Interaktionen und Modifikationen von RNAs und Proteinen
RNA-Protein Interactions II

(Modul 10-202-2208; Spezialvorlesung)

Jörg Fallmann

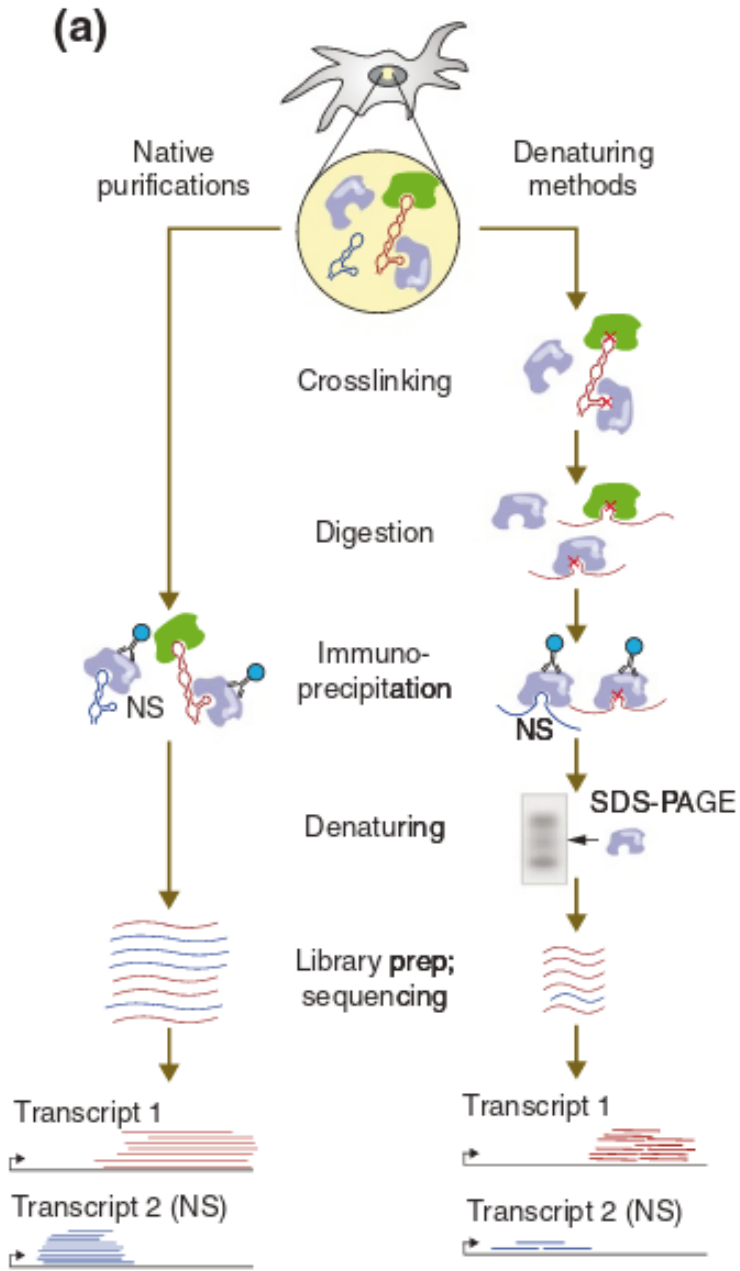
Institute for Bioinformatics
University of Leipzig

11.05.2018

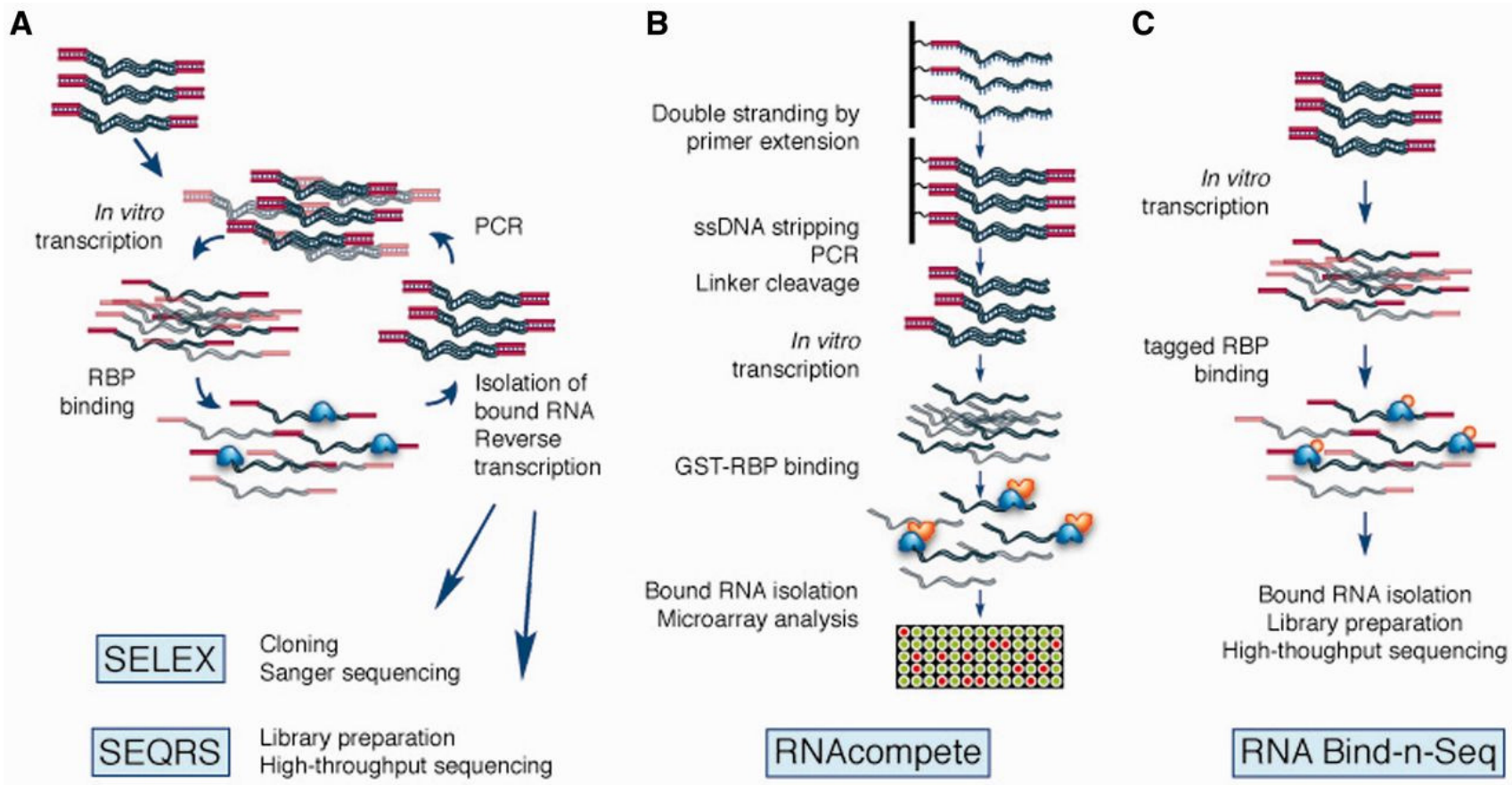
Protein-centric methods

- Purification of RNA-RBP complexes via target protein
  - Specific purification methods for protein *in vivo*
    or way to express a tagged version *in vitro* $\rightarrow$ 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



in vitro 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

SELEX

- 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 $\rightarrow$ 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

RNAcompete

- Probes binding specificities
  - Tagged RBP of interest is incubated with
    pool of $\sim$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

RNA-bind'n-seq

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

in vitro $\rightarrow$ need to express protein, not natural levels

Caveats

in vitro $\rightarrow$ need to express protein, not natural levels

Complex structure constraints can not be detected
$\rightarrow$ Oligos too small

Caveats

in vitro $\rightarrow$ need to express protein, not natural levels

Complex structure constraints can not be detected
$\rightarrow$ 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

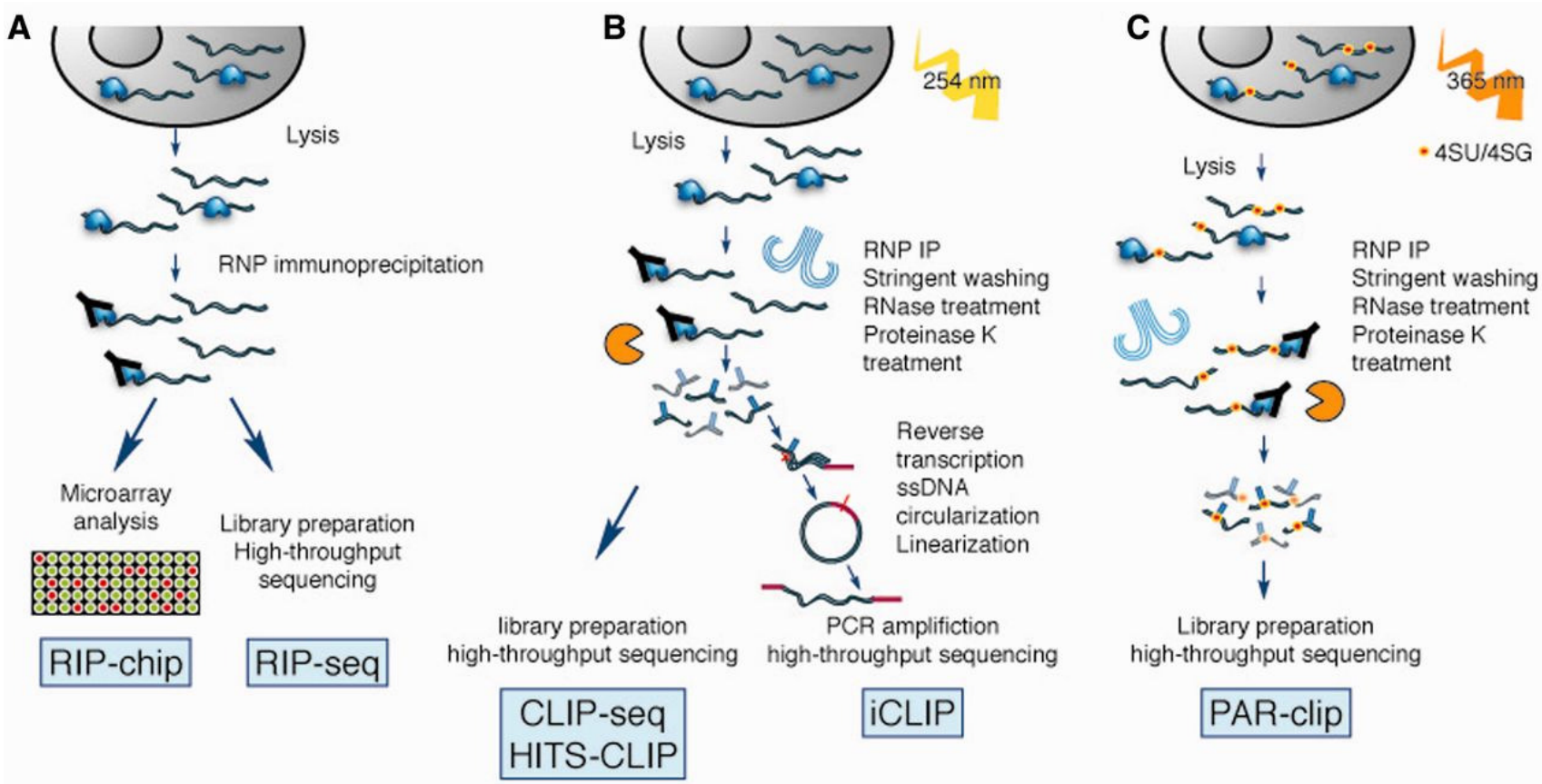
in vitro $\rightarrow$ need to express protein, not natural levels

Complex structure constraints can not be detected
$\rightarrow$ 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



RBP 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

IP-methods *in vivo* I RIP

For *in vivo* methods, native and denaturating purification methods have to be distinguished

IP-methods *in vivo* I RIP

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

IP-methods *in vivo* I RIP

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

IP-methods *in vivo* II CLIP

- Crosslink and immunoprecipitation (CLIP)
  - Denaturing
    - Crosslinking takes a snapshot of current interactions
    - Prevents RPIs in non-*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)

Basic Principle CLIP

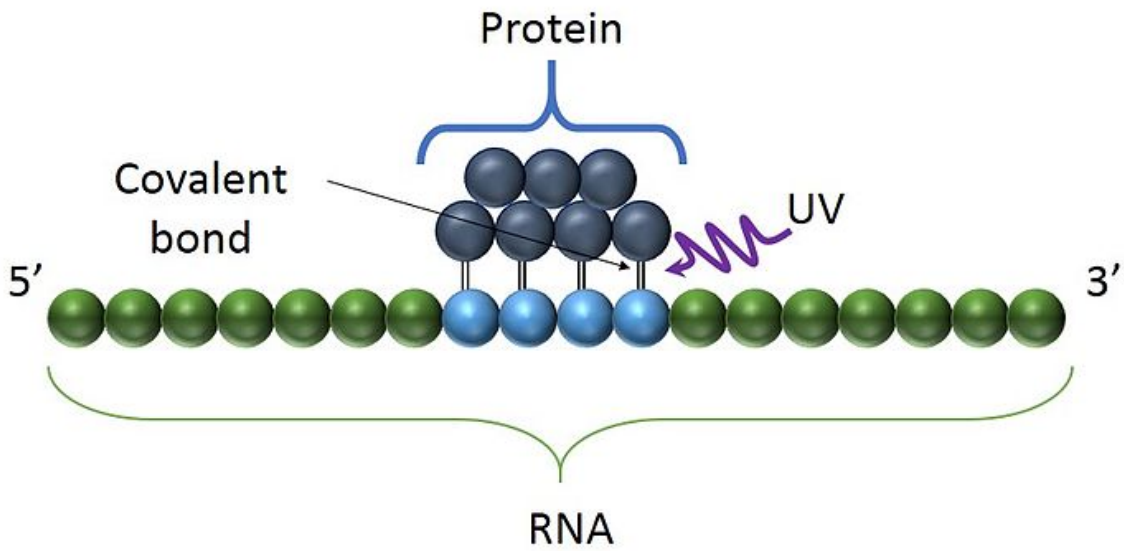


Figure 1: Basic principle of CLIP. Covalent bonds are formed between proximal proteins and RNA upon exposure to ultraviolet light. These bonds only occur at the sites of direct contact and preserve RNA-protein interactions.

CLIP and related protocols		
RIP	RNA immunoprecipitation	Lerner and Steitz, 1979
CLIP	(UV) Crosslinking and immunoprecipitation	Ule et al., 2003, 2005
Fractionation CLIP	CLIP from nucleus, cytosol. and polysomes	Sanford et al., 2008
HITS-CLIP	High-throughput sequencing of RNA isolated by CLIP	Licatalosi et al., 2008; Chi et al., 2009
CLIP-seq	CLIP coupled with high-throughput sequencing	Yeo et al., 2009
CRAC	UV crosslinking and analysis of cDNAs	Granneman et al., 2009
PAR-CLIP	Photoactivable ribonucleoside-enhanced CLIP	Hafner et al., 2010; Garzia et al., 2017
iCLIP	Individual-nucleotide resolution CLIP	König et al., 2010
CLAP	Crosslinking and affinity purification	Wang et al., 2010
4SU-iCLIP	4SU-mediated crosslinking followed by iCLIP	Huppertz et al., 2014
urea-iCLIP	iCLIP with denaturing purification	Huppertz et al., 2014
BrdU CLIP	Bromodeoxyuridine UV CLIP	Weyn-Vanhentenryck et al., 2014
FAST-iCLIP	Fully automated and standardized iCLIP	Flynn et al., 2015
irCLIP	Infrared-CLIP	Zarnegar et al., 2016
eCLIP	Enhanced CLIP	Van Nostrand et al., 2016
seCLIP	Single-end eCLIP	Van Nostrand et al., 2017c
uvCLAP	UV crosslinking and affinity purification	Aktaş et al., 2017
FLASH	Fast ligation of RNA after some sort of affinity purification for high-throughput sequencing	Aktaş et al., 2017
Fr-iCLIP	Fractionation iCLIP	Brugiolo et al., 2017
sCLIP	Simplified CLIP	Kargapolova et al., 2017
dCLIP	Denaturing CLIP	Rosenberg et al., 2017
Further applications of CLIP		
CLASH	Crosslinking, ligation, and sequencing of hybrids	Kudla et al., 2011
hiCLIP	RNA hybrid and iCLIP	Sugimoto et al., 2015
PAPERCLIP	Poly(A) binding protein-mediated mRNA 3′ end retrieval by CLIP	Hwang et al., 2016
cTag-PAPERCLIP	"Conditionally" tagged PAPERCLIP	Hwang et al., 2017
m5C-miCLIP	Cytosine-5 methylation iCLIP	Hussain et al., 2013
m6A-miCLIP	N6-methyladenosine iCLIP	Linder et al., 2015

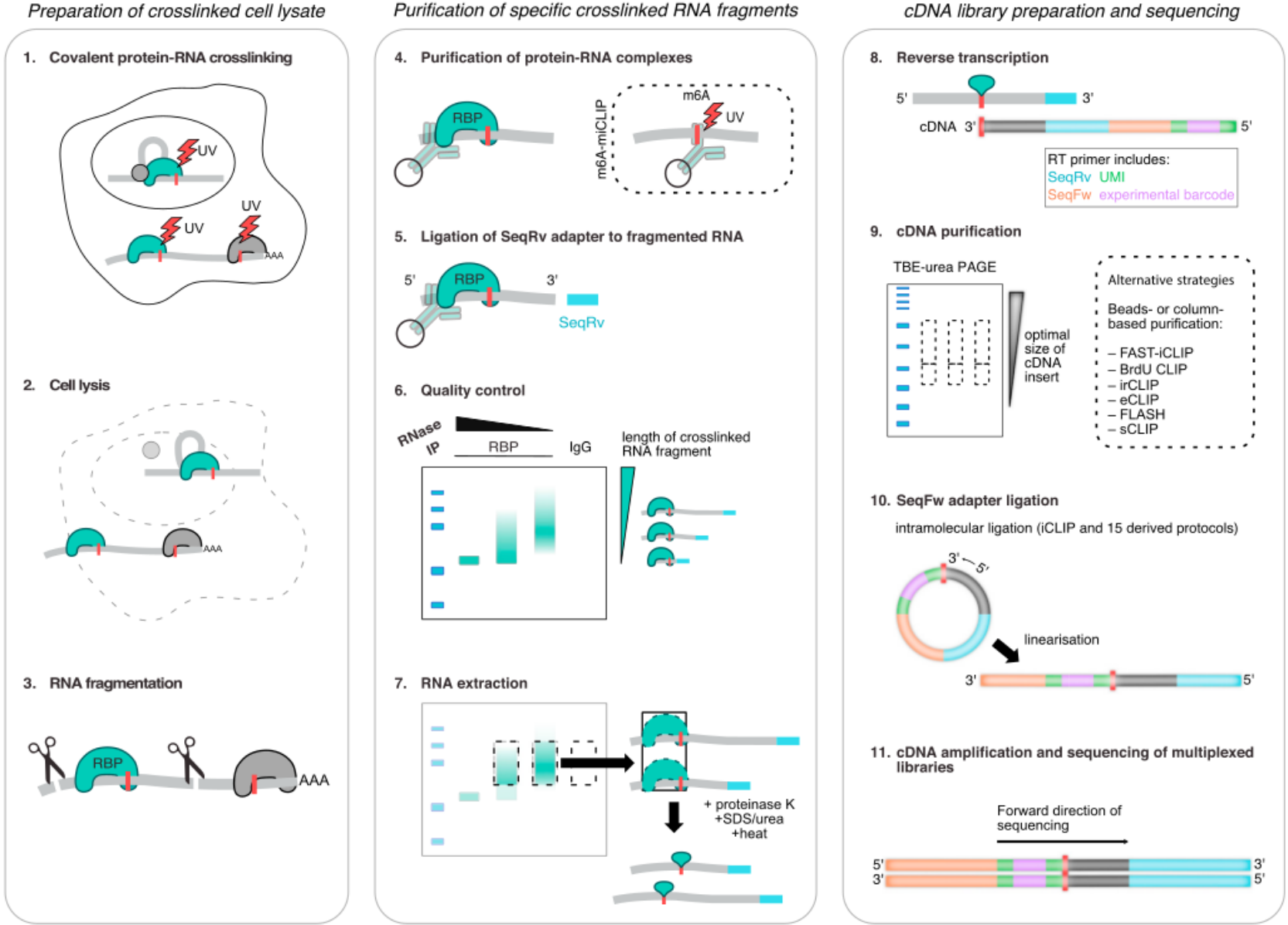


Figure 1. The Core Steps of iCLIP and Other Variants of CLIP

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.

CLIP

A bandwidth of experimental designs are available, each with certain advantages and limitations

Different protocols $\rightarrow$ different use cases

CLIP

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)

CLIP

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 $\rightarrow$ UMI if possible

Caveats

PCR can introduce artefacts $\rightarrow$ UMI if possible

CLIP-Seq variants are not bias free

Caveats

PCR can introduce artefacts $\rightarrow$ 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 $\rightarrow$ varies between cell types, is considered low

Caveats

PCR can introduce artefacts $\rightarrow$ 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 $\rightarrow$ varies between cell types, is considered low

PAR-CLIP $\rightarrow$ bonds at nt analog $\rightarrow$ tags enriched at locations repeats of that base

Caveats

PCR can introduce artefacts $\rightarrow$ 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 $\rightarrow$ varies between cell types, is considered low

PAR-CLIP $\rightarrow$ bonds at nt analog $\rightarrow$ tags enriched at locations repeats of that base

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

Caveats

PCR can introduce artefacts $\rightarrow$ 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 $\rightarrow$ varies between cell types, is considered low

PAR-CLIP $\rightarrow$ bonds at nt analog $\rightarrow$ tags enriched at locations repeats of that base

cl only at sites where nt and aromatic side chains in close proximity $\rightarrow$ 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 $\rightarrow$ can not be crosslinked $\rightarrow$ not detectable by CLIP-Seq

Bioinformatics

No matter which method, output are reads

Bioinformatics

No matter which method, output are reads

RIP reads $\rightarrow$ whole target sequence, unless breaks

CLIP reads $\rightarrow$ region around/downstream of crosslink

Bioinformatics

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 $\rightarrow$ 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

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 $\rightarrow$ 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

iCLIP $\rightarrow$ amino acid tag at cl causes termination of RT
$\rightarrow$ pinpoint interaction site

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

iCLIP $\rightarrow$ amino acid tag at cl causes termination of RT
$\rightarrow$ pinpoint interaction site

eCLIP $\rightarrow$ iCLIP with size matched input $\rightarrow$ 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(rate)s 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

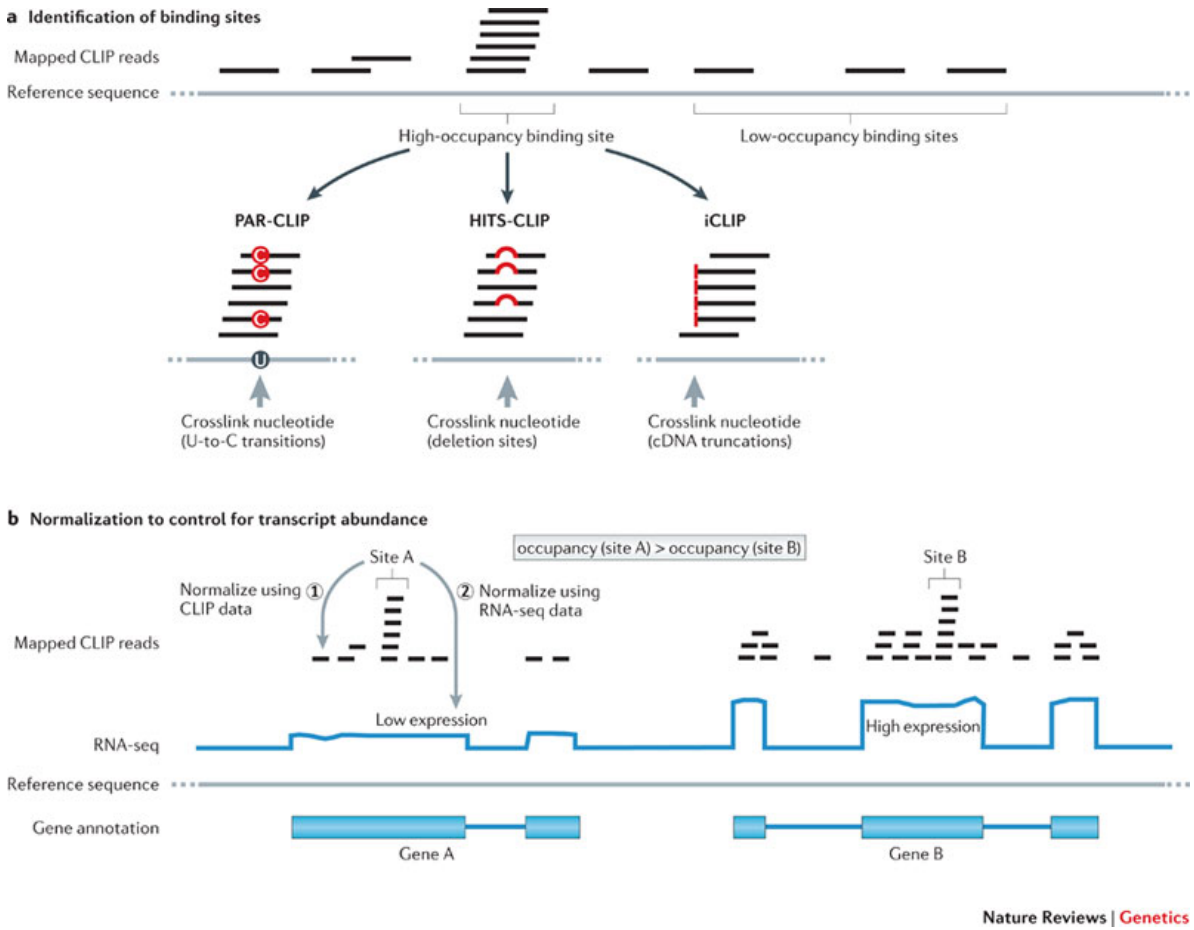
- Mutations can be used to identify interaction sites
- Transition(rate)s can be used to distinguish signal from noise

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

Analysis



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 $\rightarrow$ 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 $\rightarrow$ 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 $\rightarrow$ 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) $\rightarrow$ 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 $\rightarrow$ 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) $\rightarrow$ 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 $\rightarrow$ 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

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

Binding motif prediction

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

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

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 $\rightarrow$ 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

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

PWM

Alignments can be used to generate Position Weight Matrices (PWM) $\rightarrow$ assign each position in a sequence a probability for containing a certain nucleotide

PWM

Alignments can be used to generate Position Weight Matrices (PWM) $\rightarrow$ 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

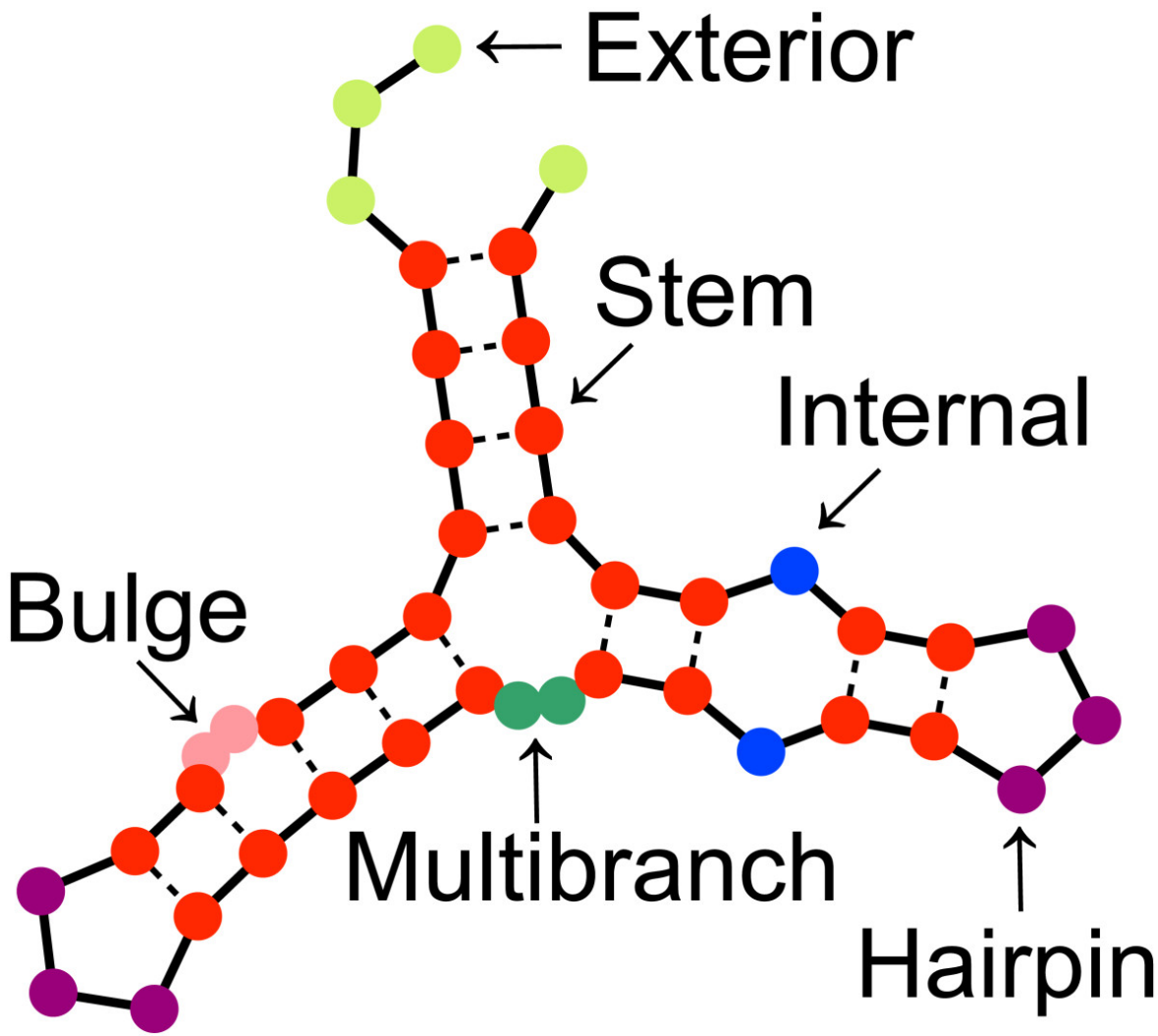
PWM

Alignments can be used to generate Position Weight Matrices (PWM) $\rightarrow$ 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
$\rightarrow$ 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

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

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?

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 $\rightarrow$ most likely secondary structure is less important than accessibility derived from ensemble of structures

Accessibility

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

McCaskill algorithm $\rightarrow$ exhaustive calculation of bp probabilities from Bolzmann ensemble of structures in thermal equilibrium

Accessibility

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

McCaskill algorithm $\rightarrow$ exhaustive calculation of bp probabilities from Bolzmann ensemble of structures in thermal equilibrium

Be aware $\rightarrow$ such predictions are made on local rather than global scale $\rightarrow$ 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

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

Take home

Secondary structure influences binding potential
$\rightarrow$ binding influences structure ensemble

Inaccessible BS require energy to unfold
$\rightarrow$ 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.)
$\rightarrow$ 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

Flora C.Y. Lee and Jernej Ule (Feb. 2018). "Advances in CLIP Technologies for Studies of Protein-RNA Interactions". en. In: *Molecular Cell* 69.3, pp. 354–369. ISSN: 10972765. DOI: 10.1016/j.molcel.2018.01.005. URL: http://linkinghub.elsevier.com/retrieve/pii/S1097276518300054 (visited on 04/20/2018)

Matthias W. Hentze et al. (Jan. 2018). "A brave new world of RNA-binding proteins". en. In: *Nature Reviews Molecular Cell Biology* 19.5, pp. 327–341. ISSN: 1471-0072, 1471-0080. DOI: 10.1038/nrm.2017.130. URL: http://www.nature.com/doifinder/10.1038/nrm.2017.130 (visited on 05/07/2018)

K. B. Cook et al. (Jan. 2015). "High-throughput characterization of protein-RNA interactions". en. In: *Briefings in Functional Genomics* 14.1, pp. 74–89. ISSN: 2041-2649, 2041-2657. DOI: 10.1093/bfgp/elu047. URL: http://bfgp.oxfordjournals.org/cgi/doi/10.1093/bfgp/elu047 (visited on 11/23/2015)

Julian König et al. (2011). "Protein-RNA interactions: new genomic technologies and perspectives". In: *Nature Reviews Genetics* 13.February, pp. 77–83. ISSN: 1471-0056. DOI: 10.1038/nrg3141. URL: http://www.nature.com/nrg/journal/v13/n2/abs/nrg3141.htmlhttp://discovery.ucl.ac.uk/1344586/

Colleen A. McHugh et al. (2014). "Methods for comprehensive experimental identification of RNA-protein interactions". In: *Genome Biol* 15, p. 203. URL: http://www.biomedcentral.com/content/pdf/gb4152.pdf (visited on 11/23/2015)

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)