Report from Dagstuhl Seminar 17252

Computational Challenges in RNA-Based Gene Regulation: Protein-RNA Recognition, Regulation and Prediction

Edited by

Rolf Backofen[1], Yael Mandel-Gutfreund[2], Uwe Ohler[3], and Gabriele Varani[4]

1 **Universität Freiburg, DE,** backofen@informatik.uni-freiburg.de
2 **Technion – Haifa, IL,** yaelmg@tx.technion.ac.il
3 **Max-Delbrück-Centrum – Berlin, DE,** uwe.ohler@mdc-berlin.de
4 **University of Washington – Seattle, US,** varani@chem.washington.edu

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 17252 "Computational Challenges in RNA-Based Gene Regulation: Protein-RNA Recognition, Regulation and Prediction".

Seminar June 18–21, 2017 – http://www.dagstuhl.de/17252
1998 ACM Subject Classification I.5 Pattern Recognition, I.5.1 Models, Statistical, I.5.2 Design Methodology, Feature evaluation and selection, I.5.3 Clustering, J.2 Physical Sciences and Engineering, Chemistry, J.3 Life and Medical Sciences, Biology and genetics, Health
Keywords and phrases Machine learning, algorithms, genomics analysis, gene expression networks, big data analysis, quantitative prediction, proteins, RNA
Digital Object Identifier 10.4230/DagRep.7.6.86

1 Executive Summary

Rolf Backofen
Yael Mandel-Gutfreund
Uwe Ohler
Gabriele Varani

All living organism must be able to differentially regulate the expression of genes encoded in their genome. Genes are first transcribed into RNA, which are either translated to proteins or functionally active as non-coding RNAs. Beside the direct regulation of the transcription of DNA into RNA, an important additional layer is the direct regulation of RNAs by RNA binding proteins (RBPs). This layer of regulation controls cellular decisions as part of gene expression networks composed of both proteins and RNAs. While being a dark matter of the cell for a long time, recent years have shown the development of sophisticated high throughput experimental technologies that greatly increased our understanding of protein-RNA recognition and regulation. Nevertheless, the quantitative molecular understanding of the transcriptome-level processes remains very limited. Especially complexity **(both in the form of data as in the required computational approaches)** limits the exploitation of these advances towards a quantitative understanding of post-transcriptional regulation.

Computational Challenges in RNA-Based Gene Regulation: Protein-RNA Recognition, Regulation and Prediction, *Dagstuhl Reports*, Vol. 7, Issue 06, pp. 86–108
Editors: Rolf Backofen, Yael Mandel-Gutfreund, Uwe Ohler, and Gabriele Varani
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

The objective of the seminar to discuss urgently needed computational approaches allowing to exploit the wealth of new data. More specifically, the seminar focused on

- addressing major computational challenges in this field
- mining the extensive genomic information on RNA and associated proteins
- investigation of RNA-protein interactions on an atomic level
- quantitative prediction of cellular regulatory networks and their dynamics.

2 Table of Contents

3 Introduction

3.1 Seminar Format

This seminar arose from the organizer's joint experience that an integration of new experimental tools with new computational approaches is needed to understand RNA-based gene regulation at the level of sophistication and depth required to answer fundamental biological and biomedical questions. It is not sufficient for structural biologists or biochemists to only understand in-depth what other structural biologists are doing: the greatest opportunity for progress lies in the joint analysis of data from different experimental approaches by new computational tools. Experimentalists often communicate with each other in different venues, but it is rare for computational biologists and experimentalists to meet.

In the seminar, we were able to bring together structural and genome biologists, who have developed powerful experimental methods to investigate RNA-protein interaction across genomes, with computational biologists who seek to model and develop predictive tools based on the confluence of these experimental advances. Individual sessions covered different aspects of RNA-based regulation and usually consisted of a mixture of presentations from experimental and computational groups. Discussions where very vivid, fostering the exchange of ideas between the experimental and computational biologists. Discussions also catalyzed the development of new and improved technologies to merge experimental analysis with new computational techniques. The seminar was attended by many leading scientists in the field, further stimulating discussions on how to better interact (possibly on the level of funding opportunities) for a better understanding of the rules of protein-RNA recognition.

3.2 Advances and challenges in studying protein-RNA interactions

The most significant advance in the field of RNA-protein interactions in the last 10 years has been the development of genome-wide approaches to determine the RNA population targeted by an RNA-binding protein (RBP) and to establish its specificity using computational approaches. These approaches, termed CLIP-seq, have now been introduced in several variants, requiring specific computational analysis pipelines. In this session, the participants reported on the one hand about new and large-scale experimental approaches to determine binding sites of RNA-binding proteins and their dynamic behavior. On the other hand, sophisticated computational tools for analyzing CLIP-seq data were shown.

3.3 Identifying new RBPs

Great advances have recently been made in the development of high-throughput screens to identify novel RBPs in cultured cell lines and in tissues. While the technology has contributed dramatically to the field of RNA-protein interactions, increasing the number of identified RBPs and suggesting novel cellular mechanisms for these proteins, these high throughput technologies generate many false positive and false negative results. The participants discussed new approaches for the purification of RNPs, which are complexes of proteins and RNA, as well as specific binding properties. Here, computational approaches allow to investigate multiple functions for these RBPs, showing the flexibility of this layer of regulation.

3.4 Integrative analysis of protein-RNA data

The genome-wide CLIP-methods and their analysis as discussed in session 1 provides the raw data for investigating RNA-centered gene regulation at the level of individual binding sites. However, only an integrative analysis of these data can determine the modes of regulation in more detail. The major topic was how to combine the information of several single CLIP-experiments for exploring RBP-based regulation. The participants discussed how an integrated map of RBP-based regulation based on several CLIP experiments can be constructed, and how advanced machine learning approaches such as matrix factorization can be used for an integrative analysis of multiple CLIP data sets.

3.5 Exploring the world of non-coding RNAs

The majority of the human genome is transcribed into non-coding RNAs, which are not further translated into proteins. Several non-coding RNAs play an important role as assembly platform for proteins. This process is mediated via RNA-protein interactions. For that reason, we decided to invite participants that report about novel findings on non-coding RNA, showing how RNA structure evolves and which effect single mutations (SNP) can have on the structure of non-coding RNAs. In addition, we discussed non-coding RNAs with atypical features.

3.6 Inferring RNA binding specificity

The genome-wide approaches reported in Session 1 (CLIP-seq) determine *in vivo* interactions. These experiments, however, do not determine the biophysical affinity of binding. Here, *in vitro* approaches such as SELEX or RNAcompete are the method of choice, which require their own type of computational analysis. The participants discussed progress both on the experimental side, as well as on the analysis of these types of data. An interesting discussion evolved about how to compare the results of these *in vitro* approaches with the results of *in vivo* CLIP experiments.

3.7 RBPs and small RNAs

In the preceding sessions, we talked in great detail about RNA-protein interactions. However, another mode of regulation is given by direct interactions between RNAs themselves. Very often, these RNA-RNA interactions are mediated by RNA-binding proteins, a well-known example being microRNAs. The interaction of microRNAs with its target RNA is mediated by Argonaute, which is an RNA-binding protein. The participants discussed new large-scale approaches for determining RNA-RNA interactions, and how one can model the combined effect of RNA-RNA and RNA-protein interactions.

3.8 Conclusions

The community around genome-wide determination protein-RNA interaction was very interdisciplinary from the beginning as the whole area started with the availability of CLIP

data, which was challenging both from the experimental as well as from the computational side. However, as there are only few venues where computational and experimental people meet on a specific topic, this workshop was actually the second event where this community could meet. For that reason, it was well appreciated by all participants and had initiated very interesting interdisciplinary collaborations that would not have been formed without this meeting.

4 Overview of Talks

4.1 The solution structure of FUS bound to RNA reveal a bipartite mode of RNA recognition with both sequence and shape specificities.

Frédéric Allain (ETH Zürich, CH)

Joint work of Fionna Loughlin, Peter Lukavsky, Sebastien Campagne, Stefan Reber, Martino Colombo, Eva-Maria Hock , Oliver Muhlemann, Magdalini Polymenidou, Marc-David Ruepp, Frédéric Allain

FUsed in Sarcoma (FUS) is a multifunctional hnRNP which regulates transcription, pre-mRNA splicing, miRNA biogenesis, DNA damage and can drive subcellular phase transitions in RNP granule formation. Two neurodegenerative diseases Amyotrophic Lateral Sclerosis (ALS) and Fronto-Temporal Lobar Degeneration (FTLD) show neuropathological protein aggregates containing FUS and it is hypothesized that mis-regulation of RNA processing could play a major role in these diseases.

FUS consists of a N-terminal prion-like region and an C-terminal RNA binding region including several RNA binding domains (RBD), an RNA Recognition Motif (RRM), and a zinc-finger (ZnF) domain interspersed between RGG repeats. FUS associates with a large variety of RNAs including pre-mRNA, pri-miRNA, lncRNA from CCND1 promoter and TERRA. Results from a dozen in vivo (CLIP) and in vitro (SELEX, binding assays) experiments suggest that RNA binding by FUS is very complex potentially combining structural, sequence and nonspecific interactions with additional influences from post-translational modifications and protein partners.

We present the solution structures of FUS RRM and ZnF domains bound to RNA. The ZnF shows a sequence-specific recognition for a single-stranded NGGU motif and this interaction accounts for the preference for GU-rich motifs found in several CLIP based experiments. The FUS RRM structure was solved bound to a stem-loop RNA and revealed an unusual shape-specific binding mode. The RRM binds the 3' side of the RNA loop using the β-sheet and C-terminal helix with very limited sequence-specificity and the RNA stem using a α_1-β_2 extension unique to FUS, EWS and TAF15 that inserts in the major groove. Furthermore, the RGG repeats between the two RBDs significantly increase the general affinity and unfold the 5'part of the RNA loop. The RRM and ZnF can coordinate in binding a bi-partite RNA site and mutation of the RNA binding surfaces of both the ZnF and RRM domains are sufficient to abolish FUS function in splicing. These results revealed why RNA recognition by FUS has been so difficult to decipher.

4.2 RNA recognition in Syncrip-mediated exosomal loading of miRNAs

Andre Dallmann (Institute of Structural and Molecular Biology, University College London, London, UK and Humboldt University, Berlin, DE)

Joint work of Andres Ramos, Evangelos Christodoulou, Fruzsina Hobor, Ian Taylor, Marco Tripodi, Neil Ball, Roksane Ogrodowicz, Andre Dallmann
Main reference unpublished

Exosomal-mediated miRNA transfer is an important mechanism of inter-cellular communication that regulates gene expression in a broad range of cells and tissues, and has been implicated in cancer, neurological diseases and cardiomyopathies. Syncrip/hnRNPQ is a highly conserved RNA binding protein that regulates mRNA metabolism, transport and translation. It was recently reported that Syncrip plays an essential role in the exosomal partitioning of a group of miRNAs. Here we analyse the molecular basis of the selectivity of his mechanism. We show that Syncrip contains a cryptic RNA binding domain which recognises a short sequence in the miRNA targets. We also discuss how this domain cooperates with the other RNA binding domains of the protein and how the RNA-binding mode of Syncrip allows the protein to recognise selectively a group of miRNA with different sequences. Finally, we discuss other examples where the combined use of bioinformatics, structural biology and functional assays was used to dissect the specificity of multi-domain proteins.

4.3 PTex – a novel method for unbiased purification of crosslinked RNPs

Benedikt Beckman (Humboldt University, Berlin, DE)

Joint work of Erika Urdaneta, Carlos Vieira, Matthias Selbach, Benedikt Beckman

RNA-binding proteins (RBPs) play important roles in post-transcriptional regulation of gene expression, e.g. as key proteins in splicing, RNA-transport or translation. In the past years, discovery of RNA-protein complexes (RNPs) has been driven by high-throughput techniques such as CLIP-Seq or mRNA interactome capture, leading to deeper insights into RNA targets bound by specific proteins or the discovery of hundreds of novel mRNA-binding proteins, respectively. Common to all of these methods is the utilisation of UV light to induce covalent crosslinks between RNA and proteins prior to purification.

Here, we present PTex (Phenol-Toulol EXtraction), a novel technique for the unbiased purification of in vivo UV-crosslinked RNPs. Our method consists of a fast extraction using different organic compounds thereby removing non-crosslinked RNA and proteins. Purification with PTex relies solely on physicochemical differences between crosslinked and "free" RNA/protein and thus enables analysis of non-poly(A) RNA interacting proteins, e.g. for interactors of ribosomal or tRNA. PTex can be performed in 2.5–3 hrs.

We tested PTex in human HEK293 cells and analysed the purified proteins by mass spectrometry. PTex-purified proteins (up to 3000, FDR 0.01) are highly enriched in known RBPs from mRNA interactome capture experiments as well as ribosomal proteins and tRNA-interacting RBPs.

4.4 Towards fully flexible modeling of protein-RNA complexes

Janusz Bujnicki (Int. Inst. of Molecular and Cell Biology – Warsaw, PL) and Michal Boniecki

Joint work of Janusz Bujnicki, Michal Boniecki
URL http://genesilico.pl

Macromolecular complexes composed of proteins and nucleic acids play fundamental roles in many biological processes, such as the regulation of gene expression, RNA splicing and protein synthesis. Our group has developed computational tools for protein and RNA 3D structure prediction, which covered approaches for template-based and template-free modeling. These tools, including the GeneSilico metaserver and Frankenstein Monster approach for proteins, and ModeRNA and SimRNA for RNA have been validated in CASP and RNA Puzzles experiments, respectively. Recently, we combined our approaches for protein and RNA modeling, and developed SimRNP, a method for modeling of protein-RNA complex structures.

4.5 Matrix factorization-based integrative analysis of multiple protein-RNA data sets

Tomaž Curk (University of Ljubljana, SI)

Joint work of Martin Stražar, Marinka Žitnik, Blaž Zupan, Jernej Ule, Tomaž Curk
Main reference Martin Strazar, Marinka Zitnik, Blaz Zupan, Jernej Ule, Tomaz Curk: "Orthogonal matrix factorization enables integrative analysis of multiple RNA binding proteins", in Bioinformatics, Vol. 32(10), pp. 1527–1535, 2016.
URL http://dx.doi.org/10.1093/bioinformatics/btw003

We have presented an integrative orthogonality-regularized nonnegative matrix factorization (iONMF) to integrate multiple data sources and discover non-overlapping, class-specific RNA binding patterns of varying strengths. The orthogonality constraint halves the factor model and outperforms other NMF models in predicting RBP interaction sites on RNA. We have integrated a large data compendium, which includes 31 CLIP experiments on 19 RBPs involved in splicing (such as hnRNPs, U2AF2, ELAVL1, TDP-43 and FUS) and processing of 3'UTR (Ago, IGF2BP). The integration of multiple data sources improves the predictive accuracy of retrieval of RNA binding sites. The key predictive factors of protein–RNA interactions were the position of RNA structure and sequence motifs, RBP co-binding and gene region type. We report on a number of protein-specific patterns, many of which are consistent with experimentally determined properties of RBPs.

4.6 Identification of transcriptomic regulatory elements from CLIP-Seq data

Philipp Drewe (Max-Delbrück-Centrum – Berlin, DE)

Joint work of Philipp Drewe-Boss, Hans-Hermann Wessels, Uwe Ohler
Main reference Philipp Drewe-Boss, Hans-Hermann Wessels, Uwe Ohler: "omniCLIP: Bayesian identification of protein-RNA interactions from CLIP-Seq data", in bioRxiv, Cold Spring Harbor Laboratory, 2017.
URL http://dx.doi.org/10.1101/161877

High-throughput immunoprecipitation methods to analyze RNA binding protein-RNA interactions and modifications have great potential to further the understanding of post-transcriptional gene regulation. Due to the differences between individual approaches, each of a diverse number of computational methods can typically be applied to only one specific sequencing protocol. Here, we present a Bayesian model called omniCLIP that can be applied to data from all protocols to detect regulatory elements in RNAs. omniCLIP greatly simplifies the data analysis, increases the reliability of results and paves the way for integrative studies based on data from different sources.

4.7 RNA binding properties of a metabolic enzyme

André Gerber (University of Surrey – Guildford, GB)

Joint work of Ana M. Matia-González, Valentina Iadevaia, Emma E. Laing, André P. Gerber

RNA binding proteins (RBPs) are essential for post-transcriptional regulation of gene expression with widespread implications in development and disease. We recently described the repertoire of hundreds of RBPs that interact with polyadenylated mRNAs in the unicellular yeast Saccharomyces cerevisiae and in the multicellular nematode Caenorhabditis elegans, respectively [1]. Thereby, we found that most proteins comprising the mRNA-binding proteomes (mRBPomes) were evolutionarily conserved, including components of early metabolic pathways such as glycolysis, which indicated an ancient origin of the mRBPomes. To further investigate the RNA-binding properties of metabolic enzymes, we are currently focusing on phosphofructokinase (Pfk), which is a key enzyme of glycolysis. In the yeast S. cerevisiae, the enzyme is a hetero-octamer comprised of Pfk1 and Pfk2 subunits. We found that both paralogous subunits can independently interact with mRNAs. Furthermore, Pfk2p specifically interacts with PFK2 mRNA; and one point of interaction could be located to a conserved sequence element in the 3'UTR of PFK2 mRNA. Furthermore, we showed that enzyme-RNA interactions depend on the metabolic state of the cells, indicating cross-talk with cellular metabolism. Furthermore, we found that interaction of PFK with its own mRNA is conserved between yeast and human. We speculate that conserved enzyme-mRNA interactions could relate to an ancient mechanism for the post-transcriptional coordination of metabolic pathways, integrating metabolic activity into dynamic gene regulatory responses.

References

1 Matia-González, A. M., Laing, E. E, Gerber, A. P. (2015) Conserved mRNA-binding proteomes in eukaryotic organisms. Nat. Struct. Mol. Biol. 22(12), 1027-33.

4.8 Large-scale elucidation of RNase III targets and cleavage patterns

Hanah Margalit (The Hebrew University of Jerusalem, IL)

Joint work of Hanah Margalit, Yael Altuvia, Liron Argaman, Amir Bar, Ehud Karavani

The family of Ribonuclease III (RNase III) enzymes spans all kingdoms of life, with its most famous representatives the eukaryotic enzymes Drosha and Dicer. A founding member of the family is RNase III of Escherichia coli, which was discovered about 50 years ago and extensively studied. RNase III cleaves double-stranded RNA, typically generating two cleavages in the two sides of a stem structure while leaving 2-nt 3' overhangs. At present, 23 target genes of RNase III are known in E. coli, all of them were discovered in specific experiments focusing on particular RNAs. We reasoned that the availability of E. coli genome and the opportunities that the large-scale RNA-seq technique offers open the way for transcriptome-wide mapping of RNase III targets in E. coli. To this end we developed Cleave-seq, a method for global mapping of the cleavage sites of any endoribonuclease of interest at a nucleotide resolution, taking advantage of the chemical group at the 5' end of the cleavage product. Exploiting the characteristic 5'-phosphomonoester termini of RNase III cleavage products, we constructed sequencing libraries deigned in a way that the start position of a RNA sequenced segment should be a 5'P position, and hence, generated by an RNase that leaves 5'P ends. To identify 5'P ends generated by RNAse III we applied the protocol to RNA extracted from E. coli MG1655 WT strain and a mutant strain lacking active RNase III, and mapped the sequenced fragments to the genome. Positions where the start of sequenced fragments mapped in the WT but not in the mutant ($p \leq 0.05$ by DESeq2) were identified as RNase III cleavage sites. We re-discovered 20 out of the 23 known targets, and extended the set of cleavage sites in E. coli to $\tilde{4}00$ putative sites, residing in RNAs derived from a variety of genomic entities (CDSs, 5' and 3' UTRs, intergenic regions and antisense to annotated transcripts). 30 % of the targets had a pair of cleavage sites that followed the known cleavage rules, while the rest presented only one cleavage site. We are now investigating these single sites in search of new, additional, cleavage rules that might dictate RNase III cleavage of these targets. Especially intriguing is the possibility that base-pairing of regulatory small RNAs with these target RNAs generates substrates for RNAse III cleavage.

4.9 Kinetics of RNA-protein interactions in cells

Eckhard Jankowsky (Case Western Reserve University – Cleveland, US)

RNA-binding proteins (RBPs) often interact with many different RNAs at sometimes large numbers of binding sites. It is widely accepted that the kinetics by with a given RBP interacts with each of its binding sites is critical for the impact of the RBP on RNA metabolism. While kinetics of RBP binding to RNA can be readily measured in vitro, it has not yet been possible to determine kinetic parameters of RBP interactions with individual RNA sites in cells. Here, we report an approach to measure kinetic parameters by which the RNA binding protein Dazl from Mouse interacts with thousands of RNA binding sites in cells. We describe

how a combination of time-resolved UV-laser crosslinking, next generation sequencing and kinetic modeling can be utilized to determine on- and off rate constants, as well as fractional occupancy of Dazl to thousands of RNA binding sites in cells. Our data enable quantitative, biophysical descriptions of RNA-protein interactions in cells.

4.10 Modeling the combined effect of RBPs and miRNAs in post-transcriptional regulation

Hilal Kazan (Antalya International University, TR)

Joint work of Saber HafezQorani, Atefeh Lafzi, Ruben G de Bruin, Anton Jan van Zonneveld, Eric P van der Veer, Yesim Aydin Son, Hilal Kazan
Main reference Saber HafezQorani, Atefeh Lafzi, Ruben G. de Bruin, Anton Jan van Zonneveld, Eric P. van der Veer, Yeşim Aydın Son, Hilal Kazan: "Modeling the combined effect of RNA-binding proteins and microRNAs in post-transcriptional regulation", in Nucleic Acids Research, Vol. 44(9), p. e83, 2016.
URL http://dx.doi.org/10.1093/nar/gkw048

Recent studies support that RNA-binding proteins (RBPs) and microRNAs (miRNAs) function in coordination with each other to control post-transcriptional regulation (PTR). However, the majority of research to date has focused on the regulatory effect of individual RBPs or miRNAs. Here, we mapped both RBP and miRNA binding sites on human 3'UTRs and utilized this collection to better understand PTR. We show that the transcripts that lack competition for HuR binding are destabilized more after HuR depletion. We also validate this finding for PUM1(2) by knocking down PUM1 and PUM2 in HEK293 cells and measuring genome-wide gene expression changes. Next, to find potential cooperative interactions, we identified the pairs of factors whose sites co-localize more often than expected by random chance. We show that transcripts where the sites of PUM1(2) and its interacting miRNA form a stem-loop are stabilized more upon PUM1(2) depletion. Finally, using dinucleotide frequency and counts of regulatory sites as features in a regression model, we achieved an AU- ROC of 0.86 in predicting mRNA half-life in BEAS-2B cells. Our results suggest that combining the effects of RBPs and miRNAs lead to more accurate models of PTR.

4.11 RNA genotype-phenotype mapping

Grzegorz Kudla (University of Edinburgh, GB)

Recent advances in synthetic biology and next-generation sequencing allow experimental investigation of long-standing, fundamental questions in molecular biology and evolution, such as: Which mutations influence function? How do mutations influence function? How do the effects of mutations depend on location within molecular structures? How do the effects depend on environmental conditions and genetic background? We are studying these questions using yeast U3 snoRNA as a model system. U3 is an abundant, evolutionarily conserved noncoding RNA, which plays an essential role in ribosome biogenesis. By measuring the effects of 60,000 mutated variants of U3 on yeast growth, we found that the effects of individual mutations were correlated with evolutionary conservation and structural stability. Many

mutations had no measurable effect in an otherwise wild-type background, but were deleterious in combination with additional mutations in U3. We also found pairs of compensatory mutations, and used these to predict the secondary structure of the RNA. The effects of mutations depended on environmental conditions, and the effects of destabilizing mutations were increased at high temperature. In parallel with mutational studies, we use a high-throughput method for mapping RNA-RNA interactions, called CLASH, to investigate the folding of U3 and preribosomal RNA. By integrating these data with recent cryo-electron microscopy structural studies of preribosomes, we aim to systematically elucidate the relations between genotype, structure, molecular function, and evolutionary fitness.

4.12 Exploring global changes in protein-mRNA interactions

Markus Landthaler (Max-Delbrück-Centrum – Berlin, DE)

Joint work of Miha Milek, Koshi Imami, Neelanjan Mukherjee, Francesca De Bortoli, Ulrike Zinnall, Orsalia Hazapis, Christian Trahan, Marlene Oeffinger, Florian Heyd, Uwe Ohler, Matthias Selbach, Markus Landthaler
Main reference Miha Milek, Koshi Imami, Neelanjan Mukherjee, Francesca De Bortoli, Ulrike Zinnall, Orsalia Hazapis, Christian Trahan, Marlene Oeffinger, Florian Heyd, Uwe Ohler, Matthias Selbach, Markus Landthaler: "DDX54 regulates transcriptome dynamics during DNA damage response", in Genome Research, 10.1101/gr.218438.116, 2017
URL http://dx.doi.org/10.1101/gr.218438.116

The cellular response to genotoxic stress is mediated by a characterized network of DNA surveillance pathways. The contribution of posttranscriptional gene regulatory networks to the DNA damage response has been studied poorly. In an attempt to examine global changes in protein-mRNA interactions, we systematically identified RNA-binding proteins differentially bound to polyadenylated transcripts upon exposure of MCF-7 human breast carcinoma cells to ionizing radiation (IR). Surprisingly, about 260 proteins including many nucleolar proteins showed increased binding to mRNA in IR-treated cells. The analysis of DDX54, an RNA helicase, revealed that this protein is an immediate-to-early DDR regulator required for the splicing efficacy of its target IR-induced pre-mRNAs. Upon IR exposure, DDX54 acts by increased interaction with a well-defined class of pre-mRNAs which harbor introns with weak 3' acceptor splice sites, as well as by protein-protein contacts within components of U2 snRNP and spliceosomal B complex, resulting in reduced intron retention and higher processing rates of its target pre-mRNA transcripts. Since DDX54 promotes survival after exposure to IR, its expression may impact DDR-related pathologies. This work indicates the relevance of many uncharacterized RNA-binding proteins involved in the cellular response to DNA damage.

4.13 iCLIP determines the binding landscape of the clock-regulated RNA-binding protein AtGRP7

Martin Lewinski (Universität Bielefeld, DE)

Joint work of Katja Meyer, Tino Köster, Christine Nolte, Claus Weinholdt, Martin Lewinski, Ivo Grosse, Dorothee Staiger

Background: A key function for RNA-binding proteins in orchestrating plant development and environmental responses is well established. However, the lack of a genome-wide view on their in vivo binding targets and binding landscapes represents a gap in understanding the mode of action of plant RNA-binding proteins. Here, we adapt individual nucleotide resolution crosslinking immunoprecipitation (iCLIP) for genome-wide determining the binding repertoire of the circadian clock-regulated Arabidopsis thaliana glycine-rich RNA-binding protein AtGRP7. Results: iCLIP identified 858 transcripts with significantly enriched crosslink sites in plants expressing AtGRP7-GFP and absent in plants expressing an RNA-binding-dead AtGRP7 variant or GFP alone. To independently validate the targets, we performed RNA immunoprecipitation (RIP)-sequencing of AtGRP7-GFP plants subjected to formaldehyde fixation. 452 of the iCLIP targets were also identified by RIP-seq, thus representing a set of high-confidence binders. AtGRP7 can bind to all transcript regions with a preference for 3´untranslated regions. In the vicinity of crosslink sites, UC-rich motifs were overrepresented. Cross-referencing the targets against transcriptome changes in AtGRP7 loss-of-function mutants or AtGRP7-overexpressing plants revealed a predominantly negative effect of AtGRP7 on its targets. In particular, elevated AtGRP7 levels lead to damping of circadian oscillations of transcripts including DORMANCY/AUXIN ASSOCIATED FAMILY PROTEIN2 and CCR-LIKE. Furthermore, several targets show changes in alternative splicing or polyadenylation in response to altered AtGRP7 levels. Conclusion: We have established iCLIP for plants to identify target transcripts of the RNA-binding protein AtGRP7. This paves the way to investigate the dynamics of posttranscriptional networks in response to exogenous and endogenous cues.

4.14 Challenges in exploring the world of RNA binding proteins

Yael Mandel-Gutfreund (Technion – Haifa, IL)

Joint work of Shlomi Dvir, Chen Lesnik, Inbal Paz, Michal Amit, Yael Mandel-Gutfreund

DNA binding proteins (DBPs) and RNA binding proteins (RBPs) are the main regulators of gene expression in the cell. In the context of tight interactions between transcriptional and post-transcriptional regulation, proteins that bind both DNA and RNA, namely DNA and RNA Binding Proteins (DRBPs), are highly likely to be key players in mediating the cross talk between the different processes of gene expression pathway. Human Embryonic Stem Cells (hESCs) provide a biologically valuable and experimentally tractable model system to study the unappreciated group of dual nucleic acids binding proteins. As a first step to study the role of DRBPs in pluripotency we employed an RNA-interactome experiment on hESCs, identifying over 800 high-confidence RBPs, among them we identified 25% DRBPs. Among the proteins detected as RNA binding in our experiment more 120 proteins

were not previously shown to interact with either RNA or DNA in mammalian cells. We further differentiated the hESCs to embryoid bodies (EBs) using a spontaneous differentiation protocol and followed the RNA and protein expression levels the process. We found that a large fraction of the detected RBPs and DRBPs in the hESC interactome were significantly down regulated upon differentiation, suggesting that dual DNA and RNA binding function may play an important role in shaping the stem cell state in human cells.

4.15 Accurate identification of RBP binding sites and RNA sequence-structure motifs from CLIP-seq data

Annalisa Marsico (MPI für Molekulare Genetik – Berlin, DE)

Joint work of Sabrina Krakau, Hugues Richard, Annalisa Marsico

Interactions between RNAs and RNA binding proteins (RBPs) play essential roles in both transcriptional and post-transcriptional gene regulation. iCLIP and eCLIP techniques allow the detection of protein-RNA interaction sites at high resolution, based on diagnostic events at crosslink sites. Previous methods do not explicitly model the specfics of iCLIP and eCLIP truncation patterns and possible biases. We introduce PureCLIP, a hidden Markov model based approach, which simultaneously performs peak calling and individual crosslink site detection. It explicitly incorporates RNA abundances and, for the first time, non-specfic sequence biases. On both simulated and real data, PureCLIP is more accurate in calling crosslink sites than other state-of-the-art methods and has a higher agreement across replicates.

4.16 CLIPing STAR proteins: target specificity via compartmentalisation

Daniel Maticzka (Universität Freiburg, DE)

Joint work of Daniel Maticzka, Ibrahim Avsar Ilik, Tugce Aktas, Rolf Backofen, Asifa Akhtar
Main reference Daniel Maticzka, Ibrahim Avsar Ilik, Tugce Aktas, Rolf Backofen, Asifa Akhtar: "uvCLAP: a fast, non-radioactive method to identify in vivo targets of RNA-binding proteins", in bioRxiv, Cold Spring Harbor Laboratory, 2017.
URL http://dx.doi.org/10.1101/158410

We have developed uvCLAP (UV crosslinking and affinity purification), a method for measuring RNA-protein interactions in vivo. To test its performance and applicability we investigated binding of 15 RBPs from fly, mouse and human cells. Our results show that despite their different subcellular localizations, STAR proteins (KHDRBS1-3, QKI) bind to a similar RNA motif in vivo. Consistently a point mutation (KHDRBS1Y440F) or a natural splice isoform (QKI-6) that changes the respective RBP subcellular localization, dramatically alters target selection without changing the targeted RNA motif. Combined with the knowledge that RBPs can compete and cooperate for binding sites, our data shows that compartmentalization of RBPs can be used as an elegant means to generate RNA target specificity.

4.17 Alternative RNA structure expression and its functional roles

Irmtraud Meyer (Max-Delbrück-Centrum – Berlin, DE)

Joint work of Alborz Mazloomian, Irmtraud M. Meyer
Main reference Alborz Mazloomian, Irmtraud M. Meyer: "Genome-wide identification and characterization of tissue-specific RNA editing events in D. melanogaster and their potential role in regulating alternative splicing", in RNA Biology, Vol. 12(12), pp. 1391–1401, Taylor & Francis, 2015.
URL http://dx.doi.org/10.1080/15476286.2015.1107703

Many studies of protein-RNA interactions aim to identify new protein-RNA interaction sites, e.g. by utilising experimentally derived CLIP-data as input for a computational prediction pipeline. One important additional aim of these computational studies is often to also deduce the yet unknown characteristic binding features for new proteins. This often requires the computational modelling of the potential RNA secondary-structure context of the protein-RNA binding sites.

We here present results of a transcriptome-wide, computational study of A-to-I editing sites for several, tissue-specific RNA-seq data sets in the fruit fly *Drosophila melanogaster*. By designing a dedicated computational prediction pipeline which employs probabilistic algorithms and which utilises only transcriptome and genome sequencing data as input (rather than experimental data on protein-RNA interaction sites), we manage to reliably identify a considerable number of new protein-RNA binding sites for this particular protein. This is only possible because our prediction pipeline explicitly captures several key features of the corresponding RNA-binding protein ADAR which is responsive for these A-to-I edits: the requirement of double-stranded RNA for potential ADAR-binding sites and the typical accumulation of several editing sites in the same stretch of double-stranded RNA. Using this dedicated probabilistic prediction strategy for one particular RNA-binding protein (ADAR) in combination with a comparative approach allows us to gain considerable biological insights into the protein's potential functional role in alternative splicing and the underlying molecular mechanisms *in vivo* without having to resort to the typical assumptions made in the field.

Our main insight gained is that ADAR-binding to double-stranded RNA can result in several A-to-I edits which can induce RNA-structure changes which can in turn induce change the splicing pattern of the underlying pre-mRNA. This turns out to be of particular importance for cells of the central nervous system, not only in the fruit fly, but also the mouse and human, and proposes a dedicated RNA-structure-based molecular mechanism for tissue-specific changes of splicing patterns.

To summarize, our study shows that

- protein-specific computational studies have the potential to gain significant biological insight,
- protein-RNA binding may require a very specific RNA secondary-structure context,
- protein-RNA binding may change the RNA secondary-structure context upon binding/acting and
- modelling the RNA secondary-structure context at it is *in vivo* can be achieved by employing a comparative approach.

References

1 I. M. Meyer, In silico methods for co-transcriptional RNA secondary structure prediction and for investigating alternative RNA structure expression, Methods 120 (2017) 3–16.
2 D. Lai, I. M. Meyer, e-RNA: a collection of web servers for comparative RNA structure prediction and visualisation, Nucleic Acids Research 42 (2014) W373–6.

4.18 Inferring RNA motifs from millions of binding sites using billions of features

Quaid Morris (University of Toronto, CA) and Tim Hughes (University of Toronto, CA)

Joint work of Kate B. Cook, Shankar Vembu, Kevin Ch. Ha, Hong Zheng, Kaitlin U. Laverty, Deb Ray, Quaid Morris, Tim Hughes
Main reference Kate B. Cook, Shankar Vembu, Kevin C.H. Ha, Hong Zheng, Kaitlin U. Laverty, Timothy R. Hughes, Debashish Ray, Quaid D. Morris: "RNAcompete-S: Combined RNA sequence/structure preferences for RNA binding proteins derived from a single-step in vitro selection", in Methods, Vol. 126(Supplement C), pp. 18–28, 2017.
URL http://dx.doi.org/10.1016/j.ymeth.2017.06.024

RNA-binding proteins (RBPs) recognize RNA sequences and structures, but there are few systematic and accurate method to derive large (>12base) motifs de novo that capture an RBPs sequence and structural binding preferences. My talk will introduce RNAcompete-S, a new method that couples a single-step competitive binding reaction with an excess of random RNA 40-mers to a new computational pipeline to derive SSMs (Sequence and Structure Models) from the millions of RNA oligos in selected and background fractions of oligos.

We use Vowpal Wabbit to identify sequence and structure k-mers that are predictive of an RNA oligo being in the selected fraction and then use Gaussian mixture model clustering to group k-mers into motifs. We then combine frequently co-occurring motifs into a single, larger motifs.

RNAcompete-S confirms that HuR, QKI, and SRSF1 prefer binding sites that are single stranded, and recapitulates known 8-10bp sequence and structure preferences for Vts1p and RBMY. We also derive an 18-base long SSM for Drosophila SLBP, which has not been previously determined by selections from pure random sequence, and accurately discriminates human replication-dependent histone mRNAs. Thus, RNAcompete-S enables accurate identification of large, intrinsic sequence-structure specificities with a uniform assay.

4.19 An integrated map of RNA-binding protein mediated gene regulation

Uwe Ohler (Max-Delbrück-Centrum – Berlin, DE), Mahsa Ghanbari (Max-Delbrück-Centrum – Berlin, DE)

Joint work of Uwe Ohler, Mahsa Ghanbari, Svetlana Lebedeva, Neelanjan Mukherjee, Alina Munteanu, Hans-Hermann Wessels, Aitor Garcia, Thomas Tuschl

RNA-binding proteins (RBPs) control and coordinate each stage in the life cycle of RNAs. Recent sequencing protocols, particularly variants of in vivo RBP-RNA crosslinking and immunoprecipitation (CLIP), have in principle enabled the identification of target sites for many of these RBPs. Yet, there has been little effort to systematically reanalyze previously published data generated by different labs, both to allow for meaningful benchmarking of new algorithms e.g. for site and motif identification and for integrative analyses of these data.

Here, we examined 118 transcriptome-wide data sets representing 67 RBPs that were generated by the PAR-CLIP protocol, to characterize target RNA class binding preference, sequence and regional specificity. Furthermore, by integrating RNA metabolism measurements, we identified regulatory modules defined by subsets of mRNAs bound by specific subsets of RBPs, which may represent post-transcriptional RNA operons. This study represents

the most comprehensive investigation of empirical RBP-RNA interaction evidence and their regulatory function in a human cell line to date, and the data has potential to be a useful gold standard for future algorithmic developments.

4.20 AptaTRACE – a novel method to discover RNA sequence-structure binding motifs

Teresa Przytycka (National Center for Biotechnology – Bethesda, US)

Joint work of Phuong Dao, Jan Hoinka, Mayumi Takahashi, Jiehua Zhou, Michelle Ho, Yijie Wang, Fabrizio Costa, John J. Rossi, Rolf Backofen, John Burnett, Teresa M. Przytycka
Main reference Phuong Dao, Jan Hoinka, Mayumi Takahashi, Jiehua Zhou, Michelle Ho, Yijie Wang, Fabrizio Costa, John J. Rossi, Rolf Backofen, John Burnett, Teresa M. Przytycka: "AptaTRACE Elucidates RNA Sequence-Structure Motifs from Selection Trends in HT-SELEX Experiments", in Cell Systems, Vol. 3(1), pp. 62–70, 2016.
URL http://dx.doi.org/10.1016/j.cels.2016.07.003

RNA aptamers are short RNA molecules capable of binding, with high affinity and specificity, a specific target molecule via sequence and structure features that are complementary to the biochemical characteristics of the target's surface. While the specifics vary depending on the target, aptamers are typically identified through the Systematic Evolution of Ligands by Exponential Enrichment (SELEX) protocol. A key reason for the resurgence of interest in aptamer research relates to the emergence the HT-SELEX – a procedure which utilizes affordable next-generation sequencing technologies along with traditional SELEX. HT-SELEX calls for new scalable analytic tools for identifying sequence-structure biding motifs from HT-SELEX data. In this talk I will present AptaTRACE, a computational approach that leverages the experimental design of the HT-SELEX protocol, RNA secondary structure, and the potential presence of many secondary motifs to identify sequence-structure motifs that selected for in the HT-SELEX experiment.

AptaTRACE is not limited to the detection of a single motif but capable of elucidating an arbitrary number of binding sites along with their corresponding structural preferences. Unlike previous methods, AptaTRACE does not rely on aptamer frequency or its derivative – cycle-to-cycle enrichment. Instead, our method builds on tracing the dynamics of the SELEX process itself to uncover motif-induced selection trends. Specifically, we expect that in the initial pool the structural contexts of each k-mer are distributed according to a background distribution that can be determined from the data. However, for sequence motifs involved in binding, in later selection cycles, this distribution becomes biased towards the structural context favored by the binding interaction with the target site. Consequently, AptaTRACE aims at identifying sequence motifs whose tendency of residing in a hairpin, bulge loop, inner loop, multiple loop, dangling end, or of being paired converges to a specific structural context throughout the selection.

We applied AptaTRACE to identify nine motifs in C-C chemokine receptor type 7 targeted by aptamers in an in vitro cell-SELEX experiment. We experimentally validated two aptamers whose binding required both sequence and structural features. AptaTRACE can identify low-abundance motifs, and we show through simulations that because of this it could lower HT-SELEX cost and time by reducing the number of selection cycles required. AptaTRACE is available for download at http://www.ncbi.nlm.nih.gov/CBBresearch/Przytycka/index.cgi#aptatools.

In the second part of the talk I will describe our recently developed method, AptaBlocks, to design Aptamer – drug delivery systems.

4.21 Determination of RNA-protein interactions

Rolf Backofen (Universität Freiburg, DE)

Main reference Tuğçe Aktaş, İbrahim Avşar Ilık, Daniel Maticzka, Vivek Bhardwaj, Cecilia Pessoa Rodrigues, Gerhard Mittler, Thomas Manke, Rolf Backofen, Asifa Akhtar: "DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome", Nature, Vol. 544, pp. 115–119, Springer Nature, 2017.
URL http://dx.doi.org/10.1038/nature21715
Main reference Daniel Maticzka, Sita J. Lange, Fabrizio Costa, Rolf Backofen: "GraphProt: modeling binding preferences of RNA-binding proteins", Genome Biology, Vol. 15(1), p. R17, 2014.
URL http://dx.doi.org/10.1186/gb-2014-15-1-r17

It is becoming increasingly clear that a RNA-binding proteins are key elements in regulating the cell's transcriptome. Thus, unraveling the interaction network of the RNA-binding proteins by determining their binding sites is becoming an increasingly important topic. CLIP-seq is one of the major tools to determine binding sites but suffers from high false negative rate due its expression dependency. This critical hinders the use of public CLIP-data. We will show in several examples how use of raw public CLIPp data can lead to false biological reasoning and how advanced advanced machine learning approach can overcome this problem. I will further discuss our results from our new Nature paper, showing that the human RNA helicase DHX9 predominantly binds to IRAlu elements and such suppresses the negative effect of Alu inflation in transcripts.

4.22 Modelling the RNA lifecycle with χ-CRAC data

Guido Sanguinetti (University of Edinburgh, GB)

Gene expression responses to cellular stress involve a wide repertoire of RNA dynamic profiles. The regulation of RNA abundance is a balance of production, processing and decay, yet how these processes combine to shape RNA profiles is only partly understood. In this talk, I describe how χ-CRAC, a new technology to assay dynamic protein-RNA interactions, can be used to start deconvolving the various components of the RNA life cycle. Evidence from our recent work [1] suggests that degradation factors such as the yeast protein Nab3 respond to stress by dynamically altering their binding. We are in the process of collecting data about other processing pathways; in the talk, I describe how we are planning to integrate such data sets within a modelling framework.

References

1 van Nues, Rob and Schweikert, Gabriele and de Leau, Erica and Selega, Alina and Langford, Andrew and Franklin, Ryan and Iosub, Ira and Wadsworth, Peter and Sanguinetti, Guido and Granneman, Sander. *Kinetic CRAC uncovers a role for Nab3 in determining gene expression profiles during stress.* Nature Communications 8, 2017

4.23 Decoding regulatory protein-RNA interactions in gene regulation using integrated structural biology

Michael Sattler (TU München, DE)

Main reference Andreas Schlundt, Jan-Niklas Tants, Michael Sattler: "Integrated structural biology to unravel molecular mechanisms of protein-RNA recognition", Methods, Vol. 118-119(Supplement C), pp. 119–136, 2017.
URL http://dx.doi.org/10.1016/j.ymeth.2017.03.015

RNA plays essential roles in virtually all aspects of gene regulation. This involves the recognition of cis regulatory RNA sequences by RNA binding proteins (RBPs). Most eukaryotic RBPs are multi-domain proteins that comprise multiple structural domains to mediate protein-RNA or protein-protein interactions. Large scale biochemical approaches to map RBP-RNA interactions in vivo at a genome-wide scale combined with computational analysis provide important information about RNA sites recognized. Also, large scale approaches such RNACompete identify RNA sequence motifs that are recognized by a given RBP in vitro. However, the roles of individual RNA binding domains and other structural modules in these multidomain RNA binding proteins are not revealed by these approaches. We employ integrative structural biology to unravel the molecular mechanisms involving these regulatory RNP (ribonucleoprotein) complexes. For these studies, solution NMR-spectroscopy, SAXS/SANS and FRET experiments provide unique information on functionally important dynamics and are combined with X-ray crystallography and electron microscopy to elucidate the structural mechanisms and dynamics of regulatory RNPs. In collaboration with the Koenig group (IMB Mainz) we show how the combination and comparative bioinformatics analysis of in vitro and in vivo CLIP experiments identifies novel regulators of RBP-RNA interactions and is a useful approach to link large scale approaches in vivo with mechanistic analysis of RNA binding by RBPs using structural biology. For the heterodimeric constitutive splicing factor U2AF we found that the tandem RNA recognition motif (RRM) domains of U2AF65 adopt an equilibrium of open and closed domain arrangements in solution. RNA binding shifts this equilibrium towards the open, active domain arrangement depending on the overall RNA binding affinity. The population shift towards the open conformation induced by RNA binding quantitatively correlates with the efficiency of spliceosome assembly on corresponding pre-mRNAs (Mackereth et al Nature, 2011). Notably, the addition of the small U2AF35 subunit preshifts this equilibrium and thereby enhances RNA binding of U2AF65 also to weaker RNA ligands (von Voithenberg et al, PNAS 2016). We also found that a linker preceding one of the RRM domains in U2AF65 has an autoinhibitory role to proofread RNA binding. For a set of paralogous multidomain alternative splicing factor RBPs we show that distinct contributions and dynamic domain arrangements mediate RNA binding specificity and enagement of different cellular pre-RNAs. These results indicate that distinct molecular mechanisms of dynamic RNA recognition by multidomain RNA binding proteins represent important features for the regulation of gene expression.

References

1 Voith von Voithenberg L, Sanchez-Rico C, Kang HS, Madl T, Zanier K, Barth A, Warner LR, Sattler M, and Lamb DC Recognition of the 3' splice site RNA by the U2AF heterodimer involves a dynamic population shift. (2016) Proc Natl Acad Sci U S A 113, E7169-E7175. doi: 10.1073/pnas.1605873113

2 Huang JR, Warner LR, Sanchez C, Gabel F, Madl T, Mackereth CD, Sattler M, Blackledge M. Transient Electrostatic Interactions Dominate the Conformational Equilibrium Sampled

by Multidomain Splicing Factor U2AF65: A Combined NMR and SAXS Study. (2014) J Am Chem Soc. 136:7068-76. doi:10.1021/ja502030n

3 Mackereth CD, Madl T, Bonnal S, Simon B, Zanier K, Gasch A, Rybin V, Valcarcel J, and Sattler M. Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. (2011) Nature 475, 408-11. doi:10.1038/nature10171

4.24 ANRIL and STAIR18 – two long non-coding RNAs with atypical features

Peter F. Stadler (Universität Leipzig, DE)

Joint work of the groups of Daniel Teupser, Friedemann Horn, Steve Hoffmann, Peter F. Stadler
Main reference Lesca M. Holdt, Steve Hoffmann, Kristina Sass, David Langenberger, Markus Scholz, Knut Krohn, Knut Finstermeier, Anika Stahringer, Wolfgang Wilfert, Frank Beutner, Stephan Gielen, Gerhard Schuler, Gabor Gäbel, Hendrik Bergert, Ingo Bechmann, Peter F. Stadler, Joachim Thiery, Daniel Teupser: "Alu Elements in ANRIL Non-Coding RNA at Chromosome 9p21 Modulate Atherogenic Cell Functions through Trans-Regulation of Gene Networks", PLOS Genetics, Vol. 9(7), pp. 1–12, Public Library of Science, 2013.
URL http://dx.doi.org/10.1371/journal.pgen.1003588
Main reference Stefanie Binder, Nadine Hösler, Diana Riedel, Ivonne Zipfel, Tilo Buschmann, Christoph Kämpf, Kristin Reiche, Renate Burger, Martin Gramatzki, Jörg Hackermüller, Peter F. Stadler, Friedemann Horn: "STAT3-induced long noncoding RNAs in multiple myeloma cells display different properties in cancer", Scientific Reports, Vol. 7(1), p. 7976, Springer Nature, 2017.
URL http://dx.doi.org/10.1038/s41598-017-08348-5

ANRIL has been discovered by GWAS studies and is known to function via polycomb-based gene silencing. The "gene" comprises a plethora of isoforms with discernible effects on a large number of targets. Interestingly, a Alu-DEIN element is associated in the targetting. An interesting facet of the story is that the standard analysis of the ChIP-seq data completely miss the targeting specificity, which requires the use multi-mapping reads to account for the repetitive element right in the functionally interacting site. STAR18 is a very long, inefficiently spliced transcript located in the largest human-specific duplicated region of the genome. Dependent on the the transcription factor STAT3 for its transcription, it interacts with STAT3 in manifold ways in an autoregulatory loop.

4.25 Structure, evolution and targeting of long non coding RNAs

Gabriele Varani (University of Washington – Seattle, US)

The molecular basis for activity of long non coding RNAs and their structure-function relationship remain to be established. A commonly stated but unproven hypothesis is that their secondary structures are conserved and functional, despite low levels of primary sequence conservation. We have discovered a complex secondary structure in the functional core of Cyrano, a rare lincRNAs conserved over significant evolutionary distances, at the center of which is a strikingly conserved cloverleaf structure maintained over >420 million years of evolution. This structure provides protein interaction sites and is recognized by miR-7 in a non-canonical fashion. Structures within ncRNAs are functional, as we demonstrated for the promoter associated transcript controlling expression of the tumor suppressor E-cadherin,

where a single SNP that alters RNA structure affects differential recruitment of epigenetic enzymes to the promoter and leads to different outcome in cancer patients. These RNA structures provide unexploited targets for intervention using peptidic and small molecule chemistry, as we illustrate with targeting of the microRNA precursor coding for the oncogenic miR-21.

4.26 Large-scale prediction and analysis of RBPs

Gene Yeo (UC – San Diego, US)

I will present my lab's recent efforts in expanding the repertoire of RNA binding proteins (RBPs) using available affinity precipitation followed by mass spectrometry datasets. Our algorithm SONAR can be applied to yeast, fly and human protein-protein interaction datasets to identify novel RBPs. I will also discuss our improvements to CLIP technologies (enhanced CLIP) and large-scale CRISPR-tagging of RBPs in human cell lines.

Participants

- Frédéric Allain
ETH Zürich, CH
- Rolf Backofen
Universität Freiburg, DE
- Benedikt Beckman
HU Berlin, DE
- Janusz Bujnicki
Int. Inst. of Molecular & Cell Biology – Warsaw, PL
- Tomaž Curk
University of Ljubljana, SI
- Philipp Drewe
Max-Delbrück-Centrum – Berlin, DE
- André Gerber
University of Surrey – Guildford, GB
- Mahsa Ghanbari
Max-Delbrück-Centrum – Berlin, DE
- Tim Hughes
University of Toronto, CA
- Eckhard Jankowsky
Case Western Reserve University – Cleveland, US
- Hilal Kazan
Antalya International University, TR
- Grzegorz Kudla
University of Edinburgh, GB
- Markus Landthaler
Max-Delbrück-Centrum – Berlin, DE
- Martin Lewinski
Universität Bielefeld, DE
- Yael Mandel-Gutfreund
Technion – Haifa, IL
- Hannah Margalit
The Hebrew University of Jerusalem, IL
- Annalisa Marsico
MPI für Molekulare Genetik – Berlin, DE
- Daniel Maticzka
Universität Freiburg, DE
- Irmtraud Meyer
Max-Delbrück-Centrum – Berlin, DE
- Quaid Morris
University of Toronto, CA
- Uwe Ohler
Max-Delbrück-Centrum – Berlin, DE
- Teresa Przytycka
National Center for Biotechnology – Bethesda, US
- Andres Ramos
University College London, GB
- Guido Sanguinetti
University of Edinburgh, GB
- Michael Sattler
TU München, DE
- Peter F. Stadler
Universität Leipzig, DE
- Gabriele Varani
University of Washington – Seattle, US
- Gene Yeo
UC – San Diego, US