

Figure S1. CLIP model architectures: (a,b) eCLIP models. They differ in the way distance is used. In (a), only a single value (scalar) per input sequence is used. In (b) the distance is used as a vector along the sequence (one value per nucleotide). (c) iDeep model extended with spline transformation. Blue rectangles denote the input data modality, yellow rectangles denote the pre-processing function and white rectangles denote the neural-network layers using Keras notation. Batch normalization and dropout layers are not shown here but are specified in the Methods section.



Figure S2. Relative genes positions for 112 RBPs measured by eCLIP. RBPs are sorted by the t-test statistic comparing relative genes positions of binding to non-binding sites.



Figure S3. Training time of models shown in Figure 2b using 4 virtual threads of an Intel Core i7-6700 CPU.

	(gene TSS	transcr. TSS	start codon	exon-intron	intron-exon	stop codon	transcr. polyA	gene polyA		
0000	.75 .50 .25 .00		A	$ \land \land$			Λ		\bigwedge	UPF1	
	0.6 - 0.4 - 0.2 - 0.0 -	\square	$ \land \land $			AA	AA	An	Λ	SUB1	Highest improvement with distance
	0.6 0.4 0.2 0.2	Λ	\wedge	$ \land \land$	\wedge	AA	AA	\wedge	Λ	IGF2BP1	
	0.6 - 0.4 - 0.2 -	\land	\sim	$\sim \Lambda$	AA	A	\wedge	\wedge	\bigwedge	FXR1	
	0.6 0.4 0.2 0.2	Λ	$ \land \land$				AA	A	Λ	DDX55	
	0.6 - 0.4 - 0.2 - 0.0 -	\land	$ \land \land$				AA	A	\wedge	LARP4	
	0.6 - 0.4 - 0.2 - 0.0 -	\land	$ \land \land$	\sim			AA		\bigwedge	DDX6	
	0.6 0.4 0.2 0.0	\land	\sim				\wedge	\wedge	\bigwedge	MTPAP	
	0.6 0.4 0.2 0.0	A	\sim				Λ	Λ	\bigwedge	FTO	
	321 21 0						~			RBM22	
	0.6 0.4 0.2 0.2	A			AA		Λ_{a}	Λ	\bigwedge	DDX3X	
	0.6 0.4 0.2 0.0	Λ	\wedge	$\sim \Lambda$		AA	AA	A	Λ	PUM2	
	0.6 - 0.4 - 0.2 - 0.0 -	A	$ \wedge $	$\overline{\mathbf{A}}$		$\wedge \wedge$	Λ_{Λ}	$\Lambda_{}$	\bigwedge	NKRF	
	0.6 - 0.4 - 0.2 - 0.0 -	$ \land $		$\overline{\Lambda}$			Λ	\wedge	\bigwedge	HNRNPU	
	0.6 0.4 0.2 0.0	$ \land$		$\overline{\Lambda}$		$\wedge \wedge$	A	\wedge	\bigwedge	TARDBP	
	0.6 - 0.4 - 0.2 - 0.0 -	$ \land$		$\overline{\Lambda}$		\land	A	\wedge	\bigwedge	QKI	
	0.6 - 0.4 - 0.2 - 0.0 -	A		$\overline{\Lambda}$	$ \land \land$		Λ_{Λ}	\wedge	\bigwedge	GTF2F1	
	0.6 - 0.4 - 0.2 - 0.0 -	$ \land$					AA	Λ	\bigwedge	PCBP2	
	0.6 - 0.4 - 0.2 - 0.0 -	\square		$\overline{\Lambda}$			A	\wedge	\bigwedge	HNRNPK	
	0.6 - 0.4 - 0.2 - 0.0 -	$ \land$		$\overline{\Lambda}$			Λ	\wedge	\bigwedge	HNRNPC	
	0.6 0.4 0.2 0.2	\bigwedge		$\overline{\Lambda}$			Λ	\wedge	\bigwedge	HNRNPM	
	0.6 0.4 0.2 0.2	Λ		$\overline{\Lambda}$			\wedge	\wedge	\bigwedge	SUGP2	
	0.4 0.2 0.0	Λ					\wedge	\wedge	\bigwedge	SFPQ	
		-4 0 4	-4 0 4	-4 0 4	-4 0 4	-4 0 4	-4 0 4	-4 0 4	-4 0 4		

Signed log10 distance to landmark

Figure S4. Relative distance to all 8 considered genomic landmarks for eCLIP peaks. Only RBPs labelled in Figure 2c are shown.



Figure S5. Test-accuracy (auPR) comparing two different variants of the eCLIP DNN w/ dist model: the main model using a single scalar per input sequence (y-axis, architecture shown in Figure S1a) and the alternative model using a vector of distances alongside the sequence (x-axis, architecture shown in Figure S1b). Black represents statistically significant difference (P < 0.0001, Wilcoxon test on 200 bootstrap samples, Bonferroni correction for multiple testing). Histogram within the scatterplot shows a distribution of performance differences and the average value.



Figure S6. Relative test-accuracy (auPR) decrease when excluding individual features from the eCLIP DNN w/ dist model. A relative decrease of 1 means that the test-accuracy decreased to the accuracy of the sequence-only model (DNN) and a relative decrease of 0 means the test-accuracy remained unchanged (i.e. equals the DNN w/ dist model accuracy). Top rows in purple shows the auPR of the DNN, DNN w/ dist model and their difference. Darker red denotes a higher auPR. Relative decrease variations for the 11 rightmost RBPs should be interpreted with care since both the overall performance and the added value of including distance is low.



Figure S7. Fraction of branchpoints per position for the remaining seven features in log-odds scale (black dot, outlier shown in red) compared to the shallow NN model fit: inferred spline transformation (orange) and predicted fraction of branchpoint per position (blue).



Figure S8. Training time per epoch of models shown in Figure 4e using 4 virtual threads of an Intel Core i7-6700 CPU.