

A High-Throughput, Synthetic Biology Approach to the Study of Eukaryotic Gene Regulation

Allison Goldberg¹, Noel Goddard¹

¹Hunter College, City University of New York, NY
Systems Biology Center New York and Department of Pharmacology and Systems Therapeutics,
Mount Sinai School of Medicine, New York NY.

Understanding gene regulation in model species, such as *S. cerevesiae* (yeast), is an important step in understanding gene regulation in higher order eukaryotes. *S. cerevesiae* has not only been sequenced, but a vast amount of data and computational resources are available in the public domain. Unlike other eukaryotes, the non-coding portion of DNA in yeast is relatively small, and there is an absence of much post-transcriptional control (e.g. introns, RNAi). These two factors make the study of regulatory regions in yeast a more feasible problem.

A library of synthetic promoter variants is designed from manipulation of predicted regulatory motifs in the native sequence. Because commercially synthesized DNA is commonly limited to ~100bp, we chose to use low-cost chip-synthesized sequences as substrates. Assembly of the promoter variants begins by taking cleaved sequences (140bp) from a Microarray (provided by Agilent) and amplifying subsets of the sequences away from the larger pool (40,000). One subset contains approximately 10 sequences (or fragments) to build a single regulatory sequence (1 promoter variant). The design of the fragments includes primer sequences that flank all of the sequences belonging to a single subset. Primers are modified with ribonucleotides at positions 10 and 20 (20bp length total), thus inserting them into the PCR product. When exposed to alkali conditions, the ribonucleotides are nicked. The resulting short pieces are stripped away by denaturing the strands and separating the 10bp pieces away from the 120bp pieces. Strands are then re-natured (100bp of each piece are double-stranded with a 20bp overhang on both sides). Overhangs are blunted with T4 polymerase, leaving 100bp fragments to build regulatory sequences.

Fragments are assembled in a two-part process. First, the pool of fragments is thermo-cycled (like a PCR) to allow the fragment overlaps to extend. Ideally, fragments of different sequences exist in approximately equal concentrations, but because that is difficult to achieve, only a few full-length sequences are made in the extension step. A normal PCR is then performed only amplifying successfully assembled, full-length sequences. Partially assembled sequences are not recognized by the chosen primers.

Assembled variants are then cut at a restriction site (Not1) downstream from the promoter variant, and a reporter is then inserted. We have chosen to use both a fluorescent reporter and a sequence barcode to enable flexible downstream analysis. A particularly powerful part of the design is that the assembled regulatory sequence with the reporter is inserted directly into the chromosome, subjecting the regulatory sequence to “normal” chromatin conformation. A total of 200 variants will be assembled and expression will be measured using high-throughput sequencing technologies (barcode) or with a cell counter (fluorescent marker)