Interaktionen und Modifikationen von RNAs und Proteinen

RNA-Protein Interactions II
(Modul 10-202-2208; Spezialvorlesung)

Jörg Fallmann

Institute for Bioinformatics
University of Leipzig

11.05.2018
Protein-centric methods

Purification of RNA-RBP complexes via target protein
Specific purification methods for protein \textit{in vivo}
or way to express a tagged version \textit{in vitro} \rightarrow \text{recombinant protein}

Immunoprecipitation (IP) of the protein via specific antibodies
Most common
Quality and specificity of AB has huge impact on reliability

Co-IP’d RNA is reverse transcribed into cDNA

PCR amplification (NOT POSSIBLE WITH PROTEINS!!!)
Detect interaction partners from less starting material

Sequencing
Protein centric assays
**RBP in vitro assays**

A) SELEX and SEQRS, RNAs undergoes binding and amplification rounds, resulting pools analyzed via sequencing (SELEX) or after each round (SEQRS).

B) RNAcompete assays binding affinity of proteins with designed RNAs on microarray.

C) RNA Bind-n-Seq sequences protein concentration dependent amounts of bound RNAs.
Systematic Evolution of Ligands by EXponential enrichment
Identification of binding motifs

Randomized RNA oligos incubated with RBP of interest
Followed by reverse transcription (RT) of bound RNAs
cDNA is then PCR amplified and in vitro transcribed
Repeat → enrich high-affinity binding sites
Sequencing

SELEX enriches high-affinity motifs

- Functional binding sites with lower affinity?
- No quantitative affinity information for sub-optimal motifs

SEQRS pools are sequenced after each selection
- Gives some information on sub-optimal motifs
Probes binding specificities
Tagged RBP of interest is incubated with pool of ∼40 nt long RNAs
Designed to represent all 9-mers in a compact way

RNA is incubated in excess
  Competition for a limited amount of protein binding sites
  Deduct relative affinity from abundance after single-step selection

Microarray or Sequencing
Estimate binding affinity

protein of interest is *in vitro* expressed
Concentration curve of protein incubated with random RNAs of length 40nt
IP and sequencing

Ratio of protein concentration and bound RNA used to
Determine real dissociation constants ($K_d$)
Infer simple secondary structure preferences
40nt long enough to preserve basic structures
Caveats

\textit{in vitro} \rightarrow \text{need to express protein, not natural levels}
Caveats

*in vitro* → need to express protein, not natural levels

Complex structure constraints can not be detected
→ Oligos too small
Caveats

*in vitro* → need to express protein, not natural levels

Complex structure constraints can not be detected → Oligos too small

RNAcompete oligos are even designed to prevent complex structures, to represent all single-stranded 9-mers in the most compact way
Caveats

*in vitro* → need to express protein, not natural levels

Complex structure constraints can not be detected → Oligos too small

RNAcompete oligos are even designed to prevent complex structures, to represent all single-stranded 9-mers in the most compact way

Only Bind-n-Seq has the potential to be used for RNA secondary structure probing
**in vivo assays**

A) RIP assays bound RNAs after IP

B) CLIP-Seq methods, co-IP of bound RNAs after UV-crosslinking and identification of targets via NGS

C) PAR-CLIP first treats cells with modified U or G nucleoside analogs for higher crosslinking efficiency
For *in vivo* methods, native and denaturating purification methods have to be distinguished.
For "in vivo" methods, native and denaturating purification methods have to be distinguished.

**RNA immunoprecipitation (RIP)**

Native
- Preserves physiological conditions
- Native RP and PP complexes during purification
For *in vivo* methods, native and denaturating purification methods have to be distinguished.

**RNA immunoprecipitation (RIP)**

Native
- Preserves physiological conditions
- Native RP and PP complexes during purification

Be aware
- Protein can interact with RNAs not present in *in vivo* context
- Unspecific interactions with highly abundant RNAs, e.g. rRNAs can mask specific interactions with low-abundancy targets
Crosslink and immunoprecipitation (CLIP)

Denaturing

Crosslinking takes a snapshot of current interactions
Prevents RPIs in non-
\textit{in vivo} manner in later steps of purification
Short wavelength UV light $\rightarrow$ covalent bonds between aromatic AA and RNA in close proximity
Without crosslinking proteins with other proteins
AB purification, denatured in sodiumdodecylsulfate (SDS)
Figure 1: Basic principle of CLIP. Covalent bonds are formed between proximal proteins and RNA upon exposure to ultraviolet light. These bonds only occur at the sites of direct contact and preserve RNA-protein interactions.
<table>
<thead>
<tr>
<th>CLIP and related protocols</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RIP</strong></td>
</tr>
<tr>
<td><strong>CLIP</strong></td>
</tr>
<tr>
<td><strong>Fractionation CLIP</strong></td>
</tr>
<tr>
<td><strong>HITS-CLIP</strong></td>
</tr>
<tr>
<td><strong>CLIP-seq</strong></td>
</tr>
<tr>
<td><strong>CRAC</strong></td>
</tr>
<tr>
<td><strong>PAR-CLIP</strong></td>
</tr>
<tr>
<td><strong>iCLIP</strong></td>
</tr>
<tr>
<td><strong>CLAP</strong></td>
</tr>
<tr>
<td><strong>4SU-iCLIP</strong></td>
</tr>
<tr>
<td><strong>urea-iCLIP</strong></td>
</tr>
<tr>
<td><strong>BrdU CLIP</strong></td>
</tr>
<tr>
<td><strong>FAST-iCLIP</strong></td>
</tr>
<tr>
<td><strong>irCLIP</strong></td>
</tr>
<tr>
<td><strong>eCLIP</strong></td>
</tr>
<tr>
<td><strong>seCLIP</strong></td>
</tr>
<tr>
<td><strong>uvCLAP</strong></td>
</tr>
<tr>
<td><strong>FLASH</strong></td>
</tr>
<tr>
<td><strong>Fr-iCLIP</strong></td>
</tr>
<tr>
<td><strong>sCLIP</strong></td>
</tr>
<tr>
<td><strong>dCLIP</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Further applications of CLIP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CLASH</strong></td>
</tr>
<tr>
<td><strong>hiCLIP</strong></td>
</tr>
<tr>
<td><strong>PAPERCLIP</strong></td>
</tr>
<tr>
<td><strong>cTag-PAPERCLIP</strong></td>
</tr>
<tr>
<td><strong>m5C-miCLIP</strong></td>
</tr>
<tr>
<td><strong>m6A-miCLIP</strong></td>
</tr>
</tbody>
</table>
Figure 1. The Core Steps of iCLIP and Other Variants of CLIP

The majority of currently available CLIP protocols (18 out of 28; Table S1) amplify truncated cDNAs to identify the protein-RNA crosslink sites. Therefore, this schematic follows the core steps of iCLIP, a variant that was developed to amplify truncated cDNAs. The structure of RNA fragments, cDNA inserts, and sequenced reads is marked along with color-coded adapters, unique molecular identifiers (UMIs), experimental barcodes, and primers. The adapters are named as SeqRv and SeqFw according to their conventional orientations relative to the final sequenced reads. Where indicated, variations introduced by other CLIP protocols are illustrated.
A bandwidth of experimental designs are available, each with certain advantages and limitations

Different protocols → different use cases
A bandwidth of experimental designs are available, each with certain advantages and limitations

Different protocols → different use cases

Main ones currently

iCLIP (individual nucleotide resolution)
PAR-CLIP (PhotoActivatable-Ribinucleoside-enhanced)
eCLIP (enhanced)
CLASH (crosslinking, ligation, and sequencing of hybrids)
A bandwidth of experimental designs are available, each with certain advantages and limitations

Different protocols $\rightarrow$ different use cases

Main ones currently

- iCLIP (individual nucleotide resolution)
- PAR-CLIP (PhotoActivatable-Ribinucleoside-enhanced)
- eCLIP (enhanced)
- CLASH (crosslinking, ligation, and sequencing of hybrids)

For dsRNA

- often crosslink poorly
- stringent denaturing plus epitope tagging

Maybe RIP/CLASH is better method
Caveats

PCR can introduce artefacts
Caveats

PCR can introduce artefacts → UMI if possible
Caveats

PCR can introduce artefacts → UMI if possible

CLIP-Seq variants are not bias free
Caveats

PCR can introduce artefacts → UMI if possible

CLIP-Seq variants are not bias free
  certain nt and aa are preferentially crosslinked by UV-light
crosslink efficiency varies between proteins
just as incorporation rate of nt analogs → varies between cell
types, is considered low
Caveats

PCR can introduce artefacts → UMI if possible

CLIP-Seq variants are not bias free
  
  - Certain nt and aa are preferentially crosslinked by UV-light
  - Crosslink efficiency varies between proteins
  - Just as incorporation rate of nt analogs → varies between cell
    types, is considered low

PAR-CLIP → bonds at nt analog → tags enriched at locations
  repeats of that base
Caveats

PCR can introduce artefacts → UMI if possible

CLIP-Seq variants are not bias free
  certain nt and aa are preferentially crosslinked by UV-light
  crosslink efficiency varies between proteins
  just as incorporation rate of nt analogs → varies between cell
types, is considered low

PAR-CLIP → bonds at nt analog → tags enriched at locations
  repeats of that base

cl only at sites where nt and aromatic side chains in close
  proximity → even if nt analog is incorporated, cl only if analog
  is close to actual binding site
Caveats

PCR can introduce artefacts → UMI if possible

CLIP-Seq variants are not bias free
- certain nt and aa are preferentially crosslinked by UV-light
- crosslink efficiency varies between proteins
- just as incorporation rate of nt analogs → varies between cell types, is considered low

PAR-CLIP → bonds at nt analog → tags enriched at locations repeats of that base
- cl only at sites where nt and aromatic side chains in close proximity → even if nt analog is incorporated, cl only if analog is close to actual binding site

Conceptual problem if interacting amino-acid side chains not aromatic → can not be crosslinked → not detectable by CLIP-Seq
No matter which method, output are reads
No matter which method, output are reads

RIP reads \(\rightarrow\) whole target sequence, unless breaks

CLIP reads \(\rightarrow\) region around/downstream of crosslink
No matter which method, output are reads

RIP reads $\rightarrow$ whole target sequence, unless breaks
CLIP reads $\rightarrow$ region around/downstream of crosslink

So how would you go one from there?
How to analyze this?
How to analyze this?

RIP $\rightarrow$ what could be target, low resolution

CLIP $\rightarrow$ info on (exact) site of interaction
How to analyze this?

RIP → what could be target, low resolution
CLIP → info on (exact) site of interaction

RT misreads crosslink, or drops off completely
→ WE USE THIS
How to analyze this?

RIP $\rightarrow$ what could be target, low resolution
CLIP $\rightarrow$ info on (exact) site of interaction

RT misreads crosslink, or drops off completely $\rightarrow$ WE USE THIS

PAR-CLIP $\rightarrow$ introduced nucleotide analogs (e.g. thio-uridine) misinterpreted as guanines by RT $\rightarrow$ T2C transitions in reads $\rightarrow$ pinpoint interaction sites
How to analyze this?

RIP → what could be target, low resolution
CLIP → info on (exact) site of interaction

RT misreads crosslink, or drops off completely
→ WE USE THIS

PAR-CLIP → introduced nucleotide analogs (e.g. thio-uridine)
misinterpreted as guanines by RT → T2C transitions in reads
→ pinpoint interaction sites

iCLIP → amino acid tag at cl causes termination of RT
→ pinpoint interaction site
How to analyze this?

RIP → what could be target, low resolution

CLIP → info on (exact) site of interaction

RT misreads crosslink, or drops off completely
   → WE USE THIS

PAR-CLIP → introduced nucleotide analogs (e.g. thio-uridine)
misinterpreted as guanines by RT → T2C transitions in reads
   → pinpoint interaction sites

iCLIP → amino acid tag at cl causes termination of RT
   → pinpoint interaction site

eCLIP → iCLIP with size matched input → control background binding
Challenges

Depending on CLIP technique used (iCLIP, HITS-CLIP, Par-CLIP etc.), downstream analysis requires specific algorithms to filter signal from noise.
Challenges

Depending on CLIP technique used (iCLIP, HITS-CLIP, Par-CLIP etc.), downstream analysis requires specific algorithms to filter signal from noise

- Mutations can be used to identify interaction sites
- Transition(rate)s can be used to distinguish signal from noise
Challenges

Depending on CLIP technique used (iCLIP, HITS-CLIP, Par-CLIP etc.), downstream analysis requires specific algorithms to filter signal from noise.

- Mutations can be used to identify interaction sites.
- Transition(rates) can be used to distinguish signal from noise.

CLIP-Seq signal is a qualitative measure for RBP targets with high resolution.
Challenges

Depending on CLIP technique used (iCLIP, HITS-CLIP, Par-CLIP etc.), downstream analysis requires specific algorithms to filter signal from noise

- Mutations can be used to identify interaction sites
- Transition(rates) can be used to distinguish signal from noise

CLIP-Seq signal is a qualitative measure for RBP targets with high resolution

Quantitative measure only for the relative amount of protein titert by it

- Indicates which RNAs are targets and which are not
- No quantitative measure of binding strength or affinity
CLIP-Seq peak finding and normalization  A) Regions with enriched signal (crosslink events) are filtered from background with peak finder algorithms. B) CLIP-Seq signal of such regions depends on the amount of available transcript and total signal over transcript as well as transcript abundance can be used for normalization.
Analysis

Identify true binding sites by filtering spurious and unspecific binding
Analysis

Identify true binding sites by filtering spurious and unspecific binding

Use to identify binding motifs, structures
Analysis

Identify true binding sites by filtering spurious and unspecific binding

Use to identify binding motifs, structures

The latter can then be used for binding site predictions, given that their quality is good enough and that the protein of interest has binding preferences
Defining binding sites (BS) from CLIP-Seq experiments

Major challenge in BS prediction is missing negative control
Defining binding sites (BS) from CLIP-Seq experiments

Major challenge in BS prediction is missing negative control
Without negative control → come up with measure to distinguish true binding from background binding
Defining binding sites (BS) from CLIP-Seq experiments

Major challenge in BS prediction is missing negative control
Without negative control → come up with measure to distinguish true binding from background binding
Algorithms work on read counts in defined genomic regions or sequence stretches derived from data directly
Defining binding sites (BS) from CLIP-Seq experiments

Major challenge in BS prediction is missing negative control
Without negative control → come up with measure to distinguish true binding from background binding
Algorithms work on read counts in defined genomic regions or sequence stretches derived from data directly

Straight forward way to distinguish real binding from noise is random distribution of reads over defined region (e.g. the gene body) → calculate the probability for finding the read density observed in the experiment
Defining binding sites (BS) from CLIP-Seq experiments

Major challenge in BS prediction is missing negative control
Without negative control → come up with measure to distinguish true binding from background binding
Algorithms work on read counts in defined genomic regions or sequence stretches derived from data directly

Straight forward way to distinguish real binding from noise is random distribution of reads over defined region (e.g. the gene body) → calculate the probability for finding the read density observed in the experiment

P-values for peak regions and enrichment values between theoretical/experimental signal
More on this topic

Nr. of tools for peak detection/CLIP analysis is growing

Remaining challenges

Elimination of background from CLIP-Seq experiments
High signal does not automatically indicate strong binding and vice versa
Some regions tend to show high signal across conditions and protein of interest → suggests background binding
One might miss important binding sites with low signal due to low expression of target sites

Adequate experimental quality will always be of the essence for successful CLIP-Seq analysis
<table>
<thead>
<tr>
<th>TOOL</th>
<th>YEAR</th>
<th>EXPERIMENT</th>
<th>FOCUS</th>
<th>MAIN ADVANTAGE</th>
<th>RECOMMENDED CASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paralyzer</td>
<td>2011</td>
<td>PAR-CLIP</td>
<td>Peak detection</td>
<td>Exploits T to C mutations to Improve Signal to noise ratio</td>
<td>PAR-CLIP data</td>
</tr>
<tr>
<td>wavClusteR</td>
<td>2012</td>
<td>PAR-CLIP (BAM format)</td>
<td>Noise and false positives reduction Peak detection</td>
<td>Distinguishes between non-experimentally and experimentally induced transitions</td>
<td>PAR-CLIP data</td>
</tr>
<tr>
<td>Piranha</td>
<td>2012</td>
<td>CLIP-seq and RIP-seq (BED or BAM)</td>
<td>Noise and false positives reduction Peak detection CLIP-seq data comparison [correction for transcript abundance]</td>
<td>Corrects the reads dependence on transcript abundance</td>
<td>CLIP-seq and Transcript abundance data</td>
</tr>
<tr>
<td>mCarts</td>
<td>2013</td>
<td>CLIP-seq</td>
<td>Sites prediction on different samples</td>
<td>Considers accessibility in local RNA secondary structures and cross-species conservation</td>
<td>RBP motif</td>
</tr>
<tr>
<td>PIPE-CLIP</td>
<td>2014</td>
<td>CLIP-seq (SAM or BAM)</td>
<td>Noise and false positives reduction Statistical assessment Peak detection</td>
<td>Provides a significance level for each identified candidate binding site</td>
<td>HITS-CLIP, iCLIP</td>
</tr>
<tr>
<td>GraphProt</td>
<td>2014</td>
<td>CLIP-seq and RNAcompete</td>
<td>Peak detection Sites prediction on different samples</td>
<td>Detects RBP motif secondary structure common characteristics. It estimates binding affinities</td>
<td>RBP motifs that are NOT located within single-stranded regions</td>
</tr>
<tr>
<td>CLIPper</td>
<td>2016</td>
<td>eCLIP-seq</td>
<td>Peak detection from eCLIP data</td>
<td>Models background binding</td>
<td>eCLIP</td>
</tr>
</tbody>
</table>
Binding motif prediction

Search for preferred binding motif is routine task → identification of a motif non-trivial
Search for preferred binding motif is routine task → identification of a motif non-trivial

Motif finding → the problem of discovering motifs without prior knowledge of how the motifs look
Binding motif prediction

Search for preferred binding motif is routine task $\rightarrow$ identification of a motif non-trivial

Motif finding $\rightarrow$ the problem of discovering motifs without prior knowledge of how the motifs look

Given set of sequences, find subsequences that occur more often than expected $\rightarrow$ over-represented
Binding motif prediction

Search for preferred binding motif is routine task → identification of a motif non-trivial

Motif finding → the problem of discovering motifs without prior knowledge of how the motifs look

Given set of sequences, find subsequences that occur more often than expected → over-represented

Motif of interest will occur in many input sequences and can in principle be found by aligning the input sequences and searching for conserved regions
Binding motif prediction

Search for preferred binding motif is routine task $\rightarrow$ identification of a motif non-trivial

Motif finding $\rightarrow$ the problem of discovering motifs without prior knowledge of how the motifs look

Given set of sequences, find subsequences that occur more often than expected $\rightarrow$ over-represented

Motif of interest will occur in many input sequences and can in principle be found by aligning the input sequences and searching for conserved regions

Motifs do not have to be fully conserved, and they can even consists of sub-motifs themselves, or at least show some variability in their nucleotide content
Alignments can be used to generate Position Weight Matrices (PWM) → assign each position in a sequence a probability for containing a certain nucleotide
Alignments can be used to generate Position Weight Matrices (PWM) → assign each position in a sequence a probability for containing a certain nucleotide

From such a PWM, the frequency of a given motif in the input can be computed and compared to the background frequency (e.g. number of motifs in genes), such that a score for over-representation is derived
Alignments can be used to generate Position Weight Matrices (PWM) → assign each position in a sequence a probability for containing a certain nucleotide

From such a PWM, the frequency of a given motif in the input can be computed and compared to the background frequency (e.g. number of motifs in genes), such that a score for over-representation is derived

MEME is the most widely used algorithm for this task → Expectation maximization (EM) algorithm to find the most over-represented motifs in a set of sequences
Is structure important?

Overview of RNA secondary structure elements Loop types that occur in RNA molecules and are distinguished by *in silico* structure prediction algorithms due to their differing thermodynamic effects. One distinguishes stem loops, hairpin loops, multi loops, bulges, interior loops and exterior loops.
Improve motif prediction

RBP binding motifs can be predicted by DNA motif finders
RBP binding motifs can be predicted by DNA motif finders. Most RBPs are thought to prefer single stranded RNA (ssRNA) regions for interaction.
Improve motif prediction

RBP binding motifs can be predicted by DNA motif finders.
Most RBPs are thought to prefer single stranded RNA (ssRNA) regions for interaction.
So what can we do to improve motif prediction?
RBP binding motifs can be predicted by DNA motif finders. Most RBPs are thought to prefer single stranded RNA (ssRNA) regions for interaction. So what can we do to improve motif prediction?

Include accessibility of binding sites.
Accessibility

RNA must be accessible for most RBPs to interact → most likely secondary structure is less important than accessibility derived from ensemble of structures.
Accessibility

RNA must be accessible for most RBPs to interact → most likely secondary structure is less important than accessibility derived from ensemble of structures

McCaskill algorithm → exhaustive calculation of bp probabilities from Bolzmann ensemble of structures in thermal equilibrium
RNA must be accessible for most RBPs to interact → most likely secondary structure is less important than accessibility derived from ensemble of structures

McCaskill algorithm → exhaustive calculation of bp probabilities from Boltzmann ensemble of structures in thermal equilibrium

Be aware → such predictions are made on local rather than global scale → they are very context-sensitive

When analyzing e.g. CLIP-Seq target sites, length of the surrounding region one selects for folding has strong impact on the results
<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Input</th>
<th>Type of motif generated</th>
<th>Considers secondary structure?</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEME</td>
<td>Positive (and optionally, negative)</td>
<td>PWM</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>sequences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PhyloGibbs</td>
<td>Positive (and optionally, negative)</td>
<td>PWM</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>sequences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cERMIT</td>
<td>Rank ordered sequences</td>
<td>PWM</td>
<td>No</td>
</tr>
<tr>
<td>DRIMUST</td>
<td>Rank ordered sequences</td>
<td>IUPAC motif, possibly gapped</td>
<td>No</td>
</tr>
<tr>
<td>StructuRED</td>
<td>Positive and negative sequences</td>
<td>PWM in a hairpin loop</td>
<td>Yes, considers possible hairpin loops up to 7 bases with at least 3 paired bases</td>
</tr>
<tr>
<td>TEISER</td>
<td>Sequences and scores (e.g., stability</td>
<td>PWM in a hairpin loop</td>
<td>Yes, considers possible hairpin loops with stems 4-7 bases long and loop sizes of 4-9 bases</td>
</tr>
<tr>
<td></td>
<td>scores)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNAcontext</td>
<td>Sequences and affinity scores</td>
<td>PWM with structural context scores</td>
<td>Yes, learns the preferred structural context of each base in a motif</td>
</tr>
<tr>
<td>GraphProt</td>
<td>Positive and negative sequences</td>
<td>graph-based sequence and structure motifs,</td>
<td>Yes, models RNA structure using a graph-based encoding</td>
</tr>
<tr>
<td></td>
<td></td>
<td>can be visualized with logos</td>
<td></td>
</tr>
<tr>
<td>CMfinder</td>
<td>Positive sequences</td>
<td>structured sequence</td>
<td>Yes, SCFG-based, examines the most stable structures in the input</td>
</tr>
<tr>
<td>RNApromo</td>
<td>Positive sequences</td>
<td>structured sequence</td>
<td>Yes, SCFG-based, optimizes a motif from an initial set of substructures generated from the input</td>
</tr>
<tr>
<td>#ATS</td>
<td>Positive and negative sequences</td>
<td>IUPAC</td>
<td>Yes, scores candidate binding sites by accessibility</td>
</tr>
<tr>
<td>MEMERIS</td>
<td>Positive and negative sequences</td>
<td>PWM</td>
<td>Yes, uses accessibility as prior knowledge to guide motif finding toward single-stranded regions</td>
</tr>
</tbody>
</table>
Take home

Secondary structure influences binding potential
→ binding influences structure ensemble

Inaccessible BS require energy to unfold
→ binder can prevent structures from forming or provide the energy needed to form it

_in vivo_ RNA is in (constant) contact with binders (e.g. proteins, miRNAs, ligands, etc.)
→ all influence and are influenced by structure ensemble

_in silico_ methods to predict structures, also under constraints of interaction

There is not THE right way to analyze
Always depends on your data and the experimental context
Interesting reads


Stefanie Gerstberger et al. (Nov. 2014). “A census of human RNA-binding proteins”. In: Nature Reviews Genetics 15.12, pp. 829–845. ISSN: 1471-0056, 1471-0064. DOI: 10.1038/nrg3813. URL: http://www.nature.com/doifinder/10.1038/nrg3813 (visited on 11/24/2015)