Dear iScience Editorial Advisory Board members,

On behalf of myself and my fellow co-authors, it is my pleasure to submit the manuscript "Consenus Machine Learning for Gene Target Selection in Pediatric AML" for publication in iScience.

Pediatric acute myeloid leukemia (AML) is a deadly cancer with hetergeneous molecular presentation in the clinic. Analysis of pediatric AML RNA-seq data is challenging due to this heterogeneity and the high dimensionality of gene expression data. Machine learning can identify gene targets of disease risk from RNA-seq data, but results are dependent on properties of the respective statistical algorithm(s) used and drawing consensus from disparate algorithms remains challenging.

In the present manuscript we present, to our knowledge, the first consensus machine learning method for RNA-seq data to identify gene targets of importance for pediatric AML risk. Our consensus gene set validates prior findings implicating activity of several *HOX* pathway genes in pediatric AML risk. Our gene target set aggregates genes implicated across ablation experiment iterations with two penalized algorithms, which indicates our method can mitigate possible exclusion bias for such algorithms. This study furthers understanding of important disease risk genes for pediatric AML, and the described consensus machine learning method can be readily applied to other heterogeneous diseases. Outside of clinical biomedicine, our methods set the stage for further formalization and generalization for consensus feature selection with machine learning for high-dimensional datasets.

For review of this work, we recommend some background in biostatistics, including familiarity with the four algorithms we used (lasso, Random Forest, XGBoost, and SVM). Familiarity and experience studying pediatric cancers, especially leukemias, would also be helpful.

Thanks very much for your time, we appreciate it. My coauthors and I look forward to hearing from you.