

Keegan D. Korthauer

CONTACT	CLSB 11007 450 Brookline Ave Boston, MA 02115, USA phone: 651-274-1132	email: keegan@jimmy.harvard.edu website: kkorthauer.org github: kdkorthauer twitter: @keegankorthauer
---------	---	---

EDUCATION

- 2015 **Ph.D., Statistics**, University of Wisconsin, Madison, Wisconsin USA
Emphasis in Biostatistics
Advisor: Christina Kendzierski, Ph.D.
Dissertation: Bayesian hierarchical modeling of high-throughput genomic data with applications to cancer bioinformatics and stem cell differentiation
- 2009 **M.S., Biostatistics**, University of Minnesota, Minneapolis, Minnesota USA
Advisor: David Nelson, Ph.D.
Thesis: The impact of missing confounders on propensity stratification in observational studies
- 2007 **B.S., Biology**, University of Minnesota, Minneapolis, Minnesota USA
Summa cum laude

PROFESSIONAL POSITIONS

- 2015 - **Postdoctoral Research Fellow**, Boston, Massachusetts
Department of Biostatistics & Computational Biology, Dana-Farber Cancer Institute
Department of Biostatistics, Harvard T.H. Chan School of Public Health
Mentor: Rafael Irizarry, Ph.D.
Develop statistical & computational tools for high-throughput genomics data
- 2013 -2015 **Graduate Research Assistant**, Madison, Wisconsin
Department of Biostatistics and Medical Informatics, University of Wisconsin
Advisor: Christina Kendzierski, Ph.D.
Collaborate with clinical oncologists and developmental biologists to gain insight into biological processes
- 2010-2013 **Predoctoral Fellow, NIGMS Biostatistics Training Grant**, Madison, WI
Department of Biostatistics and Medical Informatics, University of Wisconsin
Advisors: Bret Larget, Ph.D., Christina Kendzierski, Ph.D., and Michael Newton, Ph.D.
Rotation projects:
- *Hidden Markov Modeling of Palindromic Sequences*
- *SNP-based Genomic Signature of Survival in Ovarian Cancer*
- *Analysis of Copy Number in Whole Genome Amplified Samples*
- 2009-2010 **Project Assistant**, Madison, Wisconsin
Institute for Clinical and Translational Research (ICTR), University of Wisconsin
Advisor: Mary Lindstrom, Ph.D.
Consult with ICTR investigators on design and analysis plans

- 2008-2009 **Statistics Intern**, Minneapolis, Minnesota
 Veterans Affairs (VA) Medical Center
Advisor: David Nelson, Ph.D.
Investigate the impact of missing confounders in causal inference
- 2007-2009 **Project Assistant**, Minneapolis, Minnesota
 Biostatistics Design and Analysis Center, University of Minnesota
Advisor: Cynthia Davey, M.S.
Consult with on-campus investigators on design and analysis plans

PUBLICATIONS

Pre-prints

- [1] **K. Korthauer**[†], P. K. Kimes[†], C. Duvallet[‡], A. Reyes[‡], A. Subramanian[‡], M. Teng, Chinmay Shukla, E. J. Alm, and S. C. Hicks. A practical guide to methods controlling false discovery rates. *bioRxiv*, <https://doi.org/10.1101/458786>, 2018.
- [2] **K. Korthauer**, R. A. Irizarry. Genome-wide repressive capacity of promoter DNA methylation is revealed through epigenomic manipulation. *bioRxiv*, <https://doi.org/10.1101/381145>, 2018.

Peer-Reviewed Journal Articles

- [1] D. Y. Takeda[†], S. Spisák[†], J.-H. Seo, C. Bell, E. O'Connor, **K. Korthauer**, D. Ribli, I. Csabai, N. Solymosi, Z. Szállási, P. Cejas, X. Qiu, H. Long, V. Tisza, P. V. Nuzzo, M. Rohanizadegan, M. M. Pomerantz, W. C. Hahn, M. L. Freedman. A somatically acquired enhancer of the androgen receptor is a noncoding driver in advanced prostate cancer. *Cell*, 174(2):422-432, 2018.
- [2] **K. Korthauer**, S. Chakraborty, Y. Benjamini, R. A. Irizarry. Detection and accurate False Discovery Rate control of differentially methylated regions from Whole Genome Bisulfite Sequencing. *Biostatistics*, kxy007, 2018.
- [3] C. J. Shukla, A. L. McCorkindale, C. Gerhardinger, **K. Korthauer**, M. N. Cabili, D. M. Shechner, R. A. Irizarry, P. G. Maass, J. L. Rinn. High-throughput identification of RNA nuclear enrichment sequences. *The EMBO Journal*, 37:e98452, 2018
- [4] J. Choi, S. Ye, K. Eng, **K. Korthauer**, W. H. Bradley, J. S. Rader and C. Kendzierski. IPI59: an actionable biomarker to improve treatment response in serous ovarian carcinoma patients. *Statistics in Biosciences*, 9(1):1-12, 2017.
- [5] **K. Korthauer**, L.-F. Chu, M. A. Newton, Y. Li, J. Thomson, R. Stewart and C. Kendzierski. A statistical approach for identifying differential distributions in single-cell RNA-seq experiments. *Genome Biology*, 17:222, 2016.
- [6] C. Bodelon, M. Horswill, A. K. Chaturvedi, N. Wentzensen, S. Vinokurova, **K. Korthauer**, S. T. Dunn, J. N. Sampson, M. Schiffman, M. A. Newton, J. den Boon, M. E. Sherman, P. Ahlquist, J. L. Walker, R. E. Zuna, S. S. Wang. Copy number alterations and HPV integration in cervical precancer and invasive cancer. *Carcinogenesis*, 37(2):188-196, 2016.
- [7] **K. Korthauer**, C. Kendzierski. MADGiC: a model-based approach for identifying driver genes in cancer. *Bioinformatics*, 31(10): 15261535, 2015.

- [8] Y. J. Sung, **K. Korthauer**, M. Swartz and C. Engelman. Methods for Collapsing Multiple Rare Variants in Whole Genome Sequencing Data. *Genetic Epidemiology*, 38(S1):S13-S20, 2014.
- [9] M. Wuthrich, K. Ersland, J. C. Pick-Jacobs, B. H. Gern, C. A. Frye, T. D. Sullivan, M. B. Brennan, H. I. Filutowicz, K. O'Brien, **K. Korthauer**, S. Schultz-Cherry, B. S. Klein. Limited model antigen expression by transgenic fungi induces disparate fates during differentiation of adoptively transferred T cell receptor transgenic CD4+ T cells: robust activation and proliferation with weak effector function during recall. *Infection and immunity*, 80(2): 787-797, 2012.
- [10] D. Z. Bliss, J. Lewis, **K. Hasselmann (now Korthauer)**, K. Savik, A. Lowry, R. Whitebird. Use and evaluation of disposable absorbent products for managing fecal incontinence by community-living people. *Journal of wound, ostomy, and continence nursing*, 38(3):289, 2011.
- [11] L. Wang, R. M. Mitra, **K. Hasselmann (now Korthauer)**, M. Sato, L. Lenarz-Wyatt, J. D. Cohen, F. Katagiri, J. Glazebrook. The genetic network controlling the Arabidopsis transcriptional response to Pseudomonas syringae pv. maculicola: roles of major regulators and the phytotoxin coronatine. *Molecular plant-microbe interactions*, 21(11): 1408-1420, 2008.

Book Chapters

- [1] **K. Korthauer**, J. Dawson and C. Kendzierski. Predicting cancer subtypes using survival-supervised latent Dirichlet allocation models. In *Advances in Statistical Bioinformatics: Models and Integrative Inference for High-Throughput Data*, K.-A. Do, Z. S. Qin and M. Vannucci (Eds). Cambridge University Press, 2013.

Conference Proceedings

- [1] J. Delgado, E. A. Jacobs, N. E. Adler, **K. Korthauer** and A. Fernandez. The relation between subjective socioeconomic status, depression and self-rated health in a population of African Americans, Hispanics and non-Hispanic whites with diabetes, *Journal of General Internal Medicine*, 28: S209, 2013.
- [2] J. Delgado, A. Fernandez, N. E. Adler, **K. Korthauer**, E. Jacobs. Subjective and objective socioeconomic status and control of hypertension and diabetes. *Journal of General Internal Medicine*, 27:S312-S312, 2012.

Ph.D. Dissertation

- [1] **K. Korthauer**. Bayesian hierarchical modeling of high-throughput genomic data with applications to cancer bioinformatics and stem cell differentiation, Ph.D. Thesis, University of Wisconsin, 2015.

†Denotes co-first authorship

‡Denotes equal contribution

HONORS & AWARDS

- | | |
|------|---|
| 2017 | Travel award for the Ascona Workshop 2017 on Statistical Challenges in Single-Cell Biology in Ascona, Switzerland |
| 2016 | Travel award from the Biometrics Section of the American Statistical Association to attend the Joint Statistical Meetings in Chicago, Illinois |
| 2015 | Poster award from the Regional Advisory Board of the International Biometric Society at the Eastern North American Region Meeting in Miami, Florida |
| 2014 | Travel grant from the Vilas Conference Presentation Trust |
| 2013 | Travel award from the Howard Hughes Medical Institute/NIGMS to attend the Jackson Laboratory short course on Systems Genetics in Bar Harbor, Maine |
| 2012 | Travel award for the Genetic Analysis Workshop in Stevenson, Washington |
| 2008 | Outstanding Teaching Assistant Award at the University of Minnesota |
| 2007 | Phi Beta Kappa Society |

FELLOWSHIPS & SCHOLAR- SHIPS

- | | |
|-----------|---|
| 2010-2013 | NIH/NIGMS Predoctoral Training Grant in Biostatistics, University of Wisconsin |
| 2006 | Undergraduate Research Opportunities Program project grant, University of Minnesota |
| 2003-2007 | National Merit James E. Casey Scholarship (\$24,000 USD) |

PRESENTATIONS

Invited Talks

- [1] Accurate inference of DNA methylation data: statistical challenges lead to biological insights. *Program in Quantitative Genomics Working Group Series, Harvard T.H. Chan School of Public Health*. 9 April 2019. Boston, Massachusetts.
- [2] Accurate inference of DNA methylation data: statistical challenges lead to biological insights. *Dana-Farber Cancer Institute Center for Functional Cancer Epigenetics Seminar*. 15 February 2019. Boston, Massachusetts.
- [3] Practical recommendations for controlling false discoveries in computational biology. *European Bioconductor Meeting*. 6-7 December 2018. Munich, Germany.
- [4] Detection and inference of differentially methylated regions from bisulfite sequencing. *International Conference on Advances in Interdisciplinary Statistics and Combinatorics*. 5-7 October 2018. Greensboro, NC.
- [5] Detection and inference of differentially methylated regions from bisulfite sequencing. *Bioconductor Conference*. 26-28 July 2017. Boston, MA, USA.
- [6] Exploiting heterogeneity in single-cell transcriptomic analyses: how to move beyond comparisons of averages. *Festival of Genomics California*. 19-21 September 2016. San Diego, California.

- [7] A statistical approach for identifying differential distributions in single-cell RNA-seq. *iB-RIGHT*. 1-3 November 2015. M. D. Anderson Cancer Center, Houston, Texas.

Selected Contributed Talks

- [1] *De novo* detection and accurate inference of differentially methylated regions. *Joint Statistical Meetings*. 29 July - 2 August 2018. Vancouver, British Columbia, Canada.
- [2] scDD: A Statistical Approach for Identifying Differential Distributions in Single-Cell RNA-Seq Experiments. *Joint Statistical Meetings*. 30 July - 4 August 2016. Chicago, Illinois.
- [3] Identifying driver genes from somatic mutations: an integrative model-based approach. *International Biometric Society Eastern North American Region (ENAR) Annual Meeting*. 16-19 March 2014. Baltimore Maryland.

Selected Poster Presentations

- [1] **K. Korthauer**, S. Chakraborty, Y. Benjamini, R.A. Irizarry. Detection and accurate False Discovery Rate control of differentially methylated regions from Whole Genome Bisulfite Sequencing. *ENCODE Consortium Meeting*. 3-7 February 2018. Palo Alto, California.
- [2] **K. Korthauer**. Exploiting heterogeneity in single-cell transcriptomic analyses: how to move past comparisons of averages. *Ascona Workshop on Statistical Challenges in Single-Cell Biology*. 30 April - 5 May 2017. Ascona, Switzerland.
- [3] **K. Korthauer**, R.A. Irizarry. Assessing Statistical Significance of Differentially Methylated Regions in Whole-Genome Bisulfite Sequencing Experiments. *ENCODE Consortium Meeting*. 15-16 June 2016. La Jolla, California.
- [4] **K. Korthauer**, C.K. Kendzierski. Differential Dynamics in Single-Cell RNA-Seq Experiments. *International Biometric Society Eastern North American Region (ENAR) Spring Meeting*. 15-18 March 2015. Miami, Florida.
- [5] **K. Korthauer**, C.K. Kendzierski. An integrative approach for the identification of somatic mutations that drive cancer. *Genomic Sciences Training Program Retreat*. 14 June 2013. Madison, Wisconsin.

EDITORIAL ACTIVITIES

Journal Referee

Annals of Applied Statistics, Bioinformatics, Biometrics, Biostatistics, Cell Systems, Genome Research, GigaScience, Nucleic Acids Research, Science

PROFESSIONAL SOCIETIES

2012-present Member, American Statistical Association

SOFTWARE

Developer

- [1] **dmrseq**: An R package for inference for differentially methylated regions (DMRs) from bisulfite sequencing, available on Bioconductor.

- [2] **scDD**: An R package for the identification of differentially distributed genes in single-cell RNA-seq, available on [Bioconductor](#)
- [3] **MADGiC**: An R package for the identification of cancer driver genes by integrating somatic mutation, expression, replication timing, and functional impact, available on [GitHub](#)

Contributor

- [1] **oligoGames**: An R package for the analysis of tiled massively parallel reporter assays (MPRAs), available on [GitHub](#).

EDUCATIONAL ACTIVITIES

Masters of Biostatistics (MS) Advisees

2018 Eunice Ye, Biostatistics
Academic co-advisor

TEACHING

Classroom Instruction

2017 **Teaching Assistant**, [Introduction to Data Science \(BST 260\)](#).
Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA
Formulated problem sets, solutions, and exam questions. Supervised final projects.

2014 **Instructor**, Applied Introductory Statistics for Engineers (STAT 324).
Department of Statistics, University of Wisconsin, Madison, Wisconsin
Curated course material and lectured to 120 students. Designed problem sets and exams. Supervised Teaching Assistants.

2013 **Tutor**, Introduction to Statistical Methods (STAT 301).
Greater University Tutoring Service, University of Wisconsin, Madison, Wisconsin
Designed and led weekly review sessions in a small group setting.

2007-2008 **Teaching Assistant**, Introduction to Biostatistical Methods I (PubH 6414).
Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota
Led weekly lab sections, held office hours, and graded homework and exams.

Shortcourses & Workshops

2019 **Associate Instructor**, Short Course: Statistical Methods for Functional Genomics.
28 June-11 July 2019. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

2017 **Co-organizer and instructor** for the [Healthcare Innovation Replicathon](#).
24-25 March 2017. University of Puerto Rico, San Juan, Puerto Rico.
Generated course material for a two-day hackathon-style event focusing on reproducibility and replicability in biomedical data science. Led interdisciplinary teams from basic data science skills through reproducible analyses and presentations using R and GitHub.

- 2016-2017 **Teaching Assistant**, Short Course: Statistical Methods for Functional Genomics. 23 June-6 July 2016 and 30 June-13 July 2017. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
Led students through hands-on laboratory sessions.
- 2016 **Co-organizer and instructor** of Workshop on [Dismantling the bulk: examining neuronal heterogeneity using single-cell techniques](#). 19 September 2016. Festival of Genomics California, San Diego, California.
Organized and led a one-day workshop for non-computational biologists on the analysis of single-cell RNA-sequencing data.