Machine learning approach identifies targeting strategy for selective AML dependency

SOHC SWISS ONCOLOGY & HEMATOLOGY CONGRESS

Experimental Haematology / Oncology

Enea Provenzano[1], Monika Burocziova[1], Artemis Baumann[1], Peppi Suominen[1], Tobias R. Eeksman[2], Elias Krainer[2], Jana M. Ellegast[1]

[1]University and University Hospital Zurich, Faculty of Medicine, Department of Medical Oncology and Haematology, Zurich. [2]ETH Zurich, Department of Mathematics, Zurich

University of Zurich[UZH]

USZ Universitäts Spital Zürich

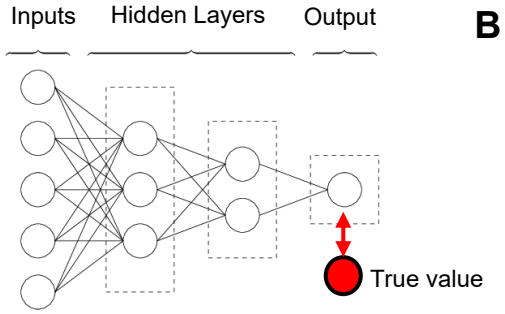
Background and objective

Acute myeloid leukaemia (AML) remains difficult to treat, with limited targeted therapies and poor outcomes. We previously identified Interferon Regulatory Factor 2 Binding Protein 2 (IRF2BP2) as selective AML dependency, demonstrating that its loss induces cell death in AML patient cells, while sparing colony-forming capacity in healthy donor-derived CD34$^+$ bone marrow cells, thus suggesting a therapeutic window for targeting *IRF2BP2*. Still, no *IRF2BP2*-specific molecules are currently available for clinical testing. To address this limitation, we developed a computational pipeline integrating machine learning with transcriptomic data to identify druggable regulators of IRF2BP2.

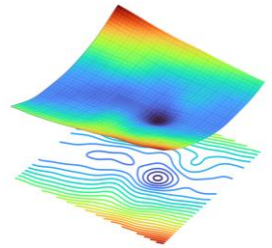
Methods

Over 90 neural network (NN) architectures were trained on publicly available AML bulk RNA-sequencing data to predict *IRF2BP2* expression. Each NN was fed 13'683 protein-coding gene transcript per million (TPM) vectors, split into training, validation, and test sets, and processed through multiple hidden layers (see schematic **Figure A**). The supervised feed-forward models minimised mean-squared error (MSE) between predicted and true IRF2BP2 values through iterative weight updates to minimise the loss function (see schematic **Figure B**). Weight adjustment was performed on the training set, while the validation set was used to monitor overfitting and assess generalisation (**Figure C**). Model performance was evaluated on the test set using the coefficient of determination (R^2) and predicted versus true values were plotted to assess accuracy (**Figure D**). The seven best-performing models were selected for further analysis (**Figure E**). To identify genes most strongly influencing *IRF2BP2* prediction, the gradient × input-TPM attribution was computed for every gene to quantify its positive or negative effect on the prediction. Attribution scores were aggregated across the seven models to produce a ranked list of genes exerting consistent regulatory influence on *IRF2BP2* expression. In parallel, a closed-form linear regression model was trained on the same dataset to rank genes by regression weight.

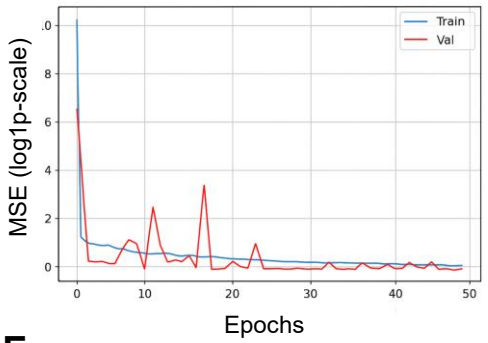
A



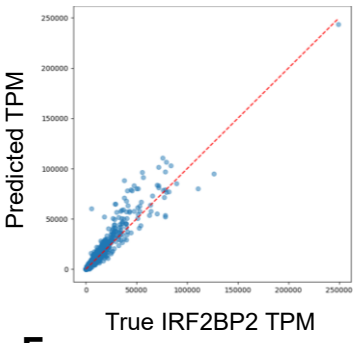
B



C



D

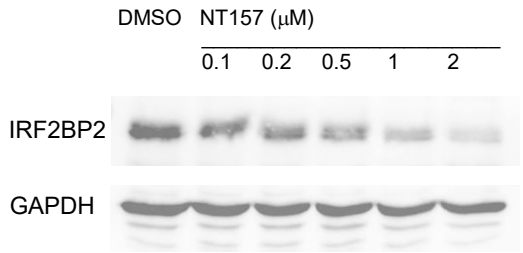


E

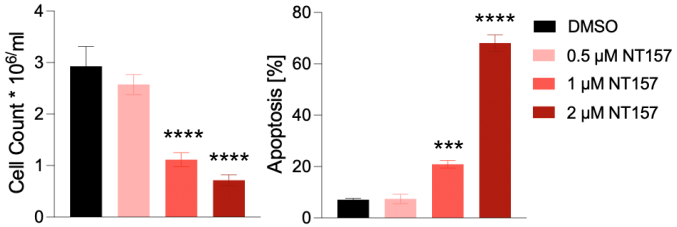
Model Name	Architecture Details	R^2 (Test Set)
h64_d0.1_bn	1 Hidden Layer, 64 nodes, 10% dropout	0.9504
h128_d0.05_bn	1 Hidden Layer, 128 nodes, 5% dropout	0.9543
h128_d0.3_bn	1 Hidden Layer, 128 nodes, 30% dropout	0.9526
h128_d0.35_bn	1 Hidden Layer, 128 nodes, 35% dropout	0.9511
h128_d0.5_bn	1 Hidden Layer, 128 nodes, 50% dropout	0.9514
h512_d0.1_bn	1 Hidden Layer, 512 nodes, 10% dropout	0.9528
h256_d0.1_bn	1 Hidden Layer, 256 nodes, 10% dropout	0.9506

F

G



H



Results

The overlap between the NN and linear model-derived genes yielded a curated set of four positive *IRF2BP2* regulators (**Figure F**).

Insulin Receptor Substrate 2 (IRS2) stood out as the best druggable predicted positive regulator of *IRF2BP2*. IRS2 functions as an adaptor protein in the insulin signalling pathway and can be targeted by the small molecule NT157.

Downregulation of IRS2 in MOLM13 cells using NT157 led to reduced *IRF2BP2* expression, as shown in the left Western blot (**Figure G**), and reproduced the phenotype of direct *IRF2BP2* perturbation with reduced cell viability due to increased apoptosis (**Figure H**).

Conclusion

Our study presents a machine learning framework to identify transcriptional regulators of cancer dependencies. Experimental validation of one predicted regulator underlines the biological relevance of our strategy. This approach may facilitate the discovery of therapeutic entry points for biologically hard-to-characterize or otherwise, intractable targets across diverse biological contexts .

Presented at SOHC 2025 from 19 – 21 November 2025