
Twenty years of technology

MELISSA J. MOORE

University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA

In looking back at previous issues of *RNA* and my own eight *RNA* papers, what strikes me, in addition to all the great science that's been published in our society journal over the past 20 years, is the evolution (or more properly, revolution!) in the methods and technologies we use. Equally striking has been the exponential increase in the amount of data in each published paper. In two decades, post-transcriptional gene regulation has evolved from a mostly qualitative science into a highly quantitative one. We've gone from spending most of our time just trying to get one experiment at the bench to work to now spending months puzzling over massive datasets generated by high throughput instruments in research cores. As I look to the future, I wonder about the technologies yet to come, how we can possibly deal with the ever increasing size of our datasets, and what will constitute a story "complete enough" to warrant publication 20 years from now.

When I entered graduate school in 1984, Sanger sequencing on large format gels and site-directed mutagenesis using the single-stranded M13 bacteriophage were state-of-the-art. Oligonucleotide synthesis was so time-consuming and expensive that ordering the wrong 17 nucleotide DNA oligo was a major catastrophe. Whole PhD dissertations were devoted to the sequencing of one gene, or the generation and characterization of just a few site-specific point mutations. By the time I arrived in 1989 as a postdoc in Phil Sharp's lab, PCR was the next new thing. I remember how our first PCR machine, shared with the adjoining Houseman lab (~35 researchers in total), was in such high demand that major yelling matches ensued when samples were left in a minute too long. By that time runoff transcription and *in vitro* splicing assays were well established, and native gels and density gradients had revealed that the pre-mRNA splicing machinery was exceptionally large and complex. My main contribution in that period was to show that T4 DNA ligase could join together long RNA molecules, enabling the synthesis of long, site-specifically modified RNAs. Eric Sontheimer in Joan Steitz's group immediately started using this method to incorporate 4-thio-U at key sites in pre-mRNAs, ushering in a decade of site-specific photocrosslinking experiments.

By the time I started my own lab at Brandeis in 1994, the age of genomics was already in full swing. With Craig Venter's 1991 publication of >170,000 expressed sequence tags (ESTs), sequenced on first generation of automated machines, came the realization that alternative splicing was much more prevalent than had previously been thought. Soon serial analysis of gene expression (SAGE; Kinzler 1995) and high-density microarrays (Brown 1995) provided tools to quantitatively measure how the abundance of thousands of mRNAs at once responded to changing cell conditions. Of course substantial completion of the human genome at the turn of the new millennium was another major milestone, with Haussler's UCSC genome browser enabling researchers to easily visualize the gDNA organization of favorite genes. The emergence of RNA-Seq (Wold 2008) coincided almost exactly with my move to UMass Medical School in late 2007. Having theretofore sworn off high-throughput methods in my own lab after an incredibly frustrating attempt to build a microarray facility at Brandeis a decade prior, my lab began making its first deep sequencing libraries within a year of my arrival at UMMS. In our case, the focus has been on mapping transcriptome-wide sites of RNP complex occupancy. Akin to ribosome profiling (Ingolia/Weissman, 2011), RNP footprinting reveals the RNA binding sites of multicomponent complexes of defined protein composition (e.g., the exon junction complex, EJC, or lariat intron-containing spliceosomes). Complementary methods developed a few years before map transcriptome-wide sites of direct RNA-protein contact (e.g., HITS-CLIP, Darnell 2008; PAR-CLIP, Tuschl 2010). So in just 20 years, we've gone from mapping one RNA-protein crosslink at a time to mapping literally millions in one shot. Major challenges for the future are how to determine which of these millions are of functional relevance and how best to integrate the hundreds of currently- and soon-to-be-available datasets to extract even more meaning.

Another area of great technological innovation over the last 20 years has been quantitative proteomics. Major innovations in the 1990's made it possible to deconvolve the spectral mess produced by trypsin digestion of complex mixtures. In collaboration with Steve Gigi, we published in *RNA* in 2002

Corresponding author: Melissa.Moore@umassmed.edu

Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.051052.115>. Freely available online through the *RNA* Open Access option.

© 2015 Moore 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/>.

one of the first complete proteomic analyses of a defined splicing complex: the human C complex. The ensuing tsunami of proteomics studies since (especially from the Lührmann lab) has given us remarkably complete splicing machinery parts lists. More recently, a UV crosslinking and mass spec mash up revealed over 1100 human proteins interacting directly with polyA+ RNA (Hentze 2012). Many are housekeeping enzymes with no apparent RNA binding domains. So now even hardcore enzymologists who've assiduously avoided RNA by studying metabolic enzymes for their entire careers are being forced to come face-to-face with the RNA world!

My last example is the evolution and proliferation of single molecule techniques. After years of trying to purify and visualize human C complex by electron microscopy (with Niko Gregorieff), I will never forget the day that Melissa Jurica showed me her first image of negative stained monodisperse particles (also published in that same 2002 *RNA* paper). Finally, I could see the massive enzyme I'd been working on for so long! That day I ran like a lunatic all over the Brandeis science complex waving that first picture for everyone to see. Although the promise of a high resolution EM structure has yet to be fulfilled (likely due the high compositional and conformational heterogeneity displayed by this incredibly complex and dynamic machine), recent emergence of new data collection methods and direct detection cameras are once again breathing life into spliceosome single particle reconstruction efforts. In the race between X-ray crystallography and EM for the structure of a fully assembled spliceosome, I would have predicted two years ago that crystallography will ultimately win. Now I'm not so sure—I just hope to see the structure in my lifetime by whatever means possible. Meanwhile, another single molecule technique has taken center stage—single molecule fluorescence (SMF) microscopy. As with mass spec and EM, my lab's first foray into SMF analysis of pre-mRNA splicing was published in *RNA* (2008). In that paper we demonstrated (in collaboration with Jeff Gelles) the possibility of performing single molecule

total internal reflectance fluorescence (SM-TIRF) microscopy in crude cell extracts. This advance crucially enabled our more recent work examining the order and dynamics spliceosome assembly, as well as the conformational rearrangements required to complete a single round of intron excision. With our new found ability to generate huge amounts of single molecule data, this more biophysical side of the lab is now facing the same problem as the bioinformatics side—how to efficiently extract meaning from these huge and incredibly rich, but noisy, datasets. I predict that algorithms developed to analyze one type of data will soon be co-opted and adapted to analyze the other. As single molecule biochemistry moves out of boutique labs and becomes democratized, and with exponentially more complex datasets following individual molecules as they move in three dimensions inside cells already being generated, the rate of progress in SMF is now limited mainly by a lack of robust high throughput data analysis tools. Here lies a major challenge for the coming decade.

Although this 20-year technological tour is decidedly skewed toward my own research interests and my penchant for biochemical and biophysical approaches (my sincere apologies for all the things and people I left out), what I hope to have conveyed is my amazement at how far and how fast we've come in the two short decades bookended by the first and this latest issue of *RNA*. No doubt, the next 20 years of *RNA* will report countless more innovations, many of which we cannot even envision right now. With their passion for driving knowledge creation at a breakneck pace, I have every confidence that the next generation of RNA scientists will continue to push methodological development and be early adopters/adapters of technologies from other fields. What I wonder about is simply how we will deal with all of that data! How much data will the average *RNA* paper 20 years hence contain and how will readers ever be able to comprehend it? I sincerely hope to be able to look back in another 20 years and once again marvel at the technological revolutions another two decades will bring.

Happy birthday *RNA*!