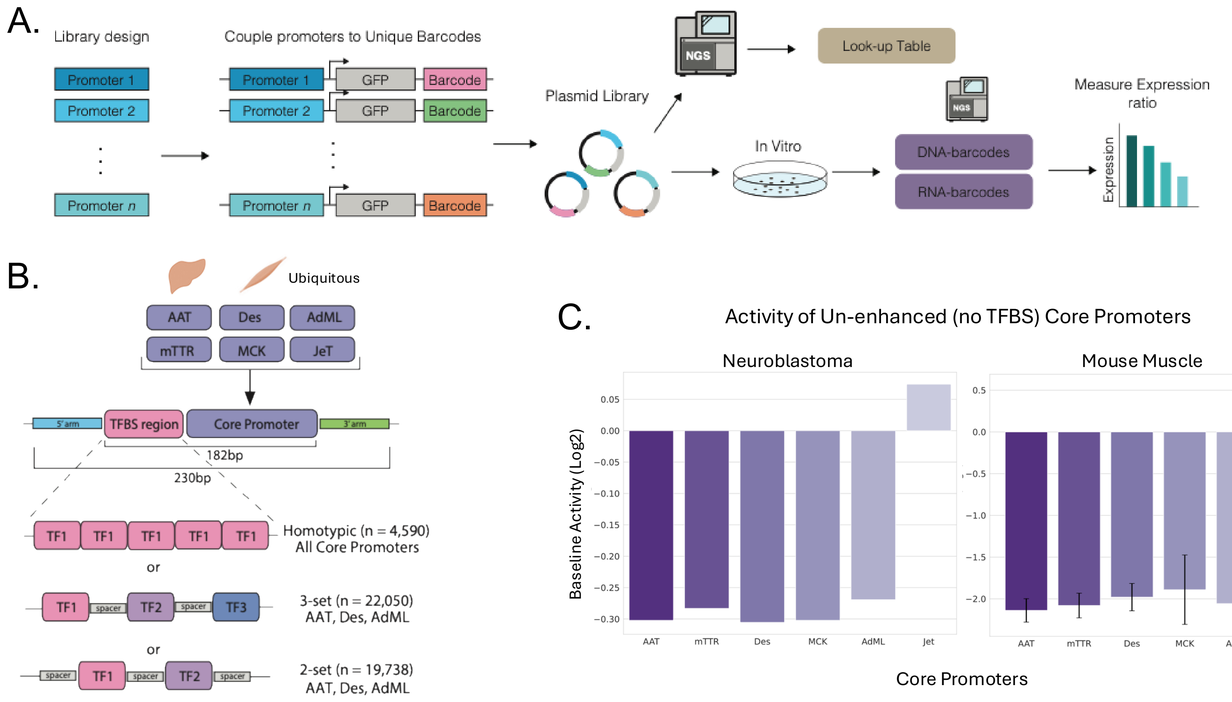
High-throughput Decoding of Transcription Factor Binding Site Grammar for Synthetic Promoter Design In Vivo

Camilla Belamarich[1], Chenjin Jin[1], Jared Sills[1], Gavin Lagani[1], Casey Lutnick[1], T.D. Barbara Nguyen-Vu[1], Josefa M. Sullivan[1], Enrico Mossotto[1]

[1]MeiraGTx New York, New York USA

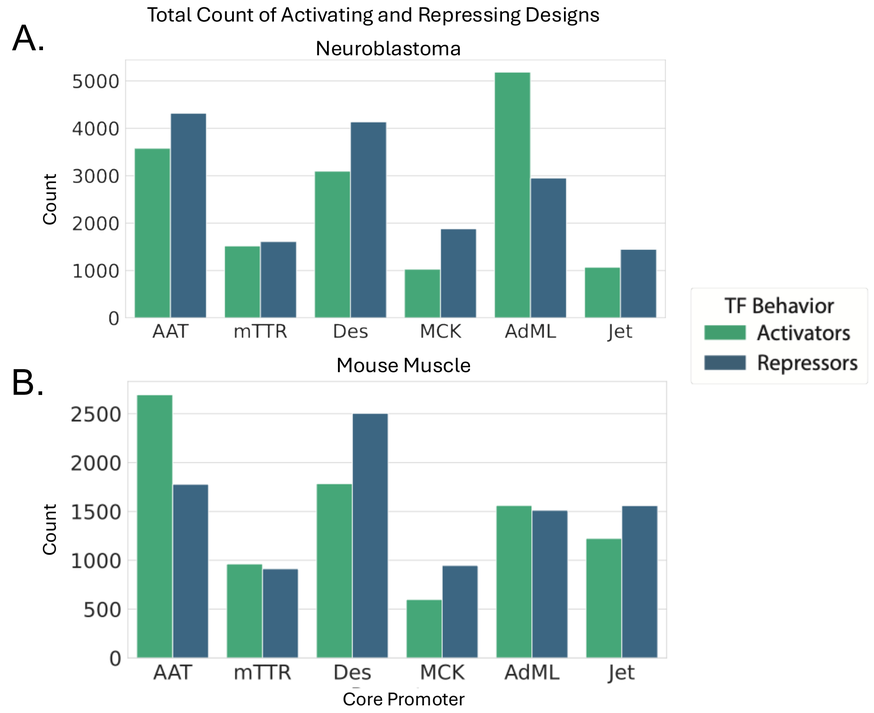
1. MPRA Designs to Inform TFBS Regulatory Grammar

MPRA screens quantify transcriptional activity of different TFBS design strategies across small synthetic promoters. (A) Massively parallel reporter assays (MPRA) measure the regulatory activity of thousands of DNA sequences simultaneously by linking each promoter design to a unique barcode and quantifying barcode expression [1] **(B)** Six core promoters (AAT, mTTR, Des, MCK, JeT, and AdML) were tested across three TFBS designs strategies resulting in 165,000 synthetic promoter sequences. Transcriptional strength was evaluated in transfected Neuro2A (mouse neuroblastoma; n=1) and transduced AAV9 via intramuscular injection in mouse gastrocnemius (n=5). **(C)** Promoter activity was calculated as RNA/DNA abundance, weighted by barcode-fragment associations and multimapping. Baseline Activity (Log2) represents the transcriptional activity of the core promoter without added TFBS.



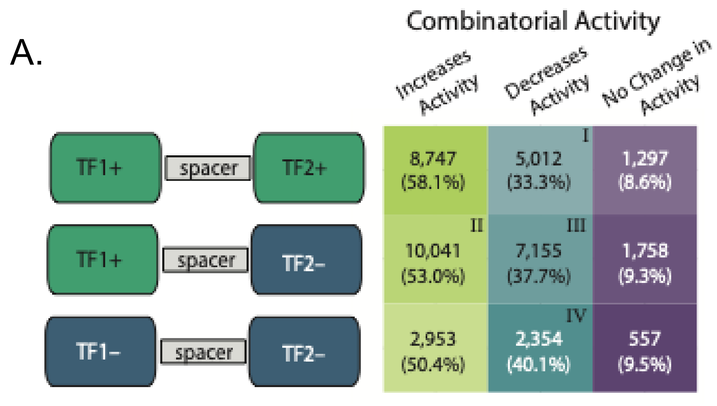
2. Characterization of TFBS Designs Reveals Activating and Repressing Effects Driven by Core Promoter

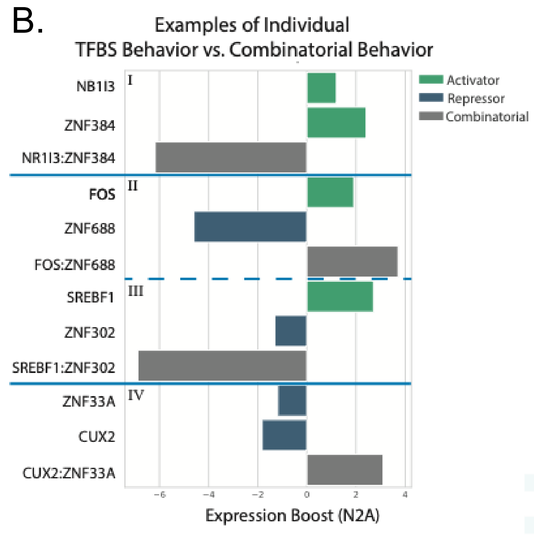
Baseline promoter activity shapes the magnitude and direction of TFBS effects. TFBS activity was quantified as Boost Activity (log2 fold-change expression relative to the parent promoter) [2]. **(A)** Designs were classified as activators, repressors, or neutral. **(B)** Boost activity was calculated independently for each animal using its corresponding baseline activity. Robust activators (n=20,909) and repressors (n=20,804) were defined by having consistent magnitude and directionality across all five animals. Excluding AAT, directionality remained consistent across tissue and magnitude was affected.

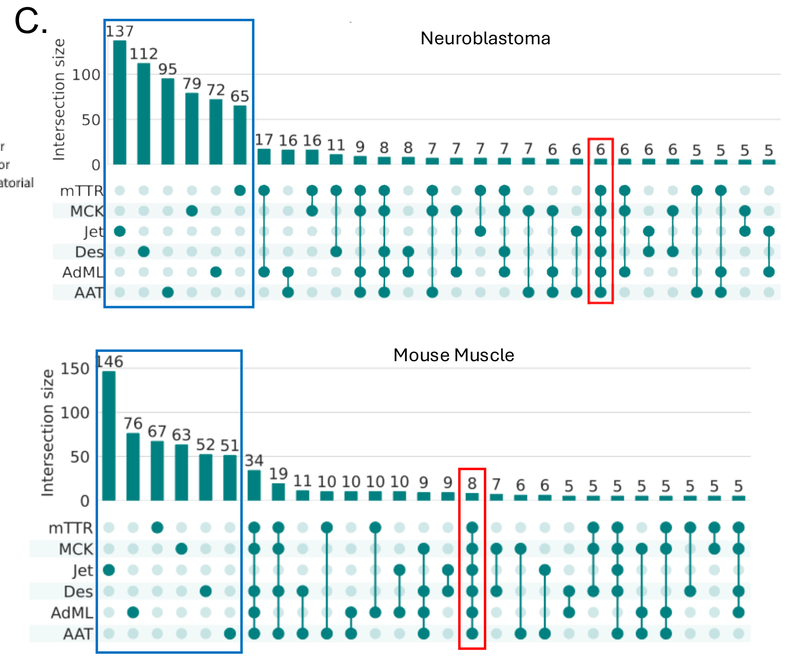


3. Promoter Context Strongly Influences TFBS function

TFBS activity is highly context-dependent, varying across tissues and promoter backgrounds. (A) Schematic of 2-set TFBS combinations, where "+" and "−" indicate activating and repressing TFs. Contingency analysis revealed the activity of an individual site is a poor predictor of its behavior in a composite context **(B)** Representative examples of each behavioral combination. Roman numerals correspond to combinations in Figure A. **(C)** UpSet plots of the top 200 activating homotypic TFBS designs across core promoters in neuroblastoma (top) and muscle (bottom), revealing limited overlap between contexts. Promoter-specific activators and promoter-independent activators are outlined in blue and red, respectively.

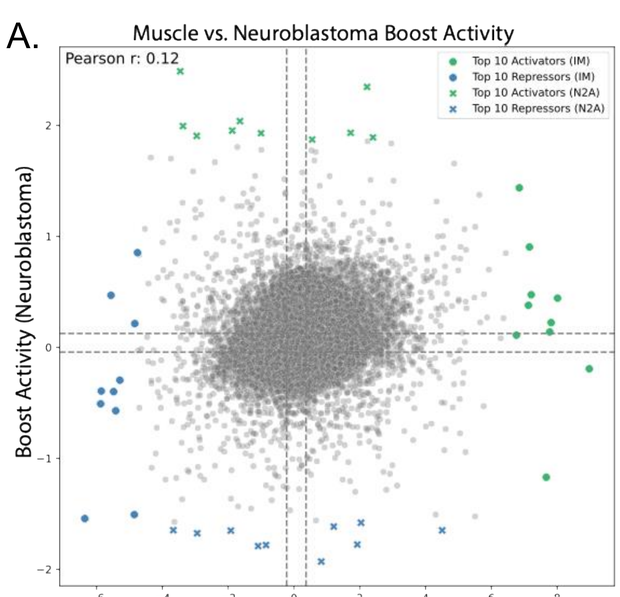


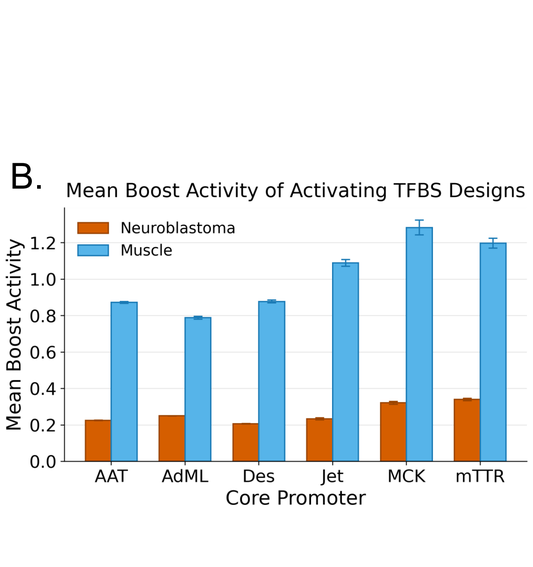




4. Cellular Context Alters the Strength of TFBS-mediated Activation

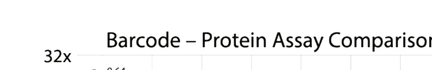
Shared TFBS designs produce distinct regulatory outcomes across models. (A) Shared TFBS designs showed poor agreement in Boost Activity between muscle and neuroblastoma (**r = 0.12**), demonstrating strong tissue-dependent effects. Activating designs showed markedly less concordance across tissues than repressing designs, with effect size varying considerably between contexts. **(B)** Mean Boost Activity of activating TFBS designs differed across core promoters and tissues, further highlighting the influence of promoter context. In muscle, the muscle-specific **MCK** promoter exhibited the greatest average enhancement, whereas in neuroblastoma, the highest average activity was observed with **MCK** and the liver-specific **mTTR** promoters.

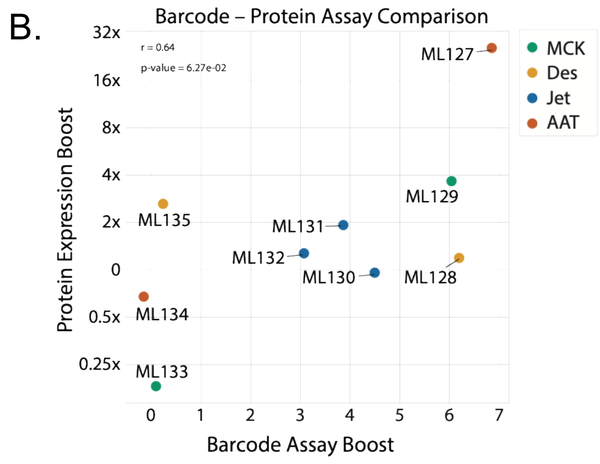


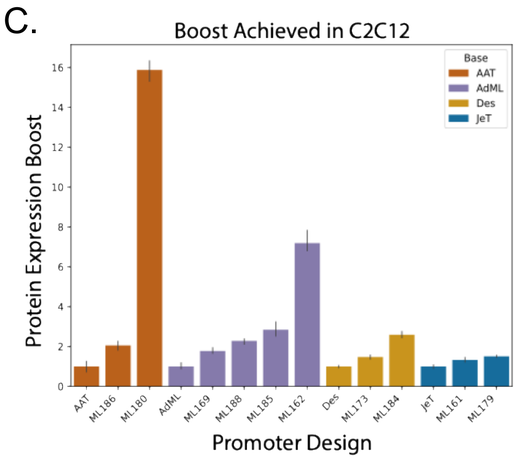


5. MPRA-discovered Designs Boost Protein Expression by up to 30-fold Across Core Promoters

TFBS-mediated activity identified by MPRA is validated at the protein level. (A) Promoter activity at the protein level was assessed using flow cytometry-based assay. Candidates were selected for validation based on performance and specificity. **(B)** The barcode assay showed good correlation with with protein assay for *in vitro* validation (r=0.64). Liver-specific AAT, which is inactive in the brain, was enhanced up to 30-fold in N2A. Already strong JeTs' performance doubled. **(C)** Core promoters were boosted up to 15x in mouse C2C12 cell lines.







6. Conclusions

- MeiraGTx has developed a predictive, experimentally validated high-throughput platform for the discovery of small, potent synthetic promoters across diverse tissues and cell types.
- TFBS in our library showed strong promoter-specific selectivity, enabling the identification of both promoter-enriching and activity-limiting sequences—key insights for designing potent synthetic promoters for gene therapy [3].
- TFBS optimization translates to larger payloads at lower viral doses with improved tissue restriction.
- The TFBS-Promoter dataset is a robust foundation for machine learning-based modelling.

References:

1. van Arensbergen, J. (2019). High-throughput identification of human SNPs affecting regulatory element activity. *Nature genetics*, *51*(7), 1160–1169.
2. Martinez-Ara et al. (2022). Systematic analysis of intrinsic enhancer-promoter compatibility in the mouse genome. *Molecular Cell, 82(13), 2519–2531.e6.*
3. Li, C., & Samulski, R. J. (2020). Engineering adeno-associated virus vectors for gene therapy. *Nature reviews. Genetics*, *21*(4), 255–272.