

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

Allison Goldberg and Noel Goddard, *Hunter College, CUNY, Department of Physics and Astronomy*



Abstract

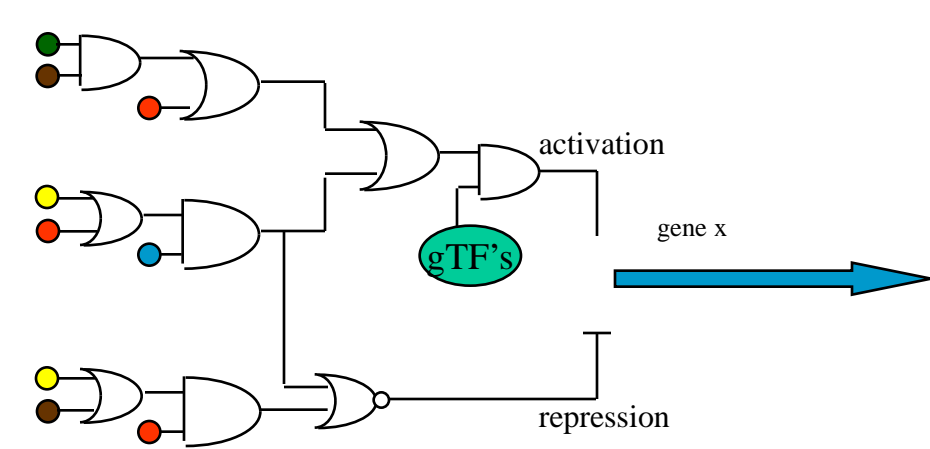
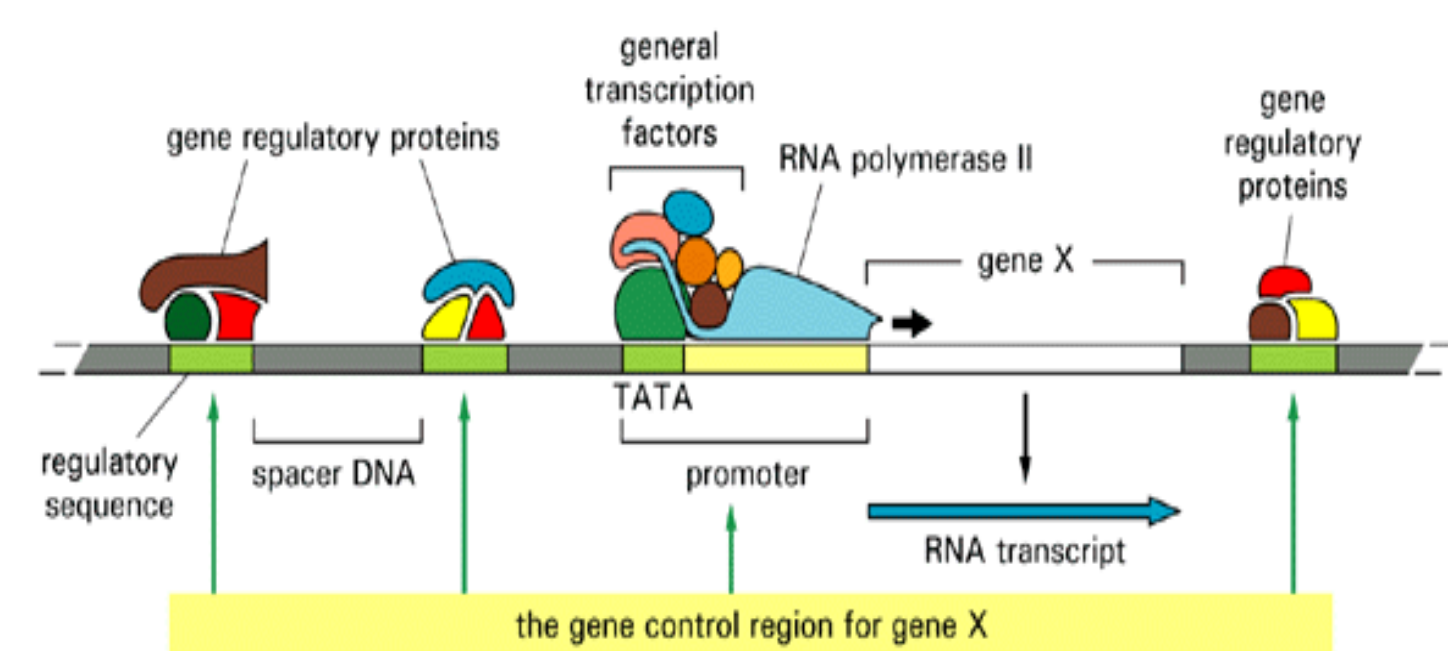
Understanding gene regulation in model species, such as *S. cerevisiae* (yeast), is an important step in understanding gene regulation in higher order eukaryotes. *S. cerevisiae* has not only been sequenced, but a vast amount of data and computational resources are available in the public domain. Unlike 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.

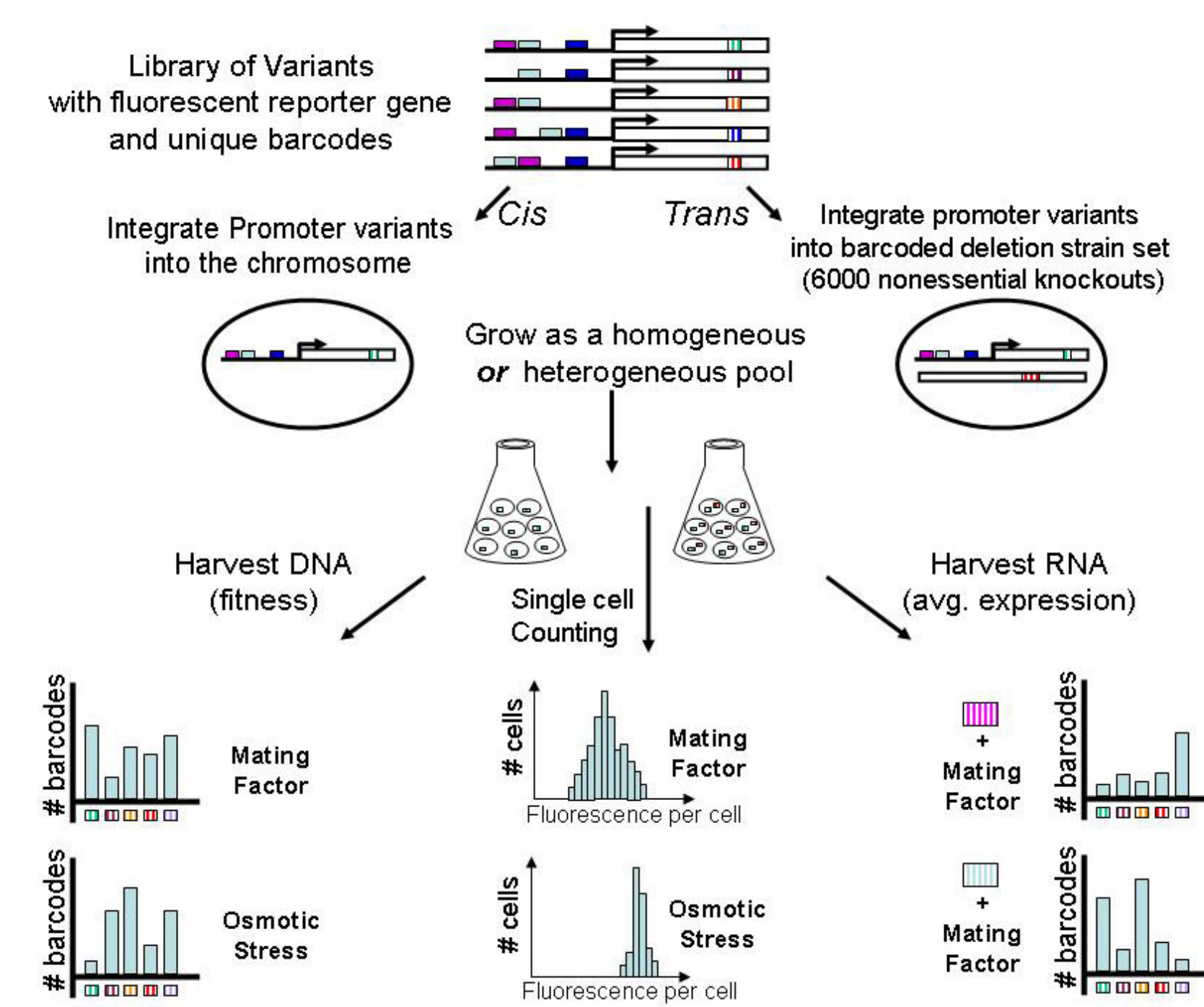
Assembled variants are then cut at a restriction site (NotI) downstream from the promoter variant, and a reporter is then inserted. We have chosen to use both a fluorescent reporter and a sequence barcode to enable flexible downstream analysis. A particularly powerful part of the design is that the assembled 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).

Gene Regulation: A Series of Logic Operators



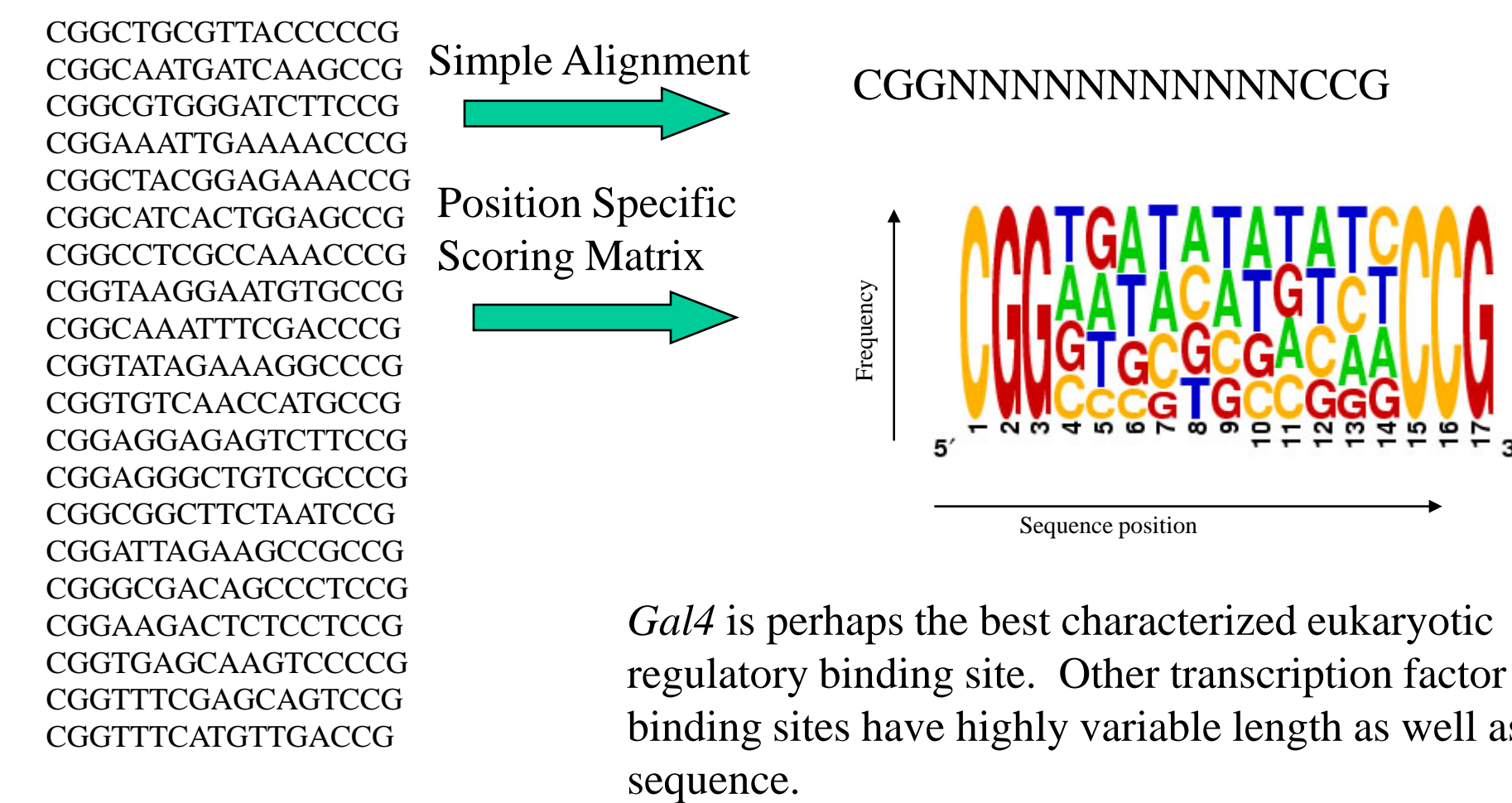
This is more applicable to development. Feedback loops are necessary to describe most response regulation.

Overview: Experimental Scheme



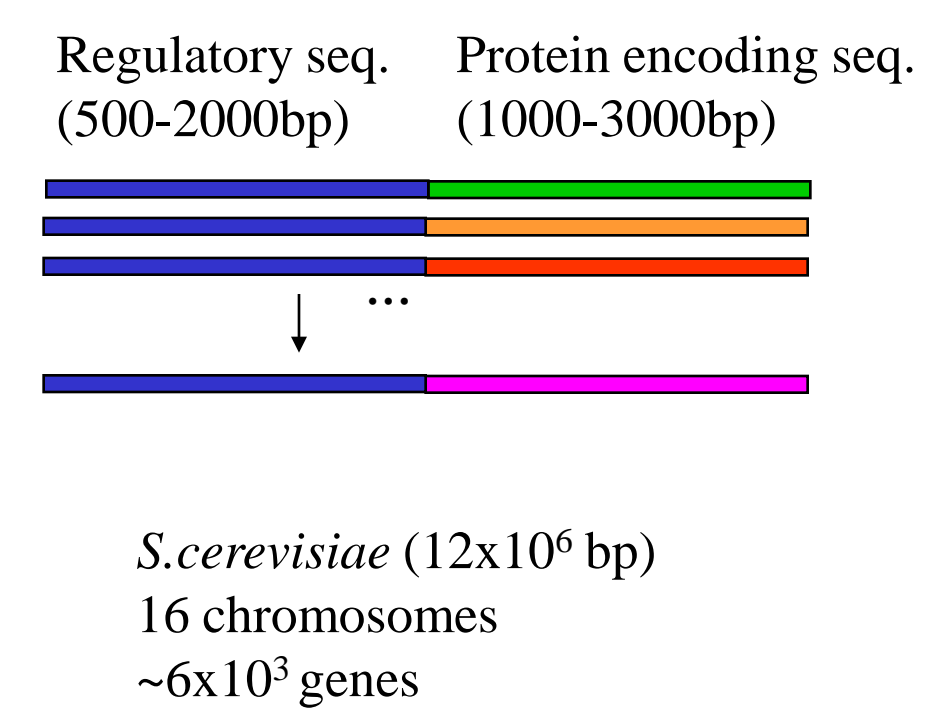
Why Is Computational Prediction Difficult?

20 of 208 *Gal4* binding Sites (1-5 copies per gene)

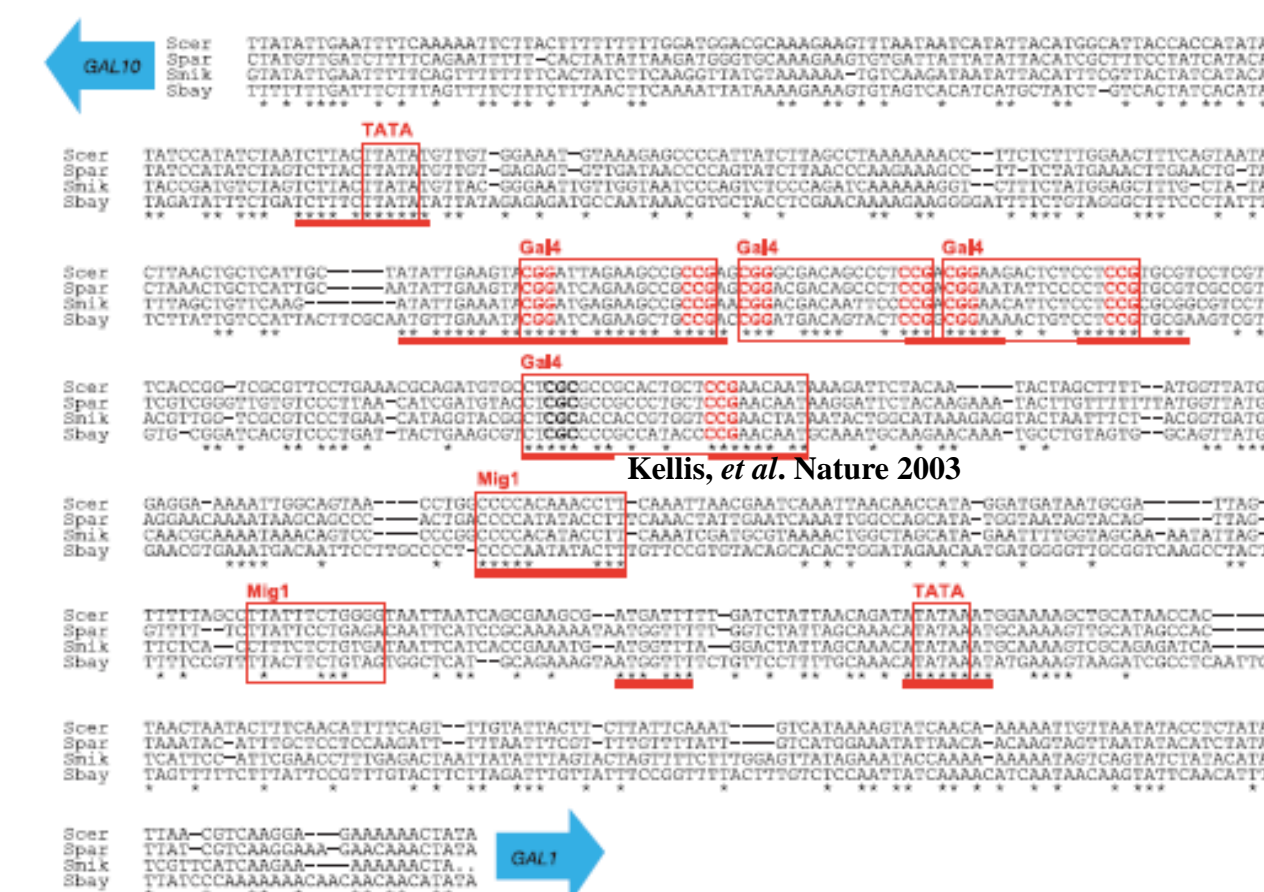


Computational Discovery of Transcription Factor Binding Motifs in the Post-Genomic Era

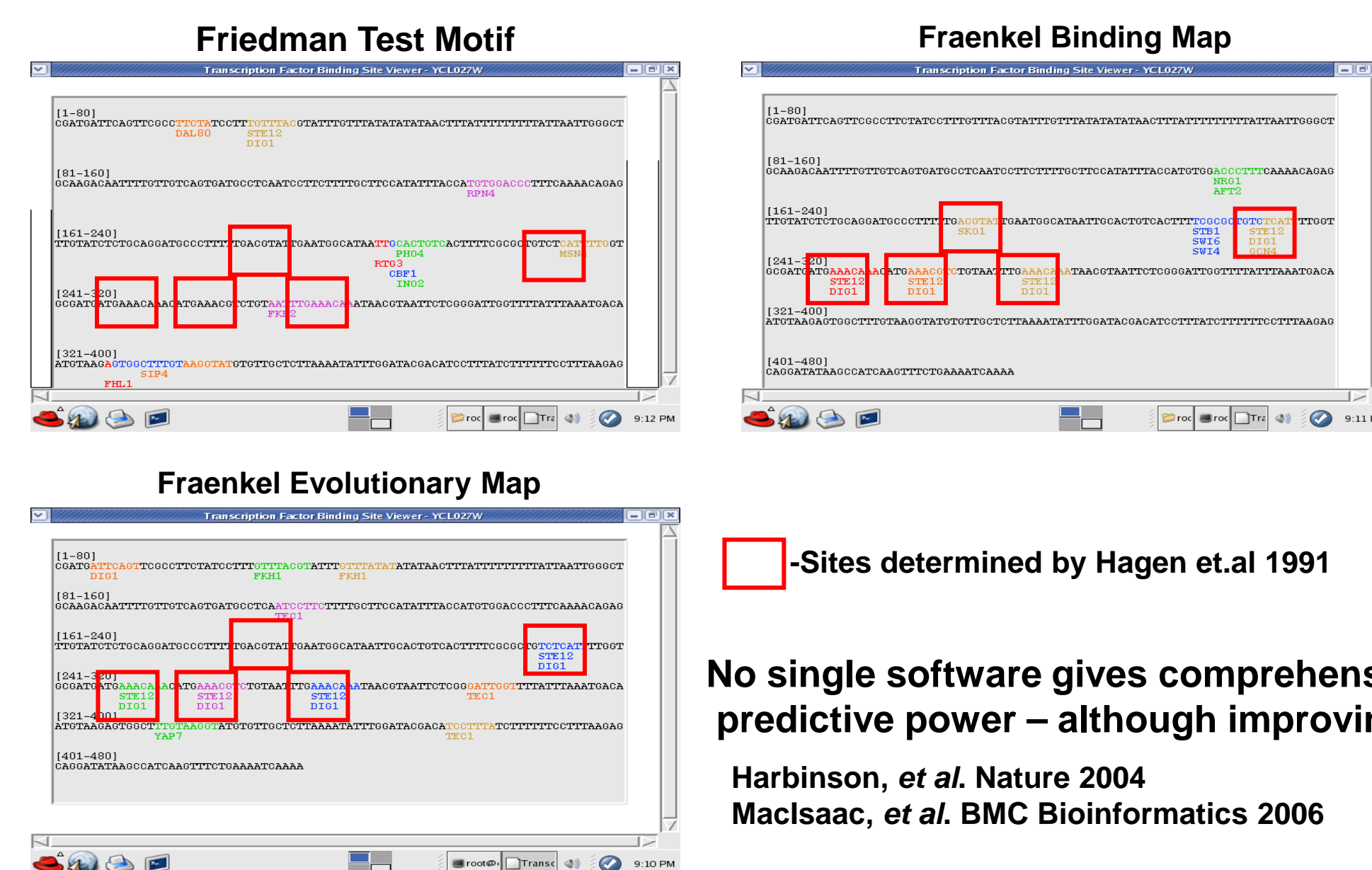
Align regulatory sequences of all genes in a single species and search for conserved sequence



Align regulatory sequences of conserved genes across closely related species and search for conserved sequence

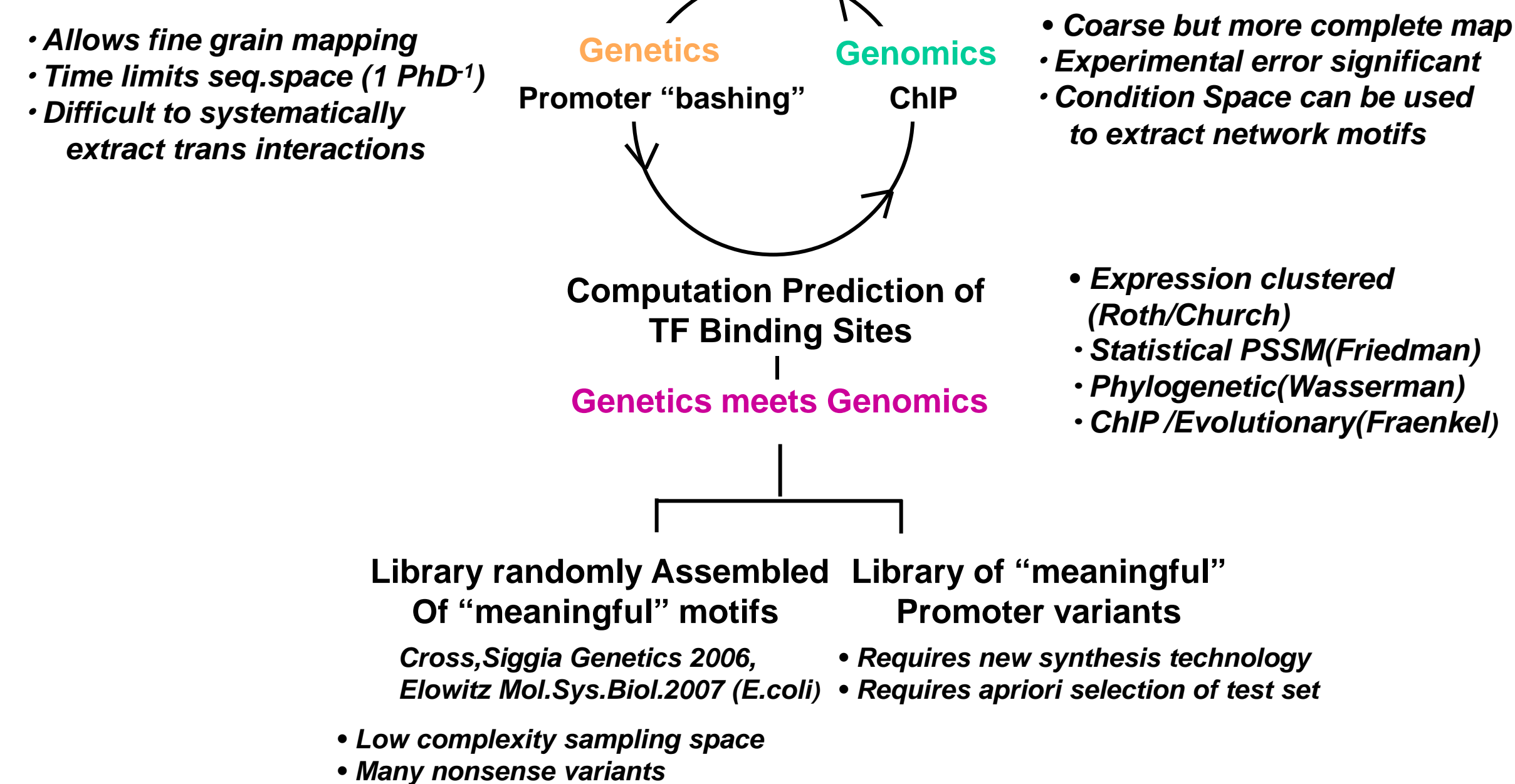


3 Predictions for FUS1 TF Motifs Using Different Motif Compendiums

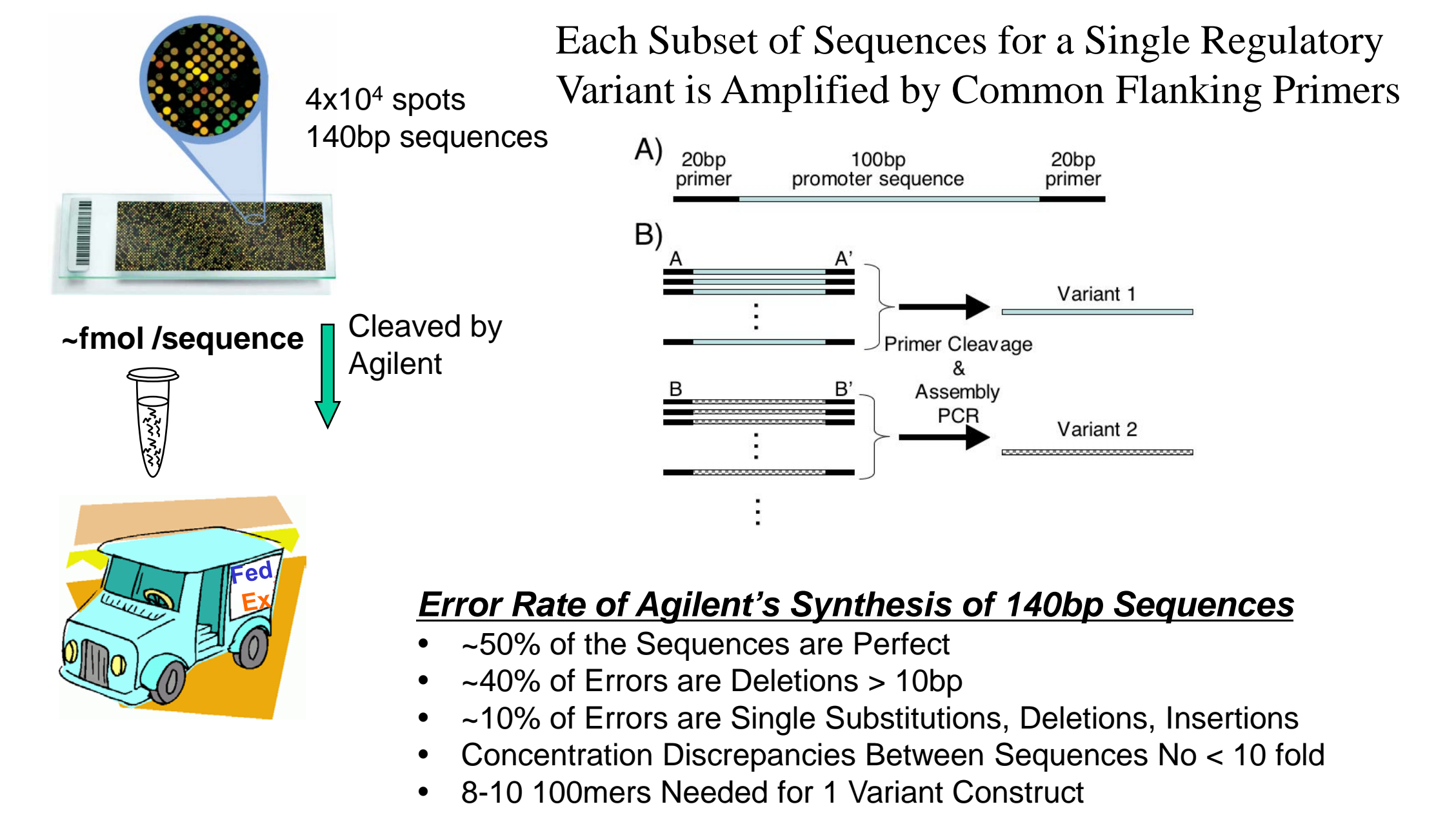


The FUS1 gene is a good candidate gene to study regulation because the dynamic range of expression is very large (~27 fold). It is not essential to the organism, therefore we can replace the gene product with a reporter, facilitating measurement. As a practical consideration, FUS1 is part of the mating pathway that is easily stimulated with commercially available pheromones. The combination of these factors allows for the measurement of small differences between our constructed regulatory variants. By utilizing computational predictions, we are able to identify possible regulatory motifs, and can use that information to design specific promoter variants for FUS1.

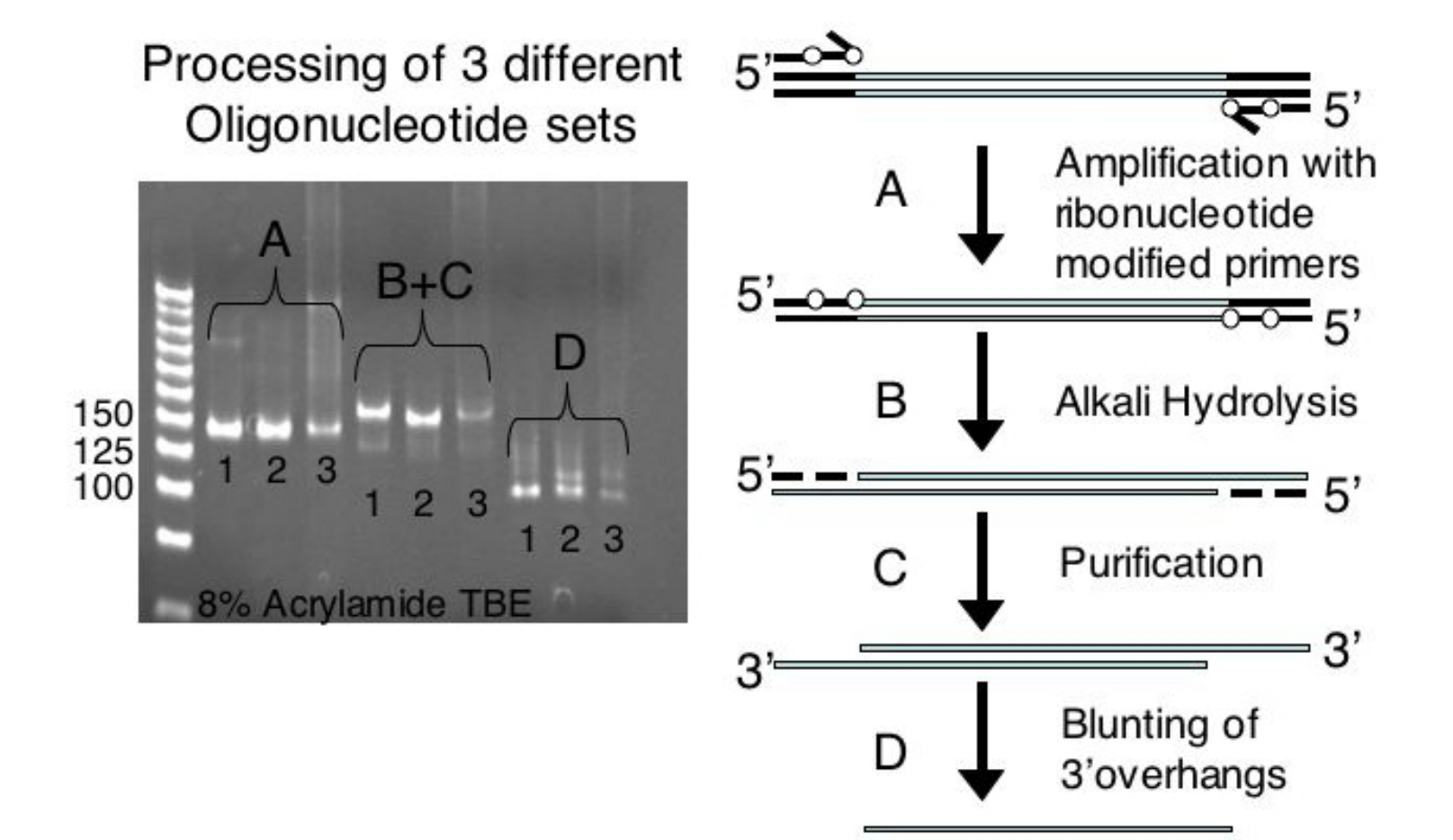
Community Progress



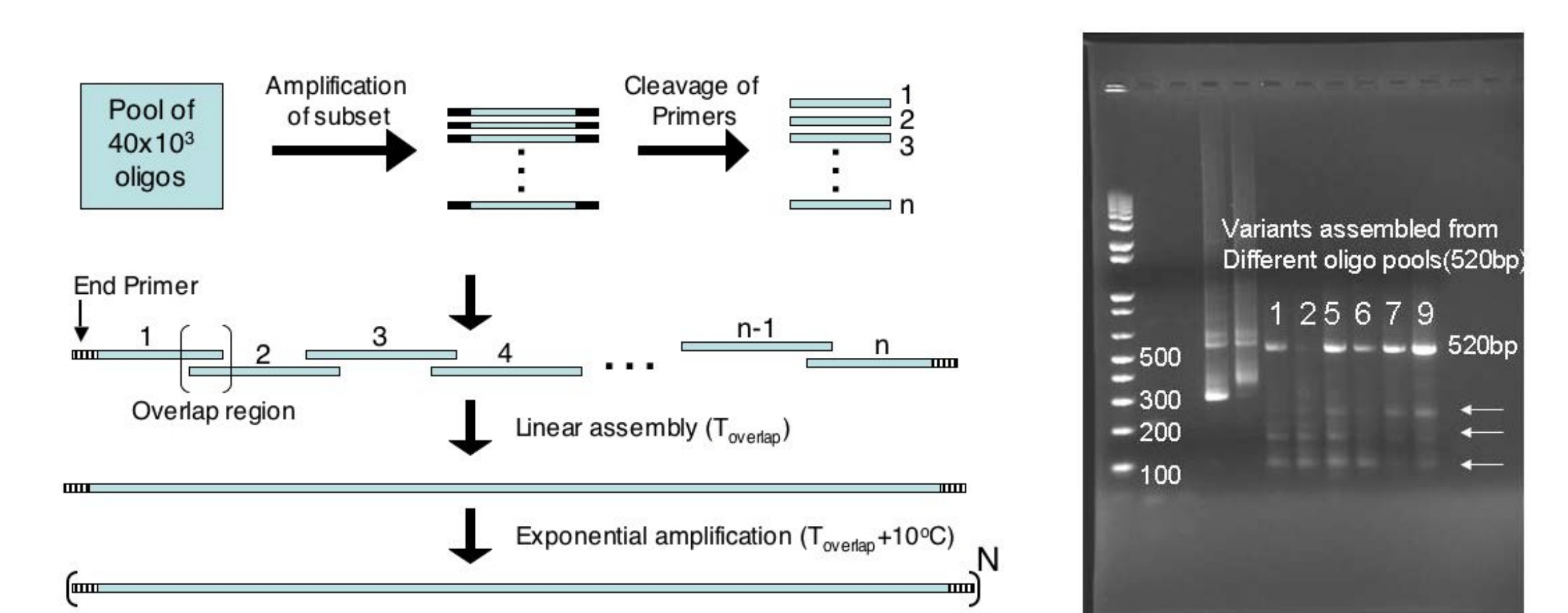
Assembly From Low Cost, High-Throughput, Solid Substrate Synthesized DNA



Primer Sequences Need to be Cleaved From the Sequence of Interest, Exposing the Ends for Assembly



100bp Fragments Are Used to Assemble the Larger Regulatory Sequences



References

Fus1 Promoter Bashing
D.C. Hagen, G.McCaffrey, and G.F.Sprague, Jr. *Mol.Cell.Biol.* 11, 2952 (1991)

Chip Based Assembly
J.Tian, H.Gong, N.Sheng, X.Zhou, E.Gulari, X.Gao & G.M.Church, *Nature* 432,1050(2004)

Polony Sequencing
J.Shendure, G.J. Porreca, N.B. Reppas, X.Lin, I J.P. McCutcheon, A.M. Rosenbaum, M.D. Wang, K.Zhang, R.D. Mitra, G.M. Church

Acknowledgements

Supported by the Systems Biology Center in New York, National Institute of General Medical Sciences P50 GM 071558

HARVARD MEDICAL SCHOOL
George Church
Fred Winston

Agilent Technologies
Emily Le Proust
Wilson Woo

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK
Dana Pe'er
Oren Litvin