How time flies! It is the 20th anniversary of RNA. For many of us, our research career is largely co-incident with the launch and progression of the journal, which has provided bonds and guiding lights for our community. Here, I share my personal experience in entering this exciting RNA world through contributing to the discovery of SR proteins in pre-mRNA splicing, and emphasize that, while a large array of principles have been established for the function of SR proteins in the past two decades, there are still many outstanding questions on the roles of SR proteins in diverse regulatory activities in mammalian cells.

A personal roller coaster in the race to discover the SR family

With the establishment of the in vitro splicing system in early ’80s, the machinery for the spliceosome was quickly shown to consist of 5 small nuclear ribonucleoprotein particles (snRNPs), two of which (U1 and U2) are directly responsible for defining the 5′ and 3′ splice sites in the pre-mRNA. But when I entered this field in 1988 as a postdoc in Tom Maniatis’s lab, we knew virtually nothing about protein factors required for pre-mRNA assembly into the spliceosome. The biochemical fractionation/reconstitution strategy, which had been effectively taken to tease apart the basal transcription machinery, was proving to be challenging. I was charged to let the mouse immune system sort out the essential factors for us by raising monoclonal antibodies against crude gel-filtration purified spliceosomes. By generating >2000 candidate monoclonal lines and testing >200, I was fortunate to land on SC35 (now renamed as SRSF2). The antibody against this non-snRNA factor potently inhibits splicing, and interestingly, also decorates the nucleus with bright “stars” known as nuclear speckles. Our first paper was accepted as an Article to Nature with only a two-sentence review to recommend publication, a once-in-lifetime experience that I have never been able to repeat. I was also desperate to further prove that SC35 protein detected on Western was a true active splicing factor by affinity purifying the protein, cutting it out of an SDS gel, extracting it with 8M urea, and renaturing it to see if it would complement the immune depleted nuclear extracts for in vitro splicing.

The initial success was quickly followed by a nightmare, as I was soon in a race to clone the gene with both Krainer and Manley groups who succeeded in biochemical purification of a protein of roughly the same size and apparently the same activity for complementing S100 (which Krainer called SF2) or inducing T antigen isoform switch (which Manley called ASF). I first tried to clone by screening a λgt11 library in bacteria; I failed of course and later realized that the antibody recognizes a phospho-epitope and bacteria do not have the right kinase. I then tried to scale up immune affinity purification and obtained pure protein for sequencing by Edman degradation, but the core facility at Harvard informed me that there was not enough protein (later, I found out why: they accidently lost my protein!). I was next forced to biochemically purify the protein using the Western as a readout. With sufficient material this time, which generated multiple tryptic peptides on HPLC, I was asked to pick candidate peaks for sequencing. Tom suggested I pick three, and if all corresponded to SF2/ASF, I should abandon the project. Among the three I picked, two were identical to SF2/ASF and the third one contained a 14 amino acid sequence with one half corresponding to SF2/ASF, but the other half appeared new. Had all three been peptides of SF2/ASF, I would be doing different things now. Too short to design degenerated PCR primers, I had to use >1000-fold degenerate oligonucleotides to screen by hybridization, which led to cloning of both SF2/ASF and SC35. Having lost the battle to clone SF2/ASF, I realized, together with Mark Roth’s work on amphibian B “snurposomes,” that SF2/ASF and SC35 are members of a larger group of proteins now known as the SR family. The rest is history. I was invited by Tim Nilsen to write the first review on SR proteins. This paper has been

Corresponding author: xdfu@ucsd.edu

Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.049858.115. Freely available online through the RNA Open Access option.

© 2015 Fu. This article, published in RNA, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.
the most highly cited one among all I have published in my career thus far.

**SR proteins in splicing control and beyond**

After thousands of papers on SR proteins contributed by the community, we now have a set of general rules for the function of SR proteins in splicing. SR proteins are essential to commit pre-mRNA to the splicing pathway by promoting U1 and U2 binding to functional 5' and 3' splice sites as well as communication between the functional splice sites during initial exon definition followed by paired interactions across the intron for spliceosome assembly. SR proteins are also involved in regulated splicing by strengthening weak splice sites and/or competing with negative splicing regulators, such as hnRNP proteins. In both constitutive and regulated splicing, SR proteins bind specific, although quite degenerate in many cases, sequence motifs, most in exons and largely purine-rich. Thus, SR proteins have been generally considered positive splicing factors and regulators by functioning through exonic splicing enhancers (ESEs).

The defining features for SR proteins are multiple Arg-Ser (RS) dipeptides in their RS domain besides RNA Recognition Motif (RRM). Relative to 12 core SR protein family members, the SR protein family has many cousins, nieces, and nephews characterized by an RS domain in combination with other protein motifs or domains. Relative to core SR proteins, however, those SR-related proteins are less understood. The RS domain in all SR proteins and related proteins is heavily phosphorylated by specific kinases, a gateway to understand the biology and regulation of these splicing regulators in cell cycle, signaling, and cancer.

SR proteins continue to fascinate us because they appear to have multiple other roles beyond the splicing control. As most splicing takes place co-transcriptionally on chromatin, and defects in the process cause genome instability, which may be connected to a wide range of biological processes. We also discovered that SR proteins have a direct role in both transcription initiation and elongation, suggesting that splicing is not simply coupled with transcription in time and space, but functionally integrated to have mutual benefits for efficient gene expression in the nucleus. SR proteins have also been implicated in multiple other RNA metabolism pathways, including mRNA nuclear export, mRNA quality control, microRNA biogenesis, and even translational regulation in the cytoplasm. These findings place SR proteins in nearly all steps along the axis of the central dogma.

**Much more is left to discover about SR proteins**

While SR proteins are among the best-characterized splicing factors, many fundamental questions remain unsolved regarding how they promote spliceosome assembly and specific interactions within the assembled spliceosome. For example, SR proteins are thought to interact with other RS domain-containing proteins, such as U1-70K and U2AF, to mediate spliceosome assembly. However, all existing data are based on yeast two-hybrid and/or in vitro pull-down assays with additional support from the FRET experiments, and thus, rigorous evidence for this textbook statement has been lacking. SR proteins are part of the spliceosome, but their specific contacts in this RNA machinery await insights from structural analysis of the spliceosome. In fact, there is no high-resolution structural information on the RS domain of any SR protein thus far, which may require both proper phosphorylation and stable interactions with its partners within a splicing complex for success crystallization.

Functional genomics has generated new insights into the in vivo function of SR proteins, such as their position-dependent and context-sensitive effects in regulated splicing, akin to many other splicing regulators. However, the binding specificity of individual SR proteins established with purified proteins does not seem to fully support their exon-central binding pattern in vivo, indicative of cooperative binding with other RNA binding proteins during splice site selection, an important direction to be further explored. As SR proteins have been implicated in diverse aspects of RNA metabolism, it is also the prime time to survey their interactions with different RNA populations in different cellular compartments, within specific complexes, and in response to depletion of potential antagonizing activities. For instance, hnRNP A/B is well known to counteract the function of SR proteins in regulated splicing, but such anticipated reciprocal relationship has not yet been substantiated by in vivo genomic evidence at the genome scale.

Now with the elucidated functions of SR proteins in transcriptional control, we have much to learn if we are to understand co-transcriptional splicing. In fact, as SR proteins are localized in nuclear speckles where many transcription factors and chromatin remodelers also reside, it appears that this nuclear domain may have a bigger role in genome organization for coordinated gene expression, rather than simply serving as storage sites for splicing factors in the nucleus. This research direction is connected to the big picture issue of the cell biology of the nucleus.

Last, but not least, many splicing factors, particularly various SR proteins, have been implicated in diseases. SF2/ASF (a.k.a. SRSF1) has well-characterized roles in oncogenesis. It is also fascinating to note that SC35 (a.k.a. SRSF2) and another RS domain-containing protein U2AF1 (a.k.a. U2AF35) are now known to be two major leukemia genes. On SRSF2, all mutations associated with Myelodysplastic Syndromes (MDS) occur at the single amino acid P95, which resides between its RRM and RS domain. It remains to be explored why this particular site is such a hot spot for mutation, which biochemical function(s) of the SR protein is disrupted, and what is the underlying mechanism for causing the disease. These
questions have connected splicers to the large disease community, which also speaks for the importance of basic research in providing the theoretical ground for understanding specific disease mechanisms and developing effective therapeutics.

Acknowledgments
This monograph is a personal reflection of SR proteins, rather than a review. I thank Dr. Manuel Ares Jr. for critical comments. Work on SR proteins in my lab has been supported by two NIH grants (GM049369 and GM052872) for the past two decades.