Toby Dylan Hocking

Curriculum Vitae

CONTACT
INFORMATION

Mail and Courier:
7260 Boyer #2
Montreal, QC
H2R 2R7 Canada

Citizenship: USA
Phone: (514) 398-3311 x 00721
E-mail: toby.hocking@mail.mcgill.ca
Webpage: http://tdhock.github.io

RESEARCH
INTERESTS

Supervised and unsupervised machine learning models for large datasets, based on convex and discrete optimization (regression, classification, ranking, clustering, changepoint detection, survival analysis). Applications to genomics, neuroscience, audio, internet, sensors, recommendation systems.

EDUCATION AND
PROFESSIONAL
EXPERIENCE

McGill University, Montreal, Canada (2014-2017).

Postdoc with Guillaume Bourque, Department of Human Genetics.
"Changepoint detection and regression models for peak detection in genomic data."

Tokyo Institute of Technology, Tokyo, Japan (2013).

Postdoc with Masashi Sugiyama, Department of Computer Science.
"Support vector machines for ranking and comparing."

École Normale Supérieure, Cachan, France (2009-2012).

Ph.D. with Francis Bach, Département d'Informatique; Jean-Philippe Vert, Institut Curie.
"Learning algorithms and statistical software, with applications to bioinformatics."

Université Paris 6, Paris, France (2008-2009).

Master of Statistics, internship at INRA with Mathieu Gautier and Jean-Louis Foulley.
"A Bayesian Outlier Criterion to Detect SNPs under Selection in Large Data Sets."

Sangamo BioSciences, Richmond, CA, USA (2006-2008).

Research Assistant with Jeff Miller in the Technology group.
"A web app for visualization and statistical analysis of experimental data."

University of California, Berkeley, CA, USA (2002-2006).

Double B.A. in Statistics, Molecular and Cell Biology; thesis in Statistics with Terry Speed.
"Chromosomal copy number analysis using SNP microarrays and a binomial test statistic."

REFEREED
PUBLICATIONS

In addition to peer-reviewed journals, I publish papers at machine learning conferences such as ICML and NIPS, which have double-blind peer reviews, and only accept about 20% of submitted papers.

Hocking TD, Goerner-Potvin P, Morin A, Shao X, Pastinen T, Bourque G. Optimizing ChIP-seq peak detectors using visual labels and supervised machine learning. *Bioinformatics* (2017) 33 (4): 491-499.

Shimada K, Shimada S, Sugimoto K, Nakatochi M, Suguro M, Hirakawa A, **Hocking TD**, Takeuchi I, Tokunaga T, Takagi Y, Sakamoto A, Aoki T, Naoe T, Nakamura S, Hayakawa F, Seto M, Tomita A, Kiyoi H. Development and analysis of patient-derived xenograft mouse models in intravascular large B-cell lymphoma. *Leukemia* 2016.

Chicard M, Boyault S, Colmet-Daage L, Richer W, Gentien D, Pierron G, Lapouble E, Bellini A, Clement N, Iacono I, Brjon S, Carrere M, Reyes C, **Hocking TD**, Bernard V, Peuchmaur M, Corradini N, Faure-Conter C, Coze C, Plantaz D, Defachelles A-S, Thebaud E, Gambart M, Millot

F, Valteau-Couanet D, Michon J, Puisieux A, Delattre O, Combaret V, Schleiermacher G. Genomic copy number profiling using circulating free tumor DNA highlights heterogeneity in neuroblastoma. *Clinical Cancer Research* 2016.

Maidstone R, **Hocking TD**, Rigaill G, Fearnhead P. On optimal multiple changepoint algorithms for large data. *Statistics and Computing* (2016). doi:10.1007/s11222-016-9636-3

Hocking TD, Rigaill G, Bourque G. PeakSeg: constrained optimal segmentation and supervised penalty learning for peak detection in count data. *International Conference on Machine Learning (ICML)*, 2015.

Suguro M, Yoshida N, Umino A, Kato H, Tagawa H, Nakagawa M, Fukuhara N, Karnan S, Takeuchi I, **Hocking TD**, Arita K, Karube K, Tsuzuki S, Nakamura S, Kinoshita T, Seto M. Clonal heterogeneity of lymphoid malignancies correlates with poor prognosis. *Cancer Sci.* 2014 Jul;105(7):897-904.

Hocking TD, Boeva V, Rigaill G, Schleiermacher G, Janoueix-Lerosey I, Delattre O, Richer W, Bourdeaut F, Suguro M, Seto M, Bach F, Vert J-P. SegAnnDB: interactive Web-based genomic segmentation. *Bioinformatics* (2014) 30 (11): 1539-1546. DOI:10.1093/bioinformatics/btu072

Hocking TD, Wutzler T, Ponting K and Grosjean P. Sustainable, extensible documentation generation using inlinedocs. *Journal of Statistical Software* (2013), 54(6), 1-20. DOI:10.18637/jss.v054.i06

Hocking TD, Schleiermacher G, Janoueix-Lerosey I, Boeva V, Cappo J, Delattre O, Bach F, Vert J-P. Learning smoothing models of copy number profiles using breakpoint annotations. *BMC Bioinfo.* 2013, 14:164. DOI:10.1186/1471-2105-14-164

Hocking TD, Rigaill G, Bach F, Vert J-P. Learning sparse penalties for change-point detection using max-margin interval regression. *International Conference on Machine Learning (ICML)*, 2013.

Hocking TD, Joulin A, Bach F, Vert J-P. Clusterpath: an Algorithm for Clustering using Convex Fusion Penalties. *International Conference on Machine Learning (ICML)*, 2011.

Gautier M, **Hocking TD**, Foulley JL. A Bayesian outlier criterion to detect SNPs under selection in large data sets. *PloS ONE* 5 (8), e11913 (2010).

Doyon Y, McCammon JM, Miller JC, Faraji F, Ngo C, Katibah GE, Amora R, **Hocking TD**, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Amacher SL. Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nature biotechnology* 26 (6), 702-70 (2008).

<div style="display:flex">
<div style="width:20%">PAPERS IN PROGRESS</div>
<div>

Hocking TD, Rigaill G, Fearnhead P, Bourque G. A log-linear segmentation algorithm for peak detection in genomic data. Preprint arXiv:1703.03352. Under review at *Annals of Applied Statistics*.

Drouin A, **Hocking TD**, Laviolette F. Max margin interval trees. Under review at *Neural Information Processing Systems (NIPS)*.

Venuto D, Spanurattana S, Sugiyama M, **Hocking TD**. Support vector comparison machines. Preprint arXiv:1401.8008.

Sievert C, Cai J, VanderPlas S, Khan F, **Hocking TD**. Extending ggplot2's grammar of graphics implementation for linked and dynamic graphics on the web.

Hocking TD, Khare A. Learning penalty functions for changepoint detection using elastic-net regularized accelerated failure time models.

</div>
</div>

Narahara M, **Hocking TD**, Bourque G, Yamada R, Setoh K, Matsuda F, Lathrop M. Transcriptomic analysis of antibody responses to seasonal influenza vaccine reveals predictive gene signatures and potential key transcription factors.

Alirezaie N, Majewski J, **Hocking TD**. A supervised machine learning method for predicting pathogenicity of genetic variants.

LIGHTLY- AND NON-REFEREED PUBLICATIONS

Hocking TD. A breakpoint detection error function for segmentation model selection and validation. Preprint arXiv:1509.00368.

Hocking TD, Bourque G. PeakSegJoint: fast supervised peak detection via joint segmentation of multiple count data samples. Preprint arXiv:1506.01286.

Hocking TD, Rigaill G. SegAnnot: an R package for fast segmentation of annotated piecewise constant signals, Preprint hal-00759129.

CONFERENCE TUTORIALS

Hocking TD, Killick R. Introduction to optimal changepoint detection algorithms, *useR* 2017.

Hocking TD, Ekstrøm CT. Understanding and creating interactive graphics, *useR* 2016.

INVITED TALKS (SELECTED)

Université Laval, Centre for Big Data Research, 2016.

McGill Barbados epigenomics workshop, 2015.

Joint statistical meeting (JSM) 2015 (declined; my collaborator Susan VanderPlas gave the talk instead).

Workshop on Machine Learning and Applications to Biology, MLAB Sapporo 2013.

Google Research, New York, 2012.

Université Rennes, 2012.

Université Angers, 2012.

INRIA (French computer science research institute), Lille, 2012.

Institut de Biologie de Lille 2011.

HONORS AND AWARDS (SELECTED)

"Mobilité entrant" travel award to work with Guillem Rigaill in Université Evry, France, 2016.

International useR conference, Best Student Poster Award, "Adding direct labels to plots," 2011.

INRIA/INRA (French computer science and agricultural research institutes), Ph.D. scholarship, 2009 (declined).

UC Berkeley, department of Statistics VIGRE research scholarship, 2001.

UC Berkeley, Cal Band George Miller scholarship, 2000.

PROFESSIONAL SERVICE

President of organizing committee for "R in Montreal 2018" conference.

Co-administrator and mentor for R project in Google Summer of Code.

Reviewer: International Conference on Machine Learning (ICML), Advances in Neural Infromation Processing Systems (NIPS), Journal of Machine Learning Research (JMLR), Artificial Intelligence Review, Journal of Computational and Graphical Statistics (JCGS), R Journal, Bioinformatics, PLOS Computational Biology, BMC Bioinformatics, IEEE Transactions on Pattern Analysis and Machine Intelligence.

Software Online (Selected)

R: clusterpath, directlabels, animint, plotly, sublogo, inlinedocs, quadmod, bams, neuroblastoma, breakpointError, SegAnnot, rankSVMcompare, gganim, animint, requireGitHub, WeightedROC, revector, PeakError, PeakSegDP, PeakSegJoint, PeakSegOptimal, memtime, namedCapture, fpop, penaltyLearning, mmit, iregnet.

Python: str.extractall in pandas, annotate_regions, SegAnnDB.