A Multimodal Deep Learning Framework for Deciphering Epigenetic Regulatory Grammar and Dysregulation in Acute Myeloid Leukemia

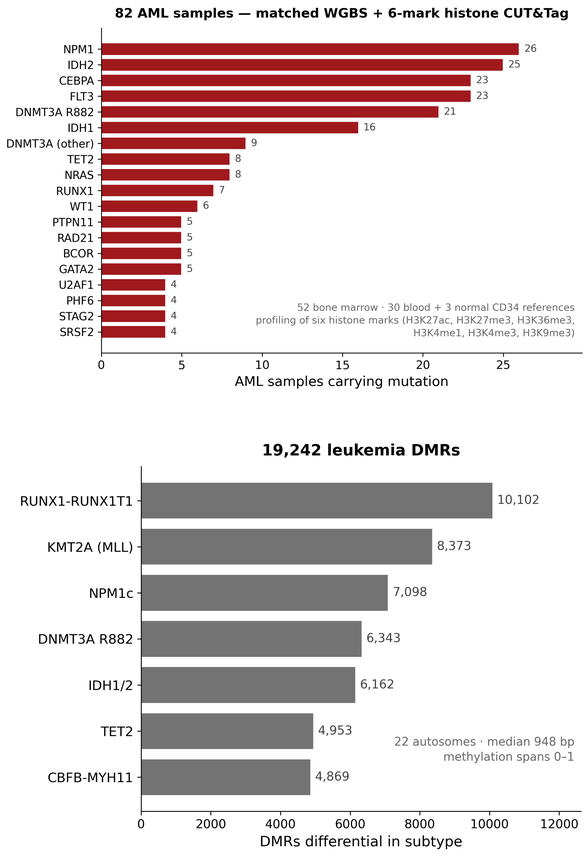
Andrew Bonney[1], Phuc Nguyen[1], Heidi Struthers[1], Emilee Kotnik[1], Mike Meers[2], Haley Abel[2], David H. Spencer[1,2,3],

[1]Division of Oncology, Department of Internal Medicine, [2]Department of Pathology and Immunology, and [3]The McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO

Objective

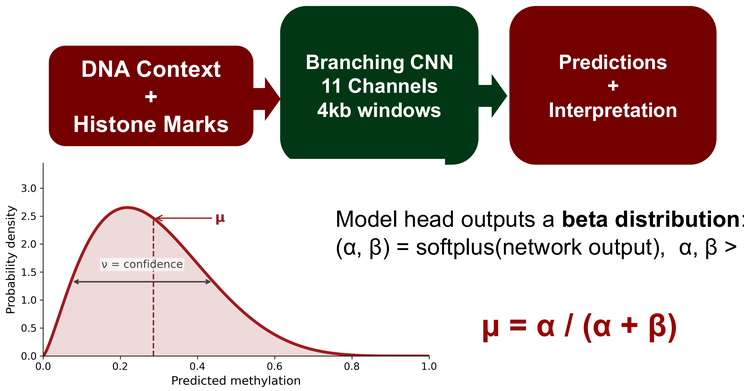
Acute Myeloid Leukemia (AML) is frequently driven by widespread epigenetic dysregulation, making it a critical area of study for understanding cancer progression. DNA methylation is significantly altered in AML, but the genomic features associated with these changes are poorly understood. To address this gap, this study aims to predict DNA methylation from local DNA sequence and histone marks using a Convolutional Neural Network (CNN) at Differentially Methylated Regions (DMRs) recurrent in a disease state, then utilize interpretable attributions to learn what drives methylation and how Acute Myeloid Leukemia (AML) rewires it.

Epigenomic Dataset



Model training sets generated from DMRs recurrent across several AML subtypes.

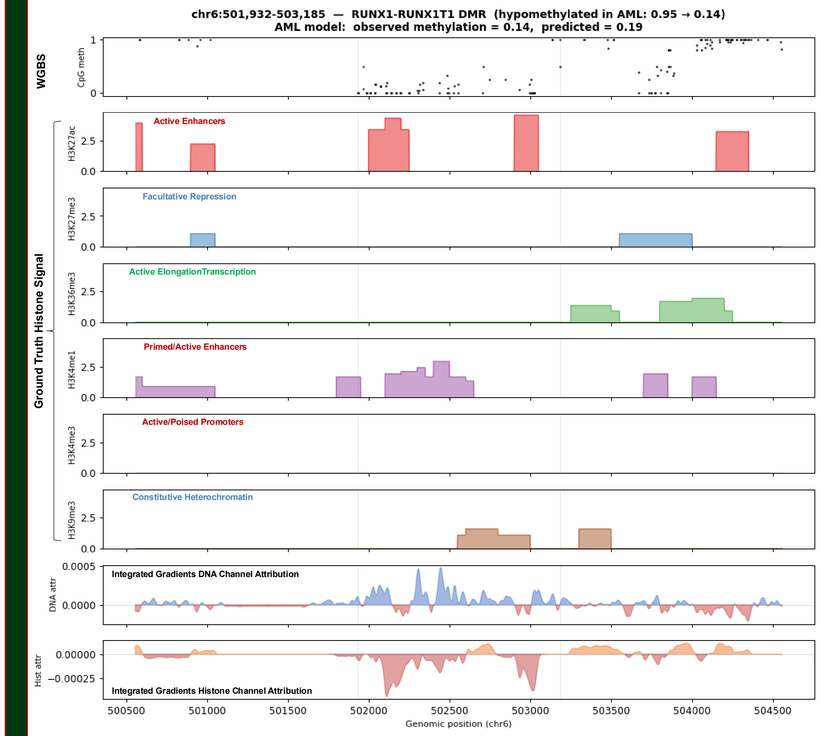
Data Pipeline



Model head outputs a **beta distribution**: (α, β) = softplus(network output), α, β > 0

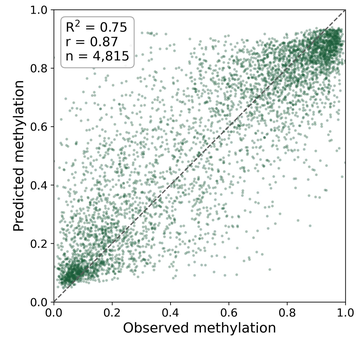
$$\mu = \alpha / (\alpha + \beta)$$

Individual DMR Attribution

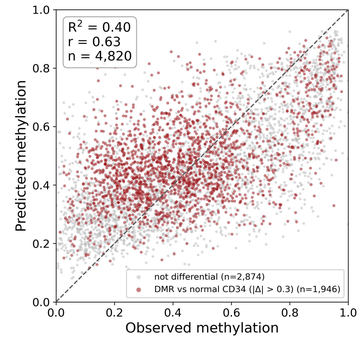


Per-base attribution at a RUNX1-RUNX1T1 DMR localizing the DNA and histone features driving its loss of methylation in AML.

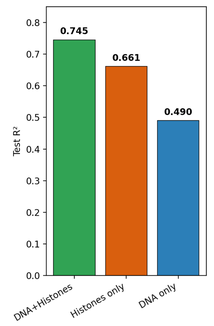
Model Architecture Tested On Normal and AML



Accurate methylation prediction from sequence and chromatin in normal CD34 progenitors.

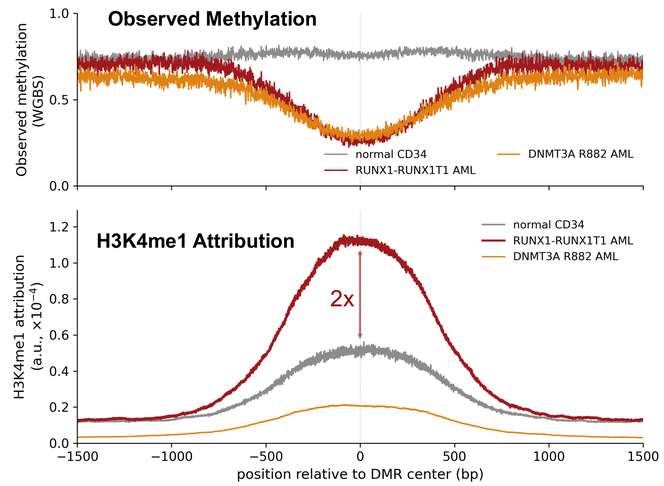


The same architecture, trained and tested in AML, predicts RUNX1 associated DMR methylation.



Modality ablation control shows that the model performs better with DNA context and Histone together.

H3K4 monomethylation predicts hypomethylation in an AML sample with RUNX1::RUNX1T1



Aggregate model attribution for H3K4me1 at hypomethylated DMRs, comparing models trained on AML samples carrying RUNX1::RUNX1T1 or DNMT3A R882 against normal CD34.

Top: observed WGBS methylation confirms these regions lose methylation in AML while staying methylated in CD34.

Bottom: magnitude of H3K4me1 attribution across the same windows demonstrating that the model pays more attention to this histone mark in RUNX1::RUNX1T1 compared to DNMT3A R882

This indicates H3K4me1 is a more important factor in the determination of local methylation fraction in this subtype compared to normal and compared to other subtypes with differing mechanisms of action.

Summary

- **A multimodal Beta-NLL CNN predicts DNA methylation more accurately than sequence or histone marks alone.**
- **Model attributions provide information about the importance of genomic features for methylation predictions at individual DMRs.**
- **Future directions include comparing chromatin states and combinatorial cross-talk between channels as well as the possibility of replacing one-hot encoded DNA with high-dimensional foundation model embeddings.**