GENE REGULATION

Breaking the silence

A large-scale effort to identify sequences involved in silencing exon and splice site usage sheds some new light on the properties of these poorly understood motifs and demonstrates the potential of an innovative splicing simulation program.

A considerable number of genes—some estimates suggest as many as 60% in the human genome—are believed to undergo alternative splicing, and this process is essential to the generation of a diverse proteome. Unfortunately, the characterization of the alternative splicing process is daunting; even though fundamental splice site determinants are well understood, far less is known about regulatory motifs enhancing or inhibiting the use of a given site.

MIT investigator Christopher Burge and his colleagues—including splicing pioneer Phillip Sharp—have dedicated considerable effort toward resolving these mysteries. "The idea," explains Burge, "is that splice sites don't contain enough information to pinpoint exon locations, and there are many decoy splice sites and 'pseudo-exons' in transcripts, so there must be additional regulatory elements that are providing this information." Several studies have identified sequences within individual exons known as exonic splice enhancers (ESEs), which seem to specifically stimulate splicing. In one recent study (Fairbrother *et al.*, 2002), Burge's group conducted a genome-wide search for ESE sequences, using a computational screen to identify ten putative ESE motifs, which were then confirmed experimentally.

But this hardly represents the whole picture—equally important are the exonic splice silencers (ESSs), which reduce the likelihood that a splicing event will take place at a 'marked' exon. These have proven particularly hard to characterize for a number of technical reasons, and only a few groups have made headway in identifying ESSs within the genome. "The examples that were known were interesting," says Burge, "and the known ESS sequences generally had little similarity to each other, suggesting that additional unknown classes of silencers were likely to exist."

Burge's group launched a new study, led by post-doc Zefeng Wang, taking an opposite approach from their ESE work—starting with an experimental screen, then applying their data to generate a computational tool for predicting pre-mRNA splicing patterns. The initial library screen employs a vector containing a spliceable minigene (**Fig. 1**); the cloning site is contained within an internal exon, flanked by two exons encoding the green fluorescent protein (GFP). For any random sequence cloned into the construct, the GFP protein resulting from complete splicing will be disrupted by the middle exon. However, if the cloned sequence contains an ESS, middle exon splicing will be suppressed and functional GFP will be expressed, allowing easy separation of the clone by FACS.

Core ESS elements are thought to be short—around 6–10 nucleotides—and the investigators began with a library of randomized decamers. 133 putative ESS sequences were identified, most of which could be grouped into one of seven broad motif classes. Three of these resembled ESSs identified in previous studies, and the nucleotide preferences for ESSs tended to differ markedly from those observed in ESEs. Working with a subset of clones, they directly confirmed splice silencing by RT-PCR, and found that the ESSs they had identified were capable of regulating splicing in alternative exon contexts and in different cell lines. The ESS decamers were then further dissected to identify a set of core hexamers, several of which occurred within previously published genomic splice-suppression mutants, and many of which were significantly enriched in alternatively spliced exons, suggesting roles in regulation.

Based on their experimental findings and on previously published data about splice regulation, the Burge group developed an algorithm called ExonScan. Unlike conventional gene prediction programs, ExonScan is designed as a splicing simulator, to predict the likelihood of individual splice events based on the presence or absence and positioning of splice sites and regulatory sequences. Large-scale tests of the algorithm, using sequence data from a selection of 1,820 human primary transcripts, demonstrated
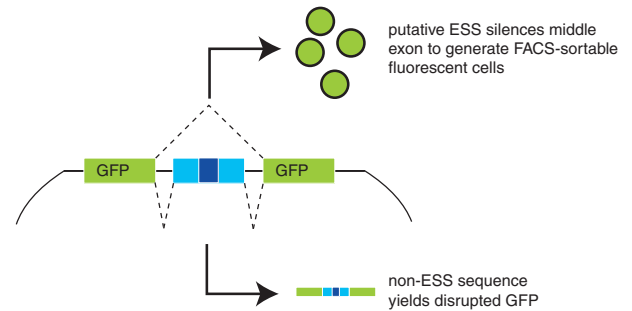

putative ESS silences middle exon to generate FACS-sortable fluorescent cells

non-ESS sequence yields disrupted GFP

Figure 1 | An overview of the Burge group's ESS-trapping strategy.

that the inclusion of ESE and ESS motif data considerably improved the accuracy of splice site predictions as compared to predictions derived exclusively from splice acceptor and donor motifs. The newly acquired ESS data especially enhanced the accuracy of specific splice site delineation, underlining the potential importance of these ESS sequences in distinguishing true exons from decoys.

The Burge group's current priorities include identifying *trans*-acting factors that bind these ESSs, and locating genes that use these elements to regulate alternative splicing. Further development of ExonScan—which is now freely accessible on the web—will also take center stage. "This is the first general splicing simulation program that's been developed for humans," says Burge. "Our goal is not strictly to come up with an algorithm that gets the highest accuracy, but more to use variations of the algorithm to explore models of what the splicing machinery is actually doing." ExonScan may also offer new opportunities for the identification of genes that may slip through the net with other standard gene prediction algorithms. "One thing we're actually thinking about is identifying spliced noncoding RNA genes by predicting their splicing patterns from the genome... [That's] a possible application where ExonScan could be used, but none of the classical gene-finders could."

Michael Eisenstein

RESEARCH PAPERS
Wang, Z. *et al.* Systematic identification and analysis of exonic splicing silencers. *Cell* **119**, 831–845 (2004).
Fairbrother, W.G. *et al.* Predictive identification of exonic splicing enhancers in human genes. *Science* **297**, 1007–1013 (2002).
WEB SITES
http://genes.mit.edu/exonscan/