A surrogate modeling framework for interpreting deep neural networks in functional genomics

Evan E Seitz[1], David M McCandlish[1], Justin B Kinney[1*], Peter K Koo[1*]

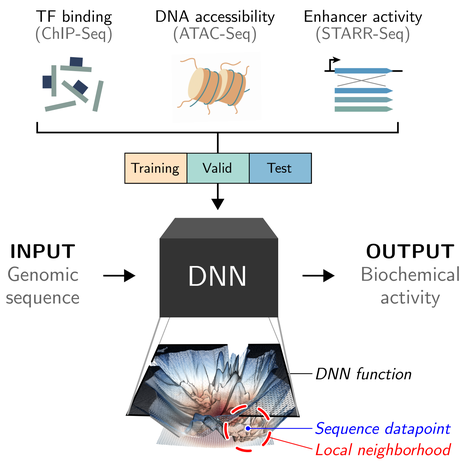
[1]Simons Center for Quantitative Biology,
Cold Spring Harbor Laboratory, 1 Bungtown Rd,
Cold Spring Harbor, 11724, NY, USA.

Cold Spring Harbor Laboratory

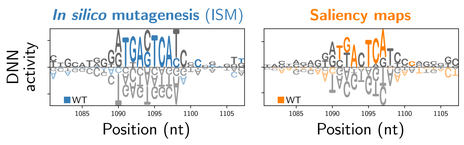
SQUID — Surrogate Quantitative Interpretability for Deepnets

1 Background

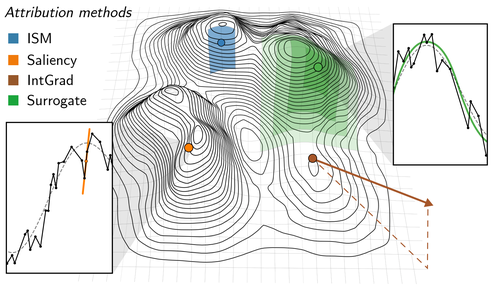
Sequence-based deep neural networks (DNNs) for functional genomics predict molecular phenotypes associated with transcriptional regulation from primary sequences. Although DNNs learn predictive functions, it remains difficult to uncover their learned features, such as binding motifs and regulatory grammars.



Attribution methods provide a feature importance score for each nucleotide in the given sequence, with a direct interpretation as single-nucleotide variant effects on model predictions.



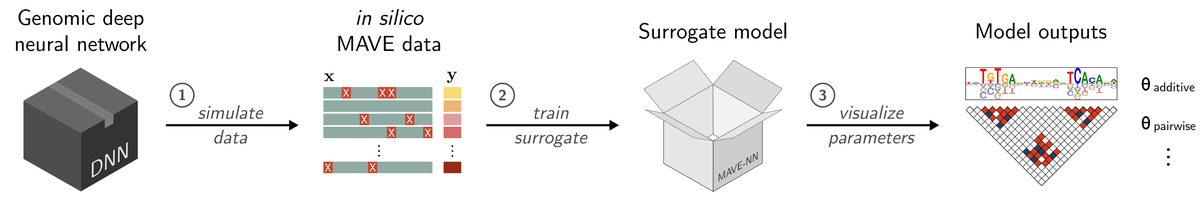
Attribution methods operate over a local neighborhood of the DNN-learned function to provide *post hoc* interpretation of DNN predictions for a specific sequence. One **limitation** is that different attribution methods provide different interpretations based on how they characterize the local function.



*Schematic of DNN-learned function over local region of sequence space, depicting the **neighborhood sizes** considered by different attribution-based approaches.*

2 SQUID: Surrogate Quantitative Interpretability for Deepnets

SQUID approximates user-defined regions of sequence space with flexible surrogate models that have mechanistically-interpretable parameters.



1. Generate *in silico* multiplex assays of variant effect (MAVE) data by mutagenizing a sequence and using the DNN as a scoring function (i.e., functional readout)
2. Fit the data with mechanistically-interpretable surrogate model designed to fit the MAVE data
3. Visualize parameters for interpretation of *cis*-regulatory mechanisms

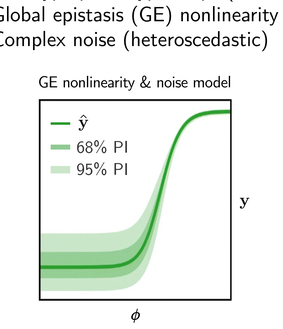
```
pip install squid-nn
```

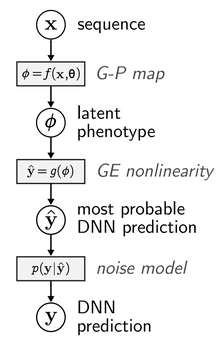
https://squid-nn.readthedocs.io

Domain-specific surrogate models

Key features of MAVE-NN:

1. Genotype-phenotype maps (additive, pairwise, higher-order)
2. Global epistasis (GE) nonlinearity
3. Complex noise (heteroscedastic)

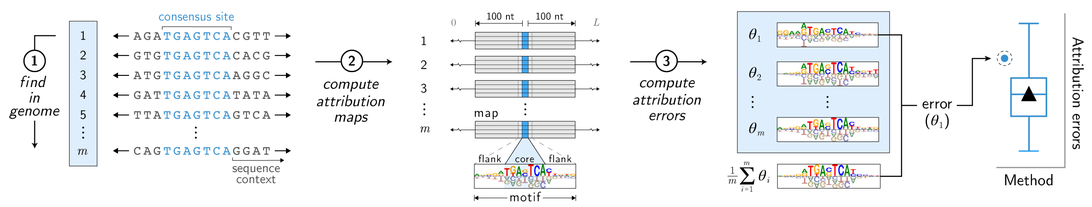


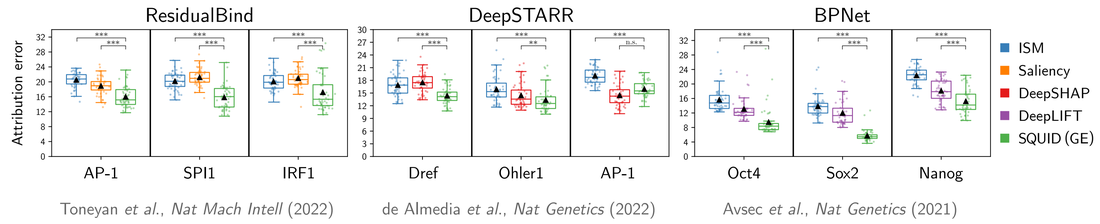


Tareen *et al.*, *Genome Biol* (2022)

3 SQUID yields more consistent motif representations across genomic loci

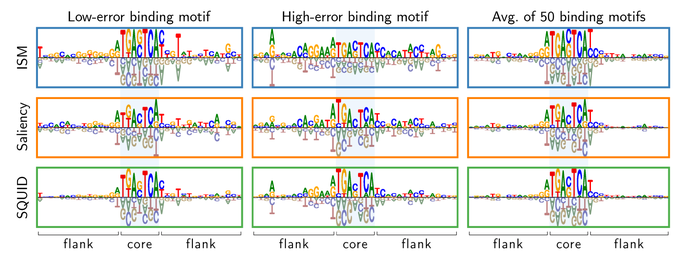
Experiment 1: Compare the consistency of binding motifs identified by different attribution methods.

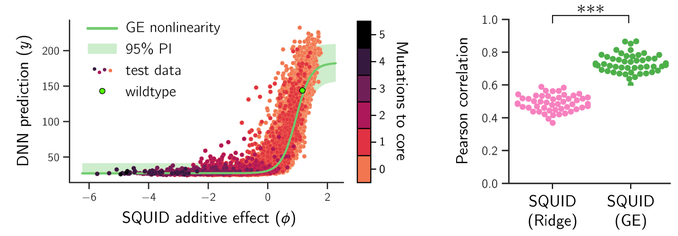




Toneyan *et al.*, *Nat Mach Intell* (2022) de Almeida *et al.*, *Nat Genetics* (2022) Avsec *et al.*, *Nat Genetics* (2021)

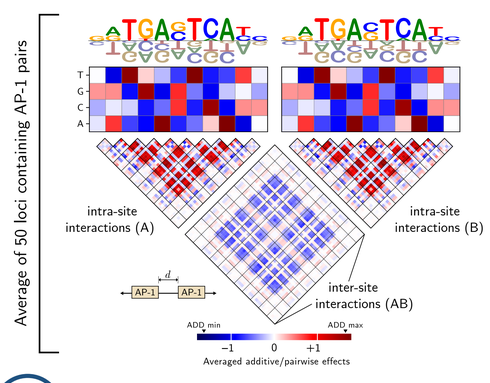
Visualizing motif consistency

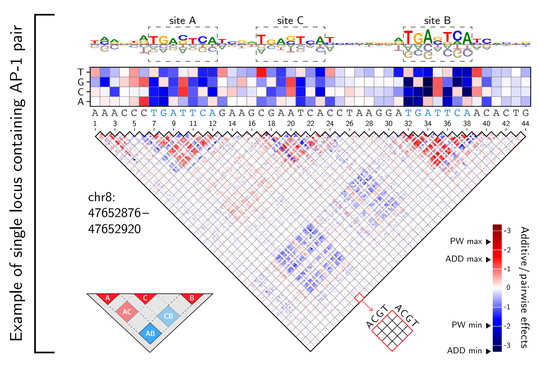




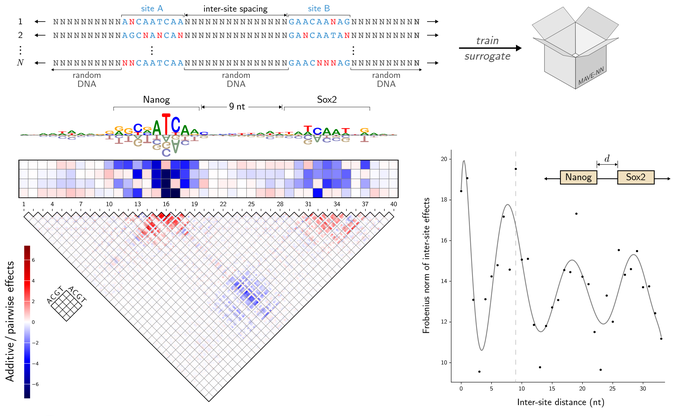
4 SQUID extensions provide different insights for interpreting DNNs

Experiment 2: Implement pairwise-interaction surrogate models to quantify the effects of pairs of binding sites.





Global DNN interpretations

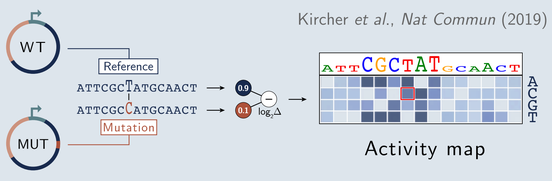


5 SQUID improves zero-shot variant effect predictions

CAGI5 challenge measures the effects of single nucleotide variants (SNVs) in massively parallel reporter assays (MPRA)

For each of the 15 300–600-bp enhancer/promoter sequences:

Kircher *et al.*, *Nat Commun* (2019)



The varient effect of each SNV is measured to form an **activity map**, similar to the attribution maps produced via ISM.

Experiment 3: Closer to *ground truth*, we next evaluated the performance of attribution-based methods to recapitulate experimental SNV measurements made in the **CAGI5 challenge** dataset.

	DNN		Average Pearson correlation				Statistical significance		
	Architecture	Task	Saliency	ISM	SQUID (Ridge)	SQUID (GE)	GE vs. Saliency	GE vs. ISM	GE vs. Ridge
Avsec *et al.*, *Nat Methods* (2021)	Enformer	DNASE	0.2977	0.4498	0.4486	**0.4801**	***	**	**
Toneyan *et al.*, *Nat Mach Intell* (2022)	Basenji-32	ATAC-seq	0.2727	0.3575	0.3700	**0.4036**	***	***	**
	ResBind-32	ATAC-seq	0.2846	0.3388	0.3567	**0.3912**	**	***	*

6 Conclusions

- SQUID leverages domain knowledge on how to characterize a regulatory genomic locus to flexibly interpret genomic DNNs
- SQUID identifies motifs more consistently across genomic loci, and yields improved variant effect predictions compared to existing attribution methods
- SQUID provides different mechanistic insights by swapping surrogate models (e.g., additive, pairwise)
- *In silico* MAVEs can be designed in different ways to address different biological questions (e.g., local, global)

bioRxiv — THE PREPRINT SERVER FOR BIOLOGY