

# THE MESSENGER

Newsletter No. 17

October 2022 — National Center of Competence in Research, RNA & Disease



**NCCR  
RNA & Disease**

National Center of Competence in Research  
The role of RNA in disease mechanisms

## Message from the director's desk



**Oliver Mühlemann**  
Director NCCR RNA & Disease

*Dear colleagues*

## Tempus fugit

We are nearly half a year into the third and final four-year phase of the NCCR RNA & Disease. As you may be aware, NCCRs, regardless of their performance, receive substantially less funding from the Swiss National Science Foundation (SNSF) for phase 3 than for the previous phases, and we are therefore very grateful that our home institutions, the University of Bern and ETH Zurich, decided to continue their generous support.

The overall reduction in funding prompted our network to adopt an internal competitive allocation of the phase 3 research funding: all PIs submitted project proposals that were then evaluated by the Scientific Advisory Board (SAB) and the Steering Committee, who decided which of the projects to include in the Pre- and Full-Proposal. The SNSF review panel found the Full-Proposal excellent and had no recommendations for changes, confirming the added-value of our internal evaluation.

In addition to the proposed research projects, which are even more disease-oriented and collaborative in phase 3, the panel appreciated the network activities, including the areas of Communication, Education, Equal Opportunities (EO) and Knowledge & Technology Transfer (KTT). I want to take the opportunity here to express my gratitude to everybody who contributed to this excellent outcome of our phase 3 Proposal, especially our SAB members, whose expertise and tireless support is highly appreciated!

I am looking forward to the exciting scientific discoveries to be made during the remaining 3.5 years and to our ongoing and new activities in other areas. Regarding network activities, after COVID-imposed postponements and cancellations, we were able to meet again in person at the retreat in Engelberg in March and at the Summer School in Saas Fee in August. I am looking forward to the upcoming retreat in January 2023, for which we will re-

turn to Kandersteg, and to the Swiss RNA Workshop that will take place just before the retreat.

On Thursday, January 26, 2023, the day before the Swiss RNA Workshop, a new KTT event will take place in Bern to which I would like to draw your attention: The first Swiss RNA Therapeutics Summit is a one day-meeting jointly organized by our KTT team and the Swiss Biotech Association with the goal to tighten the links of our researchers with biotech. Regarding EO, we are pleased to see a growing interest in the newly established parental leave support scheme.

As another teaser, I can announce that together with a professional team at the University of Bern, the production of a FLASH Mooc interactive video on RNA is being planned, which targets an interested lay audience and will be linked to our MoleCool website. For public outreach, this year's highlight was without any doubt the NCCR's exhibition "*Kosmos RNA – The Code of Life*" at the Night of Research of the University of Bern, of which impressions can be found on page 12 of this Newsletter.

The main story of this Newsletter is an interview with Prof. John Mattick (UNSW, Sydney, Australia), with whom I spent my sabbatical 6 years ago and who gave a seminar at the ETH earlier this year, sharing with us his views on the RNA Cosmos. Fresh of the press is John's book "*RNA – the Epicenter of Genetic Information*", which he co-authored with his former PhD student Paulo Amaral. The book takes you through the history of molecular and RNA biology, highlighting unexpected turns in this exciting research field.

Apropos "unexpected": New and unexpected findings were also made by some of our colleagues in the network, and they are featured as three Research Highlights in this Newsletter. I am convinced that more unforeseen and exciting discoveries will be made during phase 3 by NCCR RNA & Disease researchers. Therefore: stay tuned, take pleasure in conducting research, and enjoy reading this edition of The Messenger!

## Interview John Mattick

# "What you are liable to read in the textbooks ain't necessarily so."

Interview: Dominik Theler

**In this interview, John Mattick shares with us his view on "RNA The Epicenter of Genetic Information", which is the title of a book, which has just been published, and the past, present and future of RNA Biology and its role for life.**

*For whom and why did you write the book "RNA The Epicenter of Genetic information" with the subtitle "A New Understanding of Molecular Biology"?*

This book is not for the public but scientific peers. We wrote the book to explain how genetic information has been misunderstood. This cannot be done in a review article nor a two-hour lecture. To understand how we got to this point, one must go back to the beginning of molecular biology and follow its fascinating history with a focus on how the roles of RNA were defined.

The book was prompted by the background work done by my former student and co-author Paulo Amaral. He included an appendix documenting the early history of RNA research in his PhD thesis. I thought this was a good starting point for a book; it took three years to put it together. I am pleased with the outcome and endorsements it received, for example, from Tom Cech and Joan Steitz.

The book contains a lot of quotes. Throughout the history of Molecular Biology, progress was clumsy because, over and over again, the great and the good were skeptical of unexpected findings and opposed ideas that turned out to be correct. Prominent scientists said, "We discovered rRNA, tRNA and mRNA; the hard work has been done," genes = proteins, and the rest is now just detail.

It was the same story for epigenetics and histone modifications. David Allis faced tremendous difficulties until their importance was established. And there are many stories like this. Barbara McClintock's discovery of transposable elements, which she correctly insisted are "controlling elements", is another example. In the book, I put the quote from the American musical Porgy and Bess by George and Ira Gershwin: "It ain't necessarily so – The things that you're liable to read in the Bible – ain't necessarily so." My

adaptation is, "What you are liable to read in the textbooks ain't necessarily so." I am not talking about factual knowledge like the detailed structure of the ribosome, but the conceptual framework itself.

*"The underlying problem was to accept that the textbooks would need to be rewritten if lncRNAs were functional."*

*Regarding concepts, what causes the doubts regarding the functionality of long noncoding RNAs?*

In the early 2000s, the transcriptome projects surprisingly revealed tens of thousands of long transcripts with little or no protein-coding potential. The underlying problem was to accept that the textbooks would need to be rewritten if they were functional. These RNAs emerged out of the mist, and no conventional explanations for gene regulation could accommodate such a large army of molecules that had not been accounted for previously. A common refrain was: "They might be noise." Two arguments were used to strengthen this notion:

One argument was that these RNAs are lowly expressed and less conserved than protein-coding sequences. There are several problems with that argument, one of them being how conservation was assessed. In 2002, the mouse and human genome papers used ancient transposons common to both species to assess the rate of "neutral" evolution and found a similar degree of divergence in the rest of the genome, which they concluded is also evolving neutrally and the RNAs expressed from it must also be non-functional. This is an entirely circular argument. Nowadays, it is evident that transposable elements are major features of genome biology.



## John Mattick Biography

In 1978, John Mattick obtained his PhD from Monash University, Melbourne, Australia researching mitochondrial DNA replication and mutation. For his postdoc, he studied the organization and function of the fatty acid synthase complex at the Baylor College of Medicine, Houston, USA. As an independent group leader he worked first at the Commonwealth Scientific and Industrial Research Organisation in Sydney, Australia developing a DNA vaccine and studying bacterial type IV pilus assembly. In 1988, he moved to the University of Queensland in Brisbane, Australia where his research increasingly became focused on non-coding RNA. From 2012 to 2018, he served as the Executive Director of the Garvan Institute of Medical Research in Sydney. Parallel to studying the role of non-coding RNAs, John Mattick starting at the Garvan Institute developed several clinical sequencing initiatives including during his time in the UK as Chief Executive of Genomics England before returning to the UNSW in Sydney in 2020 as Professor of RNA Biology.

[Faculty Profile UNSW Sydney](#)

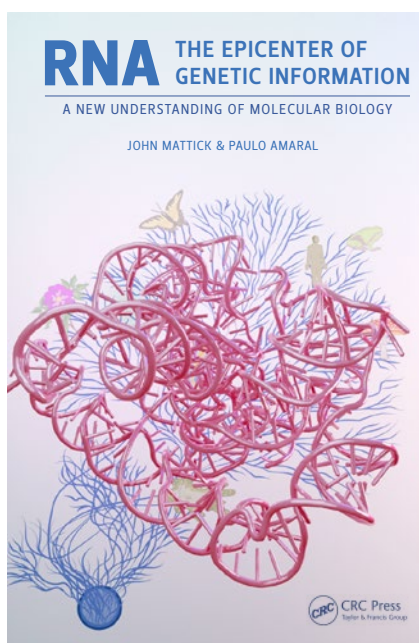
## Interview John Mattick

The second argument was that conservation imputes function and thus the rapidly evolving long noncoding RNAs (lncRNAs) are less likely to be functional. However, conservation is a relative measure, and low conservation imputes nothing. There must be lineage specificity: regulatory sequences, including promoters, evolve much more quickly than highly constrained protein-coding sequences. Protein sequences must maintain their structure for their function. On the other hand, regulatory sequences, including RNAs, have much more plastic structure-function relationships than proteins. Evolutionary developmental researchers will tell you that it is evident that phenotypic variation comes largely from regulatory sequence variation and not protein sequence variation. Thus, there is positive selection for variation in regulatory architecture, which underpins phenotypic radiation.

*Where does this fundamental resistance come from, e.g., the number of protein-coding genes between *C. elegans* and humans is similar? Should this not hint at the presence of other significant differences between the genomes of these organisms?*

The resistance to accepting the functionality of lncRNAs is fundamentally a victim of the orthodox conceptual framework of gene regulation compounded by reductionism. They did not fit and most were focussed on their gene or protein of interest, not how the system works. When the *C. elegans* genome was published and it became clear that the number of protein-coding genes was similar and that many are orthologous to those in humans, it was assumed that the combinatorics of transcription factor regulation provides more than sufficient power to enable the developmental programming of a worm or a human. However, the assertion that transcription factor combinatorics could explain everything about gene regulation and

*"We wrote the book to explain how genetic information has been misunderstood."*



diversity was vague. It was never justified theoretically, mathematically or mechanistically.

There are two interesting features of transcription factors: Nearly all of them contain intrinsically disordered regions (IDRs), and most can bind RNA. How does it mechanistically work that a transcription factor binds to different promoters in different cells at various stages of development? There is no answer to that in conventional space. However, the data show that zinc finger transcription factors have a higher affinity for RNA:DNA hybrids than they do for double-stranded DNA. RNA:DNA hybrids and triplexes occur over the genome, so a plausible explanation is be that RNA molecules select the exact binding sequence of transcription factors in a given cell at a given time. RNA regulatory networks direct where transcription factors bind to the genome for controlling transcription.

*The so called 95% "junk" has crucial functions?*

The junk idea has a long history dating back to the 1930s when theoretical biologists considered the size of genomes. They argued that the mutational load would be too high if there were the same density of protein encoding sequences in humans as you have in bacteria. A nucleotide variation in a protein that changes a codon or introduces a stop codon is often catastrophic. However, if a nucleotide in a regulatory RNA changes the regulatory architecture, which is the basis of quantitative trait variation. Back then, there was a long argument between so-called "*Mendelians*" and those working in agriculture and animal breeding who understood that quantitative trait variation is not usually a function of protein-coding mutations. This has since been confirmed by the genome-wide association studies (GWAS).

The origin story and emergence of molecular biology is muddled. The early triumphs in bacterial genetics and the complexity of animal and plant genomes complicate an intricate history. This book documents the many advances, as well as the prejudices and founder fallacies. It highlights the premature relegation of RNA to simply an intermediate between gene and protein, the underestimation of the amount of information required to program the development of multicellular organisms, and the dawning realization that RNA is the cornerstone of cell biology, development, brain function and probably evolution itself. Key personalities, their hubris as well as prescient predictions are richly illustrated with quotes, archival material, photographs, diagrams and references to bring the people, ideas and discoveries to life, from the conceptual cradles of molecular biology to the current revolution in the understanding of genetic information.

[Click here to download the Open Access pdf of the book.](#)

Cover image and book description from the book by John Mattick & Paulo Amaral "[RNA, the Epicenter of Genetic Information](#)" published under a [CC BY-NC-ND 4.0 license](#).

*"The fundamental mistake was to think that proteins transact most genetic information in complex organisms."*

## Interview John Mattick

The second argument was based on the C-value enigma. Some organisms like certain amoebae, arthropods and amphibians have much more DNA per cell than humans. The assumption was that they have variable amounts of junk, which justifies that the assumption that the human genome can also contain a lot of junk. However, the increase in noncoding sequences compared to organismal complexity suggests a massive expansion in the regulatory architecture. The only way to invalidate this proposition would be to identify downward exceptions: complex organisms with little noncoding DNA. None have been found to date.

The fundamental mistake was to think that proteins transact most genetic information in complex organisms. Most of the information is, in fact, transacted by RNAs, and most genes produce RNAs, which then organize cell fate decisions from fertilization to the adult.

*Did RNA in terms of molecular evolution passed on the heredity part to DNA and the catalytic part to proteins?*

That is a fair summary. RNA was likely the ancestral molecule because it combines the two critical functions of information storage and catalysis. Information storage was then outsourced to the more stable and easily replicable DNA, which was an intelligent move of evolution. Catalytic activities were largely outsourced to proteins because they possess more chemical versatility. The proof that RNA preceded proteins is simple: peptide bond formation in the ribosome is an RNA-catalyzed reaction.

*Moving to another RNA-catalyzed process, Splicing: You consider the discovery of introns as the biggest shock in molecular biology. Why?*

When the discovery was made that genes are not collinear with their protein products in complex organisms and that protein-coding sequences are split into bits located over vast territories, the reaction was, "Wow, what is going on here?" My big criticism is that nobody at this moment took the chance to reconsider what was really known and not known about genetic information, especially in complex organisms. Somebody once said, "The best science is done at the point of greatest surprise." If that is true, then molecular biology was found wanting because quickly and almost universally introns were condemned as another manifestation of and evidence for junk sequences in the genome.

Walter Gilbert wrote an article to rationalize the presence of introns. He suggested that they were remnants of the primordial assembly of genes where you had fragments of protein-coding sequences "exons" inter-

spersed with other RNA sequences "introns." Evolution then built proteins by removing the intervening sequences by splicing after transcription. He proposed that bacteria lost introns under the pressure of rapid replication but were retained in the slower-growing eukaryotes.

There is a significant fault with this argument. Gilbert used a colorful expression in that article, writing that animals retained "the full stigmata of their birth", which is an evolutionary non sequitur. The eukaryotes were unicellular for at least 2 billion years under the same selective pressures of rapid replication prior to the emergence of multicellularity. So, his argument tacitly suggests that there was a clade of unicellular eukaryotes sitting under a proverbial evolutionary rock, waiting for the sunny day that they would emerge as complex organisms. That does not make sense.

The more reasonable argument is that self-splicing group II introns, which exist in

another dimension to the prebiotic assembly of life. RNA-protein interactions drive phase separation. Interestingly, the most ancient codons are the ones that specify the amino acids in the IDRs.

So, the plausible scenario is that RNA has a function beyond information storage and catalysis: the ability to cooperate with primitive peptides to form phase-separated domains. These domains can then become reaction centers for biochemical and genetic evolution, producing a protocell.

Phase separation is the hidden and overlooked dimension of the organization of the cell, and of the chromatin, during development. The proportion of IDR-containing proteins has increased enormously with organismal complexity and scales with it. Nearly all proteins that control mammalian development contain IDRs. Genetic loci called enhancers – of which there are ~400,000 in the human genome – control the spatiotemporal patterns of gene expression during development by inducing chromatin rearrangements, and enhancers express lncRNAs in the cells in which they are active, likely the mechanistic basis of their function

*Staying with the spatial organization, cellular RNA localization is also far from functionally and mechanistically being understood.*

Over a decade ago, we showed specific expression of localization of particular long noncoding RNAs to unknown subcellular locations. Seeing these images was a "Wow" moment. Somewhere around 30% of lncRNAs go to the cytoplasm, while the others are retained in the nucleus. Specific cellular localization provides further evidence against the opinion that lncRNAs are just transcriptional noise. lncRNAs comprise the major information complement of the genome, they are highly alternatively spliced,

*"We do need to determine the structure-function relationships for lncRNAs."*

out of the way places in bacteria, recolonized genes after the separation of transcription and translation. The early eukaryotes were scavengers. As a scavenger, if you start engulfing things, you must protect your genome, and so transcription and translation were separated. This provided a window for group II introns to invade genes and splice themselves out before translation and led to the formation of the spliceosome. These internal segments then became the substrate for positive selection for RNA regulatory functions that were produced in parallel with the protein-coding sequences.

*On the topic of cellular compartmentalization, how does phase separation relate to the origin of life?*

Most people would agree that phase separation has been one of the most exciting developments in molecular cell biology in the last decade and has been staring us in the face since nucleoli were first observed microscopically. Phase separation also gives

*"By our estimation, there are 10 million of conserved RNA structures in the mammalian genome, with 2 million classified as high confidence."*



## Interview John Mattick

and their structure appears to be modular. If we can work out which structures in lncRNAs perform which functions, we can elucidate their mechanism and pathways.

There are thousands of publications on lncRNAs, but most are descriptive. Many labs look at long noncoding RNAs in cancer, differentiation, or something else. And then they see one changing, perturb it and report that something happens. However, that is not getting to the heart of how they work, although it is valuable because it adds to the weight of evidence of the functionality of long noncoding RNAs.

We are trying to decipher the mechanisms of lncRNA action, but I am not a big fan of getting deep down in the trench because I think you can lose your way. But we do need to determine the structure-function relationships for lncRNAs. So, I am dreaming of a new Rfam. This Rfam, like Pfam for proteins, would tell you based on the sequence that a lncRNA contains, e.g., a Polycomb binding domain. Then we can start putting some structure into understanding what lncRNAs are doing and where they are going.

*Do we have enough data to take this type of approach?*

The lack of training data is the problem. I think the field of RNA structural biology will grow and that the data will come. However, RNA structures are complex and sensitive to base changes and modifications. The way we have started to tackle this problem is to look at high confidence predictions of two-dimensional RNA structures. When we did that, we could show that almost 20% of the human genome was conserved at the level of predicted RNA structure.

We then looked in an evolutionary series at how nucleotide variations affected the predicted structure, that is, if we could find changes in predicted stems accompanied by a complementary change that would maintain the stem, in other words, co-evolution. The more depth you have in your evolutionary series, the more statistical confidence you have that your two-dimensional structure projections are correct. By our estimation, there are 10 million of conserved RNA structures in the mammalian genome, with 2 million classified as high confidence.

*You call RNA "The computational engine of the cell". How do you see the future of RNA research and which dogmas will be overturned?*

The big dogma to be overturned are the idea that genes mostly encode proteins, and that the human genome is full of junk, which is the complete reverse of the truth. To the contrary, the human genome is a highly efficient information suite. Measured in bits, it con-

tains only 825 megabytes, less than the size of Microsoft Word, yet contains the information that puts 30 plus trillion cells in the right places with all their specialized architecture and functions.

The genome contains a sophisticated program that takes a single cell and directs the development of the entire organism, e.g., to form all your bones and muscles in the right shapes and places. For me, after the brain, bones are the most fascinating organ of the body because of the variation in their architecture. Plus, bones have different densities, and all that architectural variation is derived from the same or highly similar cell types. So, most of the programming for human development has not to do with cell differentiation. It is easy to specify cell types. Instead, it is organizing the architecture, and almost nobody in molecular biology thinks about this, another example of a casualty of reductionism.

*"The big dogma to be overturned are the idea that genes mostly encode proteins, and that the human genome is full of junk, which is the complete reverse of the truth."*

Another dogma to be overturned is that transcription factors control development. They execute functions, but RNAs exercise the actual control, a huge conceptual change that may take another decade for people to accept.

At the practical level, once we understand the structure-function relationships and pathways, we can manipulate them, for example, in the case of genetic variations underpinning complex human traits and disorders. These variations primarily lie in intergenic regions, not protein-coding ones. The intergenic haplotype blocks identified by the GWAS studies are replete with lncRNAs, which are the candidates for the underlying mechanistic basis.

*"The 3' UTRs of many genes are not necessarily co-expressed with their associated protein-coding sequences but can be expressed independently."*

*Can it be that simple that it is only one or a few long noncoding RNAs, or would it not be a complex interplay of many genetic loci?* I think it will be possible to identify the best treatment for people with particular subsets of complex disorders. The GWAS data indicate that there are 50 or 100 loci that contribute, but it does not necessarily mean that one needs to reconfigure the whole network, but rather address just the part that is damaged. Once we understand the mechanistic basis of the variations underpinning complex disorders, it is likely that at least some of these damaging changes can be corrected in some way.

*What are new technological approaches most urgently needed?*

For me, this is high-resolution single-cell sequencing. That might sound odd because everyone is doing single-cell sequencing nowadays, but it is not yet high resolution, with few exceptions. Most single-cell sequencing polls the 3' ends of the abundant protein-coding transcripts. There are two problems with this. One is unknown to most people. The 3' UTRs of many genes are not necessarily co-expressed with their associated protein-coding sequences but can be expressed independently. The evidence for this comes not only from in situ hybridization and genetics. So, the fact that you get a 3' end in your sequencing data set does not mean that the actual protein is being produced.

The more significant problem is that most sequencing does not poll the splice variants. lncRNAs are almost universally alternatively spliced, potentially varying their protein cargoes and genomic targets at every one of the ~60 trillion cell fate decisions that be made during development. Therefore we need a high-resolution analysis of the transcriptional output of cells at every stage. That will be impossible in humans, but it should be possible in mice. You need to be able to sequence the entire length of all transcripts – all mRNAs, lncRNAs and small RNAs – in a cell.

## Interview John Mattick

*Where do you see problems with today's molecular biology?*

One is the lack of reproducibility of many findings in molecular cell biology. There are two big causes: One is based on self-interest, and the other is cell culture. When I was director of a medical research institute, a senior colleague asked my permission to spend five million dollars to buy a large batch of fetal calf serum. I asked him why he would want to buy such a massive amount all at once. He answered that it would provide five years stable supply of serum, which would guarantee reproducibility of experiments. I asked if he realized what he just said, namely that the experiments may not be reproducible by others; using fetal calf serum or anything undefined in experiments can give you batch effects. The only way around it is to repeat the experiments with three or five batches and see if you get the same result independent of the batch used.

The other reason is psychological. If you work on a specific gene or pathway, then it is in your interest to ensure that that gene or pathway is important because your career and grant applications depend on it. So, the design and interpretation of every experiment is subconsciously dedicated to this proposition.

*"Do what suits your soul and do your best. There is no inferiority or superiority in different career paths, and everybody is different."*

*"Above all, be curious."*

A background problem is founder fallacies and validation creep. There is an excellent article by Marc Halfon on this topic with reference to the understanding of enhancers. Early on, the idea was put forward that enhancers bind transcription factors and then loop around to bring those transcription factors into contact with the promoter of the target genes. That is the standard model, but there is no evidence whatsoever for transcription factor crosstalk, beyond the fact that enhancers cause local topological rearrangements in chromatin. It was just a conceptual proposal, but such generalizations often become founder fallacies. The proposition may have been a reasonable hypothesis, but became the conventional explanation and an article of faith, which has biased the interpretation of experimental data ever since.

*How did your research journey lead to RNA?* I did my PhD in the seventies during the period when the cloning revolution was underway. I studied DNA replication in yeast mitochondria. For my postdoc, I went to Houston to work on fatty acid synthase. Regarding RNA, a couple of months into my postdoc, I talked on Friday night over a beer to a friend who told me about introns.

So, from there on, I have been intrigued by the RNA that is transcribed but not translated. My immediate response to hearing about introns from my friend was that there might be some other form of information being transacted by these sequences. However, the data and technologies were limited in those days so exploring the idea was difficult. Nevertheless, it remained my intellec-

tual hobby and in the early nineties I started actively working on the idea that other information was being transacted by RNA, especially in complex organisms. I see biology primarily in terms of information rather than chemistry. Sure, the information is transacted by chemistry, and it is essential to understand it, but my interest is the type of information in genomes.

*What advice would you give to young researchers besides being good in bioinformatics?*

Do what suits your soul and do your best. There is no inferiority or superiority in different career paths, and everybody is different.

Find time to read and think. One of the problems with modern science, which should not be understood as a criticism of scientists, is that we do not have enough time to read and think. Thinking is crucial for planning experiments, and there is no point in executing improperly designed experiments.

Look for the things that do not make sense. The unexpected results are the most exciting, and you should follow up on them rather than just put them aside because they do not fit the current way of thinking. Unusual observations usually lead to new insights. As well, you should always be thinking to generalizability of what you do, keeping in mind that your interpretation may not be correct.

Above all, be curious.

Interview conducted on May 6, 2022.

## Research Highlights

# Surprising insights into UV Cross-Linking of Proteins and RNA

Dominik Theler

Researching RNA protein interactions at different scales and resolutions often involves UV cross-linking. However, UV-induced cross-linking between RNA and proteins is not a well-understood process despite its widespread use. The lab of Jonathan Hall joined forces those of Ruedi Aebersold, Fred Allain and Alexander Leitner at ETH Zurich to gain insights into this process.

Gaining insights was not as straightforward as it seems. Cross-linking of amino acids and nucleotides requires high local concentrations and the special environment of the protein domain. Hence, it is difficult to mimic the cross-linking reactions under conventional reaction conditions in aqueous solutions. Furthermore, cross-linking often leads to complex mixtures of reaction products that are difficult to characterize.

As a model system for cross-linking studies, the researchers turned to the RNA Recognition Motif (RRM) domain of the human splicing factor Fox-1 bound to its consensus heptanucleotide RNA binding motif. After UV irradiation and subsequent sample preparation, the cross-linked products were analyzed by mass spectrometry (MS) to pinpoint which amino acids of the RRM cross-link to which ribonucleotides. To precisely identify the cross-linking nucleotide, the RNA motifs were synthesized with different isotopic labelling schemes, in which one of the seven nucleotides

was labelled with  $^{13}\text{C}$  atoms. The MS signal of a given cross-link comprising a labelled nucleotide shows a defined mass shift with that of an unlabelled equivalent, which allows for the unambiguous identification of the involved nucleotide.

The surprising result for the Fox-1 RRM “protein-side” was that a cross-link can only occur if there is a  $\pi$ -stacking between an aromatic amino acid and guanosine or uridine. Moreover, this  $\pi$ -stacking requirement applies not only for cross-links to the stacked aromatic amino acid but also for its neighboring amino acids. “Discovering the  $\pi$ -stacking’s importance was an eye-opening moment and changed course of the project”, states Jonathan Hall, last author of the study published in *Nature Communications*.

Subsequently, the researchers looked at three published large-scale MS cross-linking datasets. Excluding cross-links to sulphur-containing amino acids, and using structural data, they looked at the percentage of cross-links in the proximity of an aromatic amino acid that is  $\pi$ -stacked to a nucleotide. In the three datasets, 42, 52 and 78 % of the cross-links involved a  $\pi$ -stacked aromatic amino acid in proximity, validating the team’s findings.

Detailed analysis of the structure-reactivity relationships of other RNA-protein complexes led to an equally surprising discovery on the

“RNA-side”. The MS data for one particular histidine showed cross-linking to uracil (U), while the 3D structure indicated  $\pi$ -stacking to cytosine (C). The most likely explanation for this discrepancy was that hydrolysis of the  $\pi$ -stacking cytosine to uridine occurs during or after cross-linking. Re-inspecting data from the Fox-1 system suggested that C to U conversions can be widespread, suggesting that not only guanosine and uridine, but also cytosine can participate in UV cross-linking reactions. The unexpected finding of the C to U conversion on the “RNA-side”, as well as the importance on the “protein-side” for  $\pi$ -stacked aromatic amino acids, should be kept in mind for analyzing CLIR-MS and CLIP data.

The publication describing these findings was published in *Nature Communications*. “We gained a lot of insights through this project but could not determine the chemical structure of the cross-linking products. Also, extending this study to sulfur-containing amino acids would contribute further important insights into this process,” says Anna Knörlein, first author of the study.

## Publication:

Knörlein et al. (2022) *Nature Communications* 13(1), 2719 (Open Access)

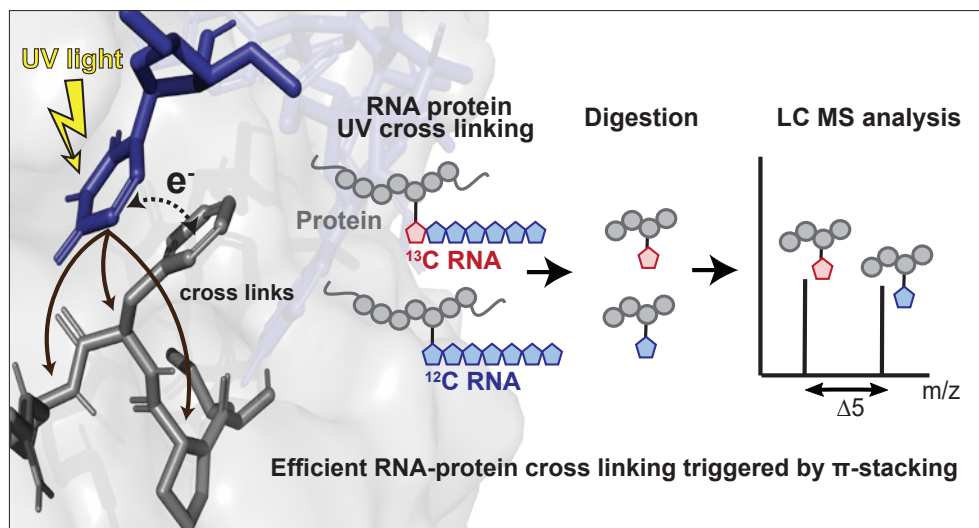


Figure kindly provided by the last author.

## Research Highlights

# A journey into non-canonical functions of Argonaute proteins

Dominik Theler

In eukaryotes, Argonaute (AGO) proteins are best known for their function in the miRNA and siRNA pathways. Besides this canonical function, Argonaute proteins were implicated in other cellular processes, primarily by experiments in cancer cell lines. Mouse embryonic stem cells (mESCs), derived from the inner cell mass of mouse blastocysts, are the only healthy cells surviving upon the depletion of most miRNA biogenesis factors. Working with this system, the Ciaudo lab (ETH Zurich), in collaboration with the Santoro lab (University of Zurich), could shed light on novel non-canonical functions of AGOs in early development.

In mESCs, only two of the four human AGOs, AGO1 & 2, are expressed. In these cells, the two AGOs have partially overlapping functions. In the absence of AGO2, AGO1 gets upregulated and loaded with miRNAs typically loaded into AGO2. But AGO1 cannot compensate for the lack of AGO2 in the extraembryonic endoderm differentiation. Hinting that AGOs have additional functions in early development besides their role in the miRNA pathway.

The journey started with the transcriptomic analysis of mESCs in which key players of the miRNA pathway were knocked out using CRISPR/Cas9 genome engineering approach. This analysis revealed that nearly 1800 genes show a changed expression only in AGO1 & 2 double-knockout cell lines but not a significant expression change in cells lacking either Dicer, Drosha or Dgcr8, indicating a miRNA-independent mode of regulation. Of these specifically differentially expressed genes (DEGs) in AGO1 & 2 double-knockout cells, over 1000 of them were downregulated and nearly 750 upregulated.

From this observation, the journey continued to elucidate how these DEGs are regulated. The first possibility investigated was the regulation by epigenetic modifications. The researchers assessed the global levels of specific histone marks. They found that overall levels of only one histone mark, H3K27me3, went drastically down in the absence of AGO1 & 2. Corresponding to this finding, several proteins forming the complex involved in its deposition showed reduced expression at the protein but not the RNA level. However, when looking at the genes most strongly regulated by this histone mark

and comparing them to the AGO1 & 2 double-knockout specific DEGs, the overlap was small.

After ruling out histone marks as the primary regulators of these DEGs, the researchers looked at changes in chromatin accessibility in the AGO1 & 2 double-knockout cell lines. Over 3000 genomic regions showed changed accessibility only in the double AGO knockout and not in the single knockouts. However, the overlap of these regions with the genomic loci of the DEGs was again small.

Next, the researchers investigated the potential role of transcription factors (TFs) and their binding sites in the context of changed chromatin accessibility. This led to the identification of five TFs with computationally predicted differential binding in the double knockout cells. Then, they looked at the correlation of the differential TF binding sites with the genomic DEGs regions and found a positive correlation. Most of this correlation was contributed by the binding sites of the two TFs, KLF4 and CTCF.

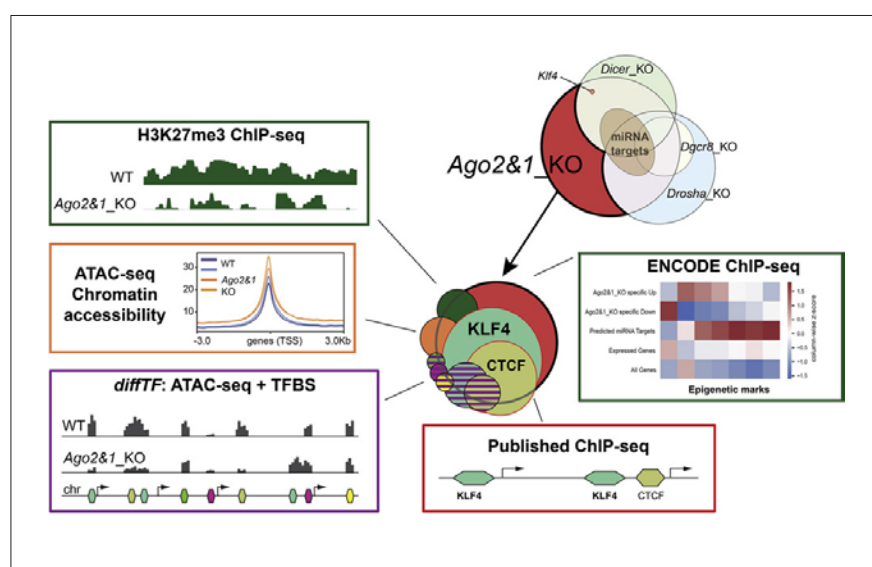
While CTCF did not display changed at protein levels in the AGO1 & 2 double-knockout cells, KLF4 levels were significantly downregulated. Analyzing KLF4-ChIP-seq data, the researchers noticed that the promoter and

enhancer regions of over 70 % of upregulated DEGs and nearly 50% of the downregulated DEGs were bound by KLF4. Intriguingly, this transcription factor is an important regulator of cellular pluripotency, which was now shown to be regulated by AGO1 & 2 in a miRNA-independent fashion. The researchers' findings were published in the journal *Stem Cell Reports*.

"It was a journey through different fields: transcriptomics, epigenetics, genome organization and transcription factors, and integrating all the data took a lot of computational biology. The project's success crucially dependent on the two co-first authors bringing in their respective expertise," says Constance Ciaudo, last and corresponding author of the study. Regarding the continuation of the journey, Madlen Müller, co-first author comments: "I would like to find out the mechanism how AGOs control KLF4 expression and function. Also, which direct functions do Argonautes have in the nucleus and which effects observed there are indirect, is an intriguing open question."

## Publication:

Müller, Schäfer et al. (2022) *Stem Cell Reports* 17(5), 1070-1080 (Open Access)



Graphical abstract from Müller, Schäfer et al. (2022) *Stem Cell Reports* published under [CC BY-NC-ND 4.0 licence](#).



## Research Highlights

# An unconventional YTH domain protein with essential functions

Dominik Theler

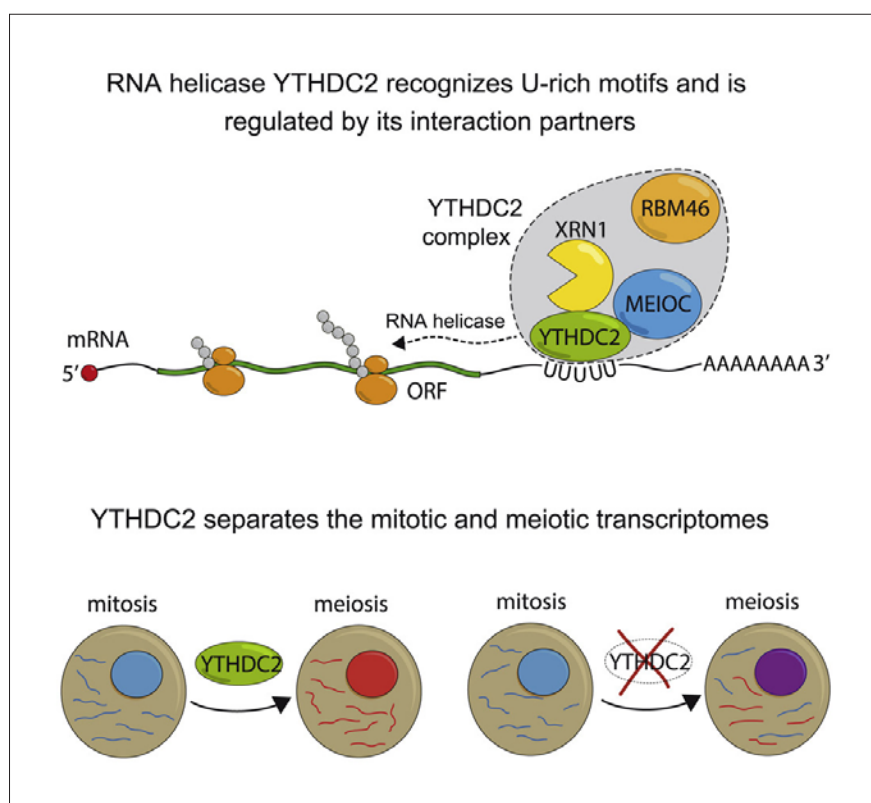
YTHDC2 is the largest member of the protein family containing YTH domains. These are RNA binding domains, most of which are m<sup>6</sup>A readers, the most frequent internal mRNA modification. Besides the YTH domain, the YTHDC2 mouse protein contains R3H, OB, RecA and Ankyrin repeat domains. Loss of YTHDC2 leads to infertile mice of both genders. The Pillai lab (University of Geneva) continued researching YTHDC2 and its function in fertility by investigating its domains' functional roles as well as its RNA and protein binding partners.

A CLIP experiment with testis tissue revealed over 30'000 YTHDC2 binding sites in thousands of transcripts. Over 90% of transcripts were mRNAs, and more than half of the sites were in the 3'UTR. The cross-linked sites contained U-rich motifs, and an examination of up- and downstream sequences did not discover an enrichment of the m<sup>6</sup>A methylation consensus motif.

These results and the weaker affinity to m<sup>6</sup>A RNA of the protein's YTH domain compared to those from other proteins prompted the question whether m<sup>6</sup>A binding is essential for YTHDC2's function in the germline. To investigate this, the researchers created knock-in mice in which a critical YTH domain residue for m<sup>6</sup>A binding was mutated. Mice homo- and heterozygous for this mutation were found to be viable and fertile and did not show any observable phenotype, especially in germline tissues. Matching these results is that the *Drosophila* YTHDC2 homolog is important for germline function but lacks a YTH domain.

Continuing with mouse genetics, the Pillai lab created mice carrying a YTHDC2 allele which contained a mutation rendering the helicase domain catalytically dead. This allele showed a dominant infertility phenotype as heterozygous cat-dead male mice failed to produce progeny when crossed with wild-type females. In addition, testicular transcriptome analysis of such a mutant revealed the downregulation of genes highly expressed in the meiotic stage of spermatogenesis.

To further study the transcriptome, the researchers applied single-cell sequencing to testicular germ cells from YTHDC2 knockout mice. They found that mutant germ cells contained a mixed transcriptome having at the same time mitosis- and meiosis-specific tran-



Graphical abstract from [Li, Krasnykov et al. \(2022\) Molecular Cell](#) published under a [CC BY 4.0 licence](#).

scripts. "Single cell sequencing allowed the distinction of cells with a mixed identity from mitotic and meiotic cells," comments Kyrlo Krasnykov, co-first author of the study published in *Molecular Cell*. This revealed that *Ythdc2* mutant germ cells start the transition from mitosis to meiosis but then get stuck and cannot complete it in the absence of YTHDC2, leading to the observed infertility phenotype.

The helicase's essential function for fertility warranted further investigation. Interestingly, the RecA helicase domain of YTHDC2 is split into two parts with two Ankyrin repeats located between them. These repeats are responsible for YTHDC2's interaction with the exoribonuclease XRN1. Studies conducted by Lingyun Li, the other co-first author of the study, showed that YTHDC2 has a low RNA unwinding activity *in vitro*, while the mutant protein lacking the ankyrin repeats showed increased activity. This helicase-breaking

function of the repeats could be released by adding recombinant XRN1, which acts as a helicase accelerator.

Overall, the researchers favor a model in which YTHDC2 is essential to degrade mitotic transcripts for cells to undergo meiosis subsequently. "For the transition from the mitotic to the meiotic state, not only a transcriptional switch is needed, but the transcriptome needs to be cleared of remaining mitotic transcripts. YTHDC2 is responsible for this transcriptome clearance, and this involves not a couple of transcripts but entire sets. We are now focused on understanding how the RNA helicase activity contributes to this process," says Ramesh Pillai, last author of the study.

## Publication:

[Li, Krasnykov et al. \(2022\) Molecular Cell 82\(9\), 1678-1690.e12 \(Open Access\)](#)

**4<sup>th</sup> NCCR RNA & Disease Summer School**

# RNA & Entrepreneurship

*Dominik Theler*

The 4<sup>th</sup> NCCR RNA & Disease Summer School on “RNA & Entrepreneurship” took place from August 22–26, 2022, in Saas-Fee, Switzerland. Close to 60 students, speakers and organizers participated in the summer school.

Lectures were given on the one hand on RNA therapies and diagnostics, and on the other hand on aspects related to forming a start-up company. Participants presented their research projects with chalk talks. Subsequently, they worked in teams with speakers and organizers to develop a business idea in the field, which was presented to the other participants.



*Investor Panel with Sam Hall & Florian Müllershausen moderated by Robert Schneider.*



*Group picture of the participants of the 4<sup>th</sup> NCCR RNA & Disease Summer School on “RNA & Entrepreneurship”.*



#### 4<sup>th</sup> NCCR RNA & Disease Summer School

In the feedback survey, participants needed to answer how likely they were before and after attending the summer school, considering either founding a start-up company or joining one as a career option. The summer school participation led in both cases to a quite increased likelihood of joining or starting a start-up. The overall feedback received on the summer school was very positive.

We want to take all organizers & speakers who contributed to the success of the summer school. Special thanks go to Frédéric Allain, Robert Schneider and Rahel Büchi.

#### Lecturers & Scientific Organizers(\*) of the 4<sup>th</sup> NCCR RNA & Disease Summer School

Frédéric Allain*	ETH Zurich, Zurich, Switzerland
James Broughton	Mammoth Biosciences, Brisbane, USA
Paul Donohoue	Caribou Biosciences, Berkeley, USA
Jonathan Hall*	ETH Zurich, Zurich, Switzerland
Matthew Hall	Lausanne, Switzerland
Samuel Hall	New York, USA
Martin Jinek*	University of Zurich, Zurich, Switzerland
Rory Johnson*	University College Dublin, Ireland
Jørgen Kjems*	Aarhus University, Aarhus, Denmark
Adrian Krainer*	Cold Spring Harbor Laboratory, Cold Spring Harbor, USA
Helen Lee	Diagnostics for the Real World, Cambridge, UK
Gene Liao	Torque Bio, Durham, USA
Kathleen McCarthy	Skyhawk Therapeutics, Waltham, USA
Barbara McClung	Caribou Biosciences, Berkeley, USA
Eric Miska	Cambridge University, Cambridge, UK
Florian Müllershausen	Novartis Venture Fund, Basel, Switzerland
Paul Nioi	Alnylam Pharmaceuticals, Cambridge, USA
Samir Ounzain	Haya Therapeutics, Lausanne, Switzerland
Steve Pascolo	University Hospital Zurich, Zurich, Switzerland
David Rees	Institute of Intellectual Property, Bern, Switzerland
Martina Roos	Sardona Therapeutics, San Francisco, USA
Robert Schneider*	New York University, New York, USA
Markus Stoffel*	ETH Zurich, Zurich, Switzerland
Thomas Tuschl	The Rockefeller University, New York, USA
Teri Willey	Indiana University, Bloomington, USA



Final Panel Discussion at the summer school on "RNA & Entrepreneurship".

## Public Outreach

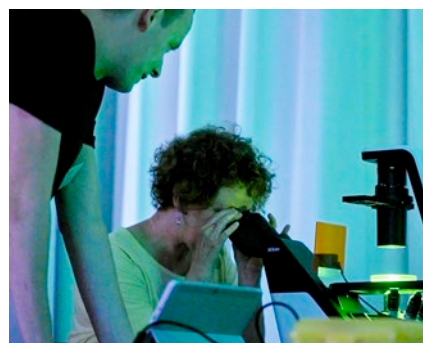
# Molecool – Kosmos RNA at the Night of Research at the University of Bern

Dominik Theler

The NCCR RNA & Disease participated with the exhibition "Kosmos RNA – The Code of Life" to the Night of Research ([Nacht der Forschung](#)) at the University of Bern. At the exhibition, the visitors could explore various experiments and posters on RNA-related topics. One of the highlights was the molecular bar, where visitors could get a drink mixed according to the recipe they translated from the RNA code to the amino acid code. In addition, our messenger RNA & ribosome roamed the event giving away snacks and

drawing attention to the exhibition. According to the University of Bern, the estimate is that 10'000 visitors attended the 2022 Night of Research.

The NCCR would like to thank the event organizers, all NCCR researchers and other persons that contributed to its exhibition, including Nikon for providing microscopes. Special thanks go to Nicole Kleinschmidt, the primary driving force of the Kosmos RNA exhibition, together with Noah Kleinschmidt, Marc Landolfo, Sofia Nasif and Karin Schranz.



## Participating Groups

Sebastian Leidel	University of Bern
Oliver Mühlemann	University of Bern
Mariusz Nowacki	University of Bern
Norbert Polacek	University of Bern
André Schneider	University of Bern
Volker Thiel	Institute of Immunology & Virology and University of Bern

## Concept & Coordination

Nicole Kleinschmidt	University of Bern
---------------------	--------------------





## Announcements

### People

We congratulate **Stefanie Jonas** (Institute for Molecular Biology & Biophysics, ETH Zurich) and **Magdalini Polymenidou** (Department of Quantitative Biomedicine, University of Zurich) who were selected as recipients of ERC Consolidator Grants!

Congratulations to **Ulrike Kutay** (Institute of Biochemistry, ETH Zurich), **Peter Scheiffele** (Biozentrum, University of Basel) and **Karsten Weis** (Institute of Biochemistry, ETH Zurich) for being awarded SNSF Advanced Grants!

We would like to welcome **Rodrigo Reis** (Institute of Plant Sciences, University of Bern) as a new associate member of the NCCR RNA & Disease! The Reis lab identifies and studies functional RNA structures involved in plant adaptation.

### NCCR RNA & Disease Retreat 2022

After a year without retreat, the 6<sup>th</sup> NCCR RNA & Disease retreat took place in Engelberg from March 21 – 23, 2022. Keynotes were given by the NCCR's Scientific Advisory Board Members Jørgen Kjems (Aarhus University, Denmark), Adrian Krainer (Cold Spring Harbor Laboratory, USA), Robert Schneider (New York University, USA) and Sarah Woodson (Johns Hopkins University, USA). A guest lecture was given by Matthias Baumgartner, who is the co-director of the University of Zurich's Research Priority Program [ITINERARE – Innovative Therapies in Rare Diseases](#).

### Swiss RNA Workshop 2023

The 23<sup>rd</sup> Swiss RNA Workshop is planned to take place on Friday, January 27, 2023, at the University of Bern. Invited keynotes will be given by **Marina Chekulaeva** (Max Delbrück Center for Molecular Medicine, Berlin, Germany) and **Eric Westhof** (Institute of Molecular and Cell Biology, CNRS, Strasbourg, France). For more information and to register visit the [Workshop's website](#).

**Frédéric Allain** (Institute of Biochemistry, ETH Zurich) **Marc Bühler** (FMI, Basel) and **Françoise Stutz** (Department of Cell Biology, University of Geneva) stepped down from the scientific organizing committee of the Swiss RNA Workshop. We would like to thank them for all their contributions to the event's success and would like to welcome their successors as scientific organizers **Maria Hondele** (Biozentrum, University of Basel) **Stefanie Jonas** (Institute of Molecular Biology & Biophysics, ETH Zurich) and **Ramesh Pillai** (Department of Molecular Biology, University of Geneva).

Join our LinkedIn [NCCR RNA & Disease Current Members & Alumni Group](#) and follow us on [LinkedIn](#) and [Twitter!](#)

Follow Molecool on [Facebook](#) and [Instagram](#).

### Swiss RNA Therapeutics Summit

The NCCR RNA & Disease together with the Swiss Biotech Association is organizing the first Swiss RNA Therapeutics Summit which unites Swiss Leaders and Talents in RNA Therapeutics. The summit will take place on Thursday, January 26, 2023, at the University of Bern the day before the Swiss RNA Workshop. The program will feature keynotes, panel discussions and company presentations. For more information visit the [Summit's website](#).

### Support grants

Please visit our webpage for more information on the [Lab exchange program](#), the [mobility grants](#) and measures in [equal opportunities](#).

### Upcoming events organized or supported by the NCCR RNA & Disease

#### > [Upcoming Speakers NCCR Seminar Series](#)

- **Clemens Plaschka** (Institute of Molecular Pathology, Vienna, Austria) 31.10.2022 Bern & 1.11.2022 Zurich
- **Mary O'Connell** (Central European Institute of Technology, Brno, Czech Republic) 14.11.2022 Bern & 15.11.2022 Zurich
- **Jeannie Lee** (Harvard Medical School, Cambridge, USA) 28.11.2022 Bern & 29.11.2022 Zurich
- **Clotilde Lagier-Tourenne** (MassGeneral Institute for Neurodegenerative Diseases, Charlestown, USA) 5.12.2022 Bern & 6.12.2022 Zurich
- **Don Cleveland** (University of California, San Diego, USA) 6.3.2023 Bern & 7.3.2023 Zurich
- **Anna Marie Pyle** (Yale University, New Haven, USA) 8.5.2023 Bern & 9.5.2023 Zurich

#### > [1<sup>st</sup> Swiss RNA Therapeutics Summit, January 26, 2023, Bern](#)

#### > [23<sup>rd</sup> Swiss RNA Workshop, January 27, 2023, Bern](#)

### NCCR RNA & Disease Internal Events

- > 7<sup>th</sup> NCCR RNA & Disease Annual Retreat, January 30 – February 1, Kandersteg

### Jobs

#### PhD program in RNA Biology

The next application deadline is December 1, 2022.

Find out more on the [PhD program website](#).

Check the [jobs's section](#) of the NCCR RNA & Disease webpage for other openings.

## IMPRINT

The National Centers of Competence in Research (NCCRs) are a funding scheme of the Swiss National Science Foundation

#### NCCR RNA & Disease

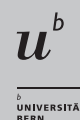
Phone: +41 31 631 38 12  
office@nccr-rna-and-disease.ch  
www.nccr-rna-and-disease.ch

#### Office Bern

University of Bern  
Departement of Chemistry and Biochemistry  
Freiestrasse 3, CH-3012 Bern

#### Office Zürich

ETH Zürich  
Institute of Biochemistry, Biochemie II  
HPP L14, Höggerbergring 64  
CH-8093 Zürich



**ETH** zürich



**Swiss National  
Science Foundation**