Enhancer identification using machine learning enables efficient enhancer RNA targeting with antisense oligonucleotide-based therapeutics

Yuchun Guo, Guanjue Xiang, Salome Manska, Gabriel Golczer, Swathi Dhanasekaran, Bryan Matthews, Jenna Williams, Brynn Akerberg, Yun Joon Jung, Yuting Liu, David Bumcrot, and Alla Sigova

CAMP4 Therapeutics Corp., Cambridge, MA, USA

Motivation

- Antisense oligonucleotides (ASOs) that specifically target functional enhancers and their associated RNAs are being pursued to modulate disease-associated gene expression for a range of therapeutic indications.
- However, identifying functional enhancers and linking them to target genes remains a major challenge.

Approach

Our enhancer-promoter interaction characterization (EPIC) model is a machine learning model for predicting functional enhancer-promoter (E-P) pairs.

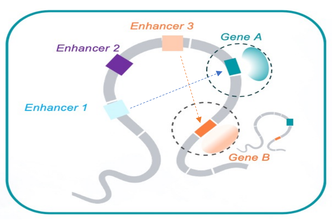
(Gasperini et al., 2019, *Cell*, 176:377-90; Xie et al., 2019, *Cell Reports*, 29: 2570-78)

K562 enhancer epigenomic and chromatin interaction features

K562 CRISPR-based enhancer perturbation screening

EPIC: a machine learning model

Infer



Enhancer 1 – Gene A: Yes
Enhancer 2 – Gene A: No
Enhancer 3 – Gene B: Yes
...

Basic features

- HiChIP.AnchorSize: AnchorSize = 5kb, 10kb, 15kb, or 20kb (n=4)
- Assay.Position.WindowSize, where Assay=ATAC, H3K27ac, H3K4me1, H3K4me3, EP300, CTCF, or Input ChIP; Position = Enh or TSS; WindowSize = 300bp, 500bp, 1kb, 2kb, or 4kb (n=7*2*5=70)
- Genomic distance (n=1)

Feature engineering

APMI = (ATAC.Enh.1kb * EP300.Enh.1kb * H3K4me1.Enh.4kb)$^{1/3}$ * HiChIP.5kb

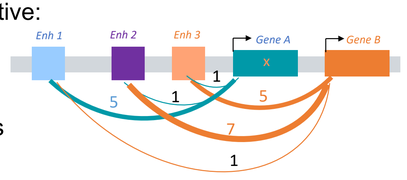
Based on APMI, we engineered a new set of features for quantifying the relative contribution of an enhancer ***e*** to a gene ***g*** from the gene perspective or enhancer perspective:

$$fracGene_{eg} = \frac{APMI_{eg}}{\sum_j APMI_{jg}}$$

where j indexes all the enhancers connected to gene ***g***.

$$fracEnh_{eg} = \frac{APMI_{eg}}{\sum_k APMI_{ek}}$$

where k indexes all the genes connected to enhancer ***e***.



	Enh1 - GeneA	Enh3 -GeneB
fracGene	$\frac{5}{1+1+5} = 0.71$	$\frac{5}{1+5+7} = 0.42$
fracEnh	$\frac{5}{1+5} = 0.83$	$\frac{5}{1+5} = 0.82$

In addition, we combined these features to form new features.

$$fracGmE_{eg} = fracGene_{eg} * fracEnh_{eg}$$

$$fracGpE_{eg} = fracGene_{eg} + fracEnh_{eg}$$

$$apmiGene_{eg} = fracGene_{eg} * APMI_{eg}$$

$$apmiEnh_{eg} = fracEnh_{eg} * APMI_{eg}$$

$$apmiGmE_{eg} = fracGmE_{eg} * APMI_{eg}$$

$$apmiGpE_{eg} = fracGpE_{eg} * APMI_{eg}$$

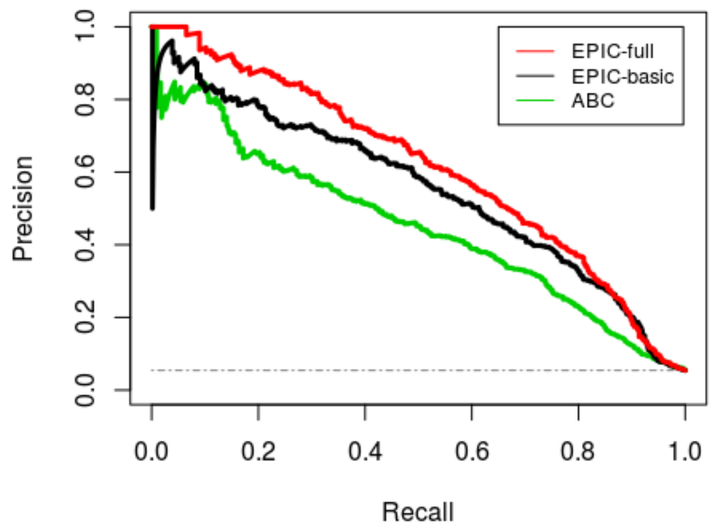
Machine learning model

- Random forest classification model trained on K562 data
- Five-fold cross-validation
- Genetic algorithm for feature selection

Results

EPIC outperforms ABC model in predicting enhancer-promoter pairs (holdout test data)

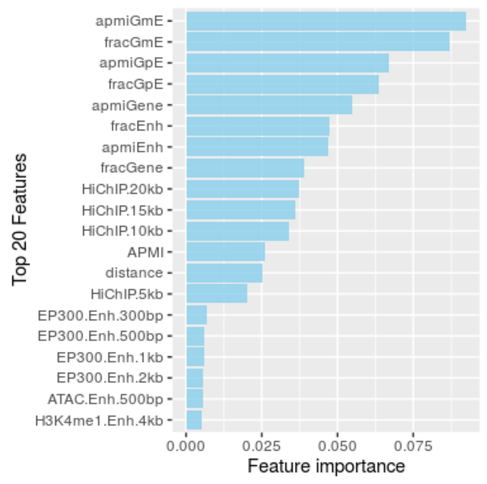
ABC = (ATAC.Enh.500bp * H3K27ac.Enh.500bp)$^{1/2}$ * HiC.5kb (Fulco et al., 2019, *Nat. Genet.*, 51:1664-9)



Model	AUPR	AUROC
EPIC-full	**0.613**	**0.918**
EPIC-basic	0.551	0.912
ABC	0.451	0.885

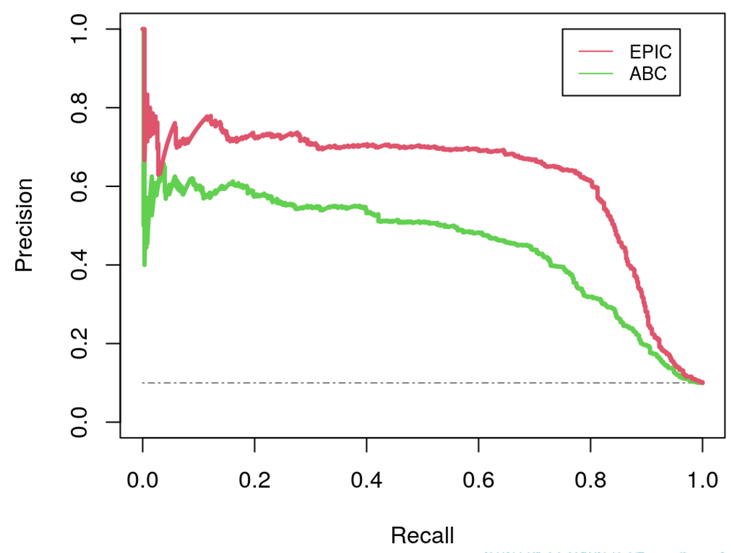
- The area under receiver operating characteristic (AUROC) curve of EPIC-full is significantly higher than that of ABC (p = 7.6e-11) (DeLong et al., 1988, *Biometrics,* 44:837-45).
- The AUROC of EPIC-full is significantly higher than that of EPIC-basic (p = 0.01), demonstrating the value of feature engineering.

Engineered features ranked highest in feature importance



EPIC outperforms ABC model when evaluated with experimental validation in a new cell type

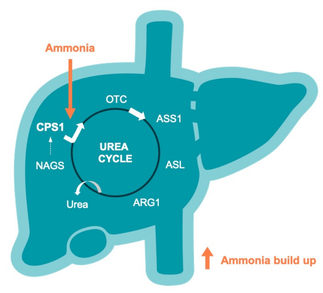
- We conducted a large scale single-cell CRISPRi-based enhancer perturbation screen in HepG2 cells.
- The significant E-P pairs derived from the screen were used to evaluate EPIC and ABC scores.



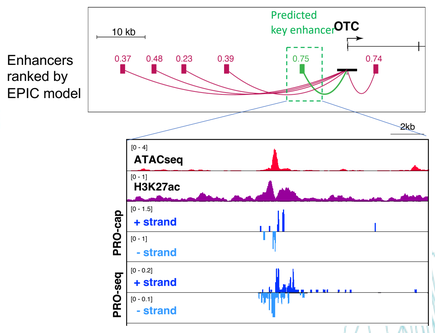
Model	AUPR	AUROC
EPIC	**0.625**	**0.910**
ABC	0.455	0.861

Developing ASOs targeting a regulatory RNA of the human ornithine transcarbamylase (*OTC*) gene

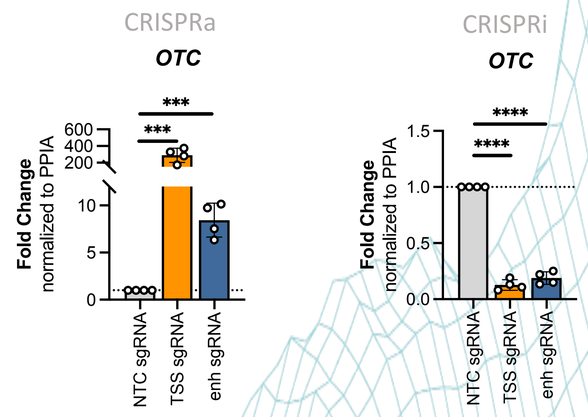
Decreased *OTC* expression results in an inadequate ammonia processing



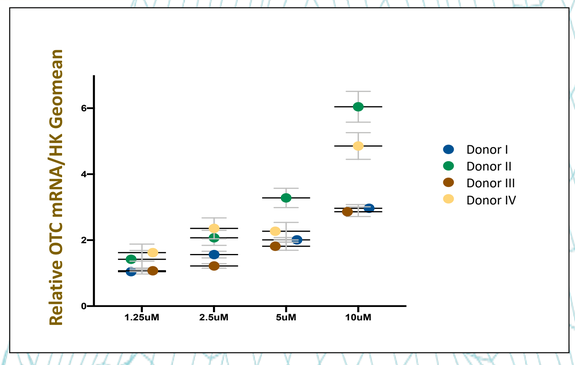
EPIC model was used to identify key *OTC* enhancer in hepatocytes



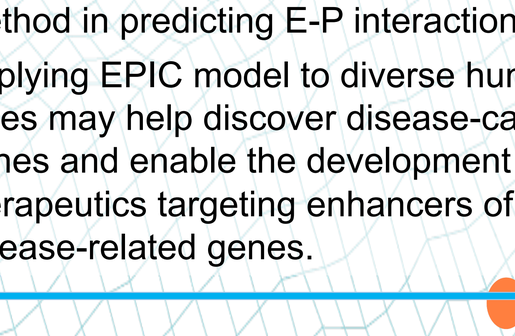
OTC enhancer targeting by CRISPRa or CRISPRi changes *OTC* mRNA levels



ASO targeting *OTC* enhancer RNA (eRNA) shows dose-dependent increase in *OTC* mRNA in hepatocytes from multiple donors



OTC eRNA-targeting ASO increases *OTC* mRNA and ureagenesis in human *OTC*-deficient hepatocytes



OTC-deficient hepatocytes: c.-106C>A variant (Allele ID 480410, late-onset OTC deficiency) - pathogenic (dbSNP: rs749748052) associated with 10-25% of normal OTC activity. *OTC* mRNA and urea levels determined after 6 days of ASO treatment (5 µM ASO). NTC = non-targeting control ASO

Conclusions

- Our EPIC model enables accurate cell-type-specific prediction of functional E-P interactions using epigenomic data.
- EPIC model outperforms an established method in predicting E-P interactions.
- Applying EPIC model to diverse human cell types may help discover disease-causing genes and enable the development of novel therapeutics targeting enhancers of disease-related genes.