Dear iScience Editorial Advisory Board,

On behalf of myself and my collaborators, it is my pleasure to submit this manuscript, *Consenus Machine Learning for Gene Target Selection in Pediatric AML Risk* for consideration to be published in iScience. Please note, we published a preprint of this work to bioarxiv at [this link.](https://www.biorxiv.org/content/10.1101/632166v1)

Pediatric acute myeloid leukemia (AML) is a deadly cancer presenting with heterogeneous molecular characteristics. This heterogeneity confounds clinical risk stratification as well as single-assay analysis, such as with RNA-sequencing data. Modern machine learning algorithms can identify important genes predicting disease risk from expression data, but confidence and interpretability of these results are largely dependent on the assumptions, strengths, and weaknesses of each respective statistical algorithm used.

To our knowledge, the present manuscript describes the first consensus machine learning method to identify important gene targets from expression data to predict clinical risk of pediatric AML. We described a consensus machine learning method leveraging feature importance estimates across a variety of algorithms. Using ablation tests, we demonstrated a gene exclusion bias for certain penalized algorithms, and we showed this bias is mitigated by our method. Our findings validated prior findings implicating *HOX* pathway genes and cofactors in AML risk. This study furthers our understanding of how machine learning can aid researchers in identifying the most important disease risk genes from genomics data. Finally, the described methods and supporting information set the stage to further fine-tune and improve consensus machine learning methods for applications beyond biomedicine.

For reviewers of this work, we recommend biostatistics and biomedical research background, including familiarity with the algorithms we used (lasso, Random Forest, XGBoost, and SVM). Experience studying genomics of pediatric cancers, especially leukemia, will also be helpful.

We greatly appreciate your time and consideration, and look forward to hearing from you.

Sincerely,

Sean Maden and the ConsensusML team

Correspondence:

Sean Maden (submitter, co-first author)

Ph.D. candidate,

Computational Biology program

Oregon Health & Science University

[maden@ohsu.edu](mailto:maden@ohsu.edu)

Ben Busby, Ph.D. (corresponding author)

Genomics Outreach Coordinator

National Library of Medicine

National Center for Biotechnology Information

ben.busby@gmail.com