CHIME: MULTIMODAL PREDICTION OF HISTONE MODIFICATION PROFILES ACROSS THE ADULT MOUSE BRAIN

Anonymous authors
Paper under double-blind review

ABSTRACT

Comprehensive profiling of histone modifications across the adult mouse brain remains experimentally costly. A recent integrative single-cell epigenomic atlas profiles over 2.5 million nuclei yet covers only nine brain regions, leaving the cerebellum, olfactory bulb, and additional midbrain areas unprofiled. Completing histone modification maps therefore requires predicting active and repressive chromatin features for cell types lacking histone measurements. We introduce CHIME, which integrates chromatin accessibility, DNA methylation, RNA expression, and DNA sequence features to predict four continuous histone modification signals at 200-bp resolution. CHIME leverages a distance-aware representation to integrate local and broader genomic context, paired with a position-preserving decoder that predicts cell-type-specific deviations from a shared histone profile. Evaluated via five-fold cross-validation across 118 cell types with test targets completely withheld, CHIME achieves a mean positional correlation of 0.737 and surpasses all evaluated baselines on every metric across all folds. Matched removals show complementary contributions from chromatin accessibility, DNA methylation, and gene expression. With reference profiles from training cell types available on each chromosome, models trained on chromosome 19 transfer without parameter updates to 18 additional autosomes, achieving a mean correlation across genomic positions of 0.734. These results support computational extension of brain epigenetic maps to cell populations lacking histone measurements.

1 INTRODUCTION

Histones are proteins around which DNA is wrapped to form nucleosomes, and chemical modifications to these proteins serve as epigenetic signals that help regulate chromatin structure and gene expression. Mapping these modifications across brain cell types helps distinguish active, primed, and repressive chromatin states, providing regulatory information that complements gene expression and chromatin accessibility (Preissl et al., 2023). In a recent study, accessible regions carrying histone H3 lysine 27 acetylation (H3K27ac) and histone H3 lysine 4 monomethylation (H3K4me1) were more enriched for experimentally validated enhancers than regions with accessibility alone. Histone H3 lysine 27 trimethylation (H3K27me3) and histone H3 lysine 9 trimethylation (H3K9me3) further revealed cell-type-specific repression and broader lineage boundaries (Wang et al., 2026).

However, comprehensive profiling of histone modifications across the adult mouse brain remains experimentally costly. Although the recent Paired-Tag atlas mapped over 2.5 million nuclei across nine brain regions (Wang et al., 2026), its limited anatomical coverage leaves histone modification landscapes uncharacterized in many key areas, such as the cerebellum, the olfactory bulb, and additional midbrain structures. In contrast, chromatin accessibility, DNA methylation, and RNA expression atlases offer much broader anatomical coverage of the adult mouse brain (Zu et al., 2023; Liu et al., 2023).

To extend the existing histone modification atlas, we investigate whether chromatin accessibility, DNA methylation, and gene expression profiles can be used to predict unmeasured histone modifi-

Code for data preprocessing, model training, and evaluation will be released upon acceptance.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

cation profiles at the cell-type level. Specifically, we study histone prediction across 118 brain cell types and four marks at 200 bp resolution, withholding all histone measurements of each test cell type and predicting them from its other measurements and DNA sequence.

Existing approaches predict epigenomic profiles from DNA sequence or incorporate chromatin accessibility and gene expression to condition predictions on a target cell type (Avsec et al., 2021; Murphy et al., 2024; Aksu & Vingron, 2026). Integrating these modalities with DNA methylation requires accounting for how each input represents molecular information. Specifically, chromatin accessibility and DNA methylation vary along the genome, whereas a cell-type-level gene expression profile characterizes the target cell type. A model must therefore combine cell-type information with genomic signals at different distances from the prediction site while preserving the local detail needed to reconstruct a histone signal.

We propose **CHIME** (*Cell-conditioned Histone Inference from Multi-omic Evidence*), a multimodal model for this task. Its distance-kernel module preserves fine central detail and aggregates distal signals in wider, direction-specific buckets, gated by gene expression and DNA sequence features. A position-preserving decoder combines this representation with local features and attention-based context to predict deviations from a shared histone profile estimated from training cell types. Sampling overlapping windows exposes each locus to varied contexts and output offsets, while a correlation loss directly supervises differences across cell types.

Our main contributions are:

- **Atlas-scale histone prediction.** We establish aligned multimodal profiles to study continuous histone prediction across 118 brain cell types, four marks, and 200 bp genomic bins, withholding all histone targets of each test cell type.
- **Distance-aware multimodal modeling.** We combine the four input modalities through local features, attention, and a gated distance-kernel representation. A position-preserving decoder predicts complete 25 kb windows from overlapping training contexts.
- **Cell-type-specific evaluation.** We evaluate both signal reconstruction and cell-type-specific variation, compare input and architectural variants, and test reference-assisted transfer to 18 additional autosomes without parameter updates.

2 Related work

Sequence-based prediction. Enformer uses long-range sequence context to predict transcriptional and epigenomic signals (Avsec et al., 2021); AlphaGenome extends sequence-based prediction across a broader range of regulatory outputs and resolutions (Avsec et al., 2026). These models learn outputs associated with the cell types represented in their training data. They motivate contextual sequence prediction, but supplying molecular measurements from a new cell population requires an additional conditioning mechanism. Our model instead treats DNA as one input to a predictor conditioned on multimodal measured cell state.

Cell-conditioned prediction. EPCOT learns transferable genomic representations from sequence and accessibility (Zhang et al., 2023). Enformer Celltyping combines sequence with local and genome-wide accessibility and separates shared from cell-type-specific histone signal (Murphy et al., 2024). Epitome uses accessibility similarity to reference cells to predict binary epigenetic events, including histone peaks (Morrow et al., 2021). Corgi conditions sequence representations on trans-regulator expression through feature-wise modulation to predict genomic signals in held-out cell types (Aksu & Vingron, 2026). These approaches establish the value of cell-state conditioning for genomic prediction. CHIME brings measured accessibility and methylation into a distance-dependent representation gated by cell-type-level gene expression and frozen DNA features, and combines this representation with position-specific local vectors for output decoding.

Epigenomic imputation. ChromImpute uses within-sample and across-sample features (Ernst & Kellis, 2015); Avocado learns multiscale latent factors for genomic positions, assays, and cell types (Schreiber et al., 2020a); eDICE contextualizes cell-type and assay embeddings using observed signals (Hawkins-Hooker et al., 2023). Corgi+ additionally uses local RNA coverage and training-mean signals, with a loss on deviations from the mean (Aksu & Vingron, 2026). Like Corgi+, CHIME is

reference-assisted: the mean histone profile of training cell types supplies shared structure, which the model refines using other modalities. Our benchmark hides all four target histone signals for test cell types while retaining their chromatin accessibility, DNA methylation, and gene expression.

Baseline comparability. We evaluate Avocado and eDICE with the available histone labels and non-histone measurements (Section 4.2). Other architectures require additional adaptation. Enformer Celltyping uses an epigenomically pretrained Enformer backbone and a human marker-promoter representation (Murphy et al., 2024); its external histone supervision differs from our training-label budget. Corgi uses a predefined trans-regulator expression panel, while Corgi+ additionally requires local RNA coverage (Aksu & Vingron, 2026). EpiGePT uses expression and binding features for 711 transcription factors (Gao et al., 2024). These inputs are not directly provided by our cell-type RNA representation. Epitome predicts the probability of peak presence (Morrow et al., 2021), and applying it to our benchmark would require additional adaptation to predict continuous histone signal intensities. Appendix B records the evaluated settings and remaining requirements.

Cell-type specificity. Schreiber et al. (2020b) showed that shared locus activity can produce misleading evidence of cross-cell-type prediction and advocated comparison with the training-cell mean and assessment of cell-type differences. We follow that principle by evaluating signal reconstruction, differences across cell types at a fixed locus, and departures from the reference along the genome. We report these metrics on complete chromosomes to distinguish reconstruction of shared signal from prediction of cell-type differences.

3 METHODS

CHIME predicts the histone modification signals of a target cell type from four input modalities: chromatin accessibility and DNA methylation measured along the genome, the gene expression profile of the cell type, and DNA sequence features. The architecture preserves position-specific local features while using broader genomic context and cell-state embeddings to condition predictions across the output window (Figure 1).

3.1 PROBLEM SETUP

Throughout this work, a *cell type* denotes a transcriptomically defined population at the subclass level of the shared brain-atlas taxonomy, with measurements aggregated within each population. Let $y_{s,p,m}$ denote the observed continuous histone signal for cell type s, genomic bin p, and mark $m \in \{\text{H3K27ac}, \text{H3K4me1}, \text{H3K27me3}, \text{H3K9me3}\}$. Cell types are partitioned into training, validation, and test sets. We denote the training and test cell-type sets by $\mathcal{S}_{\text{tr}}$ and $\mathcal{S}_{\text{te}}$, respectively. Test cell types provide chromatin accessibility, DNA methylation, and gene expression (RNA) inputs, but none of their histone targets are used to fit the model, reference profile means, or target scalers. Validation histone labels are used only for early stopping and checkpoint selection. Genomic bin coordinates are shared across all splits.

At each prediction center, the model receives approximately 1 Mb of binned accessibility and methylation, an RNA representation for each cell type, and a DNA representation of the local sequence from the frozen Evo 2 model (Brixi et al., 2026). These modalities serve distinct roles: epigenomic signals describe continuous spatial context, RNA summarizes global cell identity, and DNA provides sequence context shared across cell types.

3.2 SIGNAL DECOMPOSITION AND TARGETS

Given the validity mask $M_{s,p,m}$, we define the training-reference mean profile $\mu_{p,m}$ and the cell-type-specific deviation $e_{s,p,m}$ as

$$\mu_{p,m} = \frac{\sum_{s \in \mathcal{S}_{\text{tr}}} M_{s,p,m} y_{s,p,m}}{\sum_{s \in \mathcal{S}_{\text{tr}}} M_{s,p,m}}, \qquad e_{s,p,m} = y_{s,p,m} - \mu_{p,m}. \tag{1}$$

Within each fold, the reference profile $\mu_{p,m}$ is estimated exclusively from training cell types and held fixed for validation and unseen test cell types. The network never observes the target cell type's

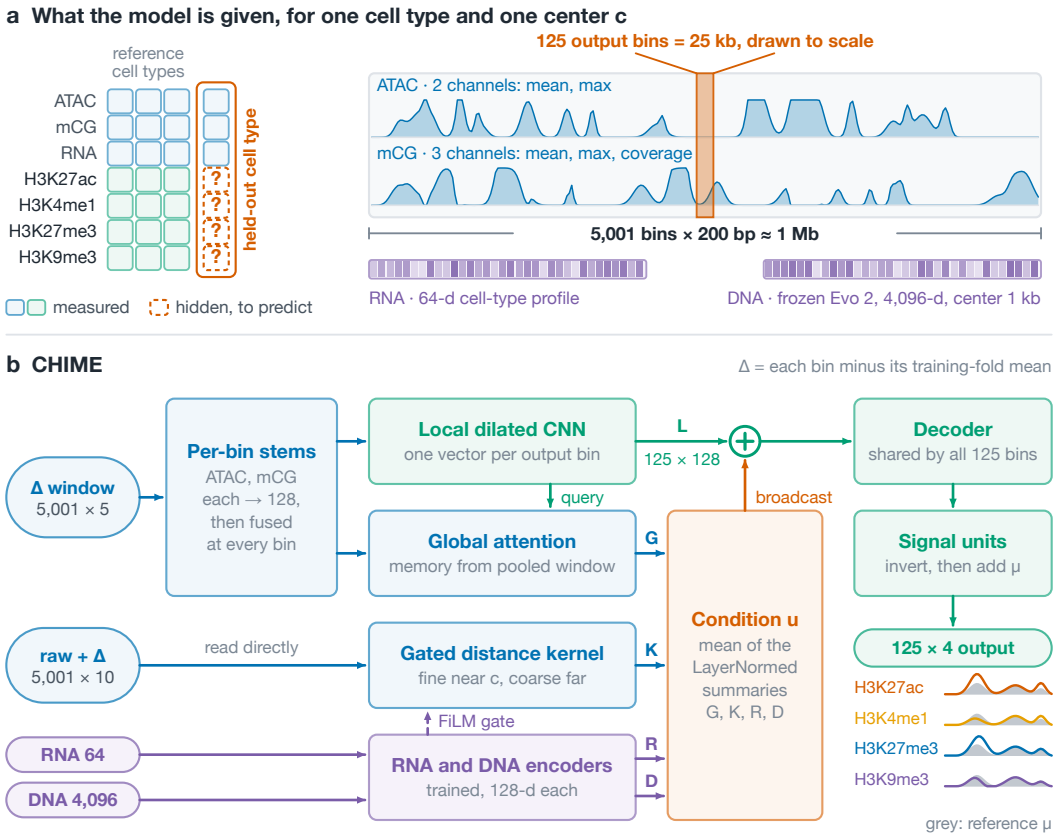

Figure 1: CHIME overview. **(a)** Accessibility, methylation, RNA, and DNA features are available for the target cell type; histone signals are hidden. **(b)** Local features preserve output positions. Broader context is summarized by global attention, whose query is the local center vector, and by a gated distance kernel that reads raw and deviation inputs directly and is modulated by the RNA and DNA encoders. Four context summaries condition every local vector; a shared decoder predicts deviations that are added to the mean over training cell types. L denotes position-specific local vectors; G, K, R, and D denote attention, distance-kernel, RNA, and DNA summaries. Pills are inputs with their tensor shapes; signal shapes are schematic.

histone measurements, and $\mu_{p,m}$ supplies the shared genomic profile, which the model refines by predicting the deviation. This decomposition gives the predictor a common genomic baseline while assigning differences among cell types to the learned component. Positive and negative deviations indicate enrichment and depletion relative to that baseline, respectively, allowing the model to represent both directions of variation at the same locus.

To stabilize training across marks with varying signal amplitudes, we transform deviations using a mark-specific scale $a_m > 0$ and an invertible soft-clipping function g. The ground-truth target is $t_{s,p,m} = g(e_{s,p,m}/a_m)$. The model decoder outputs predicted transformed deviations $\hat{t}_{s,p,m}$, from which final signal predictions are reconstructed via

$$\hat{y}_{s,p,m} = \mu_{p,m} + a_m g^{-1}(\hat{t}_{s,p,m}). \tag{2}$$

Appendix A specifies g and the training-only scale computation. Reconstructed signals are evaluated in raw signal units.

Input epigenomic signals are decomposed analogously: $\delta x_{s,p} = x_{s,p} - \bar{x}_p^{\mathrm{tr}}$, where $\bar{x}_p^{\mathrm{tr}}$ is the mean input profile across training cell types. The local and attention paths consume standardized deviations, whereas the distance-aware branch receives both standardized raw signals and standardized deviations. Scalers are fitted on training inputs and applied to validation and test sets, exposing both absolute molecular state and departure from the reference.

3.3 MULTIMODAL ENCODING

Local features and attention. The local path preserves fine spatial resolution across the output window (Figure 1b, top). Independent stems process accessibility and methylation, followed by a fusion layer to form joint positional representations. Dilated convolutions process the central output window to yield one feature vector per output position. A parallel path pools a broader window, using queries derived from local features to summarize distal epigenomic context via global attention, yielding summary $z_{s,c}^{\mathrm{att}}$.

Distance-kernel module. Collapsing broad genomic windows into single vectors risks losing fine-scale spatial resolution and signal magnitude. The proposed distance-kernel module (Figure 1b, K) preserves an exact-resolution central region and aggregates distal bins into logarithmically expanding buckets placed symmetrically upstream and downstream. For bucket b, we concatenate channel-wise mean and maximum values of raw and deviation inputs to form v_b. A shared affine transform A maps the pooled inputs to bucket features, after which bucket-specific projections and RNA/DNA gating compute

$$\begin{aligned} q_b &= \mathrm{GELU}(W_b A(v_b) + \beta_b), \qquad \alpha_b = \sigma\big(\mathrm{Gate}([z_s^{\mathrm{RNA}}; z_c^{\mathrm{DNA}}])_b\big)\,, \\ z_{s,c}^{\mathrm{dist}} &= \mathrm{Proj}([\alpha_b q_b]_{b=1}^{B}). \end{aligned} \tag{3}$$

Here z_s^{RNA} and z_c^{DNA} are the outputs of the RNA and DNA encoders (R and D in Figure 1b), B is the number of buckets, Gate generates a scalar gate per bucket, and Proj maps the concatenated gated features to the window-level summary $z_{s,c}^{\mathrm{dist}}$. Distance-specific projection weights W_b allow identical inputs to exert location-dependent influence, while gates modulate broad context based on cell state and sequence. The bucket mean summarizes the overall signal level, while the maximum retains localized high values that averaging may attenuate. Keeping upstream and downstream buckets separate preserves the direction of this context. The resulting representation complements attention by explicitly organizing molecular signals according to their distance from the output window.

3.4 OUTPUT DECODING

For an output window of J bins, let $p = c + j$ map the local output offset $j \in \{-(J-1)/2, \ldots, (J-1)/2\}$ around center c to global genomic bin p. Let $h_{s,c,j}$ denote the position-specific local feature at offset j (L in Figure 1b), and let $z_{s,c}^{(q)} \in \{z_{s,c}^{\mathrm{att}}, z_{s,c}^{\mathrm{dist}}, z_s^{\mathrm{RNA}}, z_c^{\mathrm{DNA}}\}$ denote the four window-level context summaries (G, K, R, and D). The decoding path fuses these representations via

$$u_{s,c} = \frac{1}{4}\sum_{q=1}^{4} \mathrm{LN}_q(z_{s,c}^{(q)}), \qquad \hat{t}_{s,c+j,:} = \mathrm{Dec}(h_{s,c,j} + u_{s,c}), \tag{4}$$

where LN_q indicates layer normalization and Dec is a shared multi-layer perceptron decoder (Figure 1b, right). Local vectors $h_{s,c,j}$ supply position-specific resolution, while $u_{s,c}$ modulates the entire window based on broader cell-state context.

3.5 TRAINING AND INFERENCE

An output window size of $J = 125$ corresponds to a 25 kb genomic span at 200 bp resolution. Sampling center loci across overlapping windows allows each genomic bin to be supervised under varied receptive-field offsets and surrounding contexts. During training, examples comprising a cell type s and center c are sampled with replacement. The model minimizes mean squared error over all valid targets in the window:

$$\mathcal{L}(\theta) = \frac{\sum_{(s,c)\in\mathcal{B}} \sum_{j=-(J-1)/2}^{(J-1)/2} \sum_m M_{s,c+j,m}\big(\hat{t}_{s,c+j,m} - t_{s,c+j,m}\big)^2}{\sum_{(s,c)\in\mathcal{B}} \sum_j \sum_m M_{s,c+j,m}}. \tag{5}$$

MSE penalizes errors in the transformed deviations, but small differences among cell types can contribute little to the overall objective. An additional comparison across cell types at a common locus directs supervision toward this variation, complementing the reconstruction objective. To

supervise differences among cell types directly, each batch groups training cell types at the same center loci, and a correlation term computed across cell types at each position is added:

$$\mathcal{L}_{\text{total}} = \mathcal{L}(\theta) + \lambda\, \mathbb{E}_m \mathbb{E}_{p\in\mathcal{P}_m}\Big[1 - \text{Corr}_{s\in\mathcal{B}_{p,m}}\big(\hat{e}_{s,p,m},\, e_{s,p,m}\big)\Big], \tag{6}$$

where Corr denotes Pearson correlation, $\mathcal{P}_m$ indexes valid position instances for mark m, $\mathcal{B}_{p,m}$ contains the corresponding cell types, and $\hat{e}_{s,p,m} = a_m g^{-1}(\hat{t}_{s,p,m})$. The expectations average positions within marks and then marks. The weight λ balances the two loss terms; implementation details are given in Appendix A. Since $\mu_{p,m}$ is shared across cell types, correlating deviations is equivalent to correlating reconstructed signals. At inference, adjacent non-overlapping 125-bin output tiles are assembled into whole-chromosome predictions, with each eligible genomic bin predicted once.

4 Experiments

4.1 Setup

Data and splits. We study histone prediction on mouse chr19 (mm10) in 200 bp bins, using aligned chromatin accessibility (ATAC) and 5-methylcytosine signals at CpG sites (mCG), cell-type RNA profiles, and local DNA features. Selecting this gene-dense chromosome enables computationally tractable five-fold cross-validation and systematic architectural ablations. Histone targets and RNA come from the Paired-Tag atlas (Wang et al., 2026); ATAC and mCG come from Zu et al. (2023) and Liu et al. (2023), respectively. Multimodal eligibility yields 118 cell types, assigned to five fixed test folds stratified by major cell groups. All folds share the eligible genomic positions, testing histone prediction for held-out cell types at known loci. RNA profiles are represented by principal components of cell-type gene expression. Downstream normalization, reference means, and target scaling strictly isolate training cell types. Appendix A details eligibility, splits, and preprocessing.

Training and selection. Early stopping monitors validation r_{cell}, and the final checkpoint is selected by validation r_{res}. We evaluate predictions at all valid stitched genomic positions. Training settings and the landmark evaluation protocol are given in Appendix A.

Research questions. **Q1:** How accurately does CHIME predict held-out cell types relative to baselines (Sections 4.2–4.3)? **Q2:** Which design choices support this accuracy (Section 4.4)? **Q3:** Does the mapping extend to unseen chromosomes (Section 4.5)? For Q2, we retrain variants on the same cell-type splits to compare training-center coverage, context modules, and input availability.

Metrics. For a fixed mark and genomic region, we define three mean Pearson correlations, with $\mathbb{E}$ denoting uniform averaging:

$$r_{\text{pos}} = \mathbb{E}_{s\in\mathcal{S}_{\text{te}}}\ \text{Corr}_p(\hat{y}_{s,p}, y_{s,p}), \tag{7}$$

$$r_{\text{cell}} = \mathbb{E}_p\ \text{Corr}_{s\in\mathcal{S}_{\text{te}}}(\hat{y}_{s,p}, y_{s,p}), \tag{8}$$

$$r_{\text{res}} = \mathbb{E}_{s\in\mathcal{S}_{\text{te}}}\ \text{Corr}_p(\hat{y}_{s,p} - \mu_p, y_{s,p} - \mu_p). \tag{9}$$

r_{pos} measures reconstruction across positions, r_{cell} differences across cell types at a fixed locus, and r_{res} deviations from the training reference. We use r_{cell} as the primary cell-type-specificity metric and r_{res} as a complementary measure of residual shape. Landmark metrics use cell-type/window macro aggregation; whole-chromosome metrics use all valid stitched positions. We average marks within folds and report the mean and sample SD over five folds, excluding undefined correlations.

4.2 Comparison with baselines

Comparison setting. All baseline methods and model variants are evaluated under a unified protocol using identical test folds, genomic positions, and evaluation metrics, strictly without access to test histone labels. The baseline suite includes nearest-cell-type transfer by ATAC similarity, a multilayer perceptron (MLP), a dilated convolutional network (CNN), Avocado, and eDICE. Avocado learns cell embeddings for target cell types after training (Schreiber et al., 2020a); eDICE incorporates test-cell ATAC and mCG inputs with masked histone targets during training (Hawkins-Hooker et al., 2023). Both follow established training procedures adapted to our benchmark settings (Appendix B.1).

Table 1: Baselines and model variants on complete chr19 signals (mean ± sample SD over five folds). Panels B–D share the CHIME reference in Panel A; Δr_{cell} is the paired difference from that reference. Bold denotes the highest unrounded mean in Panel A.

Model / variant	$r_{\text{pos}} \uparrow$	$r_{\text{cell}} \uparrow$	$r_{\text{res}} \uparrow$	Δr_{cell}
A. Baselines				
Nearest cell type	0.609 ± 0.034	0.064 ± 0.010	0.190 ± 0.041	-0.128 ± 0.018
MLP	0.604 ± 0.023	0.118 ± 0.011	0.164 ± 0.033	-0.074 ± 0.008
CNN	0.622 ± 0.022	0.134 ± 0.013	0.201 ± 0.031	-0.058 ± 0.007
Avocado	0.716 ± 0.022	0.129 ± 0.011	0.248 ± 0.034	-0.063 ± 0.017
eDICE	0.732 ± 0.026	0.126 ± 0.037	0.200 ± 0.032	-0.066 ± 0.037
CHIME	$\mathbf{0.737} \pm 0.028$	$\mathbf{0.192} \pm 0.020$	$\mathbf{0.271} \pm 0.028$	—
B. Training centers				
Tile-grid centers only	0.732 ± 0.027	0.143 ± 0.021	0.232 ± 0.021	-0.049 ± 0.008
C. Context modules				
w/o global attention	0.736 ± 0.027	0.163 ± 0.018	0.269 ± 0.022	-0.029 ± 0.004
w/o distance branch	0.737 ± 0.028	0.161 ± 0.017	0.267 ± 0.026	-0.031 ± 0.004
D. Input modalities				
w/o RNA	0.737 ± 0.028	0.154 ± 0.022	0.264 ± 0.027	-0.038 ± 0.007
w/o ATAC	0.727 ± 0.026	0.122 ± 0.018	0.223 ± 0.025	-0.070 ± 0.010
w/o mCG	0.734 ± 0.027	0.149 ± 0.016	0.253 ± 0.023	-0.042 ± 0.007

Average and mark-specific performance. CHIME exceeds every baseline on all three metrics in every fold (Table 1, Panel A). On the primary cell-type-specificity metric, it reaches $r_{\text{cell}} = 0.192$, compared with 0.064 for nearest-cell-type transfer and at most 0.134 for the other baselines. The margin over nearest-cell-type transfer shows that the predictions carry target-specific information beyond the histone profile of the training cell type with the most similar ATAC signal. CHIME also reaches $r_{\text{pos}} = 0.737$ and $r_{\text{res}} = 0.271$, exceeding eDICE and Avocado, the strongest baselines on r_{pos} and r_{res}, respectively. Per mark, CHIME has the highest r_{cell} for all four marks, active and repressive alike (Table 5). Mark-specific results (Appendix C) show the highest r_{pos} for H3K9me3 and the highest r_{cell} and r_{res} for H3K4me1. Center-only and tiled landmark predictions are similar; our primary whole-chromosome evaluation also covers positions outside those selected windows.

Prediction for additional cell types. Once trained, CHIME predicts a new cell type from compatible molecular inputs with fixed model weights and a shared training-cell reference. Each forward pass produces a 25 kb output window. In our benchmark, Avocado fits new cell embeddings from target ATAC/mCG, while eDICE includes target non-histone profiles during training (Appendix B.1). At inference, eDICE and nearest-cell-type transfer use individual reference-cell histone profiles; CHIME uses their mean.

4.3 Regional predictions

To evaluate spatial fidelity at individual loci, Figure 2 compares measured and predicted profiles around the *Scd4* promoter for a held-out test population, L2/3 IT CTX Glut. The profiles illustrate local reconstruction across active and repressive marks, with positional correlations of $r_{\text{pos}} = 0.942$ for H3K27ac and 0.876 for H3K4me1. For H3K27ac, the predicted and observed signal maxima align within the same 200 bp bin, about 0.6 kb upstream of the annotated *Scd4* transcription start site, and the predicted maximum reaches 92% of the measured one.

Figure 2b extends this comparison to reference-subtracted H3K27ac deviations across all 24 test cell types, ordered by mean measured deviation. CHIME captures some of the observed differences in local H3K27ac signals across cell types. For example, L2/3 IT CTX Glut (marked by arrows) shows positive deviations from the training-cell-type mean in both measured and predicted signals, whereas some other cell types show negative deviations. Evaluation across an additional 24 peak-centered windows chosen on biological grounds is reported in Table 13 (Appendix D). The two panels therefore address different aspects of prediction: the signal profiles show whether local genomic structure is reconstructed, whereas the deviation heatmaps show how the same region differs



Figure 2: Histone signals around the *Scd4* promoter, mm10 chr19:44,322,800–44,342,800 (fold 1). **(a)** Top: GENCODE M25 gene models with annotated transcript TSSs (bent arrows); the shaded interval is enlarged below. Measured and predicted signals for L2/3 IT CTX Glut; scales are shared within each mark, negative predictions are displayed as zero, and the dashed line marks the *Scd4* TSS. **(b)** Reference-subtracted H3K27ac across 24 test cell types ordered by mean measured deviation; arrows denote L2/3 IT CTX Glut. Heatmaps use independent symmetric ranges set to the 98th percentile of absolute deviations.

among cell types. Agreement in the first does not by itself establish the second, motivating their joint examination.

4.4 Design choices

Training-center coverage. Panel B of Table 1 compares sampling from all eligible genomic centers with sampling from a fixed grid aligned to the 25 kb output tiles. The grid fixes each locus's output offset; all-center sampling varies both offset and context. The main model scores higher in every fold, with mean gaps of 0.049 for r_{cell}, 0.039 for r_{res}, and 0.004 for r_{pos}. A separate

matched comparison also favors all-center sampling (Figure 4). Differences extend across output-offset groups (Appendix C.3).

Prediction with incomplete modality sets. Retrained variants lacking ATAC, mCG, or RNA inputs still predict held-out histone signals (Table 1, Panel D). Relative to the full main model, their mean r_{res} values are lower by 0.048, 0.018, and 0.006, respectively, with the largest gap for the variant lacking ATAC. These results support training predictors for datasets in which one of these modalities is unavailable.

Distance-aware and attention-based context. Variants without the distance-kernel module or global attention score below the main model on r_{cell} in all five folds (Table 1, Panel C). Their mean gaps are 0.031 and 0.029, respectively. The largest residual-correlation gap for the no-distance variant occurs on H3K9me3. A separate paired five-seed comparison shows a small mean residual-correlation gain from the distance branch, with little change in mean r_{cell} (Appendix C.1). Together, these comparisons indicate that the benefit of broader context depends on the metric and histone mark.

4.5 Chromosome transfer

The evaluations above place held-out cell types at loci represented in training. To test prediction at unseen genomic positions, we apply chr19-trained models from all five folds unchanged to 18 other autosomes. Only the position-indexed histone reference and input-deviation mean are recomputed on each chromosome from the same training cell types. Averaged over these chromosomes, $r_{\mathrm{pos}} = 0.734$, $r_{\mathrm{cell}} = 0.151$, and $r_{\mathrm{res}} = 0.247$ (Table 3). These results support transfer beyond the training loci when reference profiles are available.

5 Limitations

Our evaluation is limited to adult mouse brain cell types; transfer to other tissues, species, or disease cohorts remains untested. Models are trained on chr19, and chromosome transfer requires reference profiles from training cell types (Section 4.5). Predicted cell-type deviations are also compressed in amplitude relative to measured deviations (e.g., Figure 5), so discrete peak calling and quantitative effect-size analyses require calibration and separate validation. Extension to unprofiled brain regions also requires compatible inputs and evaluation of cell populations that may differ from those represented here.

6 Conclusion

In this work, we presented CHIME to predict four histone modification profiles at 200 bp resolution for brain cell types lacking direct histone measurements. The model integrates local and distance-aware epigenomic contexts with RNA expression and DNA sequence features to predict cell-type-specific deviations from a shared reference profile. Across 118 cell types, CHIME consistently outperforms all evaluated baselines across all three metrics in each of five folds. Zero-shot evaluations also show that models trained on chromosome 19 transfer seamlessly to 18 additional autosomes using fixed weights and chromosome-specific references.

By bridging existing multi-omic atlases with histone modification landscapes, CHIME offers a scalable framework for expanding regulatory annotations across unmeasured brain cell populations. Moreover, its ability to operate under partial modality coverage broadens its applicability to diverse real-world datasets. In downstream applications, comparing predicted active and repressive marks across cell types can help prioritize candidate enhancers, repressive domains, and cell populations for experimental follow-up. Together, high-resolution computational predictions and targeted experimental validation could accelerate the mapping of regulatory programs underlying brain cell diversity.

AI USE STATEMENT

For polishing language.

REFERENCES

Ekin Deniz Aksu and Martin Vingron. Context-aware sequence-to-function model of human gene regulation. *Nature Communications*, 17(1):6200, July 2026. doi: 10.1038/s41467-026-75527-2.

Žiga Avsec, Vikram Agarwal, Daniel Visentin, Joseph R. Ledsam, Agnieszka Grabska-Barwinska, Kyle R. Taylor, et al. Effective gene expression prediction from sequence by integrating long-range interactions. *Nature Methods*, 18(10):1196–1203, October 2021. doi: 10.1038/s41592-021-01252-x.

Žiga Avsec, Natasha Latysheva, Jun Cheng, Guido Novati, Kyle R. Taylor, Tom Ward, et al. Advancing regulatory variant effect prediction with AlphaGenome. *Nature*, 649(8099):1206–1218, January 2026. doi: 10.1038/s41586-025-10014-0.

Garyk Brixi, Matthew G. Durrant, Jerome Ku, Mohsen Naghipourfar, Michael Poli, Gwanggyu Sun, et al. Genome modelling and design across all domains of life with Evo 2. *Nature*, 652(8112): 1349–1361, March 2026. doi: 10.1038/s41586-026-10176-5.

Jason Ernst and Manolis Kellis. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nature Biotechnology*, 33(4):364–376, February 2015. doi: 10.1038/nbt.3157.

Zijing Gao, Qiao Liu, Wanwen Zeng, Rui Jiang, and Wing Hung Wong. EpiGePT: a pretrained transformer-based language model for context-specific human epigenomics. *Genome Biology*, 25 (1):310, December 2024. doi: 10.1186/s13059-024-03449-7.

Alex Hawkins-Hooker, Giovanni Visonà, Tanmayee Narendra, Mateo Rojas-Carulla, Bernhard Schölkopf, and Gabriele Schweikert. Getting personal with epigenetics: towards individual-specific epigenomic imputation with machine learning. *Nature Communications*, 14(1):4750, August 2023. doi: 10.1038/s41467-023-40211-2.

Hanqing Liu, Qiurui Zeng, Jingtian Zhou, Anna Bartlett, Bang-An Wang, Peter Berube, et al. Single-cell DNA methylome and 3D multi-omic atlas of the adult mouse brain. *Nature*, 624(7991): 366–377, December 2023. doi: 10.1038/s41586-023-06805-y.

Alyssa Kramer Morrow, John Weston Hughes, Jahnavi Singh, Anthony Douglas Joseph, and Nir Yosef. Epitome: predicting epigenetic events in novel cell types with multi-cell deep ensemble learning. *Nucleic Acids Research*, 49(19):e110, August 2021. doi: 10.1093/nar/gkab676.

Alan E. Murphy, William Beardall, Marek Rei, Mike Phuycharoen, and Nathan G. Skene. Predicting cell type-specific epigenomic profiles accounting for distal genetic effects. *Nature Communications*, 15(1):9951, November 2024. doi: 10.1038/s41467-024-54441-5.

Sebastian Preissl, Kyle J. Gaulton, and Bing Ren. Characterizing cis-regulatory elements using single-cell epigenomics. *Nature Reviews Genetics*, 24(1):21–43, 2023. doi: 10.1038/s41576-022-00509-1.

Jacob Schreiber, Timothy Durham, Jeffrey Bilmes, and William Stafford Noble. Avocado: a multi-scale deep tensor factorization method learns a latent representation of the human epigenome. *Genome Biology*, 21(1):81, March 2020a. doi: 10.1186/s13059-020-01977-6.

Jacob Schreiber, Ritambhara Singh, Jeffrey Bilmes, and William Stafford Noble. A pitfall for machine learning methods aiming to predict across cell types. *Genome Biology*, 21(1):282, November 2020b. doi: 10.1186/s13059-020-02177-y.

Zhaoning Wang, Songpeng Zu, Ethan J. Armand, Timothy H. Loe, Jonathan A. Rink, Wanying Wu, et al. An integrative single-cell epigenomic atlas annotates the regulatory genome of the adult mouse brain. *bioRxiv*, February 2026. doi: 10.64898/2026.02.07.704075.

Wen Wen, Jiaxin Zhong, Zhaoxi Zhang, Lijuan Jia, Tinyi Chu, Nating Wang, Charles G. Danko, and Zhong Wang. dHICA: a deep transformer-based model enables accurate histone imputation from chromatin accessibility. *Briefings in Bioinformatics*, 25(6):bbae459, September 2024. doi: 10.1093/bib/bbae459.

Zhenhao Zhang, Fan Feng, Yiyang Qiu, and Jie Liu. A generalizable framework to comprehensively predict epigenome, chromatin organization, and transcriptome. *Nucleic Acids Research*, 51(12): 5931–5947, July 2023. doi: 10.1093/nar/gkad436.

Songpeng Zu, Yang Eric Li, Kangli Wang, Ethan J. Armand, Sainath Mamde, Maria Luisa Amaral, et al. Single-cell analysis of chromatin accessibility in the adult mouse brain. *Nature*, 624(7991): 378–389, December 2023. doi: 10.1038/s41586-023-06824-9.

A IMPLEMENTATION AND DATA PROVENANCE

Model configuration. All folds use the architecture in Section 3. Each input spans 5,001 bins at 200 bp resolution, from which the model predicts the central 125 bins (25 kb). The dilated local convolutions cover the output window while retaining one feature vector per bin. Global attention pools the broader context into 1 kb intervals. The distance-kernel branch retains 1,251 central bins individually and groups the remaining context into 32 logarithmically expanding buckets on each side. RNA and DNA representations condition these spatial features through bucket-specific gates before fusion with the local features.

Cohort and computation. Eligibility requires a minimum cell count for each measured modality, with waivers for shortfalls confined to methylation. This admits 118 of 131 cached cell types. The earlier screens in Tables 9 and 10 use a pre-waiver cohort and a different checkpoint rule; they are not pooled with the main results. Whole-chromosome tiled inference processes 243k cell-type–position bins per second with 1966 MiB peak GPU memory on an exclusive A800 80GB, excluding frozen-sequence embedding preprocessing.

Splits and evaluation. All folds share 280,935 eligible chr19 centers; the tile-grid variant uses 2,248 centers spaced 125 bins apart. The five test folds distribute cell types within each of 15 major cell groups as evenly as possible, yielding 24, 24, 24, 23, and 23 held-out cell types. Within each fold, 19 remaining cell types are randomly allocated to validation with a fixed seed, leaving 75 or 76 for training. Validation uses 20 training-selected 200 kb windows per mark, evaluated every fifth position; test landmark evaluation uses 50 such windows per mark. The MSE-only all-center and tile-grid arms in Figure 4 share the optimization and checkpoint rules, sampled-window budgets, output size, and stitching procedure. Table 1 instead uses the main model with the correlation loss as its reference.

Statistical reporting. Paired differences include split variation and optimization noise. Two-sided paired t-tests use four degrees of freedom without multiple-comparison correction; overlapping training sets make these tests descriptive, not independent replications.

Correlation-loss implementation. The same weight is used in all folds, calibrated on the first training batch of the fold-0 screen so that the weighted correlation-loss gradient norm was 20% of the MSE gradient norm; no validation or test data entered this calibration. Its effective weight increases linearly to this value over the first 682 optimization steps. Correlations are calculated within each sampled group of 16 cell types, then averaged equally over eligible position instances within each mark and over marks with eligible positions. Eligibility requires at least eight valid cell types and target variance above 10^{-10}. For numerical stability, the implementation uses scale-normalized deviations $\hat{e}/a_m$ and e/a_m and adds 10^{-6} to each variance in the Pearson denominator. Equation 6 gives the corresponding unstabilized objective; the stabilization means the implemented value can differ slightly from ordinary Pearson correlation.

RNA provenance. The upstream PCA is fitted to the full reference-profile collection; downstream scalers and reference means are fitted within each training split. The saved RNA metadata identify Paired-Tag RNA as the expression source, with 1,111 genes and 157 cell-type profiles before the histone-model eligibility intersection. The procedure normalizes each cell to a library-size target of 10,000, applies $\log(1+x)$, averages within each cell type, standardizes genes, and applies PCA. The resulting 64-dimensional representation is reused across genomic positions for that cell type. This upstream RNA collection is larger than the histone evaluation cohort: 157 cell-type profiles exist, of which 118 pass the multimodal eligibility rule and enter the reported evaluation.

Target scaling. We scale deviations by a positive training-only mark scale a_m and apply a signed soft clip. The scale is the median, over training cell types with usable signal, of their nonzero mean histone signal for that mark. With $\tau = 10$,

$$t_{s,p,m} = g_\tau(e_{s,p,m}/a_m), \qquad g_\tau(z) = \begin{cases} z, & |z| \le \tau, \\ \operatorname{sign}(z)\big(2\sqrt{\tau|z|} - \tau\big), & |z| > \tau. \end{cases} \tag{10}$$

The transform uses AlphaGenome's soft-clipping shape on signed deviations, with a fold-specific per-mark scale rather than its per-output scale and without its 128-bp sum-resolution factor (Avsec et al., 2026). It reduces the influence of large deviations while retaining their sign.

Transform inversion. For a predicted transformed deviation v,

$$g_\tau^{-1}(v) = \begin{cases} v, & |v| \le \tau, \\ \operatorname{sign}(v)(|v| + \tau)^2/(4\tau), & |v| > \tau. \end{cases} \tag{11}$$

This inversion precedes adding the locus-specific training mean. Mark scaling, reference subtraction, and any nonlinear reporting transform are distinct operations; the results in this paper use no additional nonlinear reporting transform.

Reference subtraction. At any fixed locus and mark, $\mu_{p,m}$ is constant across test cell types. Provided the valid cell-type set is unchanged and both correlations are defined,

$$\operatorname{Corr}_s(\hat{y}_{s,p,m} - \mu_{p,m}, y_{s,p,m} - \mu_{p,m}) = \operatorname{Corr}_s(\hat{y}_{s,p,m}, y_{s,p,m}). \tag{12}$$

This follows from the shift invariance of Pearson correlation. It does not apply to across-position correlation, because the reference varies with position. The equality is also verified numerically in the archived whole-chromosome summaries. It is a property of the metric and does not imply that training on deviations leaves learned predictions unchanged.

B Baseline implementations and comparability

All main-table methods use the same 118 eligible cell types, five folds, four histone marks, and 280,935 chr19 evaluation positions. No test histone label is used for training, reference construction, or selection. The input modalities and fitting procedures differ as described below. The dHICA experiment is reported separately because its released weights include external supervision.

B.1 Implemented baselines

Nearest-cell-type transfer. For each target cell type, we select the single training cell type with the highest Pearson correlation of raw ATAC profiles across eligible chr19 positions, then copy its histone values at the same positions. Equivalently, we add that reference cell type's deviation to the training mean. We fix raw ATAC similarity and $k = 1$ without selecting among other similarity measures or neighbor counts. This baseline runs on CPU.

MLP. We flatten a 5,001-bin window of five standardized raw epigenomic channels (ATAC mean/max and mCG mean/max/coverage) and concatenate the 64-dimensional RNA and 4,096-dimensional DNA embeddings. The resulting 29,165 inputs pass through two 512-unit GELU layers with dropout 0.1 to predict 125 bins by four marks. The network has 15.45 million parameters and uses neither input deviations nor a histone reference in its output. AdamW uses learning rate 10^{-4}, weight decay 10^{-4}, and batch size 256. Training minimizes masked MSE on raw targets with 436,700 sampled windows per epoch, a 24-epoch cap, and validation-MSE checkpoint selection and early stopping (patience three), on an A40 GPU.

CNN. The CNN uses local convolutions with kernel width nine and 128 channels, followed by five residual blocks with dilation rates 1, 2, 4, 8, and 16, each containing two width-three convolutions. RNA and DNA MLP encoders condition a shared decoder that predicts the raw 125-bin output. The network has 2.58 million parameters, no attention or distance branch, and no reference added to its predictions. The dilated stack has a 125-bin receptive field; input stems add their own context. Inputs are standardized raw values, and targets remain in original units. It uses approximately the same sample and epoch budgets as CHIME, MSE-based early stopping, and a validation window-correlation checkpoint rule, on an A800 GPU. This checkpoint criterion differs from CHIME's residual-correlation selection.

Avocado. We reimplement the published factorization architecture in PyTorch (Schreiber et al., 2020a): 32 cell factors, 256 assay factors, position factors of sizes 25, 40, and 45 at 200 bp, 2 kb, and 40 kb scales, and two 2,048-unit ReLU layers. The position scales preserve the published 1:10:200 ratio at our bin size. Six assays comprise ATAC, mCG, and four histone marks, with arcsinh-transformed raw values and MSE loss. Stage one excludes test cell types entirely and observes histone only for training cell types; validation cell types contribute ATAC/mCG only. Stage two freezes the network, assay factors, and position factors and fits only new cell factors from test ATAC/mCG. Each stage uses the published 24,000-update schedule, batches of 40,000, and Adam with learning rate 10^{-3}, without validation-based selection. Runs use A40 GPUs. Position-specific factors must be fitted to extend this implementation to new genomic coordinates.

eDICE. Our PyTorch implementation follows the published cell/assay contextualization architecture (Hawkins-Hooker et al., 2023). It uses 256-dimensional embeddings, one four-head attention block with a 128-unit feed-forward layer for each axis, and a decoder with two 2,048-unit ReLU layers. Attention-block dropout is 0.1 and decoder dropout is 0.3; the model has 5.97 million parameters. At each bin, the observed tensor contains ATAC/mCG from all eligible cell types and histone from training cell types only. Training randomly hides 120 observed entries per bin and reconstructs their arcsinh-transformed values. Adam uses learning rate 3×10^{-4}, batches of 256 bins, and at most 20 epochs, with validation residual-correlation checkpoint selection and early stopping (patience three). Training takes 7–21 minutes per fold on an A40 GPU. The fixed cell vocabulary enters both learned cell embeddings and assay-input projections. Adding cell types therefore changes parameters: our evaluated procedure retrains on each expanded set, rather than applying unchanged weights. This statement concerns this implementation and protocol, not every possible transfer-learning adaptation of eDICE. Inference uses the observed training-cell histone values and target ATAC/mCG at each position.

Timing interpretation. The CHIME timing estimate uses an exclusive A800 measurement of 243,000 predicted bins per second, including tiled inference over all test cell types in fold 0. Dividing 280,935 positions by this throughput gives approximately 1.2 seconds per cell type on chr19. This is an amortized estimate, not a separate single-cell-type latency measurement, and excludes feature preparation. Initial model fitting and subsequent inference are distinct costs. No genome-wide retraining-time extrapolation is treated as a measurement. Removing attention reduces training-epoch time by 18% on an A800 80GB GPU.

B.2 Other input requirements

Table 2: Evaluation status and input requirements of related methods.

Method	Evaluated	Requirements and scope
Avocado / eDICE	Yes, main text	ATAC/mCG and training-cell histone; fitting procedures and target-cell exposure differ as specified above.
dHICA	Zero-shot, appendix	Released DNA+ATAC weights include external histone supervision; native 128 bp outputs are aggregated to 200 bp.
Enformer Celltyping	No	Human marker-promoter features and an epigenomically pretrained backbone require species and supervision controls (Murphy et al., 2024).
Corgi / Corgi+	No	A specified trans-regulator gene panel is required; Corgi+ also needs local RNA coverage, unavailable here (Aksu & Vingron, 2026).
EpiGePT	No	Expression and binding features for 711 TFs are not directly supplied by the RNA PCA inputs (Gao et al., 2024).
Epitome	No	Accessibility inputs are available, but binary peak outputs need adaptation for continuous-signal regression (Morrow et al., 2021).

B.3 PRETRAINED DHICA: A SEPARATE TRANSFER EXPERIMENT

We apply the authors' released DNA+ATAC dHICA weights without training on our histone labels (Wen et al., 2024). The model was trained on external ENCODE data and is not a comparison under the main table's training-label budget. We use mm10 sequence, approximate ATAC fold enrichment by dividing by the chromosome-wide mean, and aggregate native 128 bp predictions into 200 bp bins by base-pair overlap. For r_{res}, observations are centered by the measured training-cell mean and predictions by dHICA's predicted training-cell mean, to account for the transferred model's signal scale.

Across folds and marks, dHICA reaches $r_{\mathrm{pos}} = 0.181$, $r_{\mathrm{cell}} = 0.071$, and $r_{\mathrm{res}} = 0.096$ (Table 5). Both repressive marks have negative r_{cell} and r_{res}, and H3K9me3 has $r_{\mathrm{pos}} = -0.248$. These results characterize this pretrained transfer setting and input normalization; they do not establish the performance of dHICA retrained on this atlas. As an implementation check, we compared our driver with the authors' preprocessing and prediction pipeline on CA3 Glut at 20 aligned chr19 windows. After identical 200 bp aggregation, all four marks had Pearson correlation above 0.9999 across 11,470 bins. Correlation verifies output agreement up to scale; it is not a claim of exact numerical identity. The full zero-shot run took approximately 3.1 A40 GPU hours.

C ADDITIONAL NUMERICAL RESULTS

Table 3: Transfer to unseen autosomes with unchanged chr19 weights; the shared profile and input-deviation mean are recomputed per chromosome. Values average four marks over all valid stitched positions (mean ± sample SD over five folds); †training chromosome.

Chr.	r_{pos}	r_{cell}	r_{res}	Chr.	r_{pos}	r_{cell}	r_{res}
chr1	0.720 ± 0.028	0.149 ± 0.016	0.252 ± 0.023	chr10	0.721 ± 0.028	0.151 ± 0.016	0.246 ± 0.023
chr2	0.752 ± 0.018	0.158 ± 0.017	0.232 ± 0.020	chr11	0.729 ± 0.028	0.161 ± 0.017	0.254 ± 0.024
chr3	0.725 ± 0.028	0.144 ± 0.016	0.246 ± 0.022	chr12	0.743 ± 0.027	0.149 ± 0.016	0.252 ± 0.027
chr4	0.734 ± 0.025	0.150 ± 0.015	0.243 ± 0.024	chr13	0.751 ± 0.023	0.153 ± 0.017	0.242 ± 0.024
chr5	0.723 ± 0.028	0.150 ± 0.017	0.251 ± 0.026	chr14	0.721 ± 0.028	0.145 ± 0.015	0.246 ± 0.022
chr6	0.739 ± 0.024	0.152 ± 0.017	0.243 ± 0.023	chr15	0.722 ± 0.027	0.155 ± 0.017	0.253 ± 0.023
chr7	0.730 ± 0.028	0.141 ± 0.018	0.245 ± 0.025	chr16	0.713 ± 0.029	0.152 ± 0.017	0.252 ± 0.023
chr8	0.722 ± 0.028	0.157 ± 0.017	0.259 ± 0.023	chr17	0.815 ± 0.020	0.142 ± 0.017	0.222 ± 0.018
chr9	0.732 ± 0.023	0.162 ± 0.018	0.250 ± 0.024	chr18	0.720 ± 0.028	0.154 ± 0.016	0.255 ± 0.024
Mean	0.734 ± 0.026	0.151 ± 0.017	0.247 ± 0.023	chr19†	0.737 ± 0.028	0.192 ± 0.020	0.271 ± 0.028

C.1 SEED SENSITIVITY OF DISTANCE-AWARE CONTEXT

Distance-branch repeat. On fold 0, we compare the full architecture with a no-distance-branch variant using five paired seeds (0–4); both arms are trained with MSE alone. All arms share splits, training budgets, and checkpoint criteria. Mean r_{res} changes by $+0.0035 \pm 0.0033$ ($p = 0.08$), with a positive change in four of five seeds; mean r_{cell} changes by 0.0000 ± 0.0040. H3K27me3 r_{cell} increases in all five seeds. Table 4 includes both aggregate metrics and mark-specific follow-ups to the five-fold analyses; the selected mark results are exploratory.

Table 4: Distance-branch sensitivity on fold 0. Mean ± sample SD over five paired seeds; Δ is full minus no distance branch.

Metric	Full	No distance branch	Paired Δ	Seeds improved
Mean r_{res}	0.2359 ± 0.0022	0.2324 ± 0.0015	$+0.0035 \pm 0.0033$	4/5
Mean r_{cell}	0.1366 ± 0.0039	0.1366 ± 0.0024	0.0000 ± 0.0040	3/5
H3K27me3 r_{cell}	0.0944 ± 0.0037	0.0844 ± 0.0077	$+0.0101 \pm 0.0071$	5/5
H3K9me3 r_{res}	0.1463 ± 0.0051	0.1380 ± 0.0051	$+0.0082 \pm 0.0095$	4/5

C.2 Full baseline results

Table 5 reports all three metrics for every method and mark, with fold SDs. dHICA is a separate pre-trained transfer experiment with external supervision and a different residual reference, as specified in Appendix B.3.

Table 5: Baseline results by mark, mean ± sample SD over five folds. dHICA is evaluated zero-shot and should be interpreted separately.

Method	r_{pos}	r_{cell}	r_{res}
H3K27ac			
Nearest cell type	0.596 ± 0.038	0.075 ± 0.008	0.214 ± 0.042
MLP	0.670 ± 0.033	0.164 ± 0.021	0.251 ± 0.033
CNN	0.690 ± 0.033	0.185 ± 0.021	0.295 ± 0.036
Avocado	0.698 ± 0.025	0.146 ± 0.006	0.268 ± 0.027
eDICE	0.718 ± 0.033	0.159 ± 0.036	0.231 ± 0.016
dHICA (zero-shot)	0.452 ± 0.026	0.150 ± 0.017	0.191 ± 0.018
CHIME	0.730 ± 0.034	0.234 ± 0.023	0.344 ± 0.037
H3K4me1			
Nearest cell type	0.636 ± 0.033	0.095 ± 0.016	0.232 ± 0.041
MLP	0.663 ± 0.024	0.191 ± 0.019	0.227 ± 0.022
CNN	0.688 ± 0.024	0.219 ± 0.024	0.302 ± 0.025
Avocado	0.741 ± 0.020	0.158 ± 0.016	0.307 ± 0.022
eDICE	0.760 ± 0.024	0.183 ± 0.031	0.277 ± 0.029
dHICA (zero-shot)	0.435 ± 0.020	0.192 ± 0.012	0.267 ± 0.020
CHIME	0.767 ± 0.026	0.254 ± 0.030	0.395 ± 0.029
H3K27me3			
Nearest cell type	0.542 ± 0.034	0.041 ± 0.009	0.158 ± 0.042
MLP	0.495 ± 0.026	0.078 ± 0.016	0.109 ± 0.050
CNN	0.510 ± 0.024	0.079 ± 0.011	0.122 ± 0.044
Avocado	0.664 ± 0.030	0.104 ± 0.015	0.222 ± 0.055
eDICE	0.675 ± 0.033	0.066 ± 0.038	0.146 ± 0.052
dHICA (zero-shot)	0.085 ± 0.008	-0.041 ± 0.021	-0.031 ± 0.003
CHIME	0.676 ± 0.034	0.133 ± 0.010	0.174 ± 0.030
H3K9me3			
Nearest cell type	0.663 ± 0.035	0.044 ± 0.010	0.153 ± 0.042
MLP	0.588 ± 0.011	0.040 ± 0.012	0.069 ± 0.035
CNN	0.599 ± 0.012	0.053 ± 0.016	0.085 ± 0.028
Avocado	0.761 ± 0.015	0.109 ± 0.016	0.195 ± 0.039
eDICE	0.774 ± 0.019	0.095 ± 0.048	0.145 ± 0.042
dHICA (zero-shot)	-0.248 ± 0.007	-0.017 ± 0.010	-0.045 ± 0.009
CHIME	0.773 ± 0.018	0.145 ± 0.022	0.169 ± 0.021

C.3 Output-offset analysis

Table 6 groups genomic positions by absolute distance from their tile center. Every position is counted once under the fixed inference tiling. Values are averaged over four marks and five folds; differences are paired within folds. The tile-grid deficit occurs throughout the output window, rather than only at its boundaries. These groups describe output position, not an additional test split.

Table 6: Performance across absolute output-offset groups (200 bp bins). Differences are MSE-only variant minus the main model with the correlation loss; they combine changes in architecture or sampling with changes in training.

	Full		Tile grid Δ		No distance branch Δ	
Absolute offset	r_{cell}	r_{res}	r_{cell}	r_{res}	r_{cell}	r_{res}
0–5	0.194	0.279	−0.049	−0.039	−0.030	−0.003
6–11	0.194	0.272	−0.048	−0.040	−0.030	−0.002
12–17	0.193	0.271	−0.048	−0.038	−0.031	−0.004
18–24	0.192	0.274	−0.049	−0.039	−0.031	−0.003
25–30	0.193	0.275	−0.048	−0.038	−0.031	−0.003
31–36	0.192	0.274	−0.049	−0.042	−0.031	−0.003
37–43	0.191	0.271	−0.048	−0.038	−0.031	−0.003
44–49	0.191	0.263	−0.048	−0.035	−0.031	−0.003
50–55	0.191	0.269	−0.049	−0.039	−0.031	−0.003
56–62	0.186	0.259	−0.050	−0.038	−0.031	−0.003

C.4 Five-fold and exploratory comparisons

Evaluation scopes. Table 7 compares the complete chromosome with training-selected landmark windows. All rows use five folds. Figure 3 provides a separate per-mark view of the earlier MSE-only model.

Table 7: Evaluation-scope comparison, mean ± sample SD across five folds.

Evaluation	r_{pos}	r_{cell}	r_{res}
Full chr19 (tiled)	0.737 ± 0.028	0.192 ± 0.020	0.271 ± 0.028
Landmarks (tiled)	0.735 ± 0.026	0.199 ± 0.024	0.218 ± 0.024
Landmarks (center)	0.736 ± 0.026	0.201 ± 0.024	0.222 ± 0.025

Intermediate center densities. Table 9 gives the earlier single-fold center-density sweep that motivated the paired comparison in Table 8. It uses MSE-based checkpoint selection and center-only landmark evaluation, whereas the main five-fold model uses residual-correlation selection. Scores are similar at strides 1 and 5 and lower at strides 25 and 125. The sweep motivates a paired whole-chromosome comparison under a common training protocol; its four sampled strides do not establish a general threshold for adequate center coverage.

Processing-branch screen. A separate single-fold screen removed epigenomic processing branches from the earlier single-bin model while retaining RNA and DNA (Table 10). The local-plus-global variant has test $r_{\text{res}} = 0.1945$, compared with 0.2091 for all three branches. The local-plus-distance variant reaches 0.2105 at 513 rather than 648 seconds per epoch. These observations motivated retaining distance-aware context and examining whether all processing branches are needed. This single-bin screen is exploratory and is not a robustness test of the final 125-bin model. The final decoder requires local features, so arbitrary branch removals do not preserve its architecture; context-removal comparisons and the distance-branch seed repeat provide the relevant checks (Section 4.4, Table 4). Input-modality removal addresses a separate question about the information available to the predictor.

Reported fold SD cannot be used as the uncertainty of a difference between two models evaluated on a single fold. Such a comparison requires paired measurements for the competing models across the same folds and seeds.

Table 8: Training-center comparison on complete chr19 signals, paired across five folds. The all-center row uses the main model; the tile-grid row uses MSE alone. Their difference includes the change in training objective, unlike the matched comparison in Figure 4.

Eligible training centers	Model r_{pos}	Model r_{cell}	Model r_{res}	Folds improved
All eligible positions (ours)	0.737 ± 0.028	0.192 ± 0.020	0.271 ± 0.028	5/5
Inference tile grid only	0.732 ± 0.027	0.143 ± 0.021	0.232 ± 0.021	—

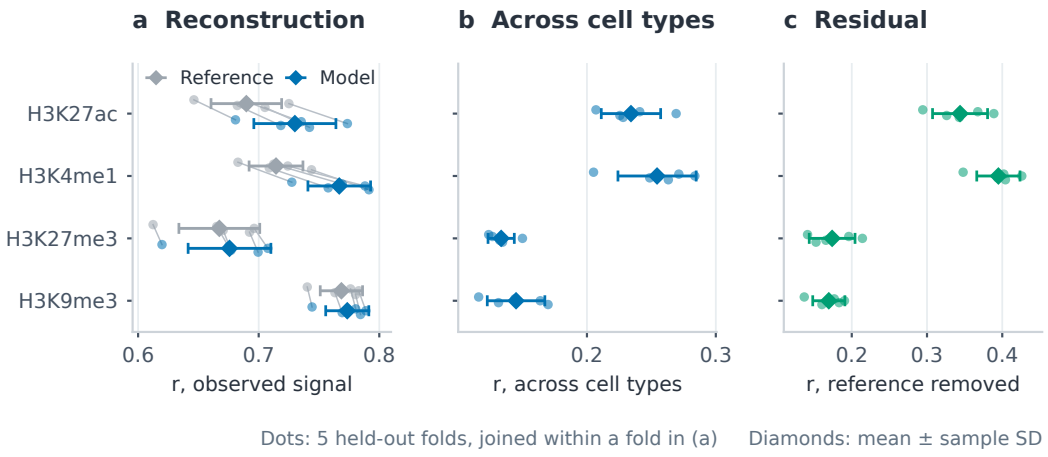

Figure 3: Five-fold prediction of complete chr19 signals by the MSE-only model. (a) Correlation with the observed signal across positions, for the training-reference mean and for the model; thin lines join paired values from the same fold. (b) Correlation across cell types at each position. (c) Correlation after the training-reference mean is removed from both prediction and observation. Dots show individual folds; diamonds and whiskers show the mean and sample SD across five folds. Active marks show higher cell-type specificity than repressive marks, and the ordering in (a) does not match the ordering in (b) or (c).

Table 9: Exploratory center-density sweep (one fold, seed 0, MSE-based stopping). Test scores use center-only predictions on 50 landmarks per mark.

Configuration	J	Center stride	Best val. r_{res}	Test r_{res}	Test r_{cell}
Single-bin reference	1	1	0.2586	0.2091	0.1604
Sample all centers	125	1	0.2547	0.2131	0.1645
Sample stride-5 grid	125	5	0.2540	0.2144	0.1649
Sample stride-25 grid	125	25	0.2273	0.1963	0.1572
Sample fixed tiles	125	125	0.2001	0.1776	0.1479

Table 10: Processing-branch screen in the earlier single-bin model (one fold). RNA and DNA are retained in every variant. Epoch times are from A800 80GB runs.

Epigenomic branches	Val. r_{res}	Test r_{res}	Test r_{cell}	s/epoch
Local + global + distance	0.2586	0.2091	0.1604	648
Distance only	0.2606	0.2022	0.1591	287
Local + distance	0.2590	0.2105	0.1583	513
Local + global	0.2368	0.1945	0.1562	638

Table 11: Whole-chromosome chr19 results by mark, mean ± sample SD across five folds.

Mark	Model r_{pos}	Model r_{cell}	Model r_{res}
H3K27ac	0.730 ± 0.034	0.234 ± 0.023	0.344 ± 0.037
H3K4me1	0.767 ± 0.026	0.254 ± 0.030	0.395 ± 0.029
H3K27me3	0.676 ± 0.034	0.133 ± 0.010	0.174 ± 0.030
H3K9me3	0.773 ± 0.018	0.145 ± 0.022	0.169 ± 0.021

Table 12: Whole-chromosome chr19 results in each held-out fold, averaged over four marks. Test cell types are disjoint across folds.

Fold	Train/val/test	Model r_{pos}	Model r_{cell}	Model r_{res}
0	75/19/24	0.6931	0.1630	0.2301
1	75/19/24	0.7400	0.1957	0.2856
2	75/19/24	0.7645	0.2167	0.3048
3	76/19/23	0.7306	0.1845	0.2673
4	76/19/23	0.7544	0.1989	0.2650

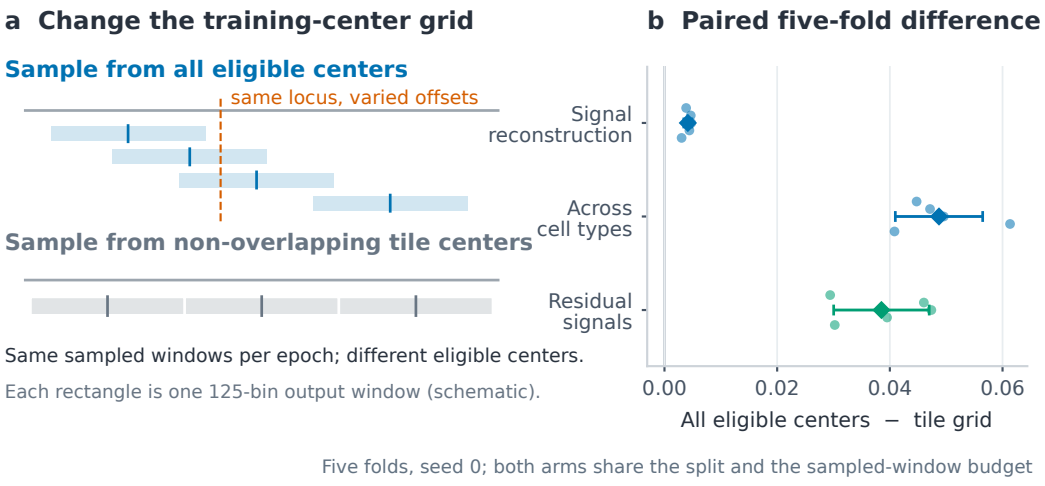

Figure 4: Training-center coverage in a matched MSE-only comparison. (a) Sampling throughout the eligible chromosome exposes a locus at different output offsets; a tile-aligned grid restricts these contexts. (b) Paired difference between the two arms on the complete stitched chr19 signal, one dot per held-out fold, diamonds and whiskers showing the mean and sample SD of the five paired differences. Both arms share the cell-type splits, the seed, the optimization settings and the 436,700 sampled windows per epoch, and differ only in the set of eligible centers. All five folds fall on the same side of zero for all three measures.

D REPRESENTATIVE REGIONS

These MSE-only model diagnostics cover 24 windows spanning four marks, three broad cell classes, and two peak-strength strata (Table 13). Selection used measured peak tables and fold-0 training cell types before inspecting predictions. Non-neuronal peak calls were mapped from finer supertype labels to the cell-type resolution used here using the atlas annotation. Windows are 200 kb long and do not overlap. Group labels describe the source of the training peak used to locate a window, not a restriction on the test cell types evaluated there.

Cell-type counts and correlation. The shared profile uses 75 training cell types. At each of the 1,000 original 200 bp bins in each selected window, $r_{\mathrm{cell}}(p)$ is calculated across all 24 fold-0 test cell types; all 24 are valid at these positions. A negative value means that predictions and measurements covary in opposite directions across these cell types at that locus. Subtracting the same training mean from every cell type does not change this correlation. Undefined values remain missing, and negative correlations are neither clipped nor removed. Table 13 reports, for each window, the mean of $r_{\mathrm{cell}}(p)$ over its 1,000 bins (r_{cell}) and r_{pos} averaged over the 24 test cell types. Figure 5 shows the unfiltered curve for the preselected H3K27ac NN high-strength window as a detailed example.

The top rows of Figure 5 show two fixed cell types, `Astro_NT_NN` and `IT_AON_TT_DP_Glut`. The heatmaps use the same cell-type order for measurements and predictions:

1: `ABC_NN`; 2: `AVPV_MEPO_SFO_Tbr1_Glut`; 3: `Astro_NT_NN`; 4: `BST_MPN_Six3_Nrgn_Gaba`; 5: `CA1_ProS_Glut`; 6: `CEA_AAA_BST_Six3_Sp9_Gaba`; 7: `CEA_BST_Ebf1_Pdyn_Gaba`; 8: `CLA_EPd_CTX_Car3_Glut`; 9: `COAa_PAA_MEA_Barhl2_Glut`; 10: `CS_PRNr_DR_En1_Sox2_Gaba`; 11: `DMH_LHA_Gsx1_Gaba`; 12: `DMH_LHA_Vgll2_Glut`; 13: `ENTmv_PA_COAp_Glut`; 14: `GPi_Tbr1_Cngb3_Gaba_Glut`; 15: `IF_RL_CLI_PAG_Foxa1_Glut`; 16: `IPN_LDT_Vsx2_Nkx6_1_Glut`; 17: `IPN_Otp_Crisp1_Gaba`; 18: `IT_AON_TT_DP_Glut`; 19: `IT_EP_CLA_Glut`; 20: `L2_3_IT_ENT_Glut`; 21: `LHA_AHN_PVH_Otp_Trh_Glut`; 22: `Lamp5_Gaba`; 23: `MB_MY_Tph2_Glut_Sero`; 24: `NDB_SI_MA_STRv_Lhx8_Gaba`.

Table 13: All pre-selected windows and their window-level correlations. Coordinates are chr19 Mb (start inclusive, end exclusive). For each of the 24 fold-0 test cell types, r_{pos} correlates predicted and measured signals across the window's 1,000 bins of 200 bp and is averaged over cell types; r_{cell} correlates the 24 cell types at each bin and is averaged over bins. These regional diagnostics are not chromosome-wide scores.

Mark	Group	Strength	Start	End	r_{pos}	r_{cell}
H3K27ac	Glut	median	5.9376	6.1376	0.723	0.140
H3K27ac	Glut	high	3.6190	3.8190	0.717	0.241
H3K27ac	Gaba	median	12.3144	12.5144	0.788	0.292
H3K27ac	Gaba	high	15.5146	15.7146	0.509	0.181
H3K27ac	NN	median	10.1074	10.3074	0.649	0.255
H3K27ac	NN	high	5.7232	5.9232	0.840	0.217
H3K4me1	Glut	median	42.0044	42.2044	0.752	0.219
H3K4me1	Glut	high	5.1730	5.3730	0.709	0.123
H3K4me1	Gaba	median	31.9030	32.1030	0.598	0.128
H3K4me1	Gaba	high	4.7592	4.9592	0.702	0.133
H3K4me1	NN	median	21.7342	21.9342	0.722	0.214
H3K4me1	NN	high	4.0952	4.2952	0.748	0.172
H3K27me3	Glut	median	11.7898	11.9898	0.570	0.149
H3K27me3	Glut	high	27.5604	27.7604	0.523	0.085
H3K27me3	Gaba	median	4.3256	4.5256	0.601	0.131
H3K27me3	Gaba	high	7.6116	7.8116	0.529	0.056
H3K27me3	NN	median	41.6744	41.8744	0.494	0.269
H3K27me3	NN	high	4.9604	5.1604	0.564	0.205
H3K9me3	Glut	median	30.7190	30.9190	0.620	0.120
H3K9me3	Glut	high	28.1586	28.3586	0.612	0.083
H3K9me3	Gaba	median	14.6696	14.8696	0.493	0.112
H3K9me3	Gaba	high	54.7078	54.9078	0.548	0.111
H3K9me3	NN	median	40.8606	41.0606	0.630	0.071
H3K9me3	NN	high	20.2396	20.4396	0.656	0.121

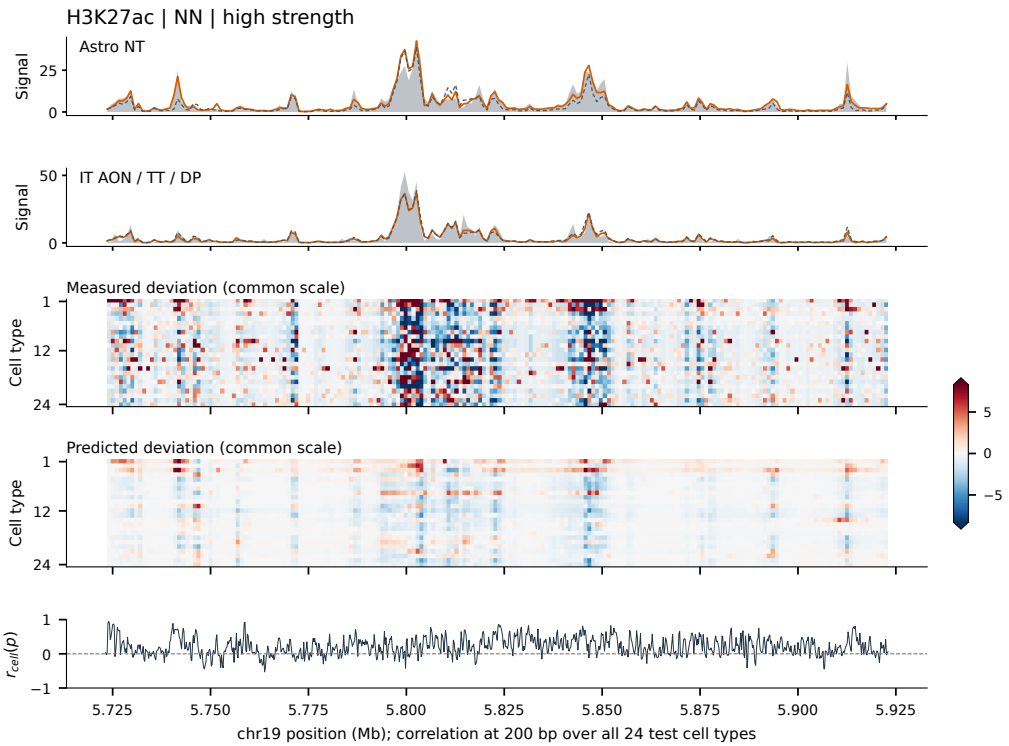

Figure 5: Individual H3K27ac diagnostic, chr19:5,723,200–5,923,200. The top rows show the two fixed cell types named above. Measured and predicted deviations share a symmetric color scale (98th percentile of their pooled absolute values, saturated beyond the limits). Bottom: raw position-wise correlation across all 24 held-out cell types at 200 bp. Signals and heatmaps are displayed at 1 kb.