

Control-independent mosaic single nucleotide variant detection with DeepMosaic

In the format provided by the
authors and unedited

Supplementary Information

Control-independent mosaic single nucleotide variant detection with DeepMosaic

Xiaoxu Yang^{1,2,*,#}, Xin Xu^{1,2,*}, Martin W. Breuss^{1,2,3}, Danny Antaki^{1,2}, Laurel L. Ball^{1,2}, Changuk Chung^{1,2}, Jiawei Shen^{1,2}, Chen Li^{1,2}, Renee D. George^{1,2}, Yifan Wang⁴, Taejeong Bae⁴, Yuhe Cheng^{5,6,7}, Alexej Abyzov⁴, Liping Wei⁸, Ludmil B. Alexandrov^{5,6,7}, Jonathan L. Sebat^{9,10,11,12}, NIMH Brain Somatic Mosaicism Network[†], Joseph G. Gleeson^{1,2,#}

1. Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA.
2. Rady Children's Institute for Genomic Medicine, San Diego, CA, USA
3. Department of Pediatrics, Section of Genetics and Metabolism, University of Colorado School of Medicine, Aurora, CO, USA
4. Department of Quantitative Health Sciences, Center for Individualized Medicine, Mayo Clinic, Rochester, MN, USA
5. Department of Cellular and Molecular Medicine, UC San Diego, La Jolla, CA, USA
6. Department of Bioengineering, UC San Diego, La Jolla, CA, USA
7. Moores Cancer Center, UC San Diego, La Jolla, CA, USA
8. Center for Bioinformatics, State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing, China
9. Beyster Center for Genomics of Psychiatric Diseases, University of California, San Diego, La Jolla, CA, USA
10. Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA
11. Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA, USA
12. Department of Pediatrics, University of California, San Diego, La Jolla, CA, USA

* These authors contributed equally to this work

Correspondence: xiy010@health.ucsd.edu and jogleeson@health.ucsd.edu

† Full membership in the Brain Somatic Mosaicism Consortium Network is listed in the Supplementary Text.

Supplementary Information Contents

Extended Data Figures

Extended Data Fig. 1: Training strategies and examples of training data for DeepMosaic.

Extended Data Fig. 2: Network model selection based on an independent gold-standard testing set.

Extended Data Fig. 3: The convolutional neural network of the DeepMosaic default model and gradient visualization with guided backpropagation for DeepMosaic default model (EfficientNet-b4).

Extended Data Fig. 4: Performance of DeepMosaic default model (EfficientNet-b4) on data hidden from training.

Extended Data Fig. 5: Performance of DeepMosaic and other mosaic variant callers on SimData2.

Extended Data Fig. 6: Sensitivity and specificity of DeepMosaic and other mosaic variant callers on BioData4.

Extended Data Fig. 7: Comparison of DeepMosaic and traditional mosaic variant calling strategies on a WGS biological dataset (BioData4).

Extended Data Fig. 8: Comparison of features of variants called by DeepMosaic and other pipelines.

Extended Data Fig. 9: Enrichment of genomic features for variants called by DeepMosaic and conventional methods.

Extended Data Fig. 10: Comparison of DeepMosaic and traditional mosaic variant calling strategies on a WES biological dataset (BioData5), and the computational resources required for WES (BioData6) and WGS (BioData4).

Supplementary tables

Supplementary Table 1: Candidate MVs detected in BioData4 (WGS) by DeepMosaic, MosaicForecast, MuTect2, Strelka2, MosaicHunter, and NeuSomatic, with orthogonal validation experiment results by MPAS and annotations for local sequence features.

Supplementary Table 2: Candidate MVs detected in BioData5 (noncancer WES) by DeepMosaic and GATKp50 pipeline, with orthogonal validation experiment results by MPAS.

Supplementary Table 3: Candidate MVs detected in BioData6 (TCGA-MC3 WES) by DeepMosaic, MuSE, MuTect, SomaticSniper, VarScan2, and Radia as well as the released gold standard from the same dataset.

Supplementary Table 4: Summaries for MV caller estimated on BioData4, BioData5, and BioData6.

Supplementary Table 5: Summaries for computational resources of DeepMosaic on WES and WGS data with regards to sequencing depth and the number of variants estimated.

Supplementary Text: Members of the Brain Somatic Mosaicism Network.

Supplementary Text: Members of the Brain Somatic Mosaicism Network

Boston Children's Hospital: August Yue Huang, Alissa D'Gama, Caroline Dias, Christopher A. Walsh, Javier Ganz, Michael Lodato, Michael Miller, Pengpeng Li, Rachel Rodin, Robert Hill, Sara Bizzotto, Sattar Khoshkhoo, Zinan Zhou

Harvard University: Alice Lee, Alison Barton, Alon Galor, Chong Chu, Craig Bohrsen, Doga Gulhan, Eduardo Maury, Elaine Lim, Euncheon Lim, Giorgio Melloni, Isidro Cortes, Jake Lee, Joe Luquette, Lixing Yang, Maxwell Sherman, Michael Coulter, Minseok Kwon, Peter J. Park, Rebeca Borges-Monroy, Semin Lee, Sonia Kim, Soo Lee, Vinary Viswanadham, Yanmei Dou
Icahn School of Medicine at Mt. Sinai: Andrew J. Chess, Attila Jones, Chaggai Rosenbluh, Schahram Akbarian

Kennedy Krieger Institute: Ben Langmead, Jeremy Thorpe, Jonathan Pevsner, Sean Cho

Lieber Institute for Brain Development: Andrew Jaffe, Apua Paquola, Daniel Weinberger, Jennifer Erwin, Jooheon Shin, Michael McConnell, Richard Straub, Rujuta Narurkar

Mayo Clinic: Alexej Abyzov, Taejeong Bae, Yeongjun Jang, Yifan Wang

Sage Bionetworks: Cindy Molitor, Mette Peters

Salk Institute for Biological Studies: Fred H. Gage, Meiyang Wang, Patrick Reed, Sara Linker

Stanford University: Alexander Urban, Bo Zhou, Xiaowei Zhu

Universitat Pompeu Fabra: Aitor Serres Amero, David Juan, Inna Povolotskaya, Irene Lobon, Manuel Solis Moruno, Raquel Garcia Perez, Tomas Marques-Bonet

University of Barcelona: Eduardo Soriano

University of California, Los Angeles: Gary Mathern

University of California, San Diego: Danny Antaki, Dan Averbuj, Eric Courchesne, Joseph Gleeson, Laurel Ball, Martin Breuss, Subhojit Roy, Xiaoxu Yang

University of Michigan: Diane Flasch, Trenton Frisbie, Hui Kopera, Jeffrey Kidd, John Moldovan, John V. Moran, Kenneth Kwan, Ryan Mills, Sarah Emery, Weichen Zhou, Xuefang Zhao

University of Virginia: Aakrosh Ratan

Yale University: Alexandre Jourdon, Flora M. Vaccarino, Liana Fasching, Nenad Sestan, Sirisha Pochareddy, Soraya Scuderi