

# Programming gene expression with combinatorial promoters

Robert Sidney Cox III<sup>1</sup>, Michael G. Surette<sup>2</sup>, Michael B. Elowitz<sup>1,3,\*</sup>

1. Division of Biology, California Institute of Technology, Pasadena, USA

2. Department of Microbiology and Infectious Diseases, University of Calgary, Calgary, Canada

3. Department of Applied Physics, California Institute of Technology, Pasadena, USA

\*email: melowitz@caltech.edu

## Abstract

Promoters control gene expression in response to one or more transcription factors. In order to understand natural—and to design synthetic—promoters, it is essential to understand the relationship between promoter function and architecture. We constructed a combinatorial library of random promoter architectures. We characterized 288 promoters containing up to three inputs from four different transcription factors. Multiple -10 and -35 boxes varied promoter strength over five decades. To analyze the functional repertoire, we defined a representation of promoter function in terms of logic and symmetry. Using these results we identified heuristics for programming gene expression with combinatorial promoters.

## Extended Abstract

We report a synthetic library-based approach for construction and analysis of modular combinatorial promoters. Here, the placement, affinity, and sequence of known binding sites<sup>1,2</sup> are systematically varied (Fig. 1), allowing us to determine the range of functions encoded by the simplest combinatorial promoters. This approach reveals fundamental features of the relationship between promoter architecture and function.

Because of the continuous nature of the output levels in each input state, Boolean logic does not accurately represent all possible promoter functions<sup>3</sup>. Therefore, we introduced an intuitive 3-dimensional parameterization for the space of promoter functions. In this scheme, we represented promoter functions with three numerical parameters that quantify dynamic range, logic type, and asymmetry (Fig. 2). We define  $r$  as the ratio of the maximum to minimum expression level. Second, the parameter  $l$  quantifies the logical behavior of the promoter: from pure OR ( $l = 0$ ) to pure AND logic ( $l = 1$ ). Third, the parameter  $a$  quantifies the asymmetry of the gate with respect to its two inputs. At  $a = 0$ , the gate responds symmetrically to either inducer, while at  $a = 1$ , the promoter responds only to a single input.

The library contained two classes of dual-input gates. The repressor-repressor (RR)-promoters contained operators for the two repressors LacI and TetR, while the activator-repressor (AR)-promoters responded to the activator AraC and one of the repressors. These two classes of dual-input gates exhibited differing, but overlapping, distributions of logical phenotypes.

Combinatorial synthesis of synthetic promoters permits systematic perturbation of promoter architecture and rapid identification of promoters that implement specific functions. The spectrum of promoter functions observed here highlights several heuristic rules for promoter design: (1) *Unlimited regulation*. Regulated promoter activity is independent of unregulated activity. (2) *One is enough*<sup>4</sup>. Full repression is possible with a single operator between -60 and +20. Activators function only upstream of -35 (*distal*), and have little effect downstream (*core* or *proximal*). (3) *Repression trend*<sup>2</sup>. The effectiveness of repression depends on the site with  $core \geq proximal \geq distal$ . Following this trend, RR-promoters may be symmetric or asymmetric. (4) *Repression dominates* activation, producing asymmetric AR-promoter logic. (5) *Separation of variables* generates SLOPE and asym-SLOPE logic only. Moving operators closer together makes the logic more AND-like.

Figure 1. Combinatorial assembly of promoters and promoter libraries. (A) The assembled sequence of a synthetic promoter. The 5' overhangs of each unit are shown in red. The RNA polymerase boxes (-10 and -35) are highlighted in yellow, and the predicted start site of transcription (+1) is capitalized. Operator colors are consistent throughout the figure. (B) Steps in promoter assembly and ligation into a luciferase reporter vector. (C) Luminescence measurements in 16 inducer conditions for the promoter shown in (A). (D) The 48 unique units used in the library. These contain operators responsive to the four TFs (indicated) in the regions *distal*, *core*, and *proximal*. Color intensity corresponds to inferred operator affinity. The promoter fragments corresponding to (A) are boxed in red.

## Figure 1.

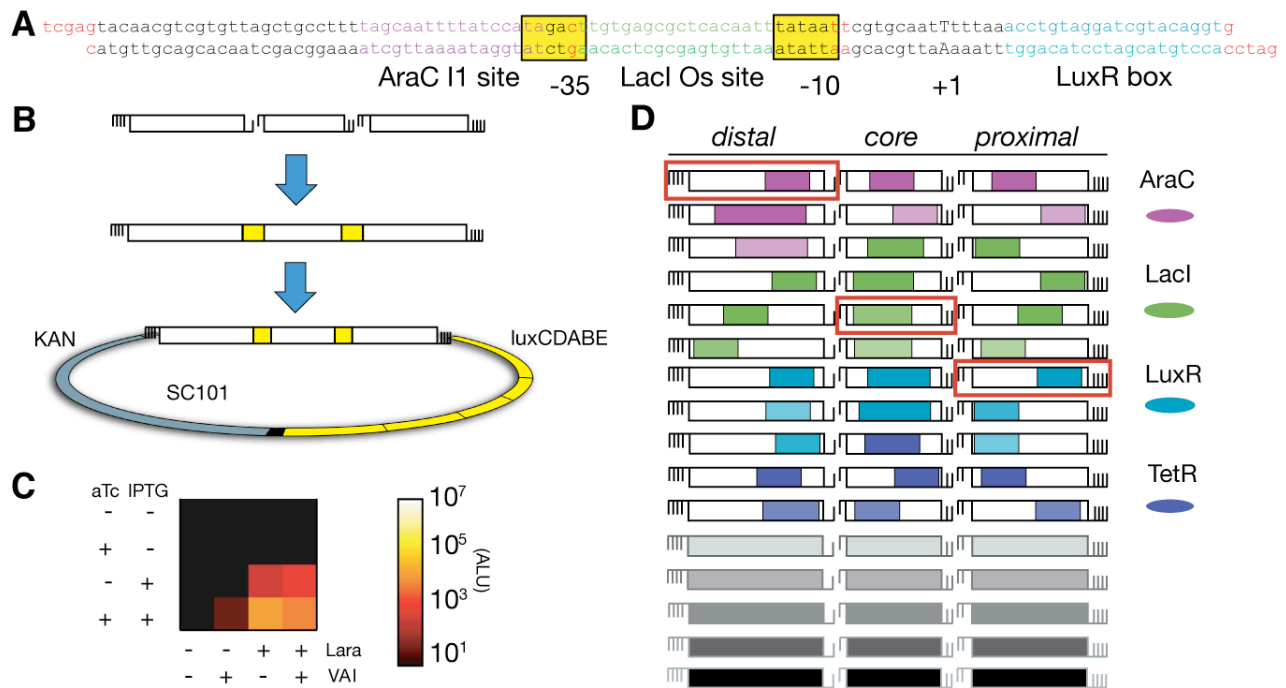
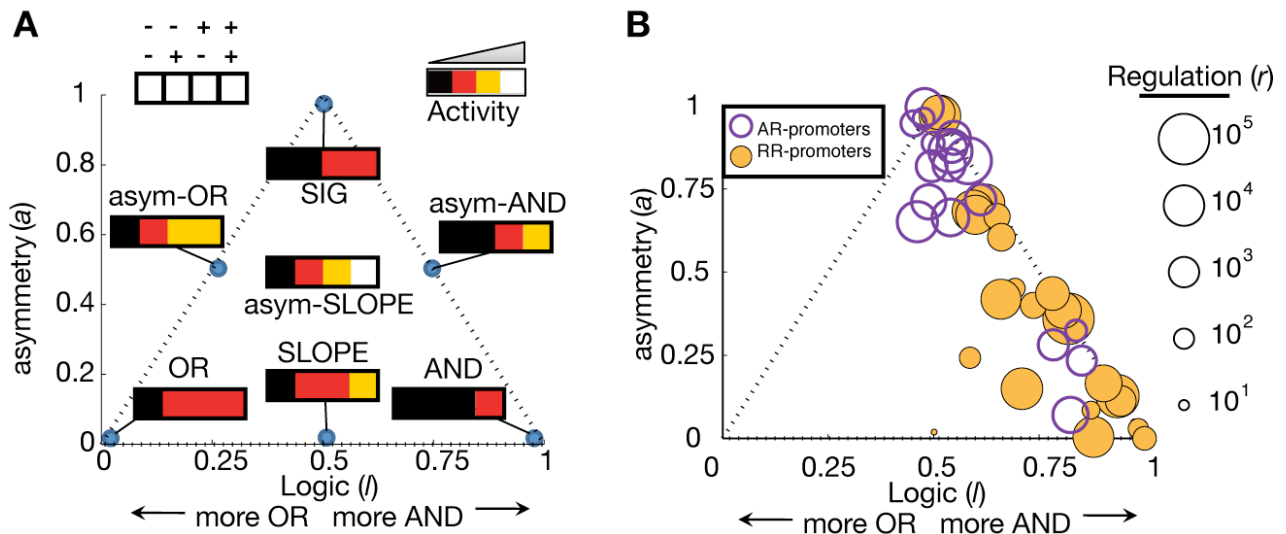


Figure 2. Dual-input gates in logic-symmetry space. (A) Diagram showing the space of allowed logical phenotypes, with the locations of Boolean and intermediate logic gates indicated. The SIG gate responds completely to one inducer and not at all to the other. The SLOPE gate represents an intermediate logical function between AND and OR, while

the asymmetric gates represent intermediates between SIG and the corresponding symmetric gate. (B) The logical phenotypes of 34 dual-input promoters exhibiting strong regulation ( $r > 10$ ). AR-promoters are shown as purple circles, RR-promoters are shown as gold disks. The diameter of each disk is proportional to the logarithm of its regulatory range,  $r$ .

**Figure 2.**



<sup>1</sup> S. Ogden, D. Haggerty, C. M. Stoner et al., *Proc Natl Acad Sci U S A* **77** (6), 3346 (1980); C. F. Beck, R. Mutzel, J. Barbe et al., *J Bacteriol* **150** (2), 633 (1982); D. K. Hawley and W. R. McClure, *Nucleic acids research* **11** (8), 2237 (1983); W. Kammerer, U. Deuschle, R. Gentz et al., *The EMBO journal* **5** (11), 2995 (1986); W. Ross, K. K. Gosink, J. Salomon et al., *Science* **262** (5138), 1407 (1993); W. C. Fuqua, S. C. Winans, and E. P. Greenberg, *J Bacteriol* **176** (2), 269 (1994); R. Lutz and H. Bujard, *Nucleic acids research* **25** (6), 1203 (1997).

<sup>2</sup> M. Lanzer and H. Bujard, *Proc Natl Acad Sci U S A* **85** (23), 8973 (1988).

<sup>3</sup> Nicolas Buchler, Ulrich Gerland, and Terence Hwa, *PNAS* **100** (9), 5136 (2003); Y. Setty, A. E. Mayo, M. G. Surette et al., *Proc Natl Acad Sci U S A* **100** (13), 7702 (2003); Mark Ptashne, *A genetic switch : phage lambda revisited*, 3rd ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2004); L. Bintu, N. E. Buchler, H. G. Garcia et al., *Curr Opin Genet Dev* **15** (2), 125 (2005); L. Bintu, N. E. Buchler, H. G. Garcia et al., *Curr Opin Genet Dev* **15** (2), 116 (2005); N. Rosenfeld, J. W. Young, U. Alon et al., *Science* **307** (5717), 1962 (2005); S. Atsumi and J. W. Little, *Proc Natl Acad Sci U S A* **103** (50), 19045 (2006); N. J. Guido, X. Wang, D. Adalsteinsson et al., *Nature* **439** (7078), 856 (2006); R. Hermsen, S. Tans, and P. R. Wolde, *PLoS Comput Biol* **2** (12), e164 (2006); A. E. Mayo, Y. Setty, S. Shavit et al., *PLoS Biol* **4** (4), e45 (2006).

<sup>4</sup> J. Collado-Vides, B. Magasanik, and J. D. Gralla, *Microbiol. Mol. Biol. Rev.* **55** (3), 371 (1991).