Discovery and engineering of hypercompact epigenetic modulators for durable gene activation

Carosso G[1], Yeo RW[1], Gainous TB[1], Jawaid Z[1], Yang X[1], Cutillas V[1], Qi LS[1,2,3,4], Daley TP[1], Hart D[1]

[1]Epicrispr Biotechnologies, South San Francisco, CA 94080, USA
[2]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA
[3]ChEM-H, Stanford University, Stanford, CA 94305, USA
[4]Chan Zuckerberg Biohub, San Francisco, CA 94080, USA

Abstract

Transcriptional and epigenetic regulators shape the chromatin microenvironment and corresponding gene expression during cellular differentiation and homeostasis. Programmable modulators of transcription provide a powerful toolkit for controlling gene dosage in therapeutic applications, but a limited catalog of functional domains constrains their robustness and durability profiles, and large cargo sizes impede clinical delivery.

To address these limitations, here we perform high-throughput screening to discover novel classes of transcriptional modulators among human, viral, and archaeal proteomes and characterize their functions in a multitude of endogenous human contexts. We identify compact, potent activators from viral proteomes with exceptional robustness across silent and expressed genes in varied cell types using distinct dCas systems. Insights from predicted 3-dimensional structures and machine learning models enabled us to rationally engineer improved activators, both in potency and persistence. Notably, engineered activators achieved mitotically durable gene activation following transient delivery.

Our discovery pipeline provides a predictive rubric for the systematic development of hypercompact activators from unannotated proteomes, yielding superior efficiency and kinetics profiles that broadly expand the epigenetic editing toolkit for therapeutic applications.
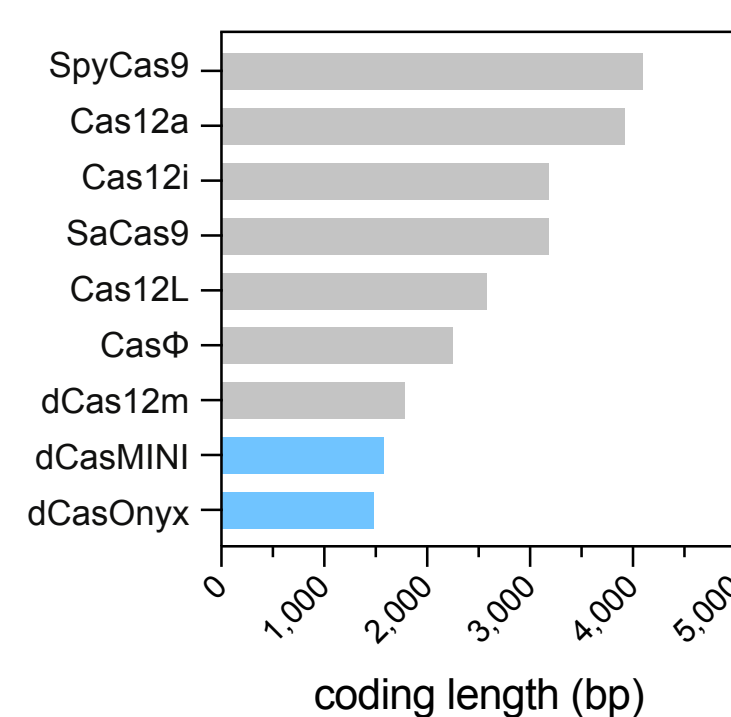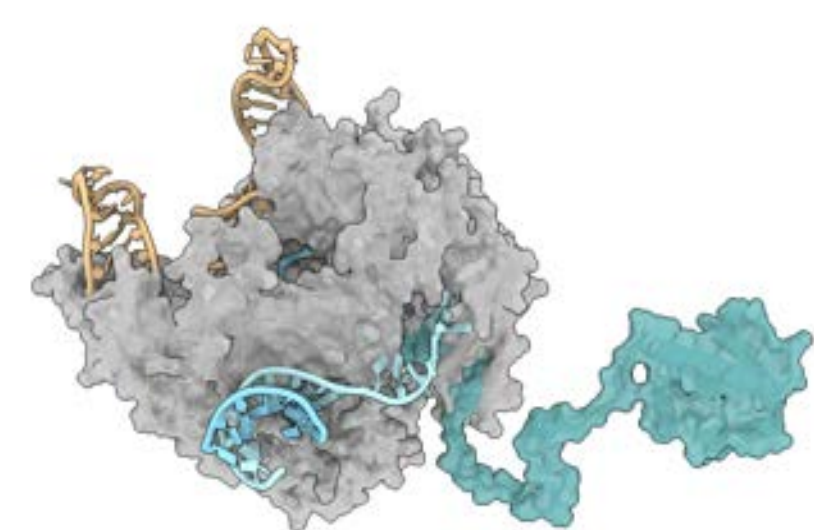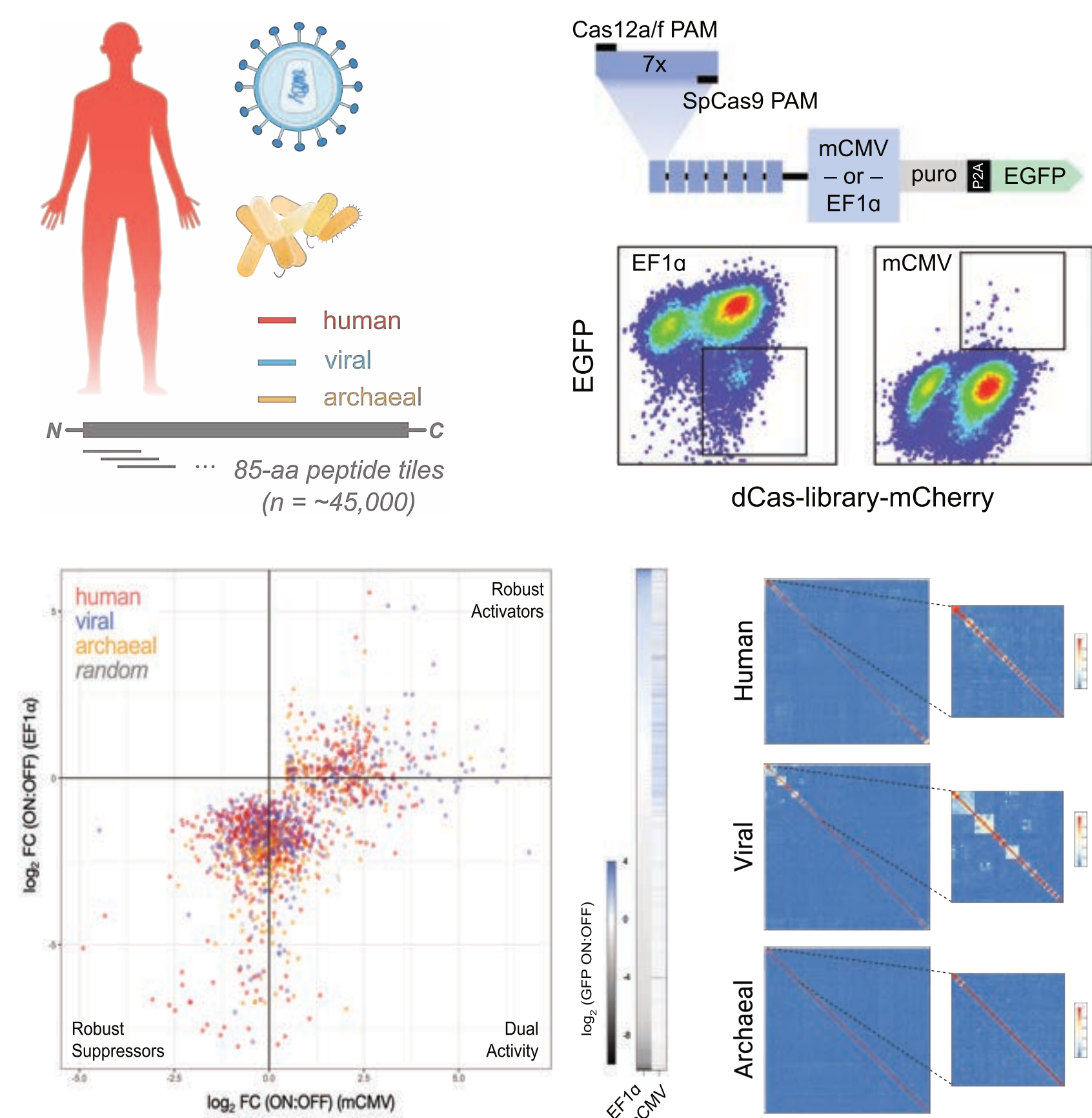
We're hiring!
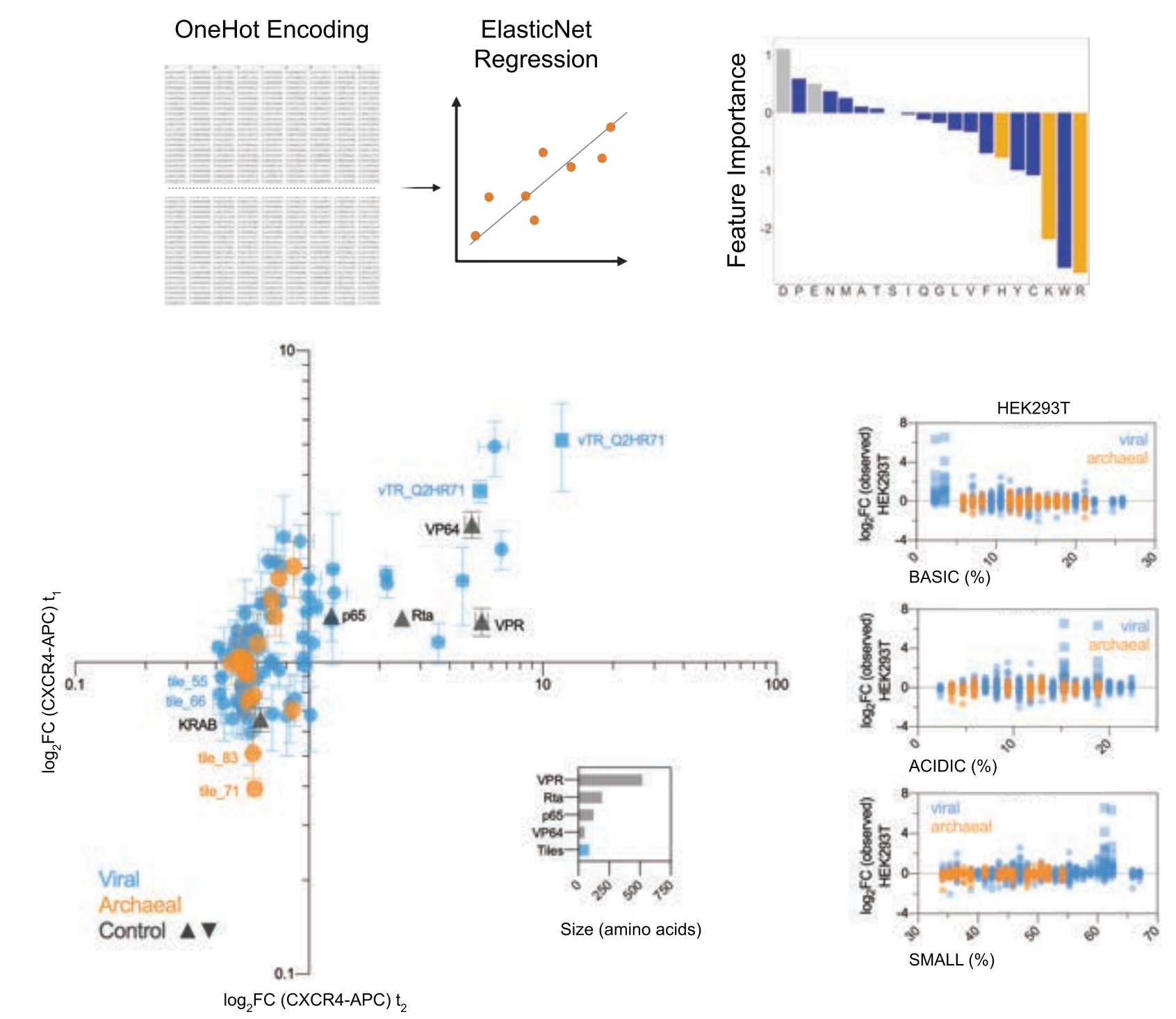
10.1101/2023.06.02.54349

bioRχiv

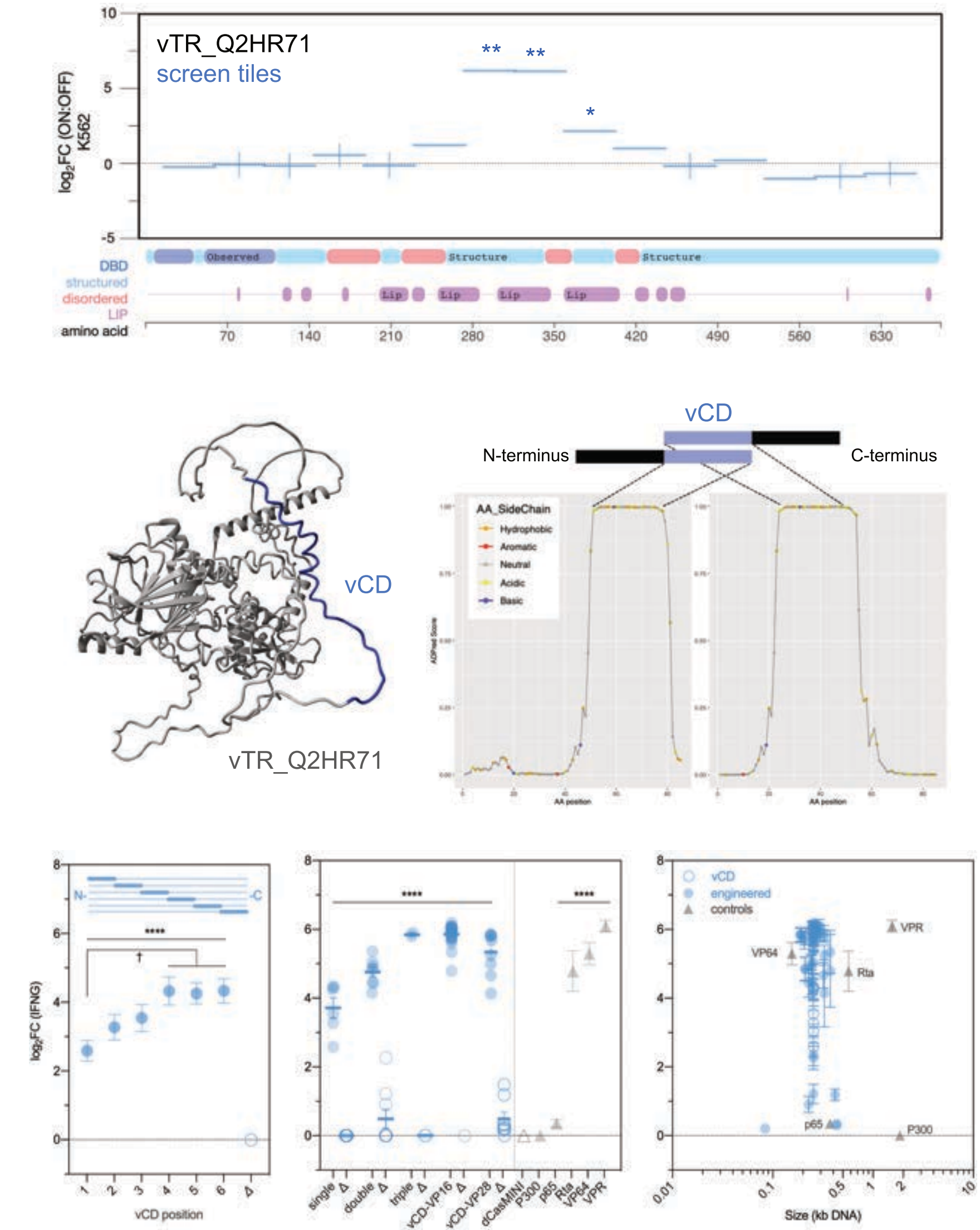## Gene Expression Modulation System (GEMS) for epigenetic engineering

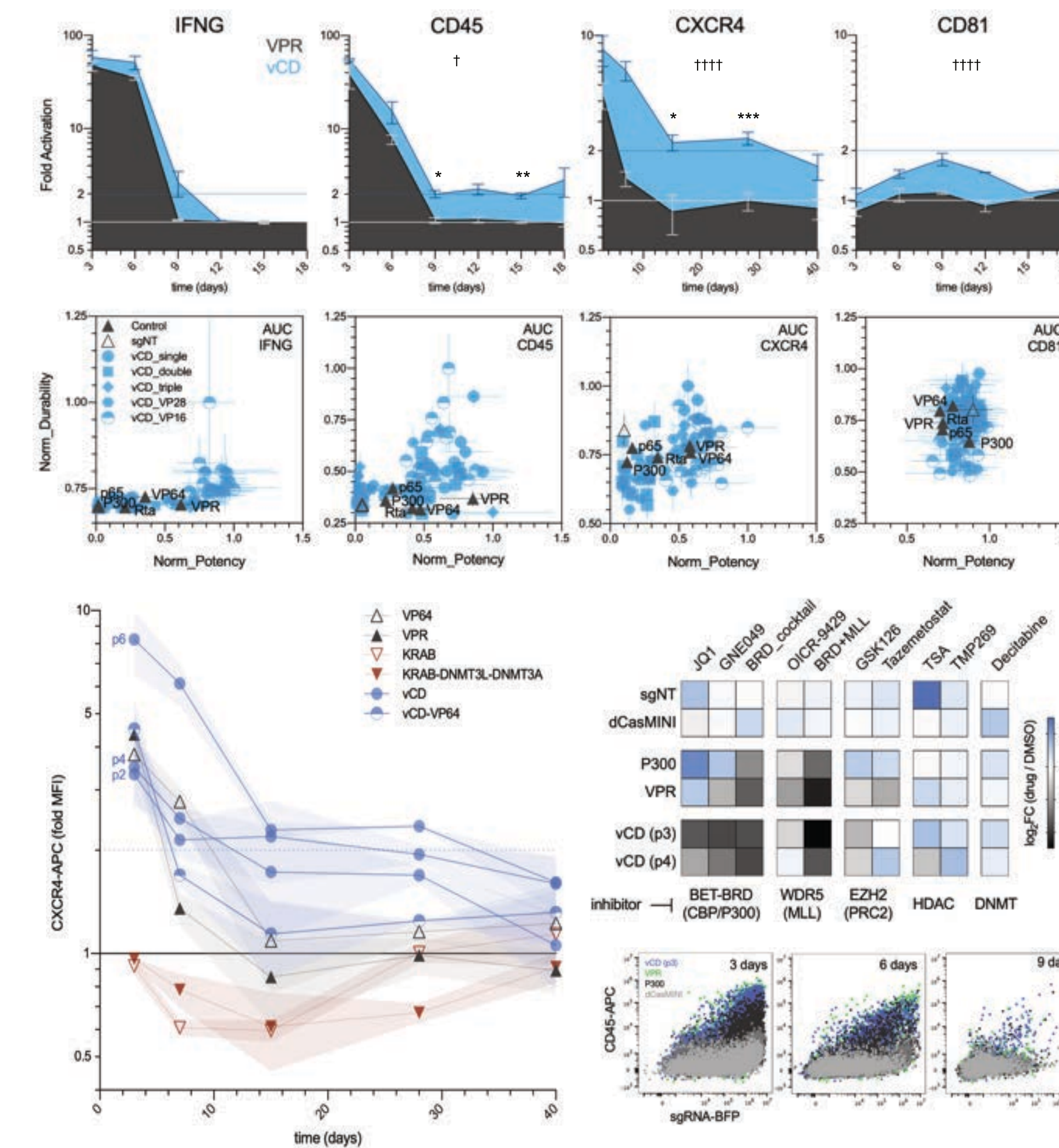## Identification of transcriptional modulators by high-throughput screening

human
viral
archaeal

Cas12a/f PAM
7x
SpCas9 PAM
mCMV
-or-
EF1α
puro EGFP

··· 85-aa peptide tiles
(n = ~45,000)

dCas-library-mCherry

EF1α   mCMV

EGFP

human
viral
archaeal
random

Robust Activators

Robust Suppressors

Dual Activity

Human
Viral
Archaeal

log₂ FC (ON:OFF) (EF1α)

log₂ FC (ON:OFF) (mCMV)

Biochemical feature importance: prediction and validation

OneHot Encoding

ElasticNet Regression

Feature Importance

D P E N M A T S I Q G L V F H Y C K W R

vTR_Q2HR71

VP64

p65 Rta VPR

KRAB

Viral
Archaeal
Control

HEK293T

BASIC (%)

ACIDIC (%)

SMALL (%)

viral
archaeal

Deep learning identifies minimal domains as a basis for engineered potency

vTR_Q2HR71
screen tiles

DBD
structured
disordered
LIP

vCD

N-terminus   C-terminus

vTR_Q2HR71

vCD
engineered
controls

VP64
Rta
p65 AB
P300

Max. Activation (3 d.p.t.): CD45, IFNG, CXCR4

P300
p65
Rta
VP64
VPR

Durable Activation (9-18 d.p.t.): CD45, IFNG, CXCR4, CD81

P300
Rta
p65
VPR
VP64

CD45
CXCR4
IFNG
CD81
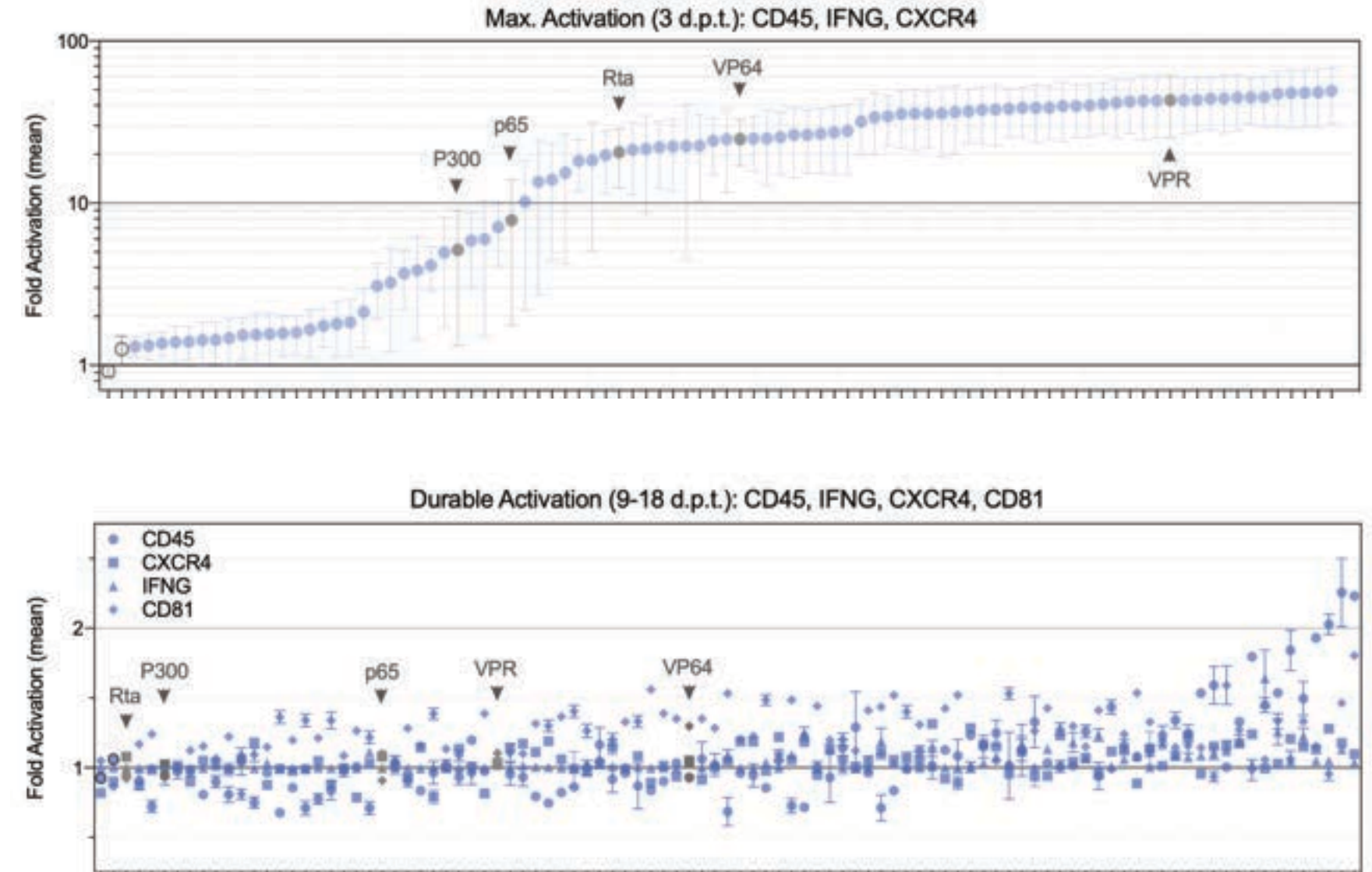
Durable activation by GEMS activators

IFNG   CD45   CXCR4   CD81

VPR
vCD

AUC IFNG   AUC CD45   AUC CXCR4   AUC CD81

VP64
VPR
KRAB
KRAB-DNMT3L-DNMT3A
vCD
vCD-VP64

BET-BRD (CBP/P300)   WDR5 (MLL)   EZH2 (PRC2)   HDAC   DNMT

Conclusions

High-throughput screening identifies known and novel transcriptional modulators from diverse species.

ML-enabled biochemical analysis provides a predictive rubric for modulator functions.

Novel activators outperform larger benchmarks in potency, context-independent robustness, and durability of activity

References

1. Xu, Xiaoshu, et al. "Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing." Molecular Cell 81.20 (2021): 4333-4345.

2. Chavez, Alejandro, et al. "Highly efficient Cas9-mediated transcriptional programming." Nature methods 12.4 (2015): 326-328.

3. Sanborn, Adrian L., et al. "Simple biochemical features underlie transcriptional activation domain diversity and dynamic, fuzzy binding to Mediator." Elife 10 (2021): e68068.

SpyCas9
Cas12a
Cas12i
SaCas9
Cas12L
CasΦ
dCas12m
dCasMINI
dCasOnyx

coding length (bp)