A computational model using the epigenome and trained on CRISPR drop-out screens can identify oncogenic dependencies in primary tumor samples

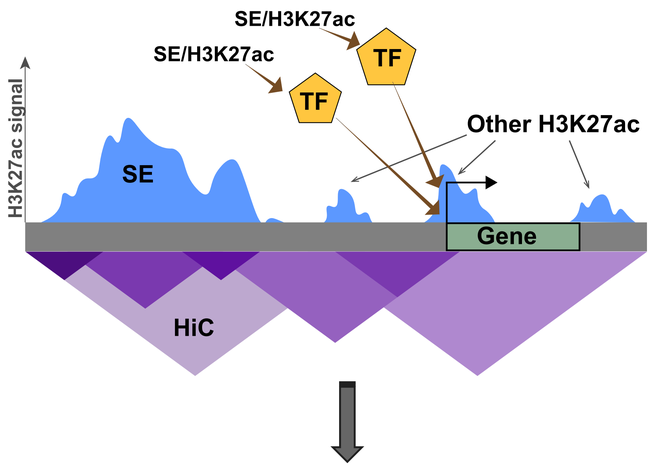
Chris Fiore, Matthew L Eaton, Brian Johnston, Matt Guenther, Cindy Collins, Mei Wei Chen, John Carulli, Eric Olson

Syros Pharmaceuticals, 35 Cambridge Park Drive, Cambridge, MA

Abstract

The epigenome can reveal which resources a cancer cell has dedicated to maintaining the expression of key genes, which are often oncogenic dependencies. Previous efforts used super-enhancers (SEs), large regions of active chromatin, to identify these oncogenic dependencies. We previously used H3K27ac ChIP-seq to identify SEs in acute myeloid leukemia (AML) and identified a SE at the RARA gene locus in a subset of AML patients that represented a novel dependency. This led to a clinical trial for SY-1425, an agonist of RARα (NCT02807558). Here we use H3K27ac ChIP-seq to expand on the concept of SEs and include other methods to quantify the effects of the epigenome on transcription, including MARGE and Activity by Contact, in order to identify oncogenic dependencies. Multiple variants and modifications of these scores are combined with additional properties of gene regulation, such as gene regulatory network (GRN) scores, into a machine learning model to predict the essentiality of genes from a genome-wide CRISPR screen in cancer cell lines. This new model, which we name PETCERF, outperforms previous individual epigenomic scores at identifying differential essential genes from H3K27ac ChIP-seq. The combination of many types of input variables was found to be important in the model predictions. Furthermore, after training the model on cell line data, we can then identify known and novel cancer dependencies as marked by the epigenome in primary tumor samples. This tool should be useful in identifying future candidate drug targets.

Many H3K27ac scores can quantify enhancers to identify oncogenic dependencies



Oncogenic dependencies
- Based on DepMap CRISPR screen
- 59 cell lines with H3K27ac ChIP-seq

Scores quantifying H3K27ac signal around genes

Super enhancers (SE)
- Size of SE linked to nearby gene by signal correlation or distance
- Published in Whyte *et al.* 2013 and McKeown *et al.* 2017

MARGE
- All nearby H3K27ac signal weighted by distance to gene
- Based on Wang *et al.* 2016

Activity by contact (ABC)
- H3K27ac weighted by strength of HiC contacts to gene
- Based on Fulco *et al.* 2019

Scores using gene regulatory network (GRN) topology to predict target (or TF) activity

Correlation networks
- Genes are linked to TFs with correlated expression

ARACNe
- From Lachmann et al. 2016
- Genes are linked to TFs with high mutual information of expression

DoRothEA
- From Garcia-Alonso *et al.* 2019
- Curated set of gene to TF links from multiple other sources and types of data

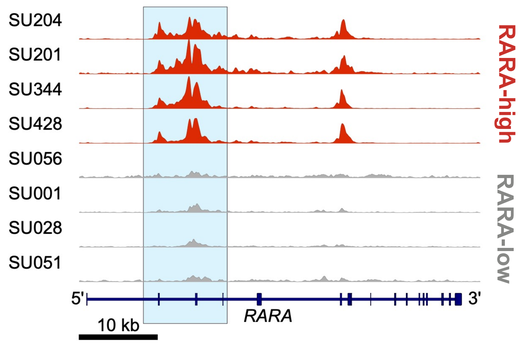
VIPER
- From Alvarez *et al.* 2016
- Determines activity of TFs from expression of downstream targets

References

W. Whyte *et al.*, *Cell* **153**, 307-309 (2013).
M.R. McKeown *et al.*, *Cancer Discov.* **7**, 1136–1153 (2017).
S. Wang *et al.*, *Genome Res.* **26**, 1417-1429 (2016).
C.P. Fulco *et al.*, *Nat. Genet.* **51**, 1664-1669 (2019).
L. Garcia-Alonso *et al.*, *Genome Res.* **29**, 1363-1375 (2019).
M. Alvarez *et al.*, *Nat. Genet.* **48**, 838-847 (2016).
A. Lachmann *et al.*, *Bioinformatics* **32**, 2233-2235 (2016).

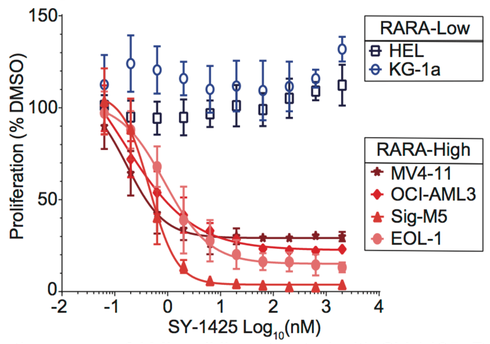
Super enhancers (SEs) can identify oncogenic dependencies

Some AML patients have a SE at the *RARA* locus



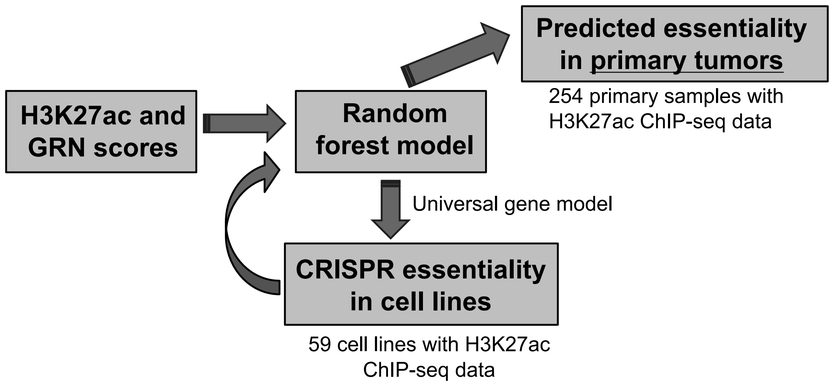
H3K27ac ChIP-seq tracks in AML patient samples, showing that an SE is present at the *RARA* locus in a subset of patients

RARA-high AML cell lines are sensitive to SY-1425

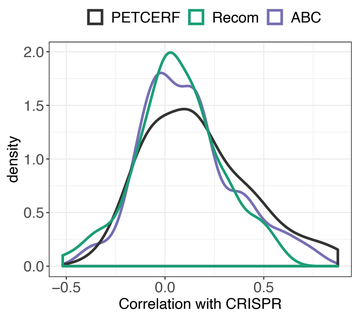


Proliferation assay of AML cell lines treated with SY-1425. RARA-high cell lines are sensitive to SY-1425, a RARα agonist, and RARA-low are insensitive.

PETCERF integrates H3K27ac and GRN scores to predict essential genes

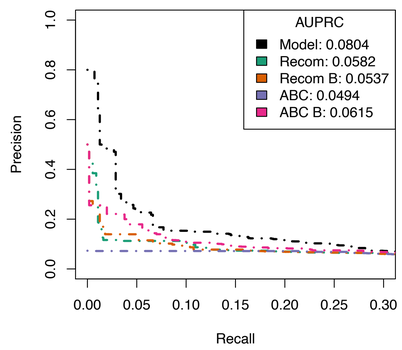


PETCERF improves correlations with CRISPR scores across cell lines by gene



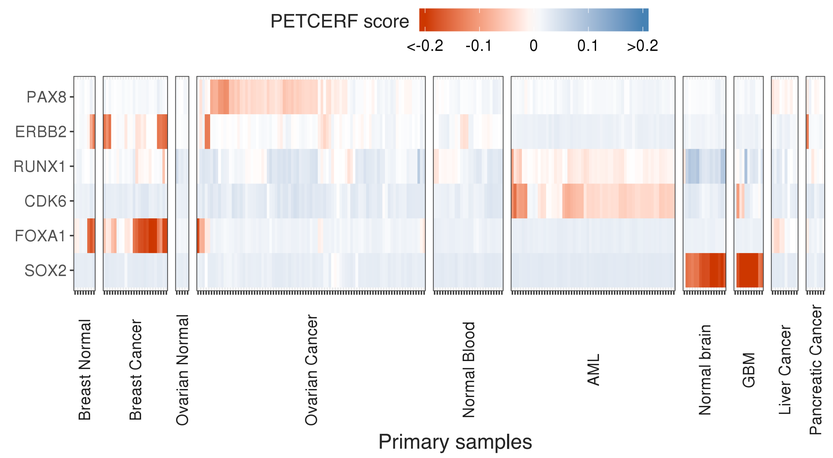
Distribution of correlation of score with CRISPR drop-out score across cell lines for each gene (only genes with high variance of each score)

PETCERF is better at selecting specific CRISPR drop-outs



Precision-recall curve for selecting specific CRISPR drop-outs across genes and cell lines (genes with high variance of each score)

PETCERF can identify oncogenic dependencies in primary tumors from H3K27ac ChIP-seq



PETCERF score (based on predicted difference of CRISPR score from mean for each gene) for a subset of known oncogenic dependencies across 254 primary human samples with H3K27ac ChIP-seq data.

Conclusions

- PETCERF uses H3K27ac ChIP-seq data to predict essential genes in cancer cells with a universal gene model
- PETCERF can identify genes that drop-out in a CRISPR screen more accurately than other scores based on H3K27ac ChIP-seq
- PETCERF can be used to identify oncogenic dependencies in primary human tumor samples