetics, Oct 2018. San Diego, CA, USA.
2. **Integrative fine-mapping of 34 complex phenotypes**
   Ruth Johnson, Gleb Kichaev, Kathryn Burch, Bogdan Pasaniuc; Annual meeting of the American Society of Human Genetics, Oct 2017. Orlando, FL, USA.
3. **Leveraging functional annotations in fine-mapping of causal variants for complex traits**
   Ruth Johnson, Gleb Kichaev, Kathryn Burch, Bogdan Pasaniuc; UCLA Undergraduate Research Poster Day, May 2017. Los Angeles, CA, USA.
   **Dean's Prize Science Award Honoring Outstanding Undergraduate Researcher*
4. **Visualizing correlated causal variants**
   Ruth Johnson, Gleb Kichaev, Bogdan Pasaniuc; Annual meeting of the American Society of Human Genetics, October 2016. Vancouver, CN.

Mentoring Experience

- 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"
  - co-author on Johnson et al. RECOMB 2020
  - 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.

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`