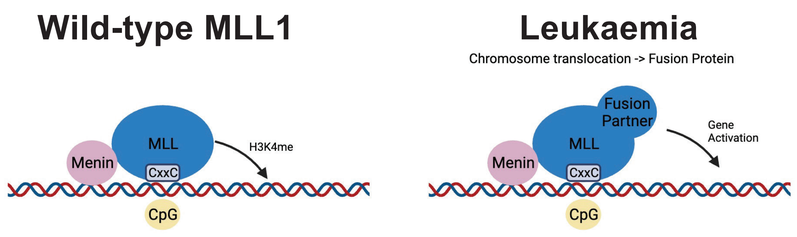
Identifying the DNA binding specificity of chromatin complexes in leukaemia

Catherine Chahrour, Alastair L. Smith, Thomas A. Milne
MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford

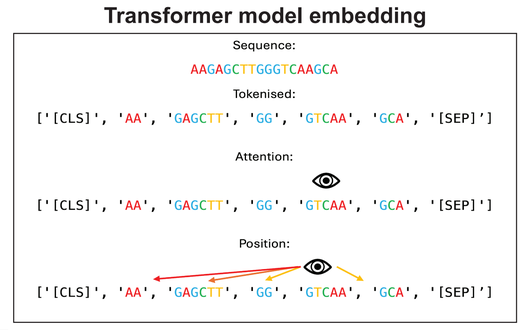
1. Introduction

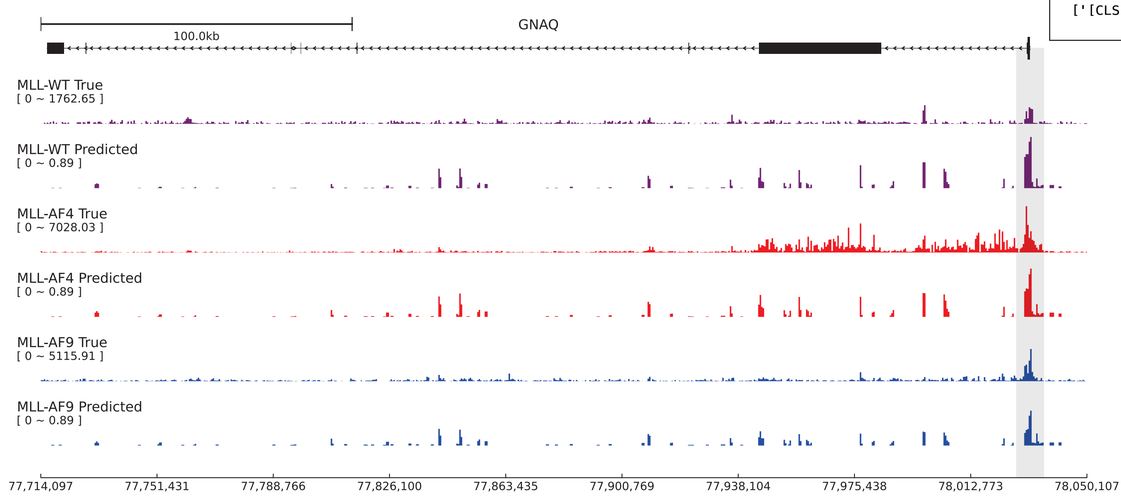


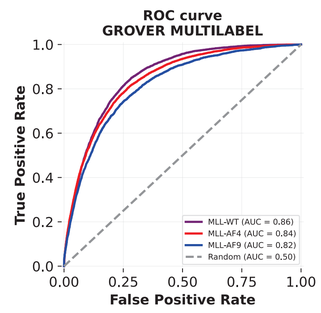
Wild-type MLL1 (KMT2A) and the oncofusion proteins MLL-AF4 and MLL-AF9 share an N-terminus that contains a CxxC domain for binding unmethylated CpG dinucleotides essential for the recruitment of MLL-WT and MLL-FPs to specific gene promoters, however, not every CpG island is bound by MLL-AF4. Moreover, MLL-AF4 also binds outside of uCpGs[1].

2. MLL Binding can be predicted from sequence

To establish if DNA sequence is important for MLL-WT or MLL-FP binding, a binary multi-label classification transformer model[2] was fine-tuned with MLL1 CUT&TAG peaks in MLL-WT (RCH-ACV), MLL-AF4 (SEM), or MLL-AF9 (THP-1) cells. MLL1 binding was predicted with high accuracy (ROC AUC > 0.82) by the model.

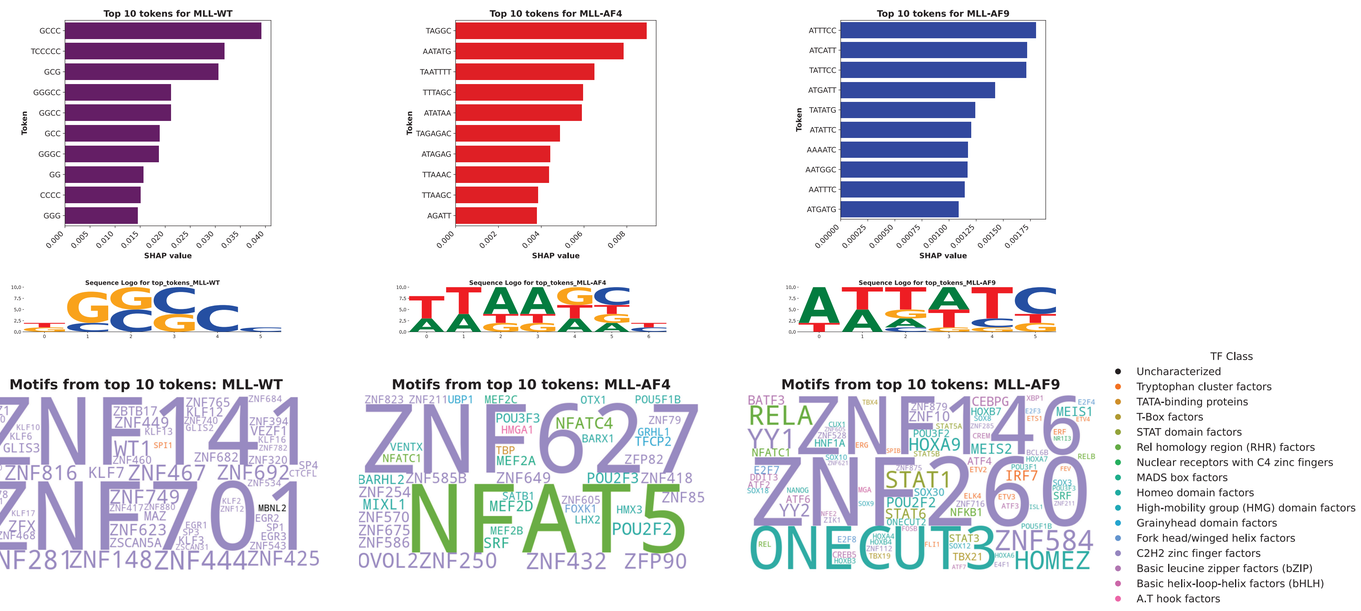






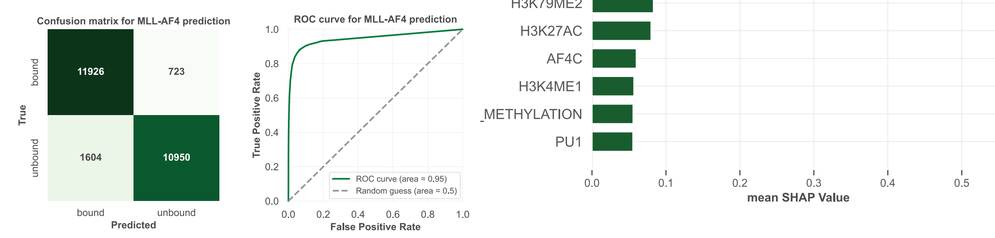
3.MLL-WT has a preference for CG rich sequences

The key tokens (K-mers) for predicting MLL-WT, MLL-AF4 or MLL-AF9 binding were extracted using SHAP[3] from the trained transformer model. Multiple sequence alignment indicated MLL-WT has a preference for CG rich sequences, whereas the most important sequence features for MLL-FPs were AT rich.

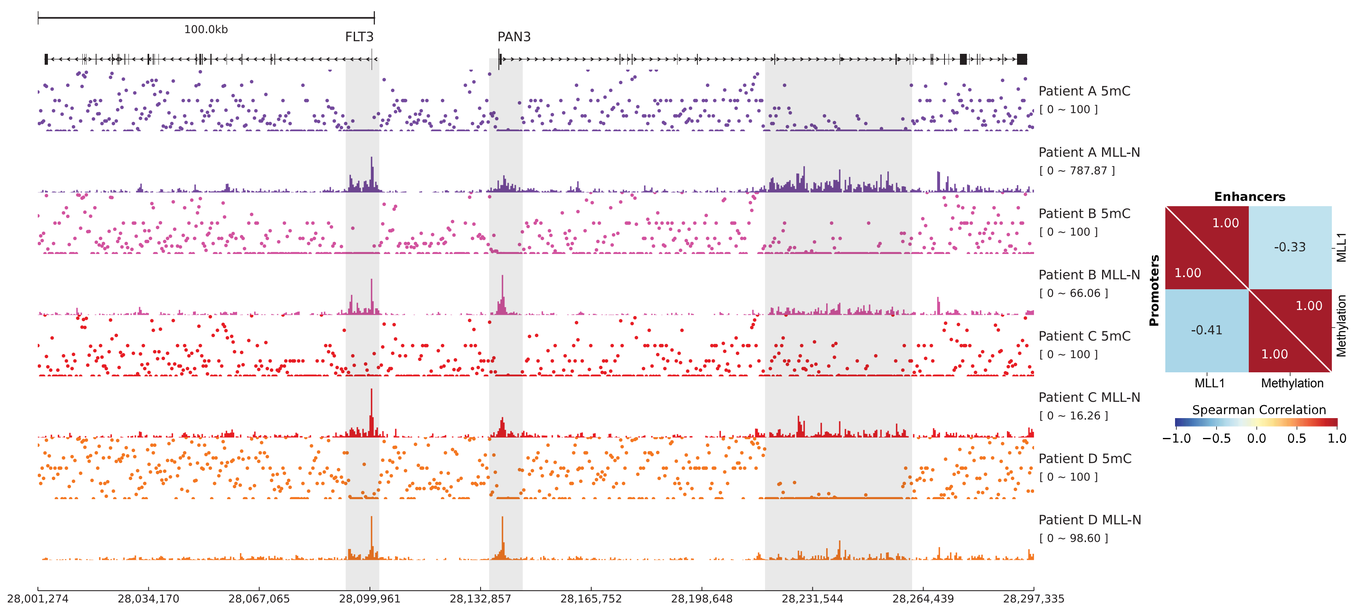


4. Epigenomic landscape influences MLL-AF4 binding

For determining which aspects of the epigenomic landscape are important for MLL-AF4 binding, a GBM classifier was trained to predict MLL-AF4 binding given a panel of 63 features. The model accurately (ROC AUC of 0.95) predicted MLL-AF4 and confirmed that uCpGs along with ERG and RUNX2 were highly important for determining MLL-AF4 binding.



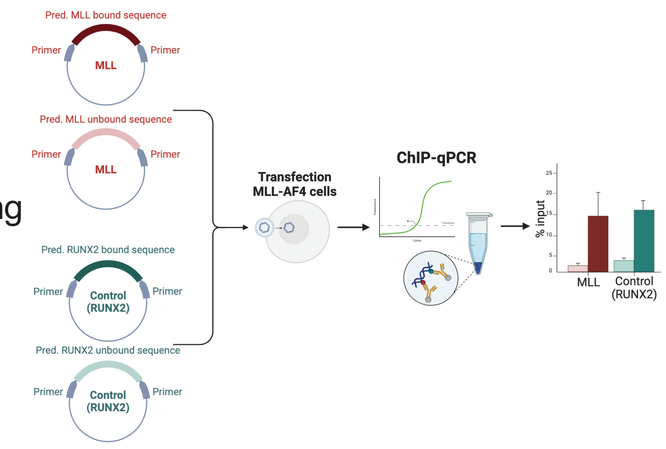
5. Methylation mirrors MLL-AF4 binding



Targeted simultaneous SNV and methylation status sequencing[4] in MLL-AF4 blast samples showed a striking inverse correlation with MLL-AF4 binding, which appears to be stronger at both promoters and enhancers compared to other inter- or intra-genic regions in MLL-AF4 cells.

6. Validation of predicted binding

Our findings enhance the understanding of how specific DNA sequences and the epigenomic landscape contribute to the unique binding patterns of MLL-AF4. Future work will combine these aspects into a unified model and validate the model predictions *in vivo*.



7. References

1. Kerry, J., et al. (2017) *Cell Reports*
2. Sanabria, M., et al. (2024) *Nature Machine Intelligence*
3. Lundberg, S and Lee, S. (2017) *NeurIPS Proc.*
4. Füllgrabe, J. (2023) *Nature Biotechnology*