Alexandr Diaz-Papkovich, PhD

github.com/diazale | alex_diaz-papkovich@brown.edu

Education

PhD, Quantitative Life Sciences (Prof. Simon Gravel)   2017 - 2023
McGill University, Montreal, QC

Thesis: "The entangled biobank: On the topology of high-dimensional human genetic data."

Master of Science, Mathematics - Probability & Statistics (Prof. Shirley Mills)   2012 - 2015
Carleton University, Ottawa, ON

Thesis: "Data Mining the Play-By-Play: Assessing and Applying NHL Performance Metrics Using Statistical Methods"

Bachelor of Mathematics, Honours Statistics, Pure Mathematics minor   2005 - 2010
University of Waterloo, Waterloo, ON

Work experience

Postdoctoral research associate   2023 - present
Data Science Institute at Brown University, Providence, RI
- Developed methods for topological research and visualization of genetic biobank data
- Studying impact of scientific articles on perceptions of genetic ancestry using Wikipedia data

Instructor   2018 - 2023
McGill University, Montreal, QC
- Led seasonal workshops on dimension reduction, statistics, and statistical programming
- Taught graduate course Epidemiology 613: Introduction to statistical software
- Taught modules in graduate course QLSC600: Foundations of Quantitative Life Sciences
- Teaching assistant for population genetics courses

Mathematical statistician   2010 - 2017
Statistics Canada, Ottawa, ON
- Implemented aspects of survey methodology including weighting and calibration, sample allocation, estimation, imputation, bootstrapping, hypothesis testing, and record linkage
- Wrote and edited reports, technical reviews, and publications, and supervised employees

Technical consultant   2008 - 2010
The Co-operators, Guelph, ON
- Created, optimized, and troubleshot databases and related applications

Publications

Alex Diaz-Papkovich, Cyril de Bodt, Michael Bleher, Kerstin Bunte, Corinna Coupette, Sebastian Damrich, Enrique Fita Sanmartin, Fred A Hamprecht, Emőke-Ágnes Horvát, Dhruv Kohli, Smita Krishnaswamy, John A Lee, Boudewijn PF Lelieveldt, Leland McInnes, Ian T Nabney, Maximilian Noichl, Pavlin G Poličar, Bastian Rieck, Guy Wolf, Gal Mishne, Dmitry Kobak. "Low-dimensional embeddings of high-dimensional data." *arXiv* (2025). 2508.15929.

Alex Diaz-Papkovich, Shadi Zabad, Chief Ben-Eghan, Luke Anderson-Trocmé, Georgette Femerling, Vikram Nathan, Jenisha Patel, Simon Gravel. "Topological stratification of continuous genetic variation in large biobanks." *bioRxiv* (2023): 2023-07*.*

Alex Diaz-Papkovich, Luke Anderson-Trocmé, and Simon Gravel. "A review of UMAP in population genetics." *Journal of Human Genetics* 66, no. 1 (2021): 85-91.

Alex Diaz-Papkovich, Luke Anderson-Trocmé, Chief Ben-Eghan, Simon Gravel. "UMAP reveals cryptic population structure and phenotype heterogeneity in large genomic cohorts". (2019) PLoS Genetics 15(11): e1008432.

Alejandro Mejia Garcia, **Alex Diaz-Papkovich**, Guillaume Sillon, Daniela D'Agostino, Anne-Laure Chong, George Chong, Ken Sin Lo, Laurence Baret, Nancy Hamel, Vincent Chapdelaine, William D Foulkes, Daniel Taliun, Adam J Shapiro, Guillaume Lettre, Simon Gravel. "Using the ancestral recombination graph to study the history of rare variants in founder populations" (2025). bioRxiv 2025.03.13.643149.

Chief Raphael Ben-Eghan, Markus Munter, Chikashi Terao, **Alex Diaz-Papkovich**, Simon Gravel, Mark Lathrop, Audrey V Grant. "Multi-ancestry fine-mapping of the chromosome 17q12-q21 asthma locus identifies independent associations implicating lymphocyte and eosinophil levels in the causal pathway." *medRxiv* 2025.08. 05.25333026.

Peyton McClelland, Georgette Femerling, Rose Laflamme, Alejandro Mejia-Garcia, Mohadese Sayahian Dehkordi, Hongyu Xiao, **Alex Diaz-Papkovich**, Justin Pelletier, Jean-Christophe Grenier, Ken Sin Lo, Luke Anderson-Trocme, Justin Bellavance, Vincent Chapdelaine, Genevieve Gagnon, Annelie De Mori, Gerardo Martinez, Kristen Mohler, Thibault de Malliard, Catherine Labbe, Marjorie Labrecque, Alexandre Montpetit, Jean-Francois Theroux, Hufeng Zhou, Simon L Girard, Julie G Hussin, Anne-Marie Laberge, Claude Bherer, Martine Tetreault, Sarah A Gagliano Taliun, Daniel Taliun, Simon Gravel, Guillaume Lettre. (2025). "A multi-ancestry genetic reference for the Quebec population." medRxiv 2025.05.14.25327536.

Jacob W Vogel, Aaron Alexander-Bloch, Konrad Wagstyl, Maxwell Bertolero, Ross Markello, Adam Pines, Valerie J Sydnor, **Alex Diaz-Papkovich**, Justine Hansen, Alan C Evans, Boris Bernhardt, Bratislav Misic, Theodore Satterthwaite, Jakob Seidlitz. "Deciphering the functional specialization of whole-brain spatiomolecular gradients in the adult brain." 121.25 *PNAS* (2024): e2219137121.

Luke Anderson-Trocmé, Dominic Nelson, Shadi Zabad, **Alex Diaz-Papkovich**, Nikolas Baya, Mathilde Touvier, Benjamin Jeffery, Christian Dina, Helene Vezina, Jerome Kelleher, Simon Gravel. "On the Genes, Genealogies, and Geographies of Quebec." *Science* 380.6647 (2023): 849-855.

Chief Ben-Eghan, Rosie Rosie, Jose Sergio Hleap, **Alex Diaz-Papkovich**, Hans Markus Munter, Audrey V. Grant, Charles Dupras, and Simon Gravel. "Don't ignore genetic data from minority populations." *Nature* 585.7824 (2020): 184-186.

Melissa L. Spear, **Alex Diaz-Papkovich**, Elad Ziv, Simon Gravel, Dara G. Torgerson, Ryan D. Hernandez. "Recent shifts in the genomic ancestry of Mexican Americans may alter the genetic

architecture of biomedical traits." *Elife* 9 (2020): e56029.

Jacob W. Vogel, Renaud La Joie, Michel J. Grothe, **Alex Diaz-Papkovich**, Andrew Doyle, Etienne Vachon-Presseau, Claude Lepage, Reinder Vos de Wael, Yasser Iturria-Medina, Boris Bernhardt, Gil D. Rabinovici, Alan C. Evans. "A molecular gradient along the longitudinal axis of the human hippocampus informs large-scale behavioral systems." *Nature Communications* 11.1 (2020): 1-17.

Razvan V. Marinescu, Neil P. Oxtoby, Alexandra L. Young, Esther E. Bron, Arthur W. Toga, Michael W. Weiner, Frederik Barkhof, Nick C. Fox, Arman Eshaghi, Tina Toni, Marcin Salaterski, Veronika Lunina, Manon Ansart, Stanley Durrleman, Pascal Lu, Samuel Iddi, Dan Li, Wesley K. Thompson, Michael C. Donohue, Aviv Nahon, Yarden Levy, Dan Halbersberg, Mariya Cohen, Huiling Liao, Tengfei Li, Kaixian Yu, Hongtu Zhu, Jose G. Tamez-Pena, Aya Ismail, Timothy Wood, Hector Corrada Bravo, Minh Nguyen, Nanbo Sun, Jiashi Feng, B.T. Thomas Yeo, Gang Chen, Ke Qi, Shiyang Chen, Deqiang Qiu, Ionut Buciuman, Alex Kelner, Raluca Pop, Denisa Rimocea, Mostafa M. Ghazi, Mads Nielsen, Sebastien Ourselin, Lauge Sorensen, Vikram Venkatraghavan, Keli Liu, Christina Rabe, Paul Manser, Steven M. Hill, James Howlett, Zhiyue Huang, Steven Kiddle, Sach Mukherjee, Anais Rouanet, Bernd Taschler, Brian D. M. Tom, Simon R. White, Noel Faux, Suman Sedai, Javier de Velasco Oriol, Edgar E. V. Clemente, Karol Estrada, Leon Aksman, Andre Altmann, Cynthia M. Stonnington, Yalin Wang, Jianfeng Wu, Vivek Devadas, Clementine Fourrier, Lars Lau Raket, Aristeidis Sotiras, Guray Erus, Jimit Doshi, Christos Davatzikos, Jacob Vogel, Andrew Doyle, Angela Tam, **Alex Diaz-Papkovich**, Emmanuel Jammeh, Igor Koval, Paul Moore, Terry J. Lyons, John Gallacher, Jussi Tohka, Robert Ciszek, Bruno Jedynak, Kruti Pandya, Murat Bilgel, William Engels, Joseph Cole, Polina Golland, Stefan Klein, Daniel C. Alexander. "The Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) Challenge: Results after 1 Year Follow-up." *arXiv preprint arXiv:2002.03419* (2020).

Invited talks

Population genetics and Wikipedia. BIRS workshop, Modeling and Theory in Population Genetics, April 2024, Banff, AB.

Insights from the topology of biobanks. Dagstuhl Seminar Series, Low-Dimensional Embeddings of High-Dimensional Data: Algorithms and Applications, March 2024, Wadern, Germany.

Topological analysis of high-dimensional human genetic data in biobanks. McGill Quantitative Life Sciences Seminar Series, February 2024, Montreal, QC.

Conference presentations

One year after the All of Us Research Project genomic release: Reflections on visualizing human genetic data in biobanks. Biology of Genomes, May 2025, Cold Spring Harbor, USA (Poster)

Genetics as a growing dimension of group identity: A study using Wikipedia data. Probabilistic Modeling in Genomics, March 2025, Cold Spring Harbor, USA (Poster)

Genetics as a growing dimension of group identity: A study using Wikipedia data. American Society of Human Genetics, November 2024, Denver, USA (Poster)

Connections between topological data analysis and identity-by-descent in biobank data. Society for Molecular Biology and Evolution, July 2024, Puerto Vallarta, Jalisco, Mexico (Poster)

Connections between topological data analysis and identity-by-descent in biobank data. The Allied Genetics Conference, March 2024, Washington D.C. (Poster)

Genetic topology groups and identifying the sources of population structure in large multi-ethnic biobanks.

Probabilistic Modeling in Genomics, March 2023, Cold Spring Harbor, USA. (Poster)

Using topological ancestry groups to describe population structure in large multi-ethnic biobanks.
American Society of Human Genetics, October 2022, Los Angeles, USA. (Contributed talk)

Density-based clustering in multi-ethnic biobanks visualizes and identifies populations and population structure related to phenotypes and environmental factors. Probabilistic Modeling in Genomics, April 2021, Virtual. (Contributed talk)

Revealing cryptic population structure and phenotype heterogeneity in large genomic cohorts using UMAP. RMGA annual conference, November 2019, Montreal, Quebec. (Contributed talk)

Revealing cryptic population structure in large genomic datasets with UMAP. EMBO Workshop: Visualising Biological Data, March 2019, Heidelberg, Germany. (Poster)

Revealing Population Structure in Large Cohorts. Closing the Genomics Research Gap, June 2018. Montreal, Quebec. (Poster)

Revealing Population Structure in Large Cohorts. 7th Annual Canadian Human and Statistical Genetics Meeting, June 2018. Harrison Hot Springs, British Columbia. (Contributed talk)

Applied Statistics and NHL Data. Ottawa Hockey Analytics at Carleton, January 2016. Ottawa, Ontario. (Contributed talk)

Data mining the play by play: A new look at player performance. Ottawa Hockey Analytics at Carleton, February 2015. Ottawa, Ontario. (Contributed talk)

Supervision

- **Abby Kuntzleman**, PhD student, 2024
  Project: Genetics as a growing dimension of group identity: A study using Wikipedia data
- **Shevaughn Holness**, PhD student, 2024
  Project: Studying gene-by-environment interaction in All Of Us
- **Vikram Nathan**, undergraduate student, 2022.
  Project: Using Genotypic and Geographic kNN to Predict Phenotypes in the UK Biobank
- **Jenisha Patel**, undergraduate student, 2020
  Project: Machine Learning Methods for Phenotype Prediction
- **Zixian (Dave) Li**, undergraduate student, 2019.
  Project: Defining Intrinsic Dimensionality for Populations Based on Dimensionality Reduction Methods

Awards

- NSERC Postdoctoral Fellowship ($140,000, 2025)
- l'Association des doyennes et des doyens des études supérieures au Québec (English: Association of Deans of Higher Education of Quebec; ADESAQ) Best Thesis in Health Sciences ($2500, 2024)
- McGill Quantitative Life Sciences Research Day, Best Oral Presentation ($400, 2023)
- Mitacs-Japan Society for the Promotion of Science Summer Program Fellowship (¥534,000, 2020 — canceled due to COVID-19 pandemic travel restrictions)
- RMGA Travel Fellowship ($1500, 2019)
- NSERC Science Exposed finalist (2019)

Research experience

McGill University and Génome Québec Innovation Centre (Dr. Simon Gravel)   2018 - 2023
- Studied high dimensional relationships in population genetics and genomics

Weber Lab (Dr. Stephanie Weber)   2018
- Developed methodology to track extra-nucleolar particles within the nuclei of *C. elegans* embryos

McGill Centre for Integrative Neuroscience (Dr. Alan Evans)   2017
- Investigated machine learning, data mining, and statistical approaches to Alzheimer's research

Skills

- Scripting, statistical computing, and scientific computing in R, Python, SAS, and Matlab
- High-performance computing
- Programming in SQL, VB.NET, Java and other languages
- Familiarity with Windows, Macintosh, Linux, Unix, and online environments
- Scripting in Bash and PowerShell
- Languages:
  - English, fluent
  - French, Government of Canada linguistic profile B/B/C (reading/written/oral)
  - Spanish, conversational
  - Russian, conversational

Activities

- Peer review for scientific journals (2020-present)
- Student representative, Quantitative Life Sciences Program Executive Committee (2018-2021)
- Host, McGill Post-Graduate Student's Society Trivia (2018-2021)
- President, McGill Quantitative Life Sciences Association (2017-2021)
- Member, Statistical Society of Canada (2016-present)
- Member-at-large, Statistical Society of Ottawa executive (2015-2018)
- Ottawa Bayesian Statistics Meetup Group (2015-2017)

Other certifications and awards

- Medal of Bravery (2018)
  - Certificate of Valour, Ottawa Police Service (2017)
- Core training in Introduction to Biosafety (2018)
- TCPS 2: CORE (2018)
- Royal Conservatory of Music Grade 6 Piano (2011)