Olga B. Botvinnik

Curriculum Vitae

Last Updated: October 10, 2024

Studying the "dark matter" of genomes since 2017. PhD in single-cell RNA-sequencing. "Grand unified theory of cells," hashing, sequencing weird creatures, STEM outreach. MIT, Broad Institute, UCSC, UCSD, Chan Zuckerberg Biohub. Open science/data/source.

Experience

Sep 2024 – **Seanome**, San Francisco, CA, Founder.

Present We build open genomics tools to unlock breakthroughs in ocean diversity.

Jan 2022 – **Bridge Bio Pharma**, Computational Genomics, San Francisco, CA, Senior Computa-Sep 2024 tional Biologist.

Analyzed next-generation sequencing data to prioritize therapeutic targets for genetic diseases with unmet needs.

- Analyzed PacBio long-read IsoSeq data of transcriptomes in cell models using Python (gffutils, pandas, scikit-learn, Biopython, sourmash in Jupyter Notebook).
- Wrote Nextflow Workflows for long- and short-read sequencing data to support gene therapy programs in capsid evolution and AAV population sequencing.
- Deployed Nextflow Workflows on Amazon Web Services Batch.
- Jul Oct Arcadia Science, San Francisco, CA, Advisor.
 - 2021 Provided advice to startup founders on data science and bioinformatics needs.
- 2017–2021 Chan Zuckerberg Biohub, Data Sciences Platform, San Francisco, CA, Bioinformatics Scientist II.
 - Contributed to Tabula Muris, Tabula Muris Senis, Tabula Sapiens, and Tabula Microcebus single-cell atlases.
 - Contributed Python and Rust code to "sourmash" software for genomic and metagenomic k-mer analyses (github.com/dib-lab/sourmash)
 - \circ Led cross-species analyses projects to study "dark matter" of transcriptomes using subsampled k-mers from translated RNA-seq reads in a reduced amino acid alphabet, to find putative homologous genes across millions of years of evolution (open draft manuscript)
 - Pioneered weekly weekly internal "Cupcakes & Coding" sessions, alternating between facilitating tutorials taught by junior programmers to share coding knowledge, and pair programming sessions (github.com/czbiohub/cupcakes)
 - Created "BioinformaticsBeyonce" Twitch channel showcasing open source bioinformatics tools and research
- 2010–2011 **Jill Mesirov Laboratory**, Broad Institute of Harvard and MIT, Cambridge, MA, Bioinformatics Research Assistant.

Created REVEALER algorithm to unveil candidate oncogenic activators, published in Nature Biotechnology. Contributed to papers published in Science Signaling and Journal of Hematology \mathscr{E} Oncology.

Education

- 2017 **Ph.D., Bioinformatics and Systems Biology**, University of California, San Diego, La Jolla, CA.
 - Dissertation: Computational analysis of single-cell alternative splicing
- 2012 M.S., Bioinformatics and Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, CA.
- 2010 S.B., Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA.
- 2010 S.B., Mathematics, Massachusetts Institute of Technology, Cambridge, MA.

Research Training

- 2013–2017 **Gene Yeo Laboratory**, *University of California*, *San Diego*, La Jolla, CA.

 Led machine-learning analyses of single-cell motor neuron differentiation mRNA-seq data, in collaboration with wet-lab researchers. Independently developed several software packages written in Python for alternative splicing analyses
- 2012–2013 **Research Rotations**, *University of California*, *San Diego*, La Jolla, CA. Worked in Profs. Trey Ideker, Gene Yeo, and Pavel Pevzner's laboratories
 - 2012 Nader Pourmand Laboratory, University of California, Santa Cruz, Santa Cruz, CA. Analyzed single-cell response of breast cancer drug resistance to paclitaxel
 - 2010 Sebastian Seung Laboratory, MIT Department of Brain and Cognitive Sciences, Cambridge, MA.
 Computed directionality of neurons in electron microscopy of rabbit retina
 - 2009 **David Gifford Laboratory**, MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA.
 - Tested whether information flow can predict gene lethality in genomic networks
 - 2008 **Sean Eddy Laboratory**, Howard Hughes Medical Institute Janelia Farm Research Campus, Ashburn, VA.
 - Used Hidden Markov Models to improve protein homology search with robust null models
 - 2007 Martha Bulyk Laboratory, Brigham and Women's Hospital, Division of Genetics, Boston, MA.
 - Analyzed DNA binding specificities of mouse homeodomain transcription factors

Awards

Fellowships

- 2017 University of Washington eScience Institute Moore/Sloan Data Science and Washington Research Foundation Innovation in Data Science Postdoctoral Fellowship (declined)
- 2014 NumFocus John Hunter Technical Fellowship for Open Source Science
- 2013–2016 National Defense Science and Engineering Graduate Fellowship Honors
 - 2016 100 Awesome Women In The Open-Source Community You Should Know, sourced.com
 - 2013 Fannie and John Hertz Foundation Fellowship Finalist
 - 2012 National Science Foundation Graduate Research Fellowship: Honorable Mention

- 2012 University of California Regents Scholarship
- 2009 Bernard M. Gordon-MIT Engineering Leadership Program
- 2008 Howard Hughes Medical Institute Janelia Farm Research Summer Scholar

Publications

Journal Articles

- [1] The Tabula Microcebus Consortium, Camille Ezran, Shixuan Liu, Jingsi Ming, Lisbeth A Guethlein, Michael F Z Wang, Roozbeh Dehghannasiri, Julia Olivieri, Hannah K Frank, Alexander Tarashansky, Winston Koh, Qiuyu Jing, **Olga B Botvinnik**, Jane Antony, Stephen Chang, Angela Oliveira Pisco, Jim Karkanias, Can Yang, James E Ferrell, Scott D Boyd, Peter Parham, Jonathan Z Long, Bo Wang, Julia Salzman, Iwijn De Vlaminck, Angela Wu, Stephen R Quake, and Mark A Krasnow. Mouse lemur transcriptomic atlas elucidates primate genes, physiology, disease, and evolution. *bioRxiv*, page 2022.08.06.503035, Aug. 2022.
- [2] Tabula Sapiens Consortium. The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science*, 376(6594):eabl4896, May 2022.
- [3] Taylor E Reiter, N Tessa Pierce-Ward, Luiz Irber, **Olga B Botvinnik**, and C Titus Brown. Protein k-mers enable assembly-free microbial metapangenomics. *bioRxiv*, page 2022.06.27.497795, June 2022.
- [4] C Ezran, S Liu, S Chang, J Ming, **Olga B Botvinnik**, and others. Tabula Microcebus: A transcriptomic cell atlas of mouse lemur, an emerging primate model organism. *bioRxiv*, 2021.
- [5] Olga B Botvinnik, Venkata Naga Pranathi Vemuri, N Tessa Pierce, Phoenix Aja Logan, Saba Nafees, Lekha Karanam, Kyle Joseph Travaglini, Camille Sophie Ezran, Lili Ren, Yanyi Juang, Jianwei Wang, Jianbin Wang, and C Titus Brown. Single-cell transcriptomics for the 99.9% of species without reference genomes. *bioRxiv*, page 2021.07.09.450799, July 2021.
- [6] Sachin J Shah, Peter N Barish, Priya A Prasad, Amy Kistler, Norma Neff, Jack Kamm, Lucy M Li, Charles Y Chiu, Jennifer M Babik, Margaret C Fang, Yumiko Abe-Jones, Narges Alipanah, Francisco N Alvarez, Olga B Botvinnik, Gloria Castaneda, CZB CLIAhub Consortium, Rand M Dadasovich, Jennifer Davis, Xianding Deng, Joseph L DeRisi, Angela M Detweiler, Scot Federman, John Haliburton, Samantha Hao, Andrew D Kerkhoff, G Renuka Kumar, Katherine B Malcolm, Sabrina A Mann, Sandra Martinez, Rupa K Mary, Eran Mick, Lusajo Mwakibete, Nader Najafi, Michael J Peluso, Maira Phelps, Angela Oliveira Pisco, Kalani Ratnasiri, Luis A Rubio, Anna Sellas, Kyla D Sherwood, Jonathan Sheu, Natasha Spottiswoode, Michelle Tan, Guixia Yu, Kirsten Neudoerffer Kangelaris, and Charles Langelier. Clinical features, diagnostics, and outcomes of patients presenting with acute respiratory illness: A retrospective cohort study of patients with and without COVID-19. EClinicalMedicine, 27:100518, Oct. 2020.
- [7] Lincoln Harris, Rohan Vanheusden, **Olga B Botvinnik**, and Spyros Darmanis. cerebra: A tool for fast and accurate summarizing of variant calling format (VCF) files. *J. Open Source Softw.*, 5(54):2432, Oct. 2020.

- Tabula Muris Consortium. Single-cell transcriptomics 20 mouse organs creates Tabula Muris. Nature. pages Oct. 2018. https://www.biorxiv.org/content/early/2017/12/20/237446.full.pdf.
- [9] D V Klopfenstein, Liangsheng Zhang, Brent S Pedersen, Fidel Ramírez, Alex Warwick Vesztrocy, Aurélien Naldi, Christopher J Mungall, Jeffrey M Yunes, Olga B Botvinnik, Mark Weigel, Will Dampier, Christophe Dessimoz, Patrick Flick, and Haibao Tang. GOATOOLS: A Python library for Gene Ontology analyses. Sci. Rep., 8(1):10872, July 2018.
- [10] Yan Song*, **Olga B Botvinnik***, Michael T Lovci, Boyko Kakaradov, Patrick Liu, Jia L Xu, and Gene Yeo. Single-Cell Alternative Splicing Analysis with Expedition Reveals Splicing Dynamics during Neuron Differentiation. *Molecular Cell*, 2017. * These authors contributed equally to this work.
- [11] Curtis A Nutter, Elizabeth A Jaworski, Sunil K Verma, Vaibhav Deshmukh, Qiongling Wang, **Olga B Botvinnik**, Mario J Lozano, Ismail J Abass, Talha Ijaz, Allan R Brasier, Nisha J Garg, Xander H T Wehrens, Gene W Yeo, and Muge N Kuyumcu-Martinez. Dysregulation of RBFOX2 Is an Early Event in Cardiac Pathogenesis of Diabetes. *Cell Reports*, 15(10):2200–2213, 2016.
- [12] Jong Wook Kim*, **Olga B Botvinnik***, Omar Abudayyeh, Chet Birger, Joseph Rosenbluh, Yashaswi Shrestha, Mohamed E Abazeed, Peter S Hammerman, Daniel DiCara, David J Konieczkowski, et al. Characterizing genomic alterations in cancer by complementary functional associations. *Nature Biotechnology*, 2016. * These authors contributed equally to this work.
- [13] Kris C Wood, David J Konieczkowski, Cory M Johannessen, Jesse S Boehm, Pablo Tamayo, **Olga B Botvinnik**, Jill P Mesirov, William C Hahn, David E Root, Levi A Garraway, et al. MicroSCALE screening reveals genetic modifiers of therapeutic response in melanoma. *Science Signaling*, 5(224):rs4, 2012.
- [14] A Goncearenco, P Grynberg, Olga B Botvinnik, Geoff Macintyre, and Thomas Abeel. Highlights from the Eighth International Society for Computational Biology (ISCB) Student Council Symposium 2012. BMC Bioinformatics, 2012.
- [15] Naomi Galili, Pablo Tamayo, **Olga B Botvinnik**, Jill P Mesirov, Margarita R Brooks, Gail Brown, and Azra Raza. Prediction of response to therapy with ezatiostat in lower risk myelodysplastic syndrome. *Journal of Hematology & Oncology*, 5(1):1, 2012.
- [16] Naomi Galili, Pablo Tamayo, Olga B Botvinnik, Jill P Mesirov, Jennifer Zikria, Gail Brown, and Azra Raza. Gene Expression Studies May Identify Lower Risk Myelodysplastic Syndrome Patients Likely to Respond to Therapy with Ezatiostat Hydrochloride (TLK199). Blood, 118(21):2779–2779, 2011.
- [17] Michael F Berger, Gwenael Badis, Andrew R Gehrke, Shaheynoor Talukder, Anthony A Philippakis, Lourdes Pena-Castillo, Trevis M Alleyne, Sanie Mnaimneh, **Olga B Botvinnik**, Esther T Chan, et al. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell*, 133(7):1266–1276, 2008.

Books

[1] P Compeau and P Pevzner. Bioinformatics Algorithms Volume 1, volume 1 of An Active Learning Approach. Active Learning Publishers LLC, 2 edition, 2015. Contributed text, figures, problems and code solutions, primarily to "Chapter 4: How Do We Sequence Antibiotics?".

Conference Posters

- [1] **Olga B Botvinnik**, Yan Song, Michael T Lovci, Boyko Kakaradov, Patrick Liu, Jia L Xu, and Gene W Yeo. Single-cell alternative splicing analysis using Expedition reveals splicing dynamics during neuron differentiation. In *RNA Society*, Kyoto, Japan, June 2016.
- [2] Olga B Botvinnik, Yan Song, Michael T Lovci, Boyko Kakaradov, Patrick Liu, Leen Jamal, and Gene W Yeo. Novel computational metrics for population-wide differences in alternative splicing in single cells. In *Stem Cell Genomics California Institute of Regenerative Medicine Retreat*, La Jolla, CA, Apr. 2016.
- [3] Olga B Botvinnik, Yan Song, Michael T Lovci, Boyko Kakaradov, Patrick Liu, Leen Jamal, and Gene W Yeo. Single-cell alternative splicing analysis using Expedition reveals splicing dynamics during neuron differentiation. In Single Cell Analysis Investigators Meeting, Bethesda, MD, Apr. 2015.
- [4] **Olga B Botvinnik**, Jonathan Kim, Wendy Lee, Paola Castro, and Nader Pourmand. Single-cell differential expression pipeline. In *Intelligent Systems for Molecular Biology* (ISMB), Long Beach, CA, June 2012.
- [5] Olga B Botvinnik, Pablo Tamayo, and Jill P Mesirov. Discovery of novel candidate oncogenic activators with REVEALER. In *Intelligent Systems for Molecular Biology* (ISMB), Vienna, Austria, June 2011.

Talks

- 2020 Biology of Genomes, Cold Spring Harbor Laboratory, Long Island, New York, Functional prediction of transcriptomic "dark matter" across species.
 Slides: speakerdeck.com/olgabot/functional-prediction-of-transcriptomic-dark-matter-across-species
- The Identity and Evolution of Cell Types, European Molecular Biology Laboratory, Heidelberg, Germany, Reference-free comparative transcriptomics.

 Slides: speakerdeck.com/olgabot/reference-free-comparative-transcriptomics
- 2019 Data Intensive Biology Summer Institute (Invited speaker), University of California, Davis, Davis, CA, Single-cell RNA-seq: To Infinity and Beyond!.
 Slides: osf.io/gdzuy/, Video: youtu.be/hAqa8DztxSU
- 2018 Biological Data Science, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, Fast approximate cell type identification via MinHash sketches of k-mers in single cell RNA-seq.

Slides: Google slides

- 2018 Current Progress in Biotechnology Seminar Series (Invited speaker), University of California, Davis, Davis, CA, If you liked it, you should have put a Seq on it: Job-seq and lessons learned.
 - $Slides: \ slideshare.net/olgabotvinnik/if-you-liked-it-you-should-have-put-a-seq-on-it-liked-it-you-should-have-put-a$
- 2017 Open Data Science Conference, San Francisco Hyatt Regency, San Francisco, CA, Co-evolution of algorithms and data in biology.

 Slides: speakerdeck.com/olgabot/co-evolution-of-algorithms-and-data-in-biology
- 2016 Festival of Genomics California, San Diego Convention Center, San Diego, CA.
- 2016 Fluidigm User Group Meeting, City of Hope Hospital, Los Angeles, CA.
- 2016 Bioiformatics and Systems Biology Bootcamp, University of California, San Diego, La Jolla, CA, Dr. You or How I Learned to Stop Worrying and Love the Ph.D.. Slides: slideshare.net/olgabotvinnik/dr-you-or-how-i-learned-to-stop-worry-and-love-the-phd
- 2016 Bioinformatics and Systems Biology Ph.D. Program Recruitment, University of California, San Diego, La Jolla, CA.
- 2015 **Bioiformatics and Systems Biology Bootcamp**, *University of California*, *San Diego*, La Jolla, CA, Dr. You or How I Learned to Stop Worrying and Love the Ph.D.. Slides: slideshare.net/olgabotvinnik/dr-you-or-how-i-learned-to-stop-worry-and-love-the-phd
- 2015 San Diego Bioinformatics User Group, University of California, San Diego, La Jolla, CA, Open-source software for single-cell and other large-scale transcriptomic datasets.

 Slides: http://nbviewer.jupyter.org/format/slides/gist/olgabot/2ee1087d74df46c842df/(same as below)
- 2015 CodeNeuro, New Museum, New York, NY, Flotilla: Data-driven conversations in biology.
 Slides: http://nbviewer.jupyter.org/format/slides/gist/olgabot/2ee1087d74df46c842df/
- 2015 **AmpNeuro**, Amplifying Neuroscience Symposium, La Jolla, CA, Open-source software for single-cell and other large-scale transcriptomic datasets.

 Slides: http://nbviewer.jupyter.org/format/slides/gist/olgabot/ba6970fbfa2babd79f55/
- 2015 Bioinformatics Exchange, University of California, San Diego, La Jolla, CA.
- 2015 Bioinformatics and Systems Biology Ph.D. Program Recruitment, University of California, San Diego, La Jolla, CA.
- 2014 RNA Club, University of California, San Diego, La Jolla, CA.
- 2014 **Bioinformatics EXPO**, *University of California*, *San Diego*, La Jolla, CA. Best Talk, 2nd place
- 2014 **PyData**, 401 Park Ave. South, New York, NY.

 Presentation: https://www.youtube.com/watch?v=IQksDvFl2_8. Slides: https://github.com/olgabot/pydata2014biodata

Teaching, Outreach, and Leadership

2018 — Core member, nf-core, nf-co.re.
Contributed to a community effort to collect a curated set of analysis pipelines built using Nextflow.

- 2017–2019 Dancer, SoulForce Dance Company, San Francisco, CA.
 Member of 11-piece hip-hop dance company. Performed in twice annual shows and assisted in beginner dance classes.
- 2015–2017 **Principal Cellist**, *UCSD Chamber Orchestra*, La Jolla, CA. First cellist out of six. Led cello section during rehearsals.
- 2016-2017 **Teaching Assistant**, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.

 Developed and led bioinformatics coursework of Single Cell Analysis Course including alignment, machine learning, Python, and basic command line tools to an audience largely with little to no programming experience. Course materials available at http://github.com/YeoLab/single-cell-bioinformatics
 - 2016 **Guest Instructor**, Quantitative Methods in Genetics and Genomics, La Jolla, CA, . Taught three weeks of git, RNA-seq and analysis methods to graduate-level UCSD course of 30 students, mostly with limited programming experience. Course materials available at http://github.com/biom262/biom262-2016
- 2015–2016 **Speaker and Co-Organizer**, *CodeNeuro*, New York, NY and San Francisco, CA.

 Presented flotilla software, taught "coding for neuroscientists" tutorial (http://github.com/codeneuro/gitgoing), and advanced data analysis tutorial
- 2015–2016 **President and Co-Founder**, *Graduate Bioinformatics Council*, La Jolla, CA. Founded graduate student council organization for UCSD Bioinformatics and Systems Biology Program. Advocated for student voices, organized "town hall" meetings, social hours, fellowship peer review, and led a team of eight vice presidents and representatives.
 - 2015 **Guest Instructor**, Quantitative Methods in Genetics and Genomics, La Jolla, CA, .

 Taught "data cleaning" and plotting course using Python to graduate-level UCSD course of 10 students, mostly with limited programming experience.
- 2013–2016 Volunteer, San Diego Science and Engineering Festival, San Diego, CA.

 Developed and demonstrated bioinformatics modules to all ages at UCSD Bioinformatics booth.
- 2013–2014 **Instructor**, Bioinformatics Algorithms, Coursera.org.

 Developed interactive curriculum for online Bioinformatics Algorithms Coursera class and textbook. Advisors: Pavel Pevzner and Phillip Compeau
- 2011–2012 **Mentor**, We Teach Science, San Jose, CA. Weekly algebra tutoring to an 8th grader
- 2011–2012 **Guest Instructor**, Pacific Collegiate School, Santa Cruz, CA.

 Created bioinformatics modules to engage students in tying genotype to phenotype for high school AP Biology
 - 2012 **Co-Chair**, Intelligent Systems for Molecular Biology Student Council Symposium, Long Beach, CA.
 - 2012 **Instructor**, Minority Access to Research Careers, Santa Cruz, CA.

 Taught inquiry-based stem cell bioinformatics curriculum to undergraduate researchers
 - 2011 **Volunteer**, *Science Club for Girls*, Cambridge, MA. Co-led after-school biology science club for a class of 16 2nd graders
- 2009–2011 **Choreographer**, $MIT\ DanceTroupe$, Cambridge, MA. Taught beginner to intermediate hip-hop choreography to fellow students
- 2008–2010 **Publicity Chair**, *MIT DanceTroupe*, Cambridge, MA.

 Designed posters and T-shirts to publicize and promote DanceTroupe concert attendance

- 2008 Social Chair, Baker House, Cambridge, MA.
 - Organized social events for students, including a popular "Dormal" event with catered dinner and jazz music performances

Software

- All software is written in Python and open source, licensed under the 3-clause BSD license, except where noted.
- anchor Categorizes alternative splicing data into "modes"—bimodal, unimodal, or uniform. http://github.com/YeoLab/anchor
- bonvoyage Transforms 1d splicing profiles into 2d space to maximize interpretability of change in signal. http://github.com/YeoLab/bonvoyage
 - dobby Dobby is a free and open source package for converting and managing plate reader fluorescence outputs, cDNA concentration files, ECHO pick lists, and creating sample sheets for Illumina sequencing. http://github.com/czbiohub/dobby
- flotilla All-in-one package to perform machine learning analyses on large-scale molecular profiling datasets such as gene expression and alternative splicing. http://github.com/YeoLab/flotilla (72 stars on GitHub)
- hermione Compare multiple distributions with horizon plots (also known as ridge plots) http://github.com/czbiohub/hermione (6 stars on GitHub)
- kvector Counts k-mers in DNA or RNA as k-mer vectors, transforms position weight matrices (PWMs) to k-mer vectors. http://github.com/olgabot/kvector (5 stars on GitHub)
- nf-core A collection of high quality Nextflow pipelines. Contributed to rnaseq, scrnaseq and wrote kmermaid pipeline http://github.com/nf-core
- outrigger Fast de novo alternative exon detection and quantification. http://github.com/YeoLab/outrigger (9 stars on GitHub)
 - qtools Submit jobs to the supercomputer cluster from within Python. http://github.com/YeoLab/qtools (11 stars on GitHub)
- poshsplice Annotates alternative splicing events with biological features such as translated protein product. http://github.com/olgabot/poshsplice (2 stars on GitHub)
- prettyplotlib Painlessly create beautiful matplotlib plots. http://github.com/olgabot/prettyplotlib (1,159 stars on GitHub)
 - pyhomer Utility functions to work with output from the HOMER motif finding program. https://github.com/olgabot/pyhomer
 - seaborn Statistical visualization library. Contributor, wrote clustered heatmap classes and function. http://github.com/mwaskom/seaborn (3,602 stars on GitHub)
 - sourmash Compute and compare MinHash signatures for DNA data sets. Contributor, wrote reduced amino acid alphabet and other protein fixes. https://github.com/dib-lab/sourmash (199 stars on GitHub)
- wasabiplot* Plot coverage and junction reads for any bam file and any region. *Derivative of SashimiPlot, and thus under the GNU General Public License (GPL). http://github.com/olgabot/wasabiplot

Mentees

- 2019 Saba Nafees, Chan Zuckerberg Biohub, San Francisco, CA, PhD Intern.
- 2018 Gerry Meixong, Chan Zuckerberg Biohub, San Francisco, CA, Undergraduate Intern.
- 2014–2017 **Jessica Lettes**, *University of California*, *San Diego*, La Jolla, CA, Undergraduate Student.
 - 2013 **Natalia La Spada**, *University of California, San Diego*, La Jolla, CA, High School Summer Student.