Essay Describing Research that won Zefeng Wang the RNA Society/Scaringe Young Scientist Award

A major surprise from human genome is that the total number of protein coding genes is much lower than previously estimated, indicating that additional genomic complexity is added at the level of RNA processing. More than 60% of human genes undergo alternative splicing, and the disruptions of splicing can cause human disease. Therefore understanding splicing regulation becomes one of the key challenges in RNA biology. My current research is to thoroughly explore the information that regulates splicing using computational and experimental approaches.

Research accomplishments (past):

My experience with RNA biology started during my dissertation research with Dr. Paul Englund studying *Trypanosoma brucei*, a protozoan parasite that causes diseases. In 1999, I was introduced to RNA interference at a Woods Hole summer course. Back then, the genetic manipulation of trypanosomes had been cumbersome, and the power of RNAi as a genetic tool was just starting to be explored. I developed an inducible RNAi system by expressing dsRNA from a stably integrated vector. This approach had allowed me to gain insight into the parasite's biology, in particular that concerning the replication of trypanosome's unique mitochondrial DNA (termed kinetoplast DNA or kDNA, a network of thousands of DNA circles that are topologically interlocked in a planar array). Using RNAi, I successfully determined the function of a DNA topoisomerase II in kDNA replication, and further clarified how the trypanosome maintains kDNA network size.

To further explore the power of RNAi as genetic tools, I collaborated with James Morris and Mark Drew to develop an RNAi vector (called pZJM) to simplify the induction of dsRNA in trypanosomes. This vector has since then become a standard reagent to silence genes in trypanosomes, and was distributed to more than 50 labs around the world. We also developed an RNAi based genetic screen with pZJM containing a library of genomic DNA fragments. This RNAi library, the first in any organism, allowed us to do forward genetic screen to find new gene(s) responsible for the biological processes of interest. We have successfully identified the genes responsible for changes on cell surface glycoprotein, and for resistance to a drug called tubercidin (a toxic analog of adenosine).

Research accomplishments (present):

I continue my journey in RNA biology as a postdoctoral fellow in Dr. Chris Burge's lab. Here I turned to the question of how splicing specificity is regulated. It had been shown in human transcripts that sequences near splice sites contain only part of the information required for accurate splicing. Further, human transcripts contain numerous 'decoy' splice sites that are almost never used. Despite this potential for sloppiness, the splicing occurs with remarkable precision, suggesting that additional information must contribute to determining the splicing specificity. The prime candidates are *cis*-elements in exons or introns that either enhance or silence the usage of adjacent splice sites. Depending on their locations and effects on splicing, these elements are defined as exonic splicing enhancer (ESE), exonic splicing silencer (ESS), intronic splicing enhancer (ISE) and intronic splicing silencer (ISS).

ESSs are exonic *cis*-regulatory elements that inhibit the use of adjacent splice sites, often contributing to alternative splicing. Although very abundant in the human genome, only a handful of ESSs had been identified by mutational analysis. Most of these examples share little similarity, suggesting that many more remain to be discovered. To systematically identify ESSs, I developed a cell-based splicing reporter system to screen a library of random decamers for ESS. This screen, which we call the Fluorescence-Activated Screen for Exonic Splicing Silencers (FAS-ESS, or FAS for short), has identified more than one hundred ESS decamers. Most ESSs can be clustered into groups to yield seven putative ESS motifs. Potential roles of ESSs in splicing were explored using various computational methods including the development of ExonScan, an algorithm that simulates splicing based on known or putative splicing-related motifs. The systematic analyses of our newly identified ESSs had lead to a new hypothesis that ESS may play a general role in altering splice site usage, which we called splice site definition. This work combined the strength of a large-scale screen and computational analysis, and a similar strategy can be used to systematically identify and analyze other regulatory elements (Wang et. al. 2004, Cell, 119: 831-845)

Correct splice site recognition is critical in pre-mRNA splicing. When testing our new hypothesis of splice site definition, I found that almost all of a disparate panel of FAS-ESS elements alter splice site choice when placed between competing sites, consistently inhibiting use of intron-proximal 5' and 3' splice sites, suggesting a general role for ESSs in splice site definition. Supporting such a role, ESSs were found to be both abundant and highly conserved between alternative splice site pairs. In addition, mutation of ESSs located

between natural alternative splice site pairs consistently shifted splicing toward the intron-proximal site. Some ESEs promoted use of intron-proximal 5' splice sites, and tethering of hnRNP A1 and SF2/ASF proteins between competing splice sites mimicked the effects of ESS and ESE elements, respectively. Further, I observed that specific subsets of ESSs had distinct effects on a multifunctional intron retention reporter, and that one of these subsets is likely preferred for regulation of endogenous intron retention events. Together, our findings provide a comprehensive picture of the activities of ESSs generally in the control of splicing decisions (Wang et.al. 2006, Mol Cell, 23:61-70).

The FAS strategy I developed can be easily adopted to screen for other *cis*-elements, and I am currently conducting a FAS-ISS screen. Since ISSs are the least understood regulatory *cis*-elements, a systematic identification and further analysis of ISS has provided new information about splicing regulation. I am also collaborating with Grace Xiao, another post-doc in our lab, to study the functional interaction of different splicing regulatory elements using the information of sequence conservation between human and mouse. The long-term goal of all these studies is to produce a set of rules for the regulation of splicing specificity, which we refer as “splicing code”.