

**BIOMEDICAL DATA SCIENCE
PRESENTS:
BIODS 260
11/17/22 1:30PM-2:50PM
MSOB X303 (SEE ZOOM DETAILS BELOW)**

Lorin Crawford
Principal Researcher, Microsoft Research
New England; Associate Professor of
Biostatistics, Brown University
<http://lorincrawford.com/>

TITLE:

Machine Learning for Human Genetics: A Multi-Scale View on
Complex Traits and Disease

ABSTRACT:

A common goal in genome-wide association (GWA) studies is to characterize the relationship between genotypic and phenotypic variation. Linear models are widely used tools in GWA analyses, in part, because they provide significance measures which detail how individual single nucleotide polymorphisms (SNPs) are statistically associated with a trait or disease of interest. However, traditional linear regression largely ignores non-additive genetic variation, and the univariate SNP-level mapping approach has been shown to be underpowered and challenging to interpret for certain trait architectures. While machine learning (ML) methods such as neural networks are well known to account for complex data structures, these same algorithms have also been criticized as “black box” since they do not naturally carry out statistical hypothesis testing like classic linear models. This limitation has prevented ML approaches from being used for association mapping tasks in GWA applications. In this talk, we present flexible and scalable classes of Bayesian feedforward models which provide interpretable probabilistic summaries such as posterior inclusion probabilities and credible sets which allows researchers to simultaneously perform (i) fine-mapping with SNPs and (ii) enrichment analyses with SNP-sets on complex traits. While analyzing real data assayed in diverse self-identified human ancestries from the UK Biobank, the Biobank Japan, and the PAGE consortium we demonstrate that interpretable ML has the power to increase the return on investment in multi-ancestry biobanks. Furthermore, we highlight that by prioritizing biological mechanism we can identify associations that are robust across ancestries---suggesting that ML can play a key role in making personalized medicine a reality for all.

SUGGESTED READINGS:

[A.R. Martin, M. Kanai, Y. Kamatani, Y. Okada, B.M. Neale, and M.J. Daly \(2019\). Clinical use of current polygenic risk scores may exacerbate health disparities. Nature Genetics. 51: 584-591.](#)

[S.P. Smith, S. Shahamatdar, W. Cheng, S. Zhang, J. Paik, M. Graff, C. Haiman, T.C. Matise, K.E. North, U. Peters, E. Kenny, C. Gignoux, G. Wojcik, L. Crawford, and S. Ramachandran \(2022\). Enrichment analyses identify shared associations for 25 quantitative traits in over 600,000 individuals from seven diverse ancestries. American Journal of Human Genetics. 109: 871-884.](#)

[P. Demetci, W. Cheng, G. Darnell, X. Zhou, S. Ramachandran, and L. Crawford \(2021\). Multi-scale inference of genetic architecture using biologically annotated neural networks. PLOS Genetics. 17\(8\): e1009754.](#)

Zoom link:

<https://stanford.zoom.us/j/92874055477pwd=aThzNmpmNEQ1L2FjV0E5ZXF5SDR1UT09&from=addon>
Password: 705300