

# JOELLE MBATCHOU, PhD

U.S. Citizen | CHICAGO IL 60612 USA | joelle.mbatchou@gmail.com | +1(773)599-2825 | linkedin.com/in/jmbatchou

## PROFESSIONAL PROFILE

- Proficient and innovative researcher with 9 years of research experience in developing cutting-edge methods to help improve understanding of genetic variation and its relationship to human health and disease
- Avid and charismatic learner who continuously strives for excellence by always meeting deadlines, providing timely solutions, and staying attentive to new developments in the field
- Eager to assist biotech companies in discovering novel targets and indications by building scalable and efficient association analysis tools and pipelines
- Developed a C++ tool which decreased computation costs at the Regeneron Genetics Center by over 50X
- Top Skills: Analytical, Highly Proficient in C++ and R, Pipeline development using WDL, Project Management

## PROFESSIONAL EXPERIENCE

### Manager, Statistical Genetics

Jan 2022 – Current

*Regeneron Pharmaceuticals, Regeneron Genetics Center | Tarrytown, New York*

- Routinely carry out statistical analyses using cloud-based computing platforms on large-scale and high-dimensional human genetics datasets containing millions of genetic variants and 100,000s of individuals
- Develop statistical methods and computational tools geared for large-scale genetic and genomics studies
- Build WDL pipelines for data sets with 100,000s of individuals from exome and whole-genome sequencing data
- Supervise a summer intern student to develop and carry out a research project

### Senior Statistical Geneticist

Nov 2019 – Dec 2021

*Regeneron Pharmaceuticals, Regeneron Genetics Center | Tarrytown, New York*

- Developed a computationally efficient whole genome regression method REGENIE for large-scale genetic association analyses which can be more than 100x faster than current state-of-the-art methods and can handle population structure and imbalanced binary traits
- Implemented REGENIE into a C++ software which was publicly released on Github
- Published the REGENIE method as first author in Nature Genetics where it was applied to UK Biobank data (>100 phenotypes, >400K individuals and >10M genetic variants)
- Presented the REGENIE method at ASHG 2020 conference (selected for platform talk)

### Graduate Student Researcher | Statistics

Oct 2013 – June 2019

*University of Chicago, Chicago, Illinois*

- Developed a computationally fast method JASPER to assess significance for a general class of association tests, including tests for high dimensional phenotypes and gene-based tests, adjusting for population structure and family relatedness
- Designed a permutation-based testing procedure BRASS for assessing significance with binary traits in structured samples for association tests with unknown exact/asymptotic distributions
- Built C/C++ software to evaluate JASPER and BRASS through simulation studies & real data applications

## EDUCATION

Ph.D Statistics, University of Chicago

June 2019

BSc Biology & Mathematical Sciences, DePaul University

June 2011

Computer Skills – Advanced: C, C++, R, Shell scripting, AWS, LaTeX | Intermediate: Git, WDL

## LEADERSHIP & AWARDS

Selected for Reviewers' Choice (top 10% scoring abstracts) - ASHG 2021

Oct 2021

Selected for platform Talk - ASHG 2020

Oct 2020

## PUBLICATIONS

- [1] Backman, J.D., Li, A. H., Marcketta, A., Sun, D., **Mbatchou, J. et al.** Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* 599, 628-634 (2021).
- [2] Akbari, P., Gilani, A., Sosina, O., Kosmicki, J. A., Khrimian, L., Fang, Y. Y., Persaud, T., Garcia, V., Sun, D., Li, A., **Mbatchou, J. et al.** Sequencing of 640,000 exomes identifies GPR75 variants associated with protection from obesity. *Science* 373(2021).
- [3] Kosmicki, J. A., Horowitz, J. E., [...], **Mbatchou, J. et al.** Pan-ancestry exome-wide association analyses of COVID-19 outcomes in 586,157 individuals. *American journal of human genetics* 108 (2021).
- [4] **Mbatchou, J. et al.** Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet* 53, 1097-1103 (2021).
- [5] **Mbatchou, J.**, Abney, M. & McPeck, M.S. Permutation methods for assessing significance in binary trait association mapping with structured samples. (2019).
- [6] Jiang, D., **Mbatchou, J.** & McPeck, M.S. Retrospective Association Analysis of Binary Traits: Overcoming Some Limitations of the Additive Polygenic Model. *Hum Hered* 80, 187-95 (2015).

## INVITED TALKS AND POSTERS

- Computationally efficient whole-genome regression framework for association analyses in large-scale biobanks, International Common Disease Alliance Virtual Scientific Plenary, March 2022
- Gene-burden tests, gene-environment interactions and time-to-event data analysis within an efficient whole genome regression framework for large-scale biobanks, The American Society of Human Genetics Annual Meeting, October 2021
- Fast and robust methods to detect gene-environment interactions in large-scale biobanks, International Genetic Epidemiology Society Conference, October 2021
- Computationally efficient whole-genome regression for quantitative and binary traits, The American Society of Human Genetics Annual Meeting, October 2020
- A generalized permutation testing method for binary trait association in structured samples, The American Society of Human Genetics Annual Meeting, October 2017
- Retrospective genetic association testing with binary traits in structured samples using a MCMC-based approach, Women in Statistics and Data Science Conference, October 2017
- Genetic association analysis of binary traits in structured samples, Eastern North American Region Spring Meeting, March 2017