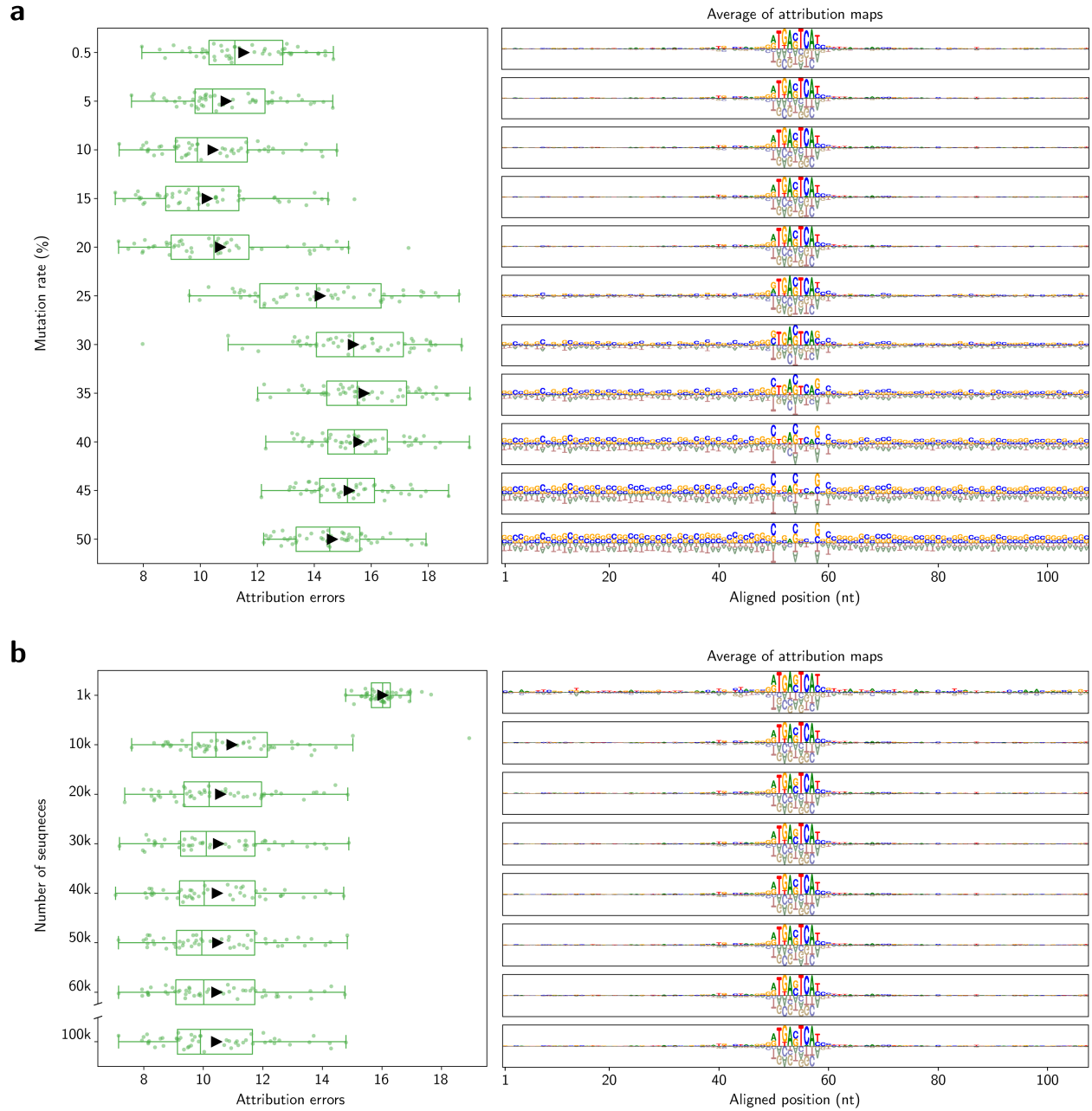
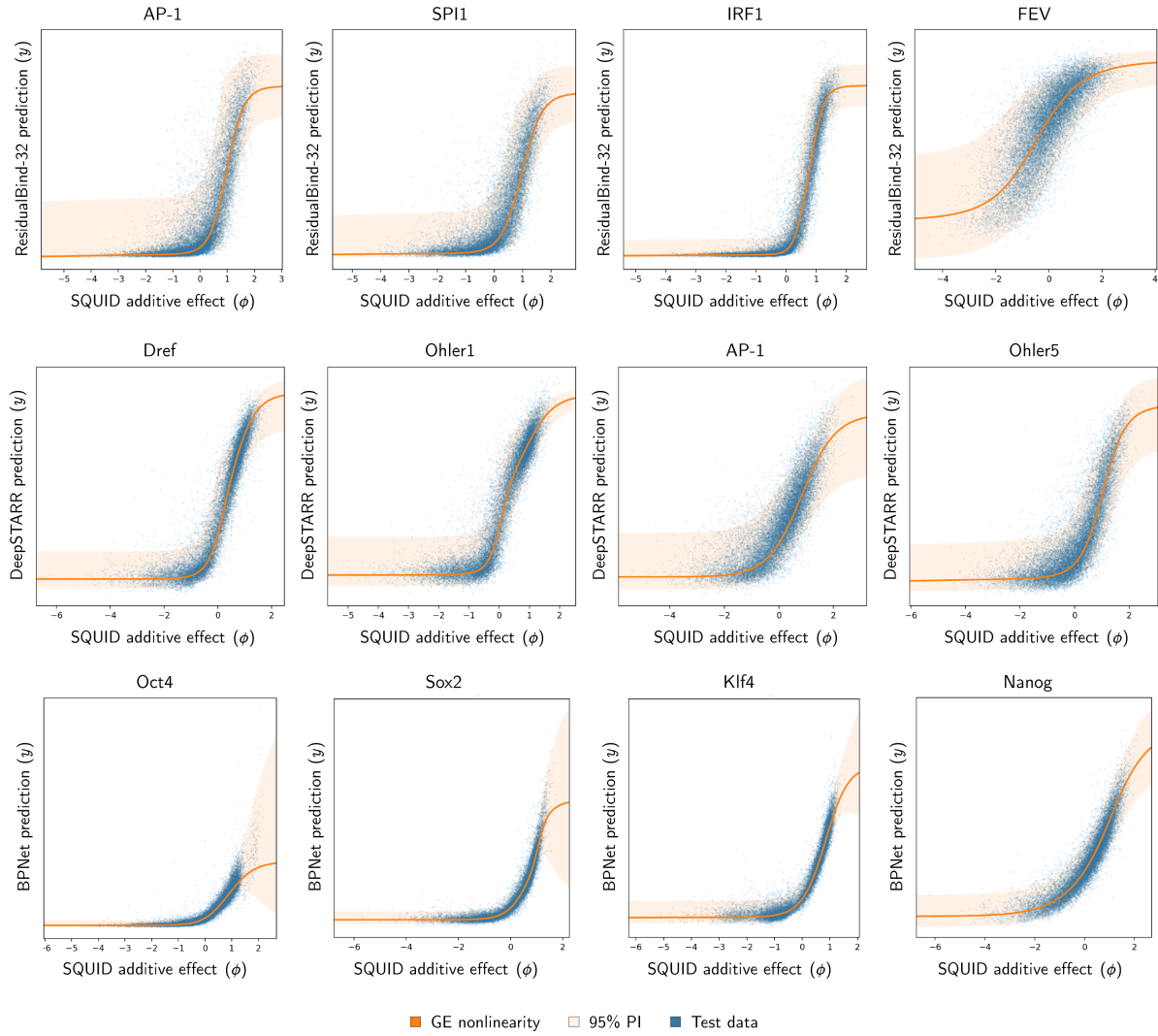


TF name	consensus seq.	DNN name	DNN task	PWM ID (database)
AP-1	TGAGTCA	ResBind-32	PC-3	MA0476.1 (Jaspar)
AP-1 / AP-1	TGANTCA ... TGANTCA	ResBind-32	PC-3	n/a
IRF1	TGAAAC	ResBind-32	GM12878	n/a
IRF1-long	AANTGAAAC	ResBind-32	GM12878	MA0050.1 (Jaspar)
SPI1	GGAAGT	ResBind-32	GM12878	n/a
FEV	CCGGAA	ResBind-32	HCT116	n/a
AP-1	TGACTCA	DeepSTARR	Dev	MA0476.1 (Jaspar)
Dref	TATCGATA	DeepSTARR	Hk	M00230 (Homer)
Ohler1	AGTGTGACC	DeepSTARR	Hk	M00232 (Homer)
Ohler5	CAGCTG	DeepSTARR	Hk	n/a
Oct4	TTTGCAT	BPNet	Oct4	n/a
Sox2	GAACAATAG	BPNet	Sox2	H12CORE.0.P.B (Hocomoco)
Klf4	GGGTGTGGC	BPNet	Klf4	n/a
Nanog	AGCCATCAA	BPNet	Nanog	H12CORE.1.P.B (Hocomoco)
Nanog / Sox2	AGCCATCAA ... GAACAATAG	BPNet	Nanog	n/a

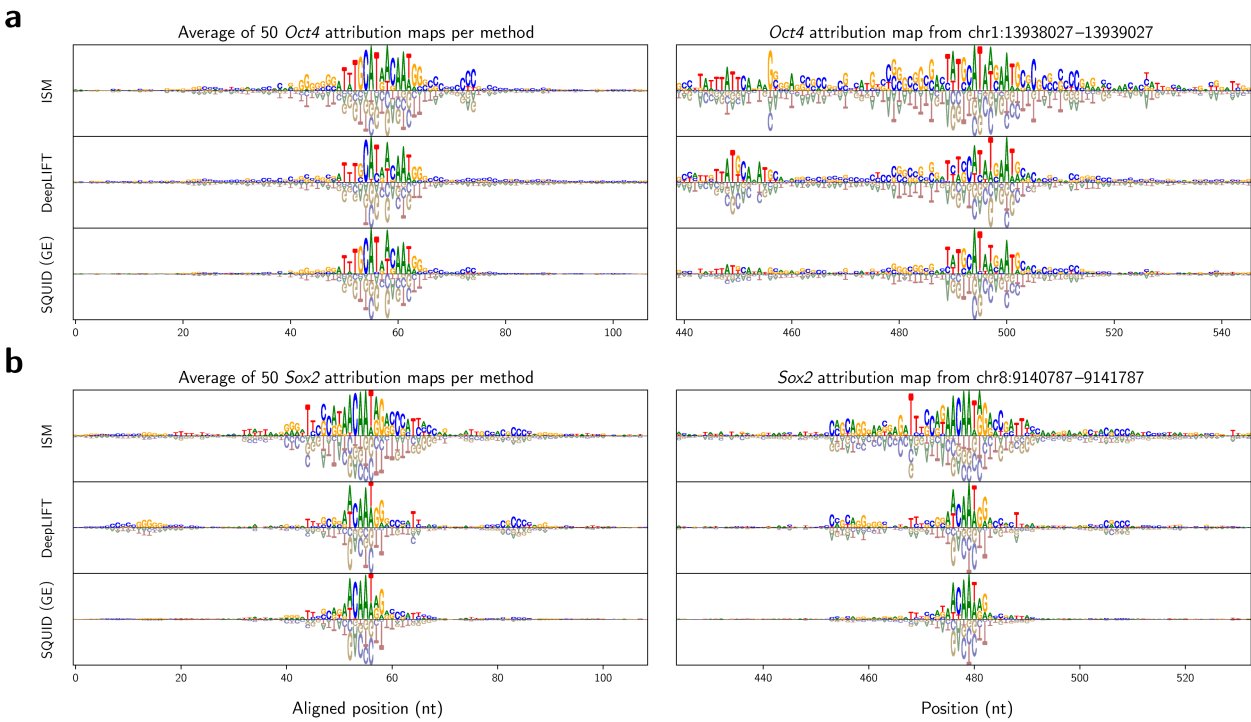
**Supplementary Table 1. TFs analyzed in our study.** Shown for each TF is the consensus sequence used, the DNN that models the TF, the DNN prediction task to which attribution methods were applied, and PWMs used to investigate weak binding sites. TF, transcription factor; DNN, deep neural network; PWM, position weight matrix.



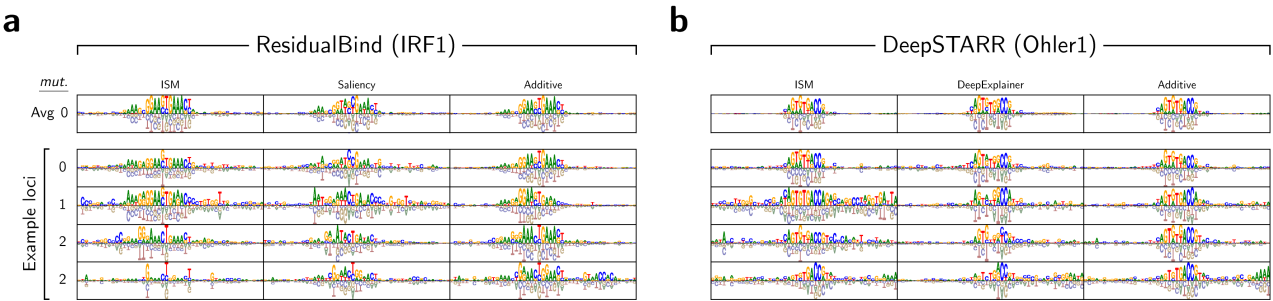
**Supplementary Figure 1. Influence of mutation rate and library size on SQUID attribution maps.** **a** and **b**, attribution errors (*left*) and average attribution maps (*right*) found for 50 SQUID attribution maps computed for the TF AP-1 as in Fig. 2, but using *in silico* MAVE libraries having **(a)** variable mutation rate  $r$  and fixed size  $N = 100,000$ , or **(b)** fixed mutation rate  $r = 10\%$  and variable size  $N$ . All SQUID attribution maps were computed using additive models with GE nonlinearities, followed by cropping these maps using a flank size of  $n_f = 50$  nt. TF, transcription factor; MAVE, multiplex assay of variant effect; GE, global epistasis.



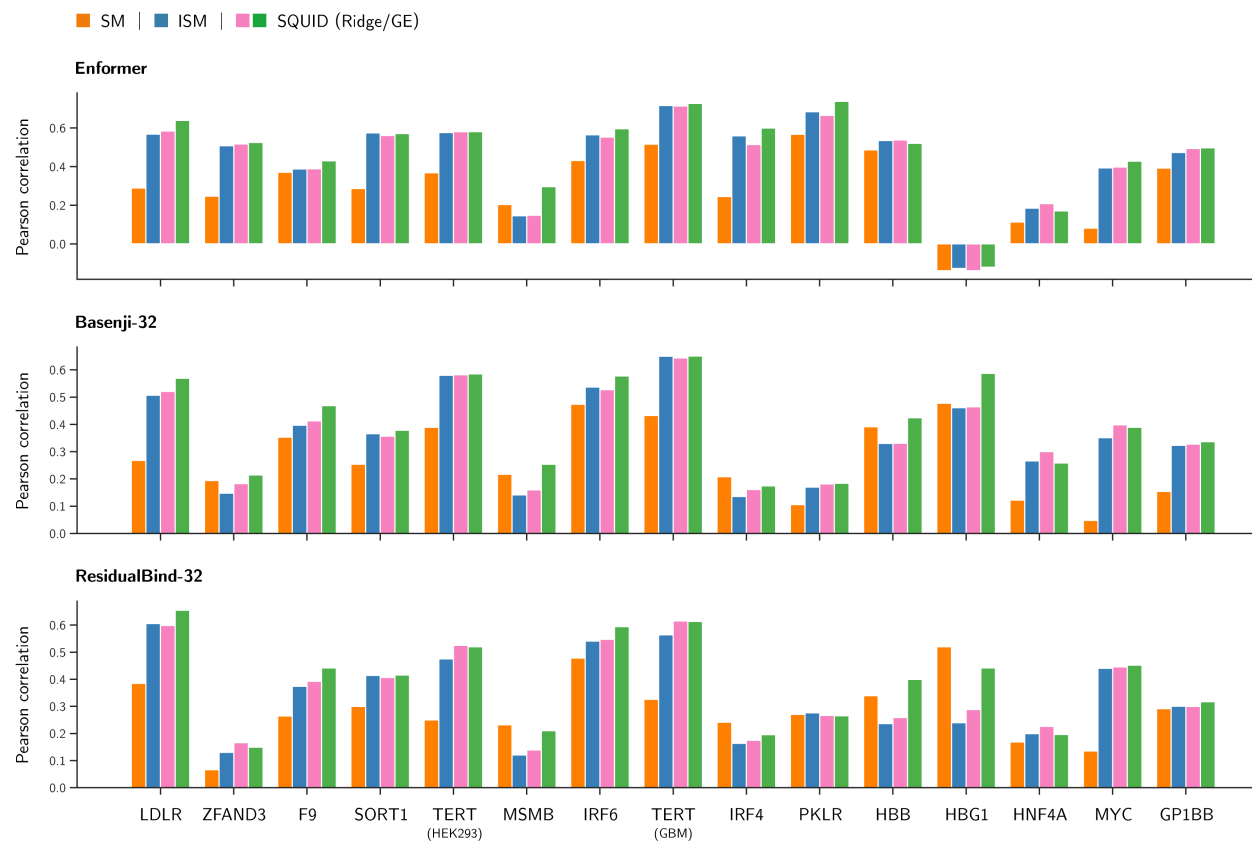
**Supplementary Figure 2. Nonlinearities and noise across DNNs and TFs.** Examples of the GE nonlinearities and heteroscedastic noise models inferred in the SQUID (GE) analyses performed for Fig. 3. Each plot shows results for a representative sequence from the 50 sequences analyzed for each combination of DNN and TF. DNN, deep neural network; TF, transcription factor; GE, global epistasis; PI, prediction interval.



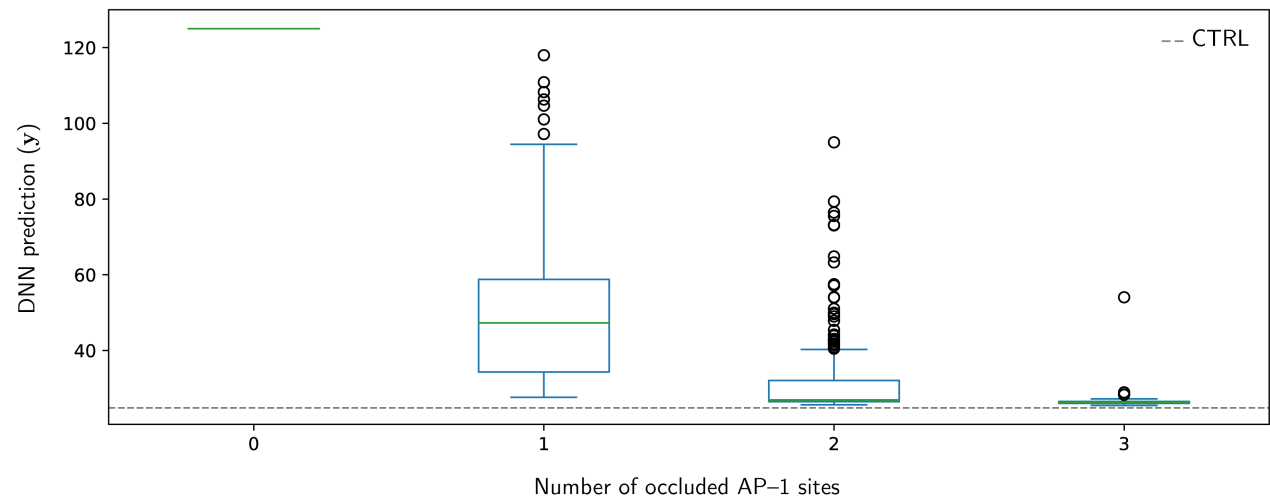
**Supplementary Figure 3. Average and example binding motifs for Oct4 and Sox2. a,** Oct4 motifs, centered on the putative binding site TTTGCAT. **b,** Sox2 motifs, centered on the putative binding site GAACAATAG. TF binding motifs are from attribution maps computed for BPNet and plotted as in Fig. 3c. TF, transcription factor.



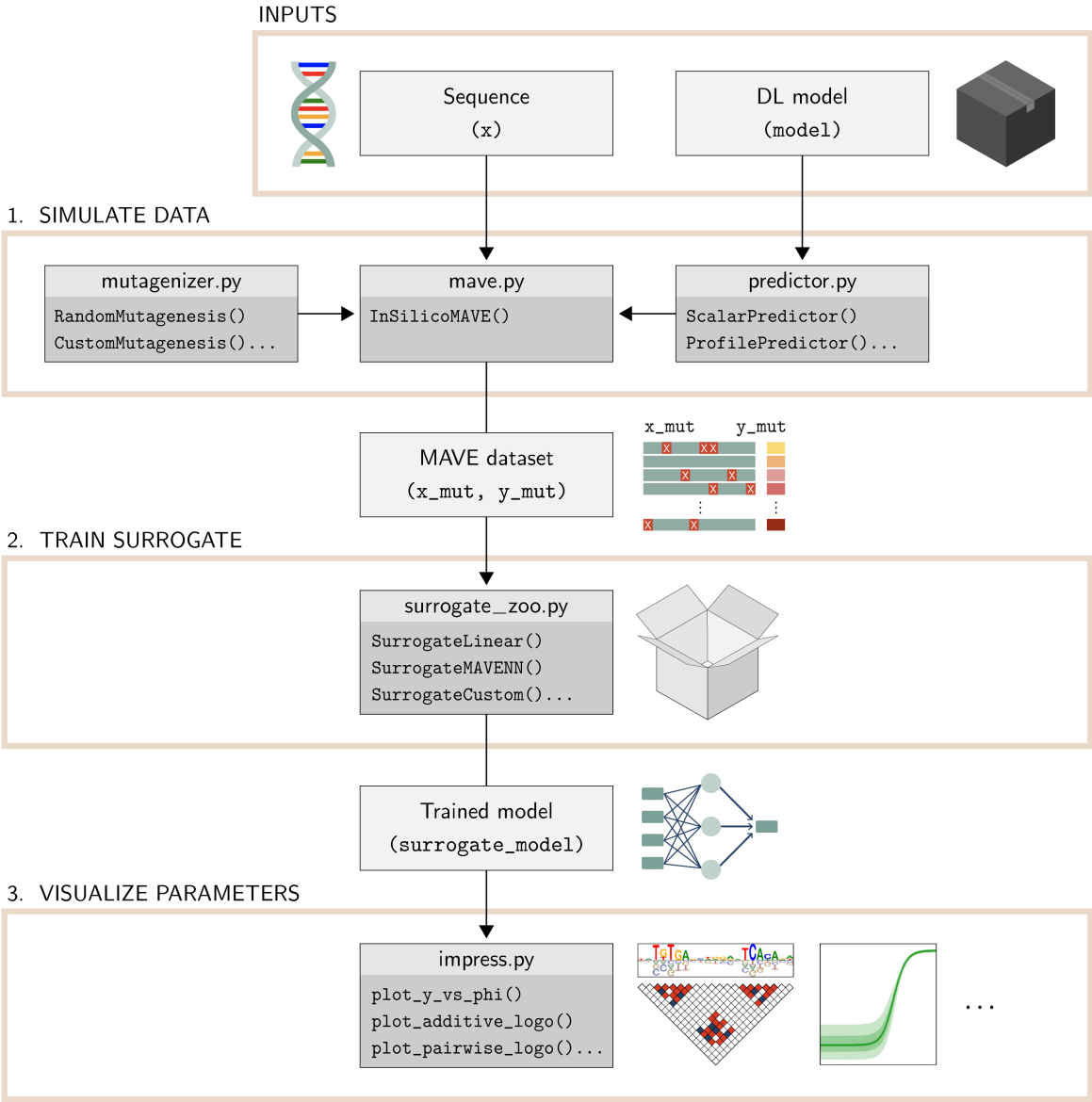
**Supplementary Figure 4. Attribution maps computed for strong and weak TF binding sites.** Top row shows the average of 50 attribution maps in the 0-mutation ensemble computed for (a) IRF1 using ResidualBind-32, and for (b) Ohler using DeepSTARR. Remaining rows show attribution maps for four representative genomic loci with the central putative binding site having varying numbers of mutations from the consensus binding site. TF, transcription factor.



**Supplementary Figure 5. Performance of attribution methods at predicting variant effects at individual loci.** Pearson correlation scores for each of the 15 disease-associated loci assayed in CAGI5, computed for the attribution methods and DNNs listed in Table 1. DNN, deep neural network.

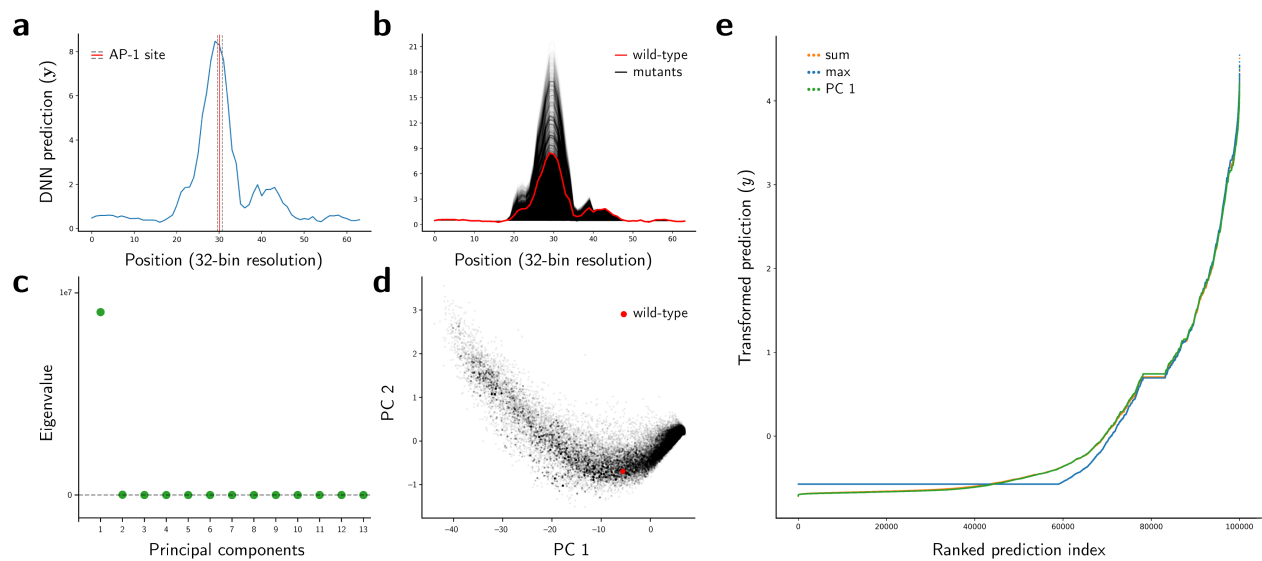


**Supplementary Figure 6. Occlusion analysis of AP-1 binding site effects.** Occlusion study based on the wild-type sequence investigated in Fig. 6d and 6e. Occlusions were performed 100 times for every combination of one, two, or three occluded motifs, with the DNN prediction taken independently for each instance. In each occluded sequence, the corresponding AP-1 core (7-mer) sites were scrambled using a uniform probability of nucleotides at each position. The baseline score (CTRL) was calculated from the median of predictions corresponding to 100 instances of a dinucleotide shuffle over the full (2048-nt) sequence. The DNN prediction rapidly approaches the genomic baseline as additional binding sites are occluded.



**Supplementary Figure 7. SQUID workflow.** Flowchart representing a typical DNN interpretation analysis pipeline using SQUID. DNN, deep neural network.





**Supplementary Figure 8. Dimensionality reduction of DNN predictions using PCA.** **a**, Example ATAC-seq profile predicted by ResidualBind-32 for a representative sequence of interest containing a putative AP-1 binding site. The profile shown was cropped to a region spanning the putative site and 30 nt of flanking DNA on either side. **b**, Profiles computed for sequences in the *in silico* MAVE dataset generated by SQUID when analyzing the sequence of interest from panel **a**. **c**, Ranked eigenvalues from a PCA analysis of the profiles in panel **b**. **d**, Projection of profiles onto the first two principal components. **e**, Scalar predictions  $y$  for three projection methods: PCA, sum, and max. PCA projections were computed by projecting profiles onto the first principal component. Sum projections were computed by summing the entries in each profile. Max projections were computed by taking the maximum entry in each profile. To aid comparisons between different projection methods, the  $y$  values for each method were centered about zero and rescaled to have unit standard deviation. The flat region observed near ranked prediction index 80,000 results from sequences in the *in silico* MAVE library that have no mutations. PCA, principal component analysis.