Twenty years ago, I was a graduate student in the physics department at MIT. I had no cell phone and could receive the entire NCBI database in the mail. I found biology fascinating, but how could one possibly make sense of it? What even were the big questions? Take Darwin’s beautiful theory of evolution. From a physicist’s perspective “survival of the fittest” is a recursive idea where the boundary condition defining “survival” and “fittest” is created by the biology doing the surviving. This intrinsically self-referential aspect was to me the root of biology’s maddening complexity. Fundamentally the key biology problems of our time: origins of life and understanding multicellular development, resist analysis due to our inability to define boundary conditions.

There is another key aspect of Darwin’s biology theory not captured by “survival of the fittest” and that is the idea of information. Just as Newton’s laws explain the universe better upon inclusion of a force due to gravity, Darwin’s theory of evolution requires the critical notion of biological memory. Biological information needs, just like a finite state Touring machine, mechanisms to “address,” “read,” “compare” and sometime “write” this information. In this picture the central dogma: DNA → RNA → protein resembles the compilation of computer code (stored at the DNA level) into an operating system (expressed at the level of RNA and protein). This biological regulatory system requires the ability to interface with its DNA code so as to respond to environmental and developmental signals making this system intrinsically self-referential and thus extremely hard to solve. Twenty years ago the importance of this puzzle was clear, but the central role of RNA was poorly appreciated by many. Twenty years ago Tim Nilsen established the RNA journal. I personally owe him a tremendous debt of gratitude for his years of steady support as an editor. This is my short story of the last 20 years from my personal perspective:

RNA to the rescue

My interest in RNA was set in motion by the announcement of Phil Sharp’s Nobel prize for the discovery of RNA splicing. I was studying in the Physics reading room when not one but two Nobel Laureates walked in. A third then joined them, whose name I only later discovered was Gobind Khorana. MIT is an intense place, but even at MIT its unusual to find three Nobel prize winners in one spot. I could not help but overhear that they were planning to welcome a new member to their “club.” Much confusion from the physicists ensued, with Khorana patiently explaining important bits of RNA biochemistry that at the time seemed quite mysterious. Only after attending Phil’s talk on his discovery did the picture clear up in my head. He used a piece of string to explain RNA splicing and while the biochemistry details were fuzzy still, RNA was clearly quite interesting stuff! Exciting to me was that the machinery used to do the splicing itself contained RNA!

This was my first real exposure to important biological machines other than the ribosome that were derived from RNA. As I started to read more about RNA I soon discovered the RNA World and later the fascinating papers by Ellington, Benner, White, Orgel and others on how such a world might have looked. Could RNA have been involved in the earliest events in evolution? If so could we make artificial life using only RNA? I wanted more and more to work in this area so being in Cambridge, I spoke with Walter Gilbert, who in turn recommended Alex Rich, who in turn pointed me to Jack Szostak. His laboratory had earlier laid the foundations for the RNA dependent telomerase field, which was to explode with interest in the near future. But more interesting to me was his development of in vitro selection together with the Gold and Joyce laboratories.

Modeled on natural selection, in vitro selection is an extremely elegant tool to study biology. It can efficiently select from extremely large populations (order 10^{15}) individual nucleic acid molecules having specific function. In contrast to biological selection this function is directly related to the boundary condition enforced by the experimentally constrained selective step. This allows an unambiguous comparison between selection and function that is so often hidden in biological systems. Perhaps I could use such a system to help explore the RNA World? Jack patiently explained to me that one of his graduate students, David Bartel, had already demonstrated that by finding an extremely efficient ribozyme
called the Class I ligase that was able to ligate itself to a short RNA primer. By the way, Jack mentioned, David was now at the Whitehead setting up his own laboratory, perhaps you might want to talk with him?

After joining David’s lab (I will always remember him asking, somewhat nervously, if I understood the concept of molarity as I made up my first MgCl₂ solution), I used in vitro selection to study RNAs ability to promote small molecule chemistry. In particular I wanted to see if RNA could at least in principle make parts of itself, so I focused on the selection of ribozymes able make nucleotides by glycosidic bond formation. Later we used the selective tools from this project to adapt David’s class I ligase into an RNA polymerase. All of this was amazing fun, but it was made all the more exciting by the fantastic biological discoveries going on at the same time.

### Rise of the RNA dependent information machines

My fondest memories of that time are the vigorous but extremely friendly conversations I had with Kelly Williams, Tom Tuschl, and Phil Zamore who were all working in Dave’s lab at the time. Kelly had been working on tmRNA a fascinating RNA able to rescue ribosomes stalled on broken messages. This RNA just happens to contain two functionalities: information (a partial ORF read out by the ribosome), together with structured RNA elements that allow it to interact with the ribosome just like a tRNA. This together with in vitro selection based splicing projects Tom had been working on with Phil Sharp, naturally triggered conversations about the potential for other RNA machines to exist.

What does it take to have a nucleic acid system generally involved in regulating cellular state? At the turn of the century key pieces of this puzzle were missing. We decided an ability to uniquely index any location of the genome would be essential, and second some mechanism would be needed to directly connect cellular state to gene expression. Largely because of the in vitro centered enthusiasms of the lab, we had all discussed that for the human genome a unique genomic address could be specified most simply by nucleic acids fragments longer than 17 nt or so. The only pity was that such a system did not seem to exist generally in nature. At the same time Breaker was creating RNA sensors that could via secondary structure rearrangements report the binding of a small molecule analyte. Such simple conceptual ideas worked so beautifully that many at the time could not help but wonder why biology was missing such golden opportunities to program itself at the level of nucleic acids.

Of course the flood gates broke in the next several years with explosive force. The discovery by Ambros in 1993 of the first miRNA was rapidly generalized using new RNA sequencing methodologies that were independent of RNA input sequence. The Bartel lab for example used adenylated DNA adapter oligos originally developed for the in vitro selection of RNA pools that were variable to their very 3' terminus. Hundreds of new small RNAs all with the similar lengths (21 to 24 nt typically) were quickly discovered. The implications were huge, and fed the flames started by Fire and Mello’s discovery of RNAi in worms. Double stranded RNA feeding providing direct evidence of how programmable a cellular state can be using only RNA as a tool. Meanwhile, Breaker’s discovery of riboswitches played a key role in changing our understanding of how RNA sensors can directly connect metabolic state to gene expression. The pace of discovery increased further with the recent connection of Koonin’s DNA repeat sequences in bacteria with the RNA programmable CRISPR machinery by Doudna and others. All of these RNA information machines give biology (and biologists!) exceedingly powerful mechanisms to program and control cellular state.

Such programable machines once understood (i.e., CRISPR and RNAi) can be exploited to more rapidly understand related RNA dependent regulatory processes. Already we know that abundant RNA dependent machines help control chromatin structure and transcriptional state, modulate RNA lifetimes, and regulate translation. New discoveries in all of these layers of RNA dependent regulation are expected in the future and will be greatly facilitated by our existing understanding of RNA processing machines. This together with the rapid growth in RNA imaging techniques and artificial RNA tools developed by in vitro selection are all expected to make the next 20 years of RNA research even more exciting than the last.