

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Love, Michael, I.

eRA COMMONS USER NAME: mikelove

POSITION TITLE: Assistant Professor

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE (if applicable)	Completion Date MM/YYYY	FIELD OF STUDY
Stanford University, Stanford, CA	B.S.	05/2005	Mathematics
Stanford University, Stanford, CA	M.S.	01/2010	Statistics
Freie Universität, Berlin, Germany	Dr. rer. nat. (PhD equiv.)	10/2013	Computational Biology
Harvard School of Public Health & Dana-Farber Cancer Institute, Boston, MA	Postdoctoral	07/2016	Biostatistics & Computational Biology

A. Personal Statement

My research focuses on developing statistical methods for analysis of high-dimensional datasets in order to answer complex biological questions and generate new hypotheses. I work in close collaboration with biologists, geneticists and clinicians through early exploration of raw data, development of appropriate statistical methods, and translation into biologically meaningful results. My training in mathematics and statistics allows me to draw on a variety of statistical techniques when approaching new high-dimensional datasets. A dominant theme among the methods I have developed is the isolation of biological signal from technical artifacts and the variation in measurements expected by change alone, having both technical and biological components. Specifically I have developed and published widely-used methods for robust detection of differential expression and transcript abundance from RNA-seq experiments, copy number state following DNA-seq, and annotation of regulatory regions using DNase-seq and ChIP-seq experiments. I have distributed all of my statistical methods in well-documented open-source software packages, and continue to maintain and provide user support of these through the Bioconductor Project. For the Seed Networks for the Human Cell Atlas, I will collaborate with the rest of the team in order to implement the dissemination and versioning of the reference cell-type catalog through R/Bioconductor and python genomic data analysis frameworks. This work leverages existing collaborations with Drs. Patro, Hicks, and Fertig on Salmon and R/Bioconductor software development.

1. **Love, M. I.**, Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550. PMID: PMC4302049.
2. **Love, M. I.**, Hogenesch, J. B., Irizarry, R. A. (2016) Modeling of RNA-seq fragment sequence bias reduces systematic errors in transcript abundance estimation. *Nature Biotechnology*, 34(12), 1287-1291. PMID: PMC5143225.
3. **Love, M.I.**, Huska, M., Jurk, M., Schöpflin, R., Starick, S., Schwahn, K., Cooper, S., Yamamoto, K., Thomas-Chollier, M., Vingron, M., Meijsing, S. (2017) Role of the chromatin landscape and sequence in determining cell type-specific genomic glucocorticoid receptor binding and gene regulation. *Nucleic Acids Research*, 45(4): 1805–1819. PMID: PMC5389550.

B. Positions and Honors

Positions and Employment

2009	Bioinformatic Analyst, University of California San Francisco, CA.
2010-13	IMPRS Doctoral Scholarship, Max Planck Institute for Molecular Genetics, Berlin, Germany.
2012-13	Visiting Scientist, European Molecular Biology Laboratory, Heidelberg, Germany.
2013-16	Postdoctoral Fellowship, T32 Cancer Training Grant, Harvard TH Chan School of Public Health and Dana-Farber Cancer Institute, Boston, MA.
2016-	Assistant Professor, Department of Biostatistics and Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Other Experience and Professional Memberships

2015-	Member, American Statistical Association
2016-	Member, ENAR, International Biometrics Society

Honors

2017	Junior Faculty Development Award, UNC-Chapel Hill, NC
------	---

C. Contribution to Science

1. Statistical tools and workflows for differential expression of RNA-seq. I have developed statistical methods, tools and workflows for the analysis of high-dimensional count data as arises from high-throughput sequencing assays, such as RNA-seq. Sequencing experiments inherently involve sampling processes that add variation in measurements on top of the inherent biological variation in the abundance of molecules for each gene across samples. Methods that appropriately model the count data and account for the heterogeneity of variance of expression values across genes have been shown to exhibit higher statistical power. I lead the development and maintain the widely used DESeq2 method and Bioconductor package, providing statistical inference and visualization for RNA-seq experiments. DESeq2 provides robust inference despite the common small per-group samples sizes ($n=3-5$), through the use of data-driven moderation of biological variance estimates. The statistical methods have been re-used in a variety of domains, including CRISPR/Cas9 assays (MAGeCK), ChIP-seq (DiffBind), chromosome conformation capture (FourCSeq) and metagenomics (PhyloSeq). I also developed a Bioconductor package, tximport, in collaboration with Dr. Mark Robinson's group (an author of edgeR), which facilitates the use of DESeq2 and other Bioconductor RNA-seq packages as statistical inference engines downstream of fast, new methods for transcript quantification, including Salmon, Sailfish and kallisto. The tximport pipeline gives the statistical benefit of working on the count scale, while correcting for differences in feature length across samples that might arise from differential isoform usage.

- a. **Love, M. I.**, Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550. PMID: PMC4302049.
- b. The DESeq2 package for differential gene expression from RNA-seq:
<http://bioconductor.org/packages/DESeq2>
- c. **Love, M. I.**, Anders, S., Kim, V., Huber, W. (2015). RNA-Seq workflow: gene-level exploratory analysis and differential expression. *F1000Research*, 4, 1070. PMID: PMC4670015.
- d. Sonesson, C., **Love, M.I.**, Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, 4, 1521. PMID: PMC4712774.

2. Transcript abundance estimation accounting for common RNA-seq technical biases. In my postdoctoral studies, I focused on accurate estimation of relative abundances at the transcript- or isoform-level, which is in many ways statistically and computationally more challenging than estimation of abundance at the gene-level. Through detailed examination of existing software packages' performance on large-scale benchmark and consortia RNA-seq datasets, I determined that common technical biases related to fragment sequence and the amplification stage of library preparation induce widespread systematic errors in abundance estimates, including misidentification of the dominant isoform of multi-isoform genes. I lead the development of a statistical solution for resolving these errors, and implemented a Bioconductor package, alpine, for visualization and correction of these biases. Through close collaboration with Dr. Rob Patro, we were able to introduce

fragment sequence bias correction to the Salmon method for transcript abundance estimation. This collaborative effort should prove to be greatly beneficial to the research community, as the Salmon software is one of the fastest and most efficient methods, and now offers the most complete set of technical bias correction available for RNA-seq data.

- a. **Love, M. I.**, Hogenesch, J. B., Irizarry, R. A. (2016) Modeling of RNA-seq fragment sequence bias reduces systematic errors in transcript abundance estimation. *Nature Biotechnology*, 34(12), 1287-1291. PMID: PMC5143225.
- b. Patro, R., Duggal, G., **Love, M.I.**, Irizarry, R.A., Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, 14(4), 417-419. PMID: PMC5600148.
- c. The Salmon software for accurate, fast, and bias-aware transcript abundance estimates using dual-phase inference:
<https://combine-lab.github.io/salmon>
- d. Teng, M., **Love, M.I.**, Davis, C.A., Djebali, S., Dobin, A., Graveley, B.R., Li, S., Mason, C.E., Olson, S., Pervouchine, D., Sloan, C.A., Wei, X., Zhan, L., Irizarry, R.A. (2016). A benchmark for RNA-seq quantification pipelines. *Genome Biology*, 17(1), 74. PMID: PMC4842274.

3. *Computational analysis of chromatin accessibility and histone modifications identifying regulatory function.*

I was the lead statistical analyst for a genome-wide investigation of binding patterns of glucocorticoid receptor (GR) across three cell types. While it is known that GR binds preferentially to accessible chromatin, we sought to identify what additional chromatin and sequence features are associated with cell-type-specific binding within the universe of open chromatin. Using cell-type-matched DNase-seq and histone post-translational modification ChIP-seq assays publicly available from the ENCODE project, I built a hierarchical model to generate posterior estimates for each chromatin and sequence features' cell-type-specific association with GR binding, while appropriately modeling experiment-specific variations within cell-types. In a related project, I collaborated with the group of Dr. Peter Robinson's group at Charité Hospital in Berlin to functionally annotate potentially pathogenic CNVs using publicly available chromatin accessibility datasets. I defined a measurement of tissue-specificity for putative regulatory regions using hundreds of DNase-seq assays of human tissues from the Roadmap Epigenomics Consortium. The set of tissue-specific regulatory regions was used to annotate potential functions and causal mechanisms for CNV found in patients with congenital disease.

- a. **Love, M.I.**, Huska, M., Jurk, M., Schöpflin, R., Starick, S., Schwahn, K., Cooper, S., Yamamoto, K., Thomas-Chollier, M., Vingron, M., Meijsing, S. (2017) Role of the chromatin landscape and sequence in determining cell type-specific genomic glucocorticoid receptor binding and gene regulation. *Nucleic Acids Research*, 45(4): 1805–1819. PMID: PMC5389550
- b. Ibn-Salem, J. *, Kohler, S. *, **Love, M. I.**, Chung, H.-R., Huang, N., Hurles, M. E., Haendel, M., Washington, N. L., Smedley, D., Mungall, C. J., Lewis, S. E., Ott, C.-E., Bauer, S., Schofield, P. N., Mundlos, S., Spielmann, M., and Robinson, P. N. (2014). Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biology*, 15(9), 423+. PMID: PMC4180961.

* denotes equal contribution

4. *Robust detection of structural variation from Exome-seq.* I developed a statistical method for the detection of copy number variants (CNV) in highly variable coverage from Exome-seq data with collaborators at the Max Planck Institute for Molecular Genetics in Berlin. The CNV detection method introduced a number of novel concepts in the area of research, including use of overdispersed count distributions into a predictive segmentation of the patient's genome, control of technical bias through modeling on GC-content and background coverage calculation, and the combination of the two stages of normalization and segmentation into a single optimization stage. I produced a Bioconductor package, *exomeCopy*, implementing the CNV detection method, which was applied to a panel of more than 400 patients with X-linked intellectual disability, detecting 10 putative disease-causing variants that were not detected by array-based screening.

- a. **Love, M. I.**, Mysickova, A., Sun, R., Kalscheuer, V., Vingron, M., and Haas, S. A. (2011). Modeling read counts for CNV detection in exome sequencing data. *Statistical Applications in Genetics and Molecular Biology*, 10(1), 52. PMID: PMC3517018.

- b. Sun, R., **Love, M. I.**, Zemojtel, T., Emde, A.-K., Chung, H.-R., Vingron, M., and Haas, S. A. (2012). Breakpointer: using local mapping artifacts to support sequence breakpoint discovery from single-end reads. *Bioinformatics*, 28(7), 1024–1025. PMID: 22302574.
- c. Hu, H., Haas, S. A., Chelly, J., Van Esch, H., Raynaud, M., de Brouwer, A. P. M., Weinert, S., Froyen, G., Frints, S. G. M., Laumonnier, F., Zemojtel, T., **Love, M. I.**, et. al (2016). X-exome sequencing of 405 unresolved families identifies seven novel intellectual disability genes. *Molecular Psychiatry*, 21(1), 133-148. PMCID: PMC5414091.
- d. The exomeCopy package for detection of CNV from Exome-seq:
<http://bioconductor.org/packages/exomeCopy>

Complete List of Published Work in MyBibliography:

<http://www.ncbi.nlm.nih.gov/myncbi/browse/collection/50822530/?sort=date&direction=descending>

D. Additional Information: Research Support

Ongoing Research Support

5 P30 ES010126-17 (Troester) 2/1/2000 3/31/21

National Institute of Environmental Health Sciences

UNC-CH Center for Environmental Health & Susceptibility - Biostatistics & Bioinformatics Facility Core

The UNC-CH Center on Environmental Health and Susceptibility brings population science, medical and biomedical researchers together to examine major issues in environmental health resulting from gene-environment interactions that affect an individual's susceptibility to disease. Role: Biostatistician

5 P30 ES010126-17 (Troester) 2/1/2000 3/31/21

National Institute of Environmental Health Sciences

UNC Center for Environmental Health and Susceptibility - Molecular Analysis Facility Core

The goal of the Molecular Analysis Facility Core (MAFC) is to support Center members, especially those that comprise the Interdisciplinary Research Groups and award recipients of the Pilot Projects Program, by ensuring access to analytical resources within the MAFC sub-core facilities, which include DNA Damage, Genomics, Biomarker Spectrometry, Metabolomics, Network Visualization, Proteomics and the Translational Pathology Laboratory). Support includes consultation and experimental design, routine sample analysis and training, development of novel analytical methods needed to advance research, and in-depth data analysis, integration, and interpretation. Role: Biostatistician

5 P01 CA142538-09 (Kosorok) 4/1/2010 3/31/20

National Cancer Institute

Statistical Methods for Cancer Clinical Trials - Project 3: Statistical/Computational Methods for Pharmacogenomics and Individualized Therapy

This research intends to develop novel and high-impact statistical and computational tools for discovering genetic variants associated with interindividual differences in the efficacy and toxicity of cancer medications and for optimizing drug therapy on the basis of each patient's genetic constitution. Role: Co-Investigator

5 R01 HG009125-02 (Nobel/Wright) 9/1/2016 6/30/19

National Human Genome Research Institute

Multi-tissue and network models for next-generation EQTL studies

The broad objective of the proposed research is to develop statistical methods and computational tools that will aid biomedical researchers investigating the genetic basis of human disease. Its central goal is to enable and enhance recent large scale initiatives to understand complex diseases through the simultaneous study of multiple human tissues. Role: Co-Investigator

2 R01 DK093757-06 (Mohlke) 8/1/2017 7/31/22

National Institute of Diabetes and Digestive and Kidney Diseases

Genetic epidemiology of rare and regulatory variants for metabolic traits

Obesity, diabetes, and metabolic syndrome are leading causes of morbidity and mortality worldwide. Metabolic traits related to these diseases have a strong inherited basis. The proposed work will identify DNA variants that influence these traits and mechanisms by which the variants regulate gene expression and alter trait levels. The results may lead to improved disease diagnosis and treatment. Role: Co-Investigator

1 UL1 TR002489-01 (Buse) 3/30/2018 2/28/23

National Center for Advancing Translational Sciences

North Carolina Translational and Clinical Science Institute (NC TraCS) - Biostatistics, Epidemiology, and Research Design (BERD)

The North Carolina Translational and Clinical Sciences Institute (TraCS) is the integrated hub of the NIH Clinical and Translational Science Awards (CTSA) program at the University of North Carolina at Chapel Hill (UNC). Over the next five years, we will apply our expertise and infrastructure to support clinical and translational research, to advance health care for North Carolinians and the national goals of NCATS.

Role: Biostatistician

Completed Research Support

5 T32 CA009337 (Parmigiani) 10/04/13-07/31/16

Training Grant in Quantitative Sciences for Cancer Research

The training program is to train students and postdoctoral fellows to be high quality quantitative researchers who are capable of conducting cutting edge methodological and collaborative research in cancer clinical trials, computational biology, cancer genomics, cancer epidemiology and population science. Also, the program trains quantitative researchers to become strong team leaders/players as well as excellent communicators in a cancer research environment, and to enable them to effectively disseminate their research results and to assume active roles in the design, analysis and interpretation of, for example, cancer genomic studies, cancer clinical trials and cancer prevention trials. This program is at the Harvard TH Chan School of Public Health and Dana-Farber Cancer Institute (DFCI).

Role: Postdoctoral Fellow

5 UL1 TR001111-05 (Buse) 9/26/2013 4/30/18

National Center for Advancing Translational Sciences

North Carolina Translational & Clinical Sciences Institute (NC TraCS) - Biostatistics Services

A national consortium of medical research institutions, funded through Clinical and Translational Science Awards, is working together and shares a common vision: to improve the way biomedical research is conducted across the country, reduce the time it takes for laboratory discoveries to become treatments for patients, engage communities in clinical research efforts, and train the next generation of clinical and translational researchers. Role: Biostatistician

OVERLAP: NONE