Discovery and engineering of hypercompact transcriptional modulators for robust and durable target 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 94158, USA

Abstract

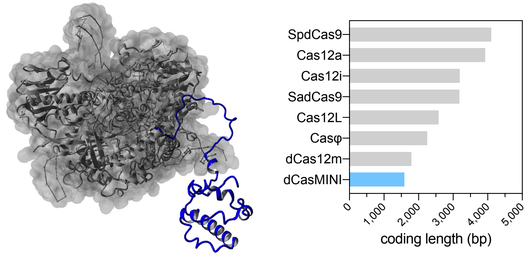
RNA-guided control of target gene expression via nuclease-dead CRISPR/dCas systems offers potential therapeutic avenues for treating a wide range of disease indications, but existing tools are limited in clinical utility by large cargo sizes and temporally transient windows of activity. Promisingly, recent work has demonstrated robust and durable gene silencing via epigenetic memory, as well as robust–but transient–gene activation using large multi-protein domain modulator fusions[1-3]. To overcome this limitation of transient gene activation, here we present our Gene Expression Modulation System (GEMS), pairing a compact and programmable dCasMINI[4] recruitment platform to a novel suite of hypercompact modulators capable of robust and durable gene activation in diverse genomic contexts.

Using human cell-based reporter platforms for dCas-mediated domain recruitment, we conducted high-throughput screens of curated human, viral, and archaeal protein-coding libraries and identified both known and novel modulator sequences capable of transcriptional activation. We then identified the predicted minimal core transactivation domains for the top screen hits, using a sequence-based deep learning tool which exploits the acidic exposure model. These resultant core domains provided a basis for iterative rational engineering to further improve activator performance, targeting endogenous human gene promoters in multiple chromatin contexts as validation. Successive engineering rounds yielded dozens of variants capable of targeted gene activation dynamics outperforming standard multi-protein domain fusions such as VPR, both in magnitude and duration of effect, despite occupying a fraction of their coding length. Investigation into the underlying epigenetic mechanisms at play demonstrated abrogation of sustained effects upon chemical disruption of CBP/P300-associated bromodomains, implicating histone acetylation as a mediator of activation persistence.

Together, the GEMS platform and its broad suite of hypercompact transcriptional modulators provide a basis for single-AAV or LNP-compatible therapeutic tools able to achieve fine-tuned, robust, and durable target gene activation.

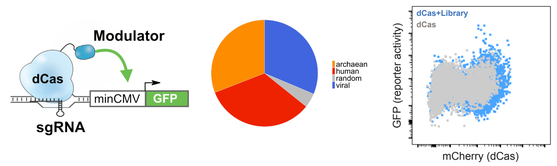
Gene Expression Modulation System (GEMS) for epigenome engineering

1. Compact, programmable DNA-binding protein (dCasMINI)[4]
2. Modulator peptide capable of activation or suppression

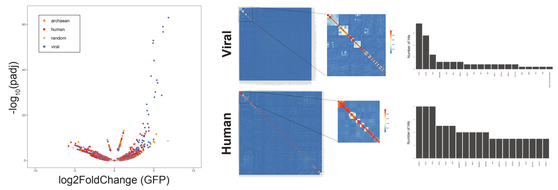


Identification of hypercompact modulators from high-throughput screening

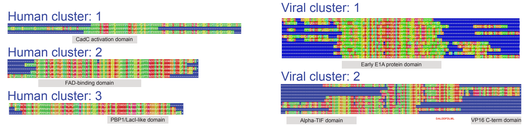
To identify novel modulator peptides we tiled across human, viral, and archaeal genomes for a library of ~45,000 85aa peptides.



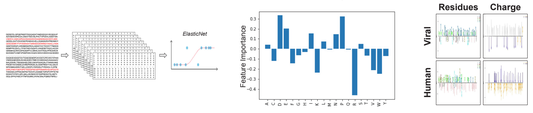
High-throughput screening identified hundreds of novel peptide tiles capable of activating GFP expression at a synthetic locus.



Screen hits share conserved sequence motifs from distinct protein families, indicating a diversity of potential activation mechanisms.

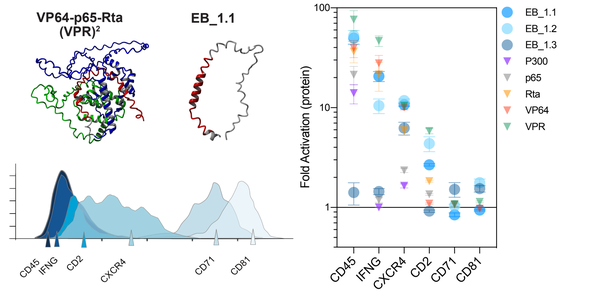


Activators are enriched for acidic and depleted of basic residues.



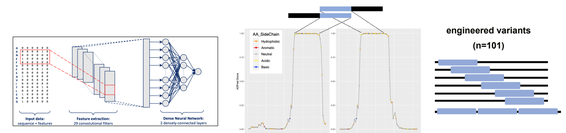
Screen hits are potent activators at multiple endogenous human loci

85aa activator tiles demonstrate comparable performance to benchmarks at human targets with diverse expression baselines.

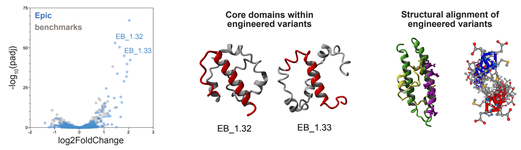


Engineered modulator variants outperform original screen tiles and benchmarks

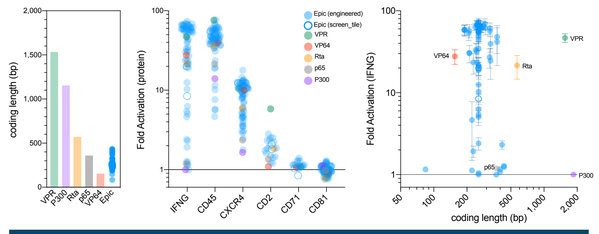
We next leveraged a deep learning model (*ADPred*[5]) to predict minimal core activation domains, and engineered a library of structural variants for subsequent screens.



Follow-up screening confirmed robust activation from novel engineered variants containing minimal core domains.

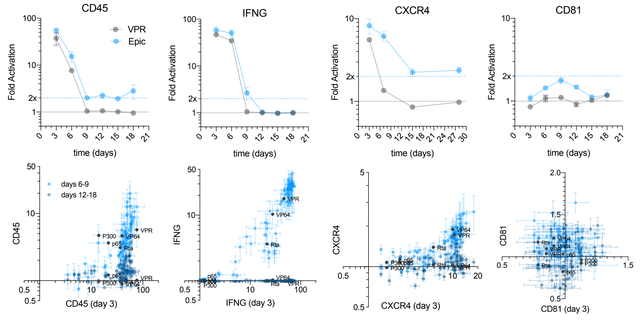


Engineered variants are hypercompact activators with comparable strength to larger-sized, established benchmarks (e.g. VPR).



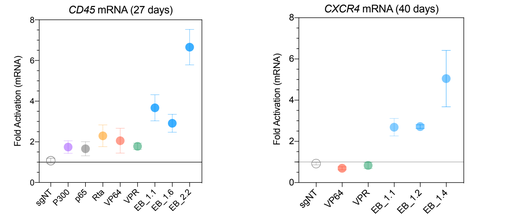
Kinetics of activation reveal persistent modulation by engineered variants

Following transient delivery, engineered variants display persistence at multiple endogenous loci while transcriptional activation by canonical modulators declines by day 9.



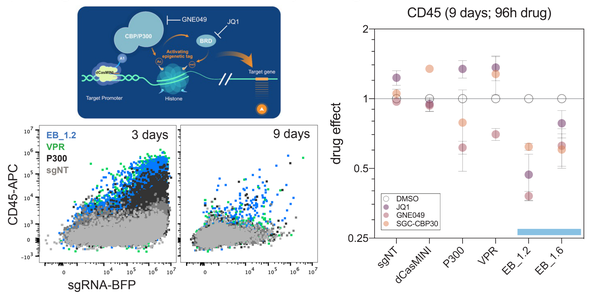
Kinetics of activation reveal persistent modulation by engineered variants (ctd.)

RT-qPCR data demonstrate target mRNA elevation past 27 days.



Epigenetic mechanisms of persistent activation

Activation by novel modulators is dependent on CBP/P300-associated BET/BRD bromodomains, implicating H3K27Ac in the propagation of activation memory through mitotic cell divisions.



Conclusions

1. High-throughput screening across human, viral, and archaeal genomes identifies known and novel transcriptional activators.
2. Novel modulators are capable of transcriptional activation at a diverse range of endogenous human target genes.
3. ML-guided engineering confers tunability and increased potency, yielding a suite of hypercompact activators with similar potency to large multi-domain fusions.
4. Engineered modulators durably maintain human target gene activation beyond the activity of larger benchmark activators.

References

(1) Nuñez, James K., et al. "Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing." Cell 184.9 (2021): 2503-2519. (2) Chavez, Alejandro, et al. "Highly efficient Cas9-mediated transcriptional programming." Nature methods 12.4 (2015): 326-328. (3) Beyersdorf, Jared P., et al. "Robust, Durable Gene Activation In Vivo via mRNA-Encoded Activators." ACS nano 16.4 (2022): 5660-5671. (4) Xu, Xiaoshu, et al. "Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing." Molecular Cell 81.20 (2021): 4333-4345. (5) Sanborn, Adrian L., et al. "Simple biochemical features underlie transcriptional activation domain diversity and dynamic, fuzzy binding to Mediator." Elife 10 (2021): e68068.