Combinatorial Screening of Transcriptional Activation Domains for Improved Activity Provides Insights into Biophysical Properties of Strong Activators

T. Blair Gainous, PhD[1], Robin W. Yeo, PhD[1], M. Zaki Jawaid, PhD[1], Aayushma Gautam[1], Lei Stanley Qi, PhD[2,3,4], Timothy P. Daley, PhD[1], Daniel O. Hart, PhD[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

CRISPR-mediated transcriptional regulation has broad potential applications in synthetic biology and gene therapy. Transcriptional activation in eukaryotes depends on the complex interplay of multiple factors, including DNA-binding transcription factors, co-activators, chromatin remodelers, and basal transcriptional machinery. These multiple inputs often exhibit a synergistic relationship, whereby the transcriptional output driven by two or more factors is greater than the sum of output driven by each factor individually. We have recently discovered hundreds of small (85aa) peptides derived from endogenous human, viral, and archaeal proteins that activate transcription to varying degrees when fused to a programmable DNA-binding dCas protein[1].

Here, we have established a platform to screen pairwise combinations of these activation domains in an inducible and reversible manner for improved activity as determined by both magnitude and duration of transcriptional output. Our screen identified ~1400 novel combinations and we find that many combinatorial activators exhibit stronger activity than their constituent parts. Consistent with our previous findings comparing viral and human activators, these combinatorial activators are strongest when one of the partner tiles is of viral origin. Using a machine learning approach, we identified a suite of partner tiles that are most predictive of strong combinatorial activators including canonical viral activators (e.g. VP64), novel viral activators (e.g. vIRF2/vIRF4), and human activators (e.g. LEUTX).

Analysis of biochemical and biophysical features revealed that strong combos have highly negative electrostatic potential and tend to fall within a "sweet spot" of structural flexibility. *In silico* structural predictions of top hits indicate that stabilizing intramolecular interactions between helices may stabilize the interaction interface of strong combos and be functionally relevant for transcriptional activation. Altogether, this work provides a novel toolkit of combinatorial transcriptional regulators with as-yet-unexplored potential, as well as a platform for the discovery of additional tools with applications in both basic research and therapeutics.

Combinatorial activators can exhibit transcriptional synergy

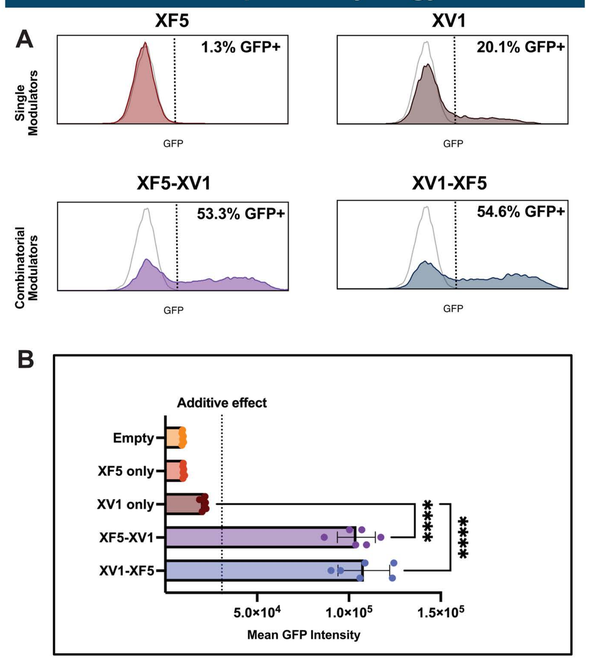

Figure 1. Activation domains identified in high throughput screens for human and viral transcriptional modulators[1] were tested as individual or tandem fusions targeted to a GFP reporter. **A)** Individual activators (top panels) show weak to moderate activation of GFP, whereas combinatorial modulators (bottom panels) show strong activation, independent of the orientation of the modulators. **B)** Combinatorial modulators activate GFP to levels greater than the sum of the individual modulators, a hallmark of transcriptional synergy.

Combinatorial activators were curated to sample diverse activities

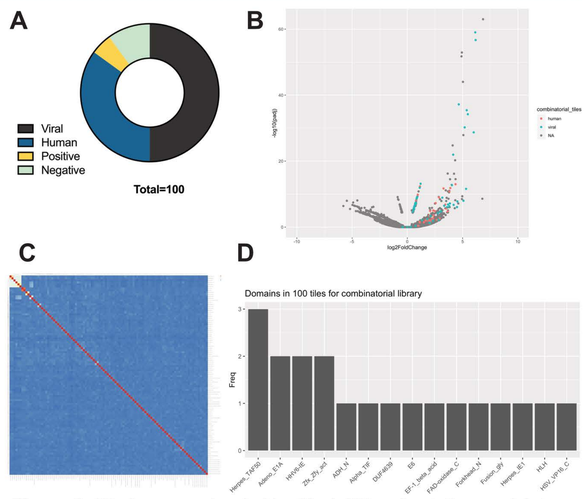

Figure 2. We have previously identified 100s of putative modulators sourced from 85aa tiles across human and viral proteomes. **A)** From this list of candidates, we curated a set of 85 modulators (50 viral, 35 human) to test in a combinatorial screen, along with positive and negative controls. **B)** The selected tiles represent weak, medium, and strong activators as identified in our previous screen. **C)** Homology clustering indicates high sequence diversity among candidate modulators. **D)** Library tiles contain a diversity of predicted protein domains.

A platform for screening combinatorial modulator potency and durability

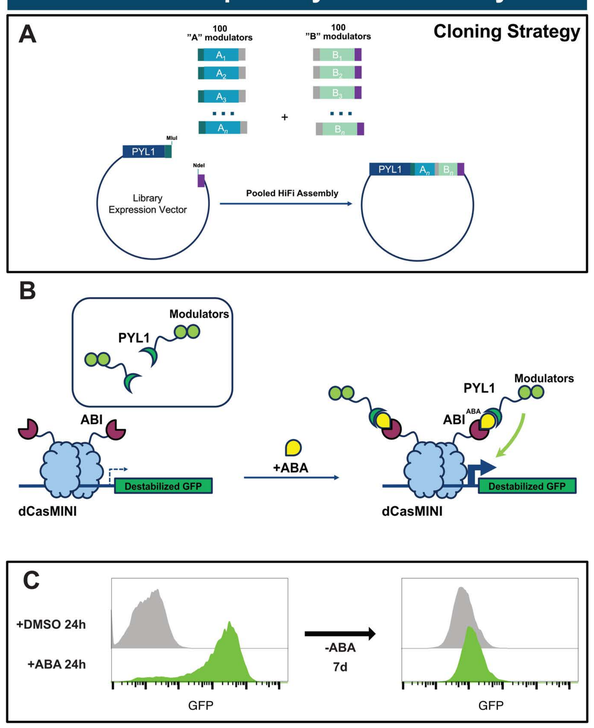

Figure 3. A) 100 candidate modulators were synthesized as A or B variants with unique codon usages and homology arms to facilitate directional cloning into a gRNA+PYL1 lentiviral expression vector. **B)** K562 cells harboring a destabilized GFP reporter were engineered to stably express a dCasMINI-ABI fusion. ABI dimerizes with PYL1 upon addition of the plant hormone abscisic acid (ABA), resulting in recruitment of modulators to the reporter locus[2]. **C)** Validation of recruitment and reversibility: K562-GFP-dCasMINI-ABI cells were transduced with lentivirus encoding a PYL1-activator fusion and a gRNA targeting the GFP reporter. After 24h ABA treatment, robust GFP activation was observed. 7 days after ABA washout, GFP expression had returned to levels comparable to DMSO control.

Combinatorial modulators exhibit a range of activation potencies

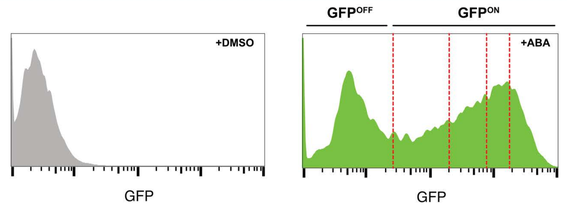

Figure 4. K562-GFP reporter cells expressing dCasMINI-ABI were transduced with the PYL1-ComboMod fusion library and treated with DMSO or ABA. GFP^{ON} cells were sorted by FACS into 4 discrete bins representing quartiles of GFP expression. Genomic DNA was harvested and Illumina libraries were prepared from the 4 GFP^{ON} bins and the GFP^{OFF} bin across 4 replicates. Modulator combos in each bin were identified by paired end sequencing and enrichment in GFP^{ON} vs GFP^{OFF} bins was determined using DESEQ2.

Combinatorial modulator hits are enriched for combos containing viral activators

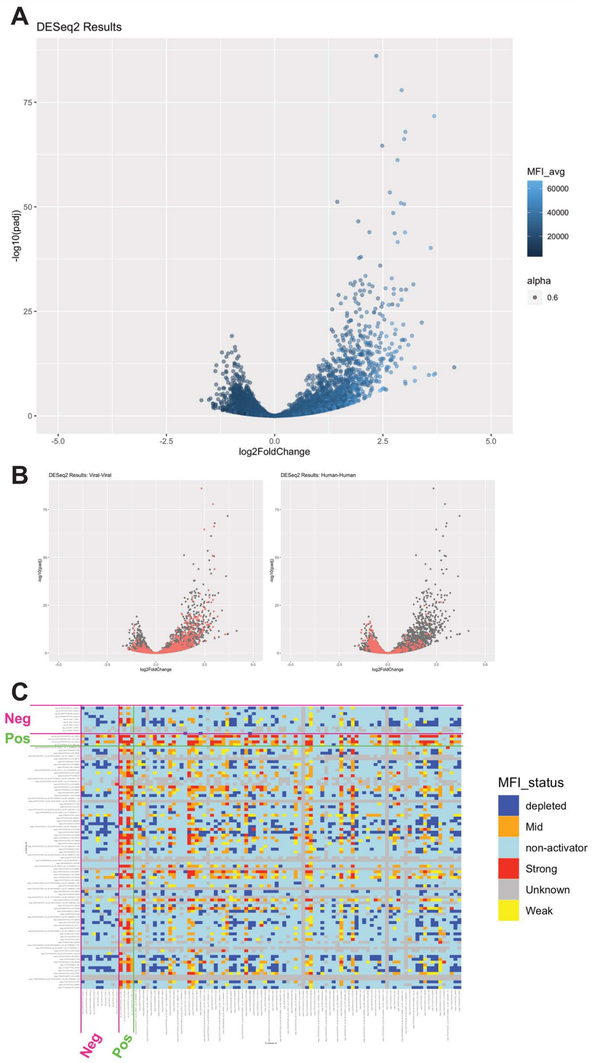

Figure 5. A) DESEQ2 analysis of ComboMod enrichment identified 1393 hits with FDR <0.05 **B)** Most strong activators consisted of 1 or more viral activators, consistent with previous findings comparing the relative strengths of viral and human activators (left); a number of human-human combinations were also identified as strong activators (right). **C)** Heatmap indicating activation status of all 100x100 combinatorial modulators. Negative and positive control combos are highlighted. Strong activation combos tend to cluster along rows or columns, highlighting the predictive nature of a small number of modulators in driving strong combos.

Feature analysis identifies biophysical properties of strong combos

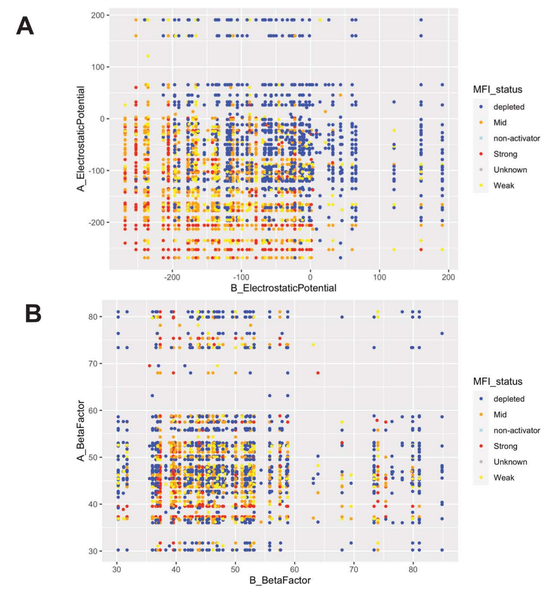

Figure 6. Analysis of biochemical and biophysical features of combinatorial modulators. **A)** Strong activation combos are enriched for negative electrostatic potential **B)** Strong activation combos generally require that one partner have a Beta factor (a measure of structural flexibility) between 35-60.

Conclusions & Future Directions

We have established and validated a platform for high throughput screening of combinations of transcriptional modulators. Our screen of 100x100 modulators of human and viral origin reinforces previous findings on the relative strengths of activation domains between these two groups. Nonetheless, we have identified a number of human-human combos that outperform their constituent modulators individually. These strong combinatorial activation domains could prove useful in therapeutic contexts where immunogenicity is of concern. Additionally, combinations of weak or moderate activators may prove beneficial in achieving tunable levels of activation.

Our initial screen failed to identify any ComboMods that achieved activation duration beyond the timeframe in which positive control activators return to baseline (7d post ABA washout). This observation could possibly be attributed to the limitations of our synthetic reporter-based assay.

We are currently testing a number of our top ComboMod hits at endogenous target genes in a variety of cell types. These studies will provide further insight into the context-dependent nature of ComboMod potency and durability.

References & Related Posters

1. Carosso, et al. "Discovery and engineering of hypercompact epigenetic modulators for durable gene activation." bioRxiv 2023.06.02.543492; doi: https://doi.org/10.1101/2023.06.02.543492

2. Gao, et al. "Complex transcriptional modulation with orthogonal and inducible dCas9 regulators" Nature Methods 13 (2016): 1043-1049.

Poster #71: Discovery and engineering of hypercompact transcriptional modulators for robust and durable target gene activation

Poster #80: EPI-321—A promising CRISPR epigenome engineering therapy for facioscapulohumeral muscular dystrophy (FSHD)

Poster #108: Gene expression modulation systems (GEMs)—Optimized CRISPR-based epigenome editing platform for versatile epigenome modulation

Poster #117: Accelerating protein engineering with machine learning—A few- shot transfer learning approach to designing novel gene activators

Poster #215: Engineering of compact ribonucleoprotein complexes for epigenome editing