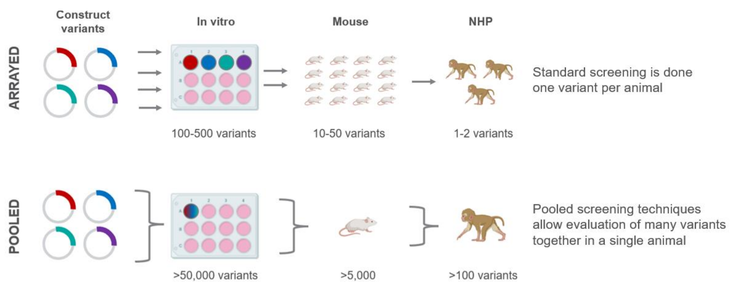
In Vivo High-throughput Pooled Screening of Transcriptional Regulatory Elements for Gene Therapy

Jessica Keenan, Jicong Cao, Soumavo Mukherjee, Anais Sidonia, and Russell Monds
Generation Bio, Cambridge, MA 02142, U.S.A.

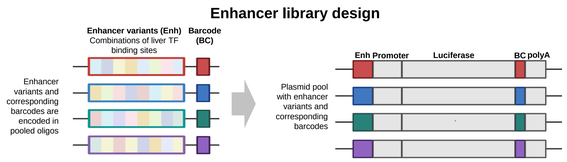
Pooled screening techniques speed the path to discovery and development



Pooled screening approaches simultaneously characterize thousands of construct variants in a single experiment[1], enabling highly efficient screening in the most relevant model systems.

In this work, we utilize a pooled screening approach to quantify expression of 6000 liver enhancer candidates in a single mouse, directly compare to expression in an *in vitro* model, and explore strategies for enhancer design and optimization.

Overview of liver enhancer pooled screening approach



A library of 6000 enhancer candidates was designed by combining binding sites for liver transcription factors (TFs) into arrays. Enhancer candidates were each assigned a unique 10 nt barcode (BC) to facilitate enhancer-specific quantification of mRNA expression. The enhancer/barcode pairs were synthesized as an oligo pool and used to generate a reporter plasmid pool.

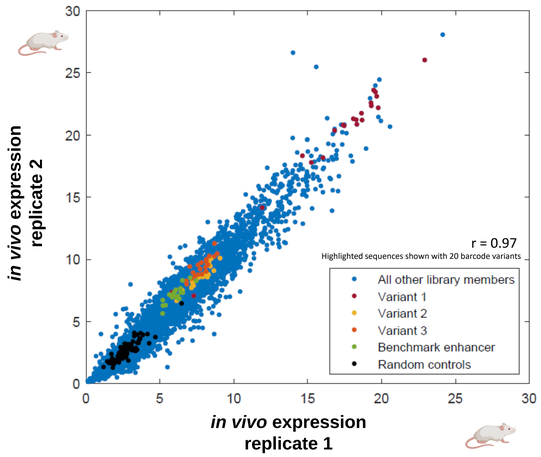
***In vivo* screening workflow**

The enhancer library was delivered to CD-1 mice via hydrodynamic tail vein injection (n = 3) and liver mRNA expression was measured by amplicon sequencing of the barcode regions. mRNA barcode reads were normalized to barcode reads from the DNA input library to control for biases in the plasmid library.

References & Acknowledgements

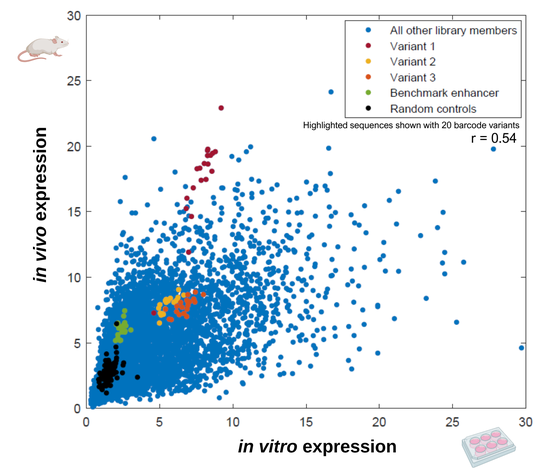
1.Smith, R. P. et al. Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. Nat Genet 45, 1021–1028 (2013).
The authors wish to acknowledge and thank Smith, et al., whose work served as a foundation for this project.

Pooled screening in a single animal identified improved enhancers from a pool of 6000 variants



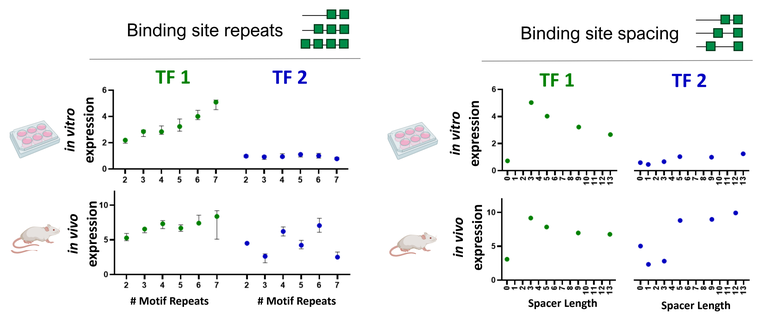
Enhancer candidates spanned a broad range of activity levels, with the top 1% of candidates showing >5-fold increase in expression compared to the median expression of 50 random size-matched sequences and >2.5-fold increase compared to a benchmark enhancer. Biological replicates were highly correlated (r > 0.96) and > 10 barcode counts were recovered for 99.8% of candidates. Control sequences showed expected rank order.

Translation between human *in vitro* and murine *in vivo* models varies between enhancers



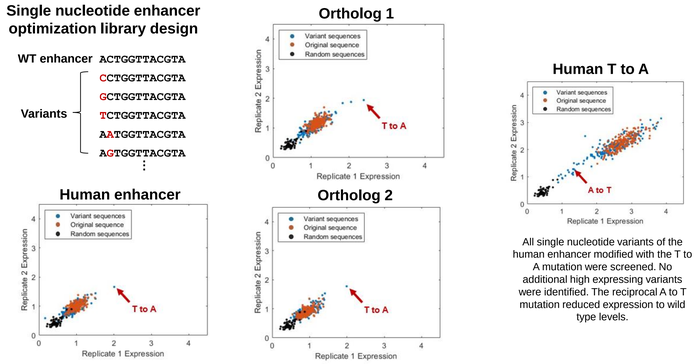
The enhancer library was also screened in the HepG2 human hepatocarcinoma cell line (r = 0.97, n = 2) and expression was compared to *in vivo* data. Many enhancers performed similarly in both model systems, but a large population displayed better performance either *in vitro* or *in vivo*, suggesting the translation between models is enhancer dependent. Notably, many enhancer variants with high expression *in vivo* may not have been chosen for additional characterization based on HepG2 data (e.g., Variant 1). Additional work is necessary to assess translation of differentially expressed enhancers to higher order model systems.

Pooled screening enables learning design rules



Within the enhancer library, a subset of variants were designed to investigate the impact of number of binding site repeats and binding site spacing on expression. TF- and model-specific preferences were observed, suggesting principles for designing enhancers with improved expression in specific models and across model systems.

Pooled screening enables high-resolution sequence optimization



All single nucleotide variants of the human enhancer modified with the T to A mutation were screened. No additional high expressing variants were identified. The reciprocal A to T mutation reduced expression to wild type levels.

A pooled screen was performed in HepG2 cells to measure expression of all single nucleotide variants of the benchmark liver enhancer and orthologs from two species (~1k sequences). A conserved single nucleotide change was identified that increased expression in the three genomic enhancers, highlighting the value of this approach for optimizing regulatory elements and elucidating sequence determinants of function.

Conclusions

- Pooled screening enables highly efficient characterization of construct expression *in vivo*
- A library of 6k enhancer variants was screened in a single mouse, resulting in identification of enhancer variants with improved expression
- Translation between human *in vitro* and murine *in vivo* models varied between enhancers, highlighting the importance of screening in appropriate model systems
- Increased screening efficiency enables learning design principles and high-resolution optimization of regulatory elements