Dissecting Regulatory Interactions of RNA and Protein

Combining Computation and High-throughput Experiments in Systems Biology
Marvin Jens

Dissecting Regulatory Interactions of RNA and Protein

Combining Computation and High-throughput Experiments in Systems Biology
Springer Theses

Recognizing Outstanding Ph.D. Research

For further volumes:
http://www.springer.com/series/8790
Aims and Scope

The series “Springer Theses” brings together a selection of the very best Ph.D. theses from around the world and across the physical sciences. Nominated and endorsed by two recognized specialists, each published volume has been selected for its scientific excellence and the high impact of its contents for the pertinent field of research. For greater accessibility to non-specialists, the published versions include an extended introduction, as well as a foreword by the student’s supervisor explaining the special relevance of the work for the field. As a whole, the series will provide a valuable resource both for newcomers to the research fields described, and for other scientists seeking detailed background information on special questions. Finally, it provides an accredited documentation of the valuable contributions made by today’s younger generation of scientists.

Theses are accepted into the series by invited nomination only and must fulfill all of the following criteria

- They must be written in good English.
- The topic should fall within the confines of Chemistry, Physics, Earth Sciences, Engineering and related interdisciplinary fields such as Materials, Nanoscience, Chemical Engineering, Complex Systems and Biophysics.
- The work reported in the thesis must represent a significant scientific advance.
- If the thesis includes previously published material, permission to reproduce this must be gained from the respective copyright holder.
- They must have been examined and passed during the 12 months prior to nomination.
- Each thesis should include a foreword by the supervisor outlining the significance of its content.
- The theses should have a clearly defined structure including an introduction accessible to scientists not expert in that particular field.
Marvin Jens

Dissecting Regulatory Interactions of RNA and Protein

Combining Computation and High-throughput Experiments in Systems Biology

Doctoral Thesis accepted by Humboldt University of Berlin, Germany

Springer
Parts of this thesis have been published in the following articles:

- **Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR**

- **Circular RNAs are a large class of animal RNAs with regulatory potency**

- **doRiNA: a database of RNA interactions in post-transcriptional regulation**

**Author Contributions**

The work presented in this thesis is the result of extensive collaborations. This paragraph strives to clarify the contributions of the author.

All plots and analyses in the introduction (Chap. 1) are the sole work of the author.

The PAR-CLIP pipeline (Sect. 2.3) is mostly the work of the author: the false-positive filter was proposed by Nikolaus Rajewsky, some code was provided by Sebastian Mackowiak for collapsing reads, and by Andranik Ivanov and Zhuo “Minnie” Fang for rendering HTML representations of clusters.

Section 3 presents a joint publication by Svetlana Lebedeva, Marvin Jens, Kathrin Theil, Björn Schwanhäusser, Matthias Selbach, Markus Landthaler and Nikolaus Rajewsky. First authorship is equally shared between SL and MJ. The PAR-CLIP experiments were carried out by SL with support from ML, Mass-Spectrometry was performed by BS, supervised by MS. The remaining experiments are by SL and KT. All computational analyses by MJ. NR designed and supervised the project. NR, SL, and MJ wrote the manuscript.

The RNA competition theory (Sect. 4) is the work of the author, who acknowledges a lot of help from Prof. Johannes Berg to get started. It is quite likely that the formalism—or an equivalent thereof—is known in chemistry, the application to transcriptome regulation and the corresponding analysis—to the best knowledge of the author—is not.
Section 5 presents a joint publication by Sebastian Memczak, Marvin Jens, Antigoni (Anna) Elefsinioti, Francesca Torti, Janna Krüger, Agnieszka Rybak, Luisa Maier, Sebastian D. Mackowiak, Lea H. Gregersen, Mathias Munschauer, Alexander Loewer, Ulrike Ziebold, Markus Landthaler, Christine Kocks, Ferdinand le Noble, and Nikolaus Rajewsky.

First authorship is equally shared between SM, MJ, AE, and FT. SM performed many experiments, assisted by LM. MJ and AE carried out most of the computation, with contributions from NR and SDM. In particular, MJ contributed the final implementation of the circRNA detection from sequencing data, which was jointly conceived by NR, AE, and MJ. MJ also performed the intersection with known transcripts, annotation and conservation analysis of coding exon-derived circRNAs and contributed the pipeline code to study AGO PAR-CLIP data. FT performed the circRNA validation experiments. AR performed all Northern blot experiments. LHG and MM contributed AGO PAR-CLIP experiments, supervised by ML. HEK293 ribominus data provided by the ML lab, originally produced by Kerstin Baethge. CK designed and carried out the single molecule experiments, in part, together with AL. UZ performed the mouse experiments. JK contributed the zebrafish experiments, supervised by FLN. NR designed and supervised the project. NR and MJ wrote the paper.
Supervisor’s Foreword

Marvin joined my lab in spring 2009. He had studied physics, with a particular interest in statistical mechanics that could also be applied to problems in biology. About 4 years later, he finished his PhD thesis, having published several highly visible and in fact already now highly cited papers in close collaboration with several experimental biologists. It is my pleasure to try to summarize his remarkable achievements.

Back in 2009, my lab had become interested in RNA binding proteins (RBPs). We had worked for long years on understanding the function of microRNAs, small non-coding RNAs which bind target messenger RNAs to regulate protein production. However, it became clear that the function of miRNAs could only be understood by also analyzing RBPs because miRNAs and RBPs compete for binding on the same substrates (usually the untranslated regions downstream of the coding sequence of messenger RNAs). We started to analyze biochemical data, which revealed the RNA binding sites of RBPs on a transcriptome wide scale and at nucleotide resolution. These “CLIP” methods (Cross-Linking RNA: proteins followed by ImmunoPrecipitation and next generation sequencing of the attached RNA) had been developed Darnell and Tuschl labs and opened the door to systematically identify the targets of RBPs in vivo. We used and later applied ourselves “PAR-CLIP”, the CLIP method developed by Markus Hafner and Markus Landthaler in the Tuschl lab. The reason for using PAR-CLIP was simple- first of all, Markus Landthaler had become a group leader at my institution the MDC (Max Delbrück Center for Molecular Medicine) and could thus collaborate and advise us how to set up PAR-CLIP in our lab. Second, PAR-CLIP has several nice features which allowed us to assess the quality of the output (very large amounts of sequencing reads). This was particularly important since we had little experience with CLIP.

A particular project in my lab was focused on HuR, a human RBP that was predicted to interact with miRNAs. Svetlana Lebedeva, another PhD student in the lab, had started to PAR-CLIP HuR, and Marvin jumped into analyzing these data. For this, he developed an entire computational pipeline and the statistical controls to use at each step- an important and complicated problem since we understood little about the problems and biases which invariably are part of any kind of high-throughput method. This pipeline is now used by many colleagues in the lab and by all colleagues on campus who are doing CLIP experiments.
Marvin worked then together with Svetlana to interpret the HuR data, and in the end they discovered that HuR had a previously unknown function in mRNA processing. We did not find much about interactions with miRNAs, but this is usually how science goes, at least in our lab- we often do not find what we expected to find. Marvin and Svetlana published the HuR data as co-first authors (Lebedeva and Jens et al. 2011, now a highly cited paper). During these years, Marvin started to think more about competition effects. One reason for this was the HuR project- for example, did a boost in expression of a HuR target “take away” or “sponge” HuR from other targets? Marvin started to apply some thought and methods borrowed from his Statistical Mechanics knowledge to address this question. The resulting quantitative model and predictions are part of his PhD thesis, and we are currently summarizing these thoughts in a review (Jens and Rajewsky, unpublished). Competition effects in the cytoplasma have recently become a heated topic of discussion in the scientific community, and I am personally looking forward that Marvin’s clear discussion of this topic will be published.

Meanwhile another project had excited people in the lab, and Marvin was in the center of the initial, important observation: Antigoni Elefsinioti (a Postdoc in the lab) discovered a noncoding RNA in the human genome that was highly bound by miRNAs (according to CLIP data). Marvin realized that this RNA had a very unusual feature—it contained 74 sites which looked like binding sites for a miRNA, immediately suggesting that this noncoding RNA may function like a miRNA “sponge”. We then noticed that the noncoding RNA was known (work by the Kjems lab in Aarhus) to be covalently closed at the head and the tail ends, i.e. it is expressed in circular, single-stranded form. At that time very little was known about circular RNAs, and we decided to try to systematically detect circular RNAs in the genome. The idea was simple-RNA sequencing reads that would map where the head and tail meet should be indicative of a head-to-tail fusion event. Marvin, in the beginning together with Antigoni, implemented this idea and once again developed statistical controls for this problem. Marvin also computationally analyzed the resulting large numbers of new circular RNAs, and to be brief, the results were published visibly together with Antigoni and two experimental PhD students in my lab (Memczak, Jens, Elefsinioti, Torti et al 2014). This project was an intense collaboration with the le Noble lab of the MDC (who did all of the zebrafish experiments) and collaborations with the Landthaler and Loewer labs (of the Berlin Institute for Medical Systems Biology, MDC).

Marvin was also been involved in more projects, some of them published, some of them still unpublished. However, I hope that my brief summary has helped to give a flavor of the high-level multidisciplinary work that Marvin has carried out, the outstanding intellectual and technical quality of his results, and the excitement that his work has sparked. I think the “Springer Award for Outstanding PhD Theses in Systems Biology” is a wonderful recognition for these achievements.

Berlin, Spring 2014 Prof. Nikolaus Rajewsky
Abstract

In eukaryotic cells, mRNA levels do not directly translate into protein levels due to post-transcriptional regulation of mRNA processing, transport, stability, and translation. These fundamental processes are controlled and carried out by hundreds of RNA-binding proteins (RBPs) and small RNAs, such as microRNAs (miRNA), which predominantly bind to specific elements located in the untranslated regions (UTRs) of target mRNAs.

Thus, combinatorial action of these post-transcriptional regulators on mRNAs is hypothesized to constitute a so-called “post-transcriptional regulatory code,” at least in part encoded by mRNA sequence features, which determines mRNA fate in the cell. A major goal for understanding gene regulation is to decipher this regulatory code. This task requires not only the experimental detection of RBP-mRNA interactions, but also a quantitative understanding of their interplay in the cell.

The computational analysis of experimental (PAR-CLIP) binding data of the human RBP HuR, demonstrates the possibilities of PAR-CLIP to elucidate endogenous RBP-mRNA interactions with high resolution, transcriptome-wide. We mapped thousands of HuR binding sites and discovered previously unknown functions of HuR in the regulation of splicing and the biogenesis of miR-7. Functionality of the binding sites was assessed by recording changes in gene expression after HuR knock down with RNA-seq, on the level of mRNAs, and on the protein level by state-of-the-art mass spectrometry-based proteomics. The low specificity of the HuR binding motif and the widespread, but typically small impact of HuR depletion on mRNA stability, motivated the study of RNA competition effects. A large pool of cellular RNAs introduces competition among binding sites for RNA binding factors with low specificity. This sheds new light on the function of HuR and predicts that stable and abundant RNAs, harboring binding sites for post-transcriptional regulators, can act as sponges and inhibit other regulators by sequestration. While this mode of post-transcriptional regulation through competitive binding has been proposed and validated for artificial miRNA sponges, all known naturally occurring miRNA sponges appear limited by low expression or copy number of binding sites.

In contrast, we have found that CDR1as, a human circular RNA antisense to the CDR1 gene and highly expressed in neural tissues, harbors 63 strong and conserved seed matches for the ancient animal miRNA miR-7. We have