

I

Our research centers on elucidating how gene regulation works at the level of DNA sequence. Most genes have regulatory sequences that determine when and to what level they are to be transcribed. Given a set of coregulated genes, it should be possible to identify those DNA sequences responsible for their regulation.

We know something about these regulatory motifs. First, they are often repeated upstream of the genes they regulate. Second, sometimes they occupy specific positions with respect to the start of transcription. Third, some of them need to be present in a specific orientation in order to function. Fourth, these motifs sometimes occur in combinations with (or without) specific spacing. We have incorporated the first three points into our algorithms and are currently working on the fourth point. **Figure 1** shows how some motifs are organized in front of a set of zebrafish genes thought to be coordinately regulated. Notice how some motifs are only present in front of a subset of the genes.

Our current efforts are directed at promoter identification, although we are also interested in identifying enhancers that certainly also play a role in gene regulation. Sequences that are identified by

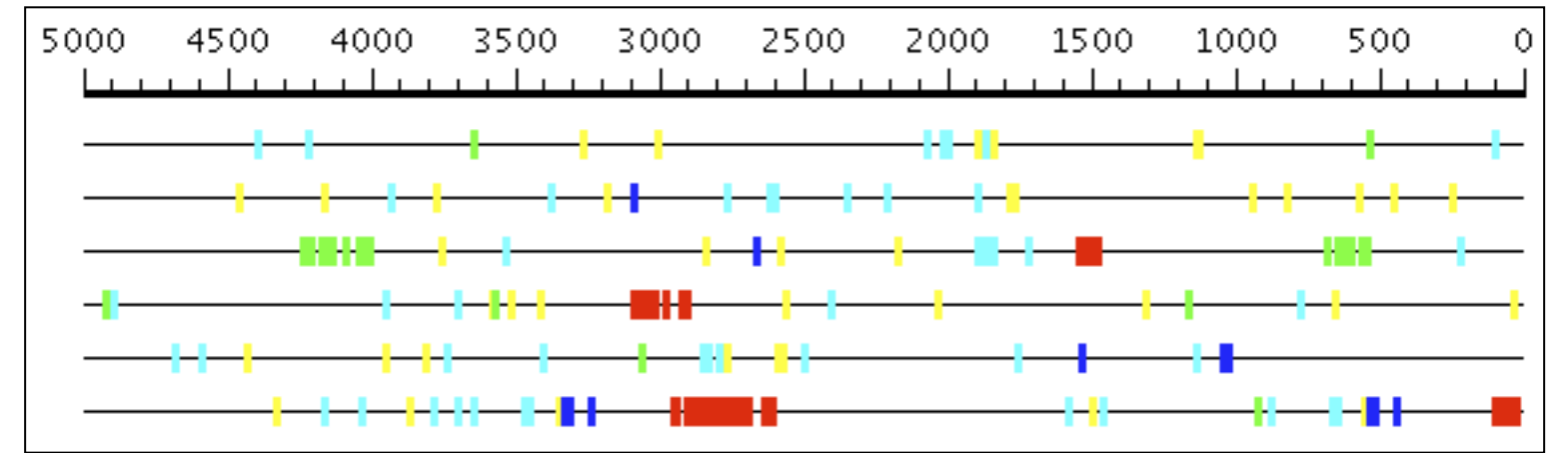


Figure 1. Regulatory motif locations in coregulated genes

our algorithms are most likely transcription factor binding sites, but they may also be involved in chromatin remodeling or other functions.

There are three types of regulatory motifs that need to be identified:

- non-degenerate: ACGGCT
- degenerate: AWGGST
- degenerate multipart: AWCCnnnnnnnGGRT

We developed different algorithms to find each of these motif types and combined them into a single program called **SCOPE** (**S**uite for **C**omputational identification **O**f **P**utative cis-**E**lements). BEAM finds non-degenerate motifs, PRISM finds degenerate motifs containing the non-degenerate output of BEAM. SPACER finds degenerate motifs that have “spacer” regions internally. SCOPE runs all three of these algorithms and merges the output to produce a list of the best scoring motifs of any type.