Interaktionen von RNAs und Proteinen

Sonja Prohaska

Computational EvoDevo
Universitaet Leipzig

May 4, 2017

Studying RNA-protein interactions

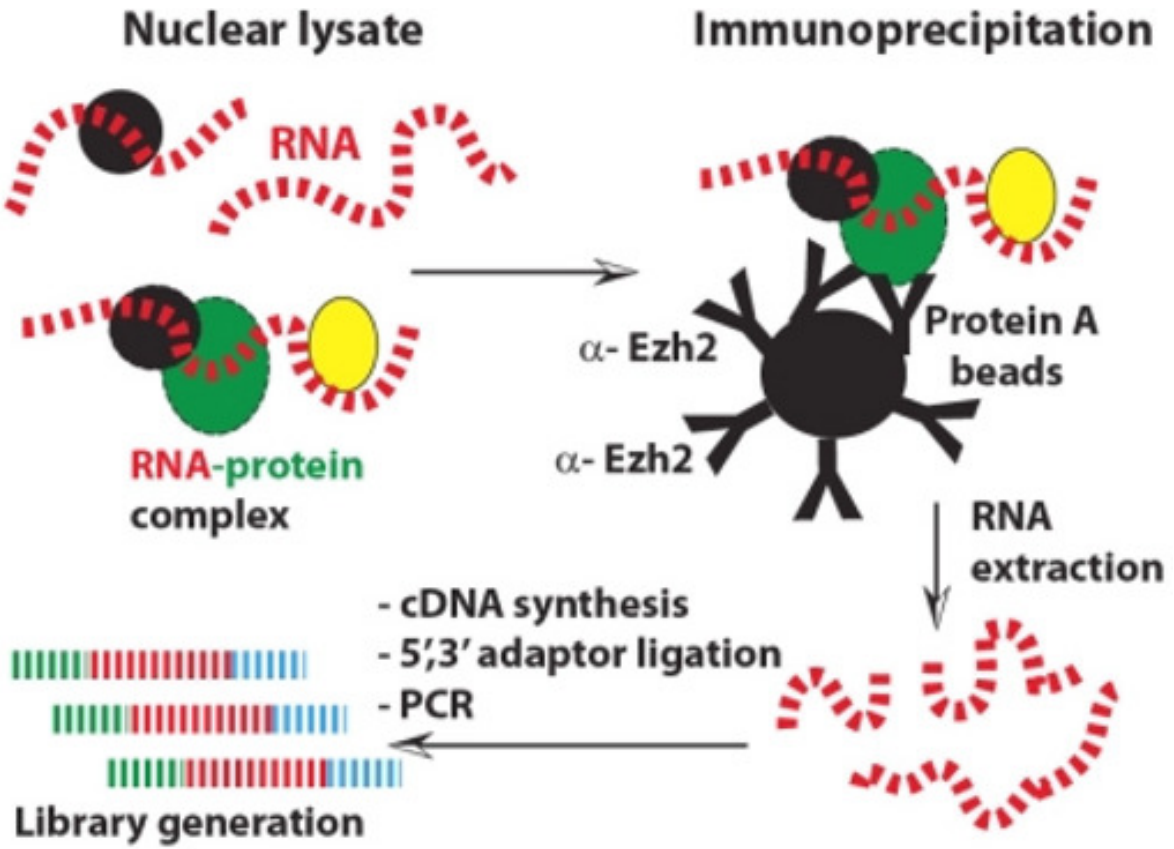
- **Given**: target protein known to bind to RNA
- **problem**: find binding partners and binding sites

experimental high-throughput techniques

- based on immunoprecipitation (IP)
  use antibodies against target protein
- **RIP-seq**: full RNAs associated with protein
- **iCLIP** and **HITS-CLIP**: RNA binding sites of proteins
- **PAR-CLIP**: nucleotide resolution RNA-protein interactions

Find RNAs bound by target protein

RIP = **R**NA **I**mmuno-**P**recipitation



key feature: no crosslinking

problem: may yield indirect interactions

Analysis of RIP-seq data

what we want: full sequences of the protein-bound RNAs (all!)
what we get: reads

task for the bioinformatician

- as in transcriptome sequencing
- (polish reads: remove adapters and barcodes)
- **map reads** to the genome/transcriptome
  (allow for indels and spliced-reads)
- **cluster reads**
- compare location of clusters to genome annotation
- identify the bound RNA (if possible)
- difficulty: alternative transcripts

CLIP-seq methods – general idea

CLIP = **C**ross-**L**inking **I**mmuno-**P**recipitation

- **UV light** is used to **crosslink** RNA and protein *in vivo*
- **stringent purification**: immunoprecipitation, SDS-PAGE, transfer to nitrocellulose
- proteinase K digests protein but
- leaves **1-2 amino acids at the UV crosslinked sites**
- reverse transcriptase (RT) makes cDNA
- cDNA often truncates at UV crosslinked sites
- ... methods differ here ...
- sequencing
- bioinformatic analysis

CLIP-seq methods – general idea

Basic principle of all CLIP methods

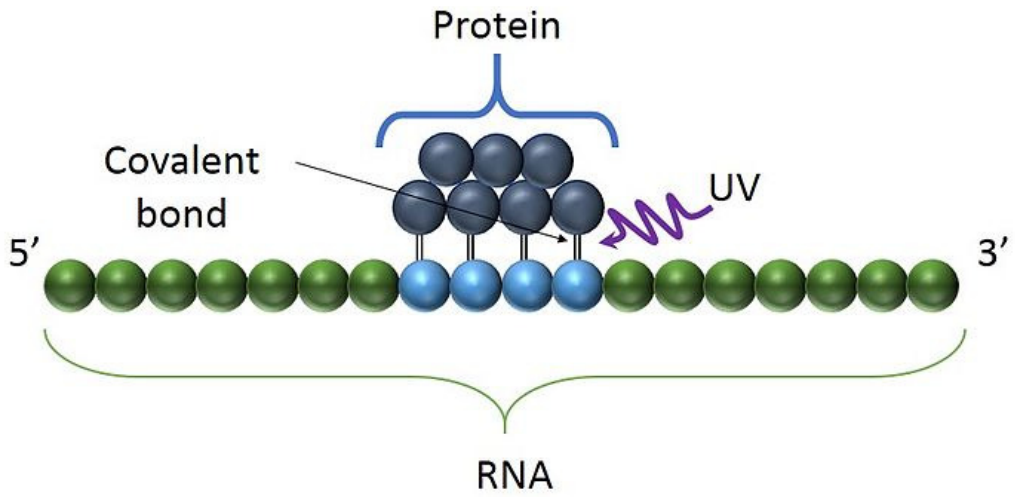


Figure 1: Basic principle of CLIP. Covalent bonds are formed between proximal proteins and RNA upon exposure to ultraviolet light. These bonds only occur at the sites of direct contact and preserve RNA-protein interactions.

Find RNA binding sites of target protein

HITS-CLIP = **HI**gh-**T**hroughput **S**equencing of RNA isolated by **C**ross-**L**inking **I**mmuno-**P**recipitation

- 254nm UV crosslinking *in vivo*
- cell lysis
- requires $3'$- and $5'$-ligated adapters (for amplification)
- revers transcriptase (RT) makes cDNA
- stalling of the RT at UV crosslink sites $\rightarrow$ no amplification ($5'$-adapter missing)
- read through at UV crosslink sites $\rightarrow$ **induces mutations**

- **key feature**: crosslink-induced mutation site (CIMS)
- **Problem**: high sequencing depths required
- **Problem**: destinguish CIMS from technical errors

Find RNA binding sites of target protein

HITS-CLIP = **HI**gh-**T**hroughput **S**equencing of RNA isolated by **C**ross-**L**inking **I**mmuno-**P**recipitation

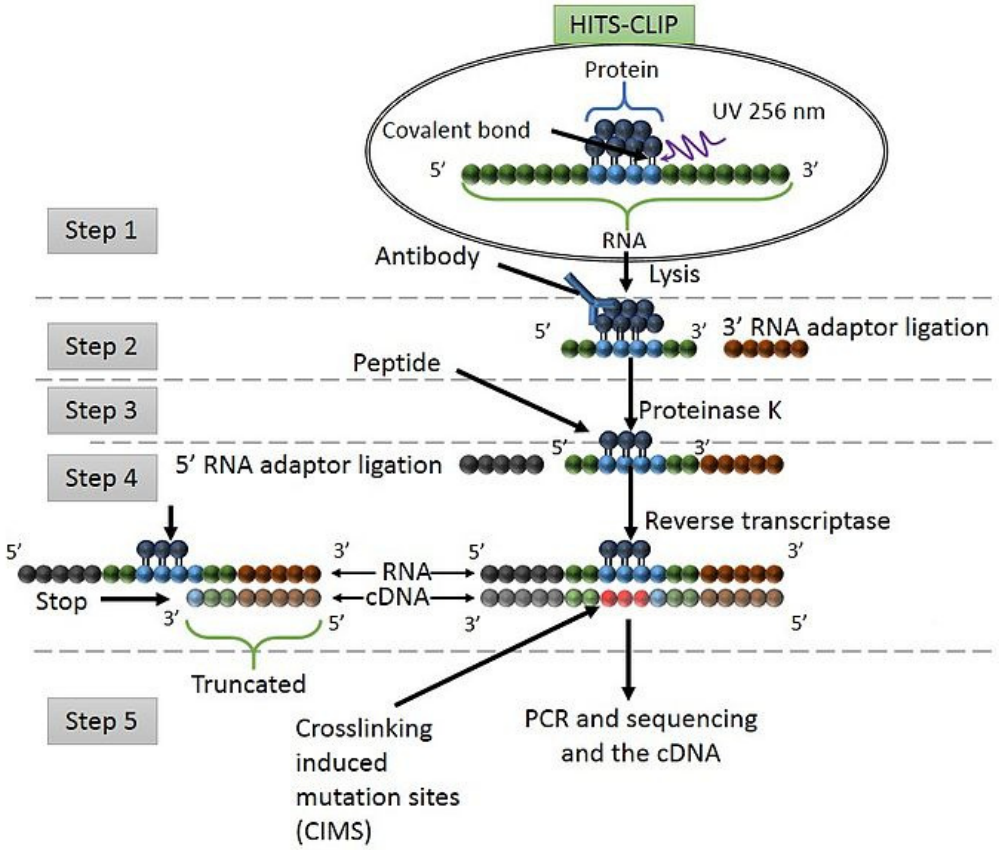


Figure 2: HITS-CLIP

Step 1
HITS-CLIP begins with the in-vivo cross-linking of RNA-protein complexes using ultraviolet light. The cell is lysed and the protein of interest is isolated using immunoprecipitation.

Step 2
Washing is performed to remove free RNA, and RNA adaptors are ligated at the 3′ ends.

Step 3
Proteinase K digestion is performed. This leaves a peptide at the cross-link site that modifies the chemical structure of the nucleotide.

Step 4
5` RNA adaptors are ligated and cDNA is synthesized using reverse transcription.

Step 5
PCR and sequencing of the cDNA.

key feature: crosslink-induced mutation site (CIMS)

Find RNA binding sites of target protein

iCLIP = **i**ndividual-nucleotide resolution **CLIP**

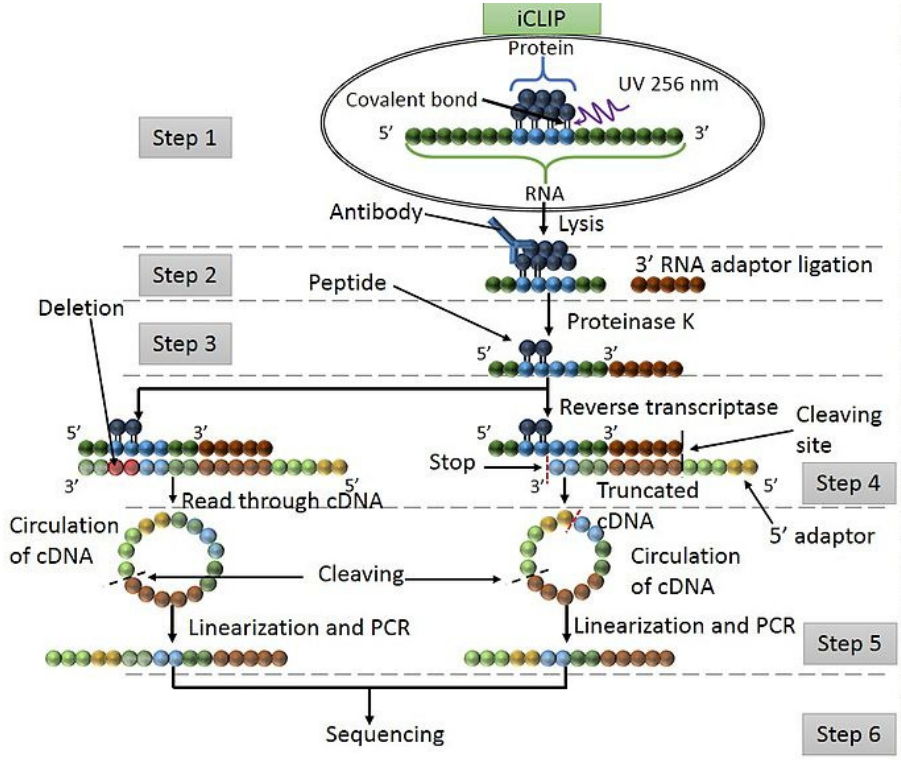
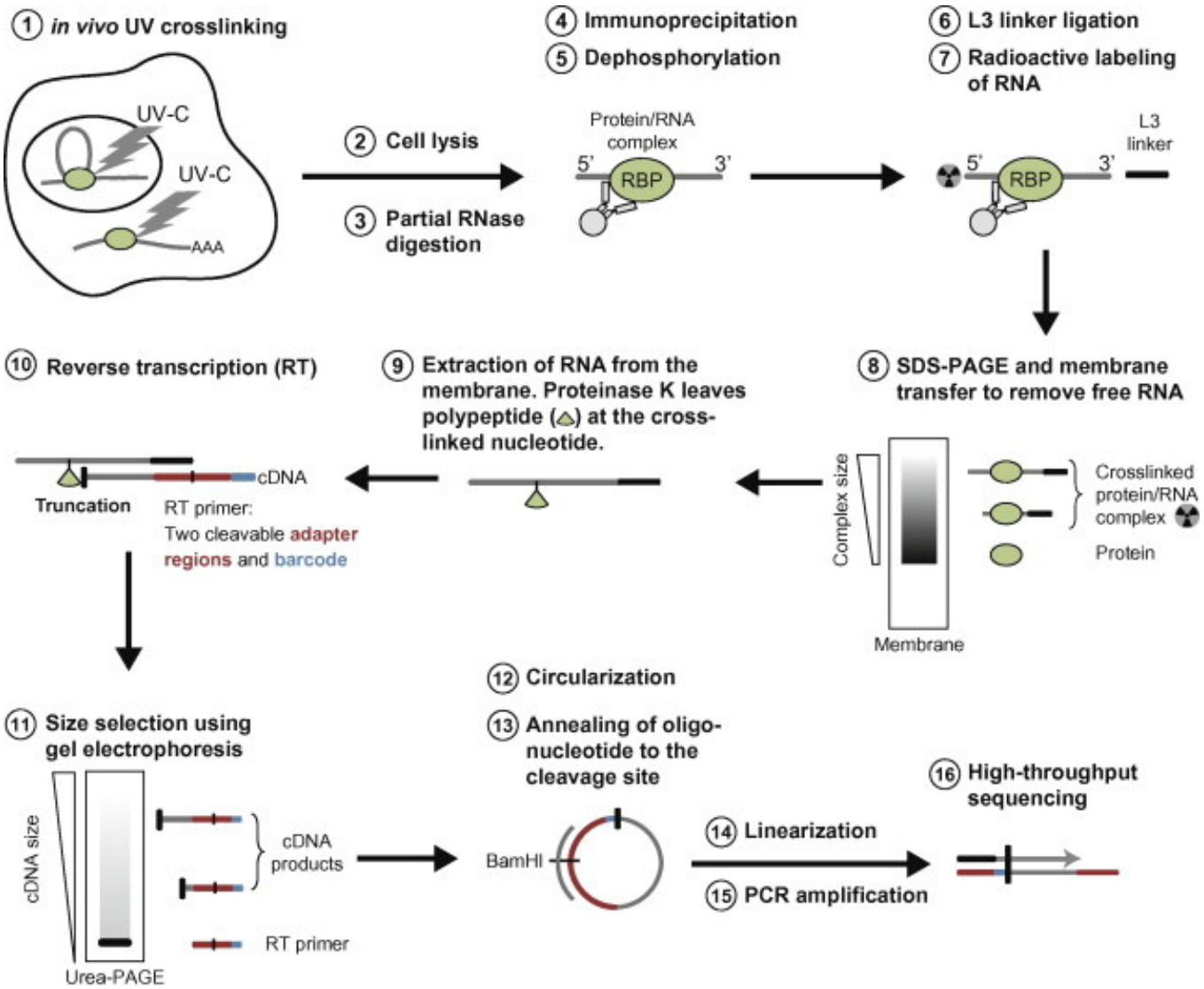


Figure 3: iCLIP
Step 1
Irradiation of the cells with UV light catalyses covalent bond formation between proteins and RNA in direct contact. The cell is lysed, and the protein of interest is isolated using immunoprecipitation.
Step 2
Washing is performed to remove free RNA, and RNA adaptors are ligated at the 3′ ends.
Step 3
Proteinase K digestion is performed. This leaves a peptide at the cross-link site that modifies the chemical structure of the nucleotide.
Step 4
Reverse transcription PCR is performed. This results in both truncated cDNAs and cDNAs that are read through the cross-link sites.
Step 5
Adapters are added to the 5' cDNA ends via circularization. Restriction enzyme cleavage is performed to linearize the cDNAs, allowing both the truncated and read through cDNAs to be sequenced. The position of cDNA truncation allows RNA-Protein interaction sites to be determined at high resolution.
Step 6
Sequencing.

key feature: revers transcription terminates at crosslinks
key insight: sites of trancation are sites of crosslinking

just another graphics for iCLIP



Single nucleotide contacts of target protein and RNA

PAR-CLIP = **P**hoto-**A**ctivated **R**NA **CLIP**

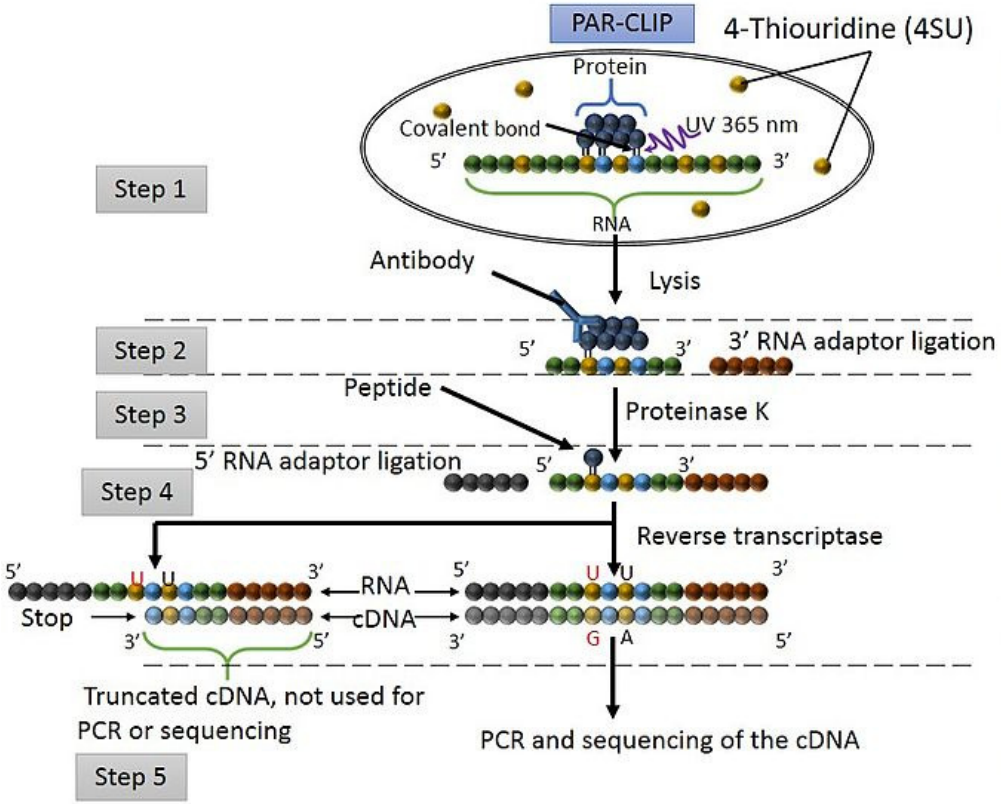


Figure 1: PAR-CLIP

Step 1
The cell is initially treated with 4- thiouridine (4-SU), a thiol-modified ribonucleoside that is incorporated during RNA transcription. Irradiation of the cells with UV light catalyses covalent bond formation between proteins and RNA in direct contact. The cell is then lysed, and the protein of interest is isolated using immunoprecipitation.

Step 2
After stringent washing to remove the nonspecific RNAs, adaptor sequences are ligated to the 3′ ends of the RNA fragments.

Step 3
Proteinase K digestion is then performed in order to remove protein from the RNA-protein complexes. This step leaves a peptide at the cross-link site, allowing for the identification of the cross-linked nucleotide.

Step 4
After ligating RNA linkers to the RNA 5′ ends, cDNA is synthesized using reverse transcription.

Step 5
PCR and sequencing of the read through cDNA.

key feature: photo-activatable nucleotides crosslinks

key insight: nucleotide transitions indicate crosslinks

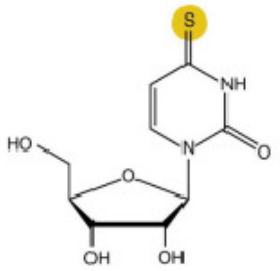
Single nucleotide contacts of target protein and RNA

PAR-CLIP = **P**hoto-**A**ctivated **R**NA **CLIP**

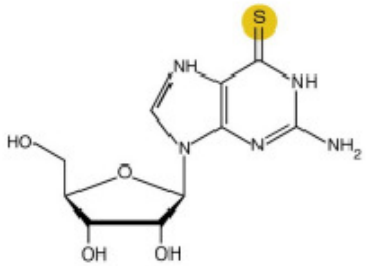
- **photo-activated nucleotides** 4SU and 6SG
- fed to cells, incorporated in nascent RNA
- reverse transcriptase inserts G opposite of 4SU

  = T → C transition
- reverse transcriptase inserts T opposite of 6SG

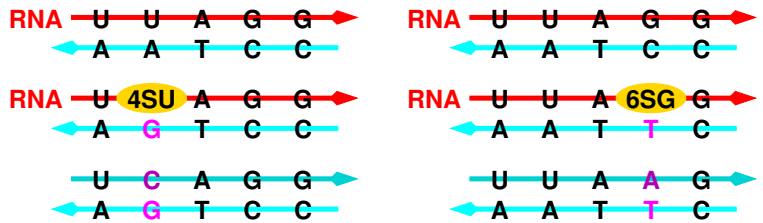
  = G → A transition
- **bioinformatics**: search read clusters with significantly high T > C mismatch frequency



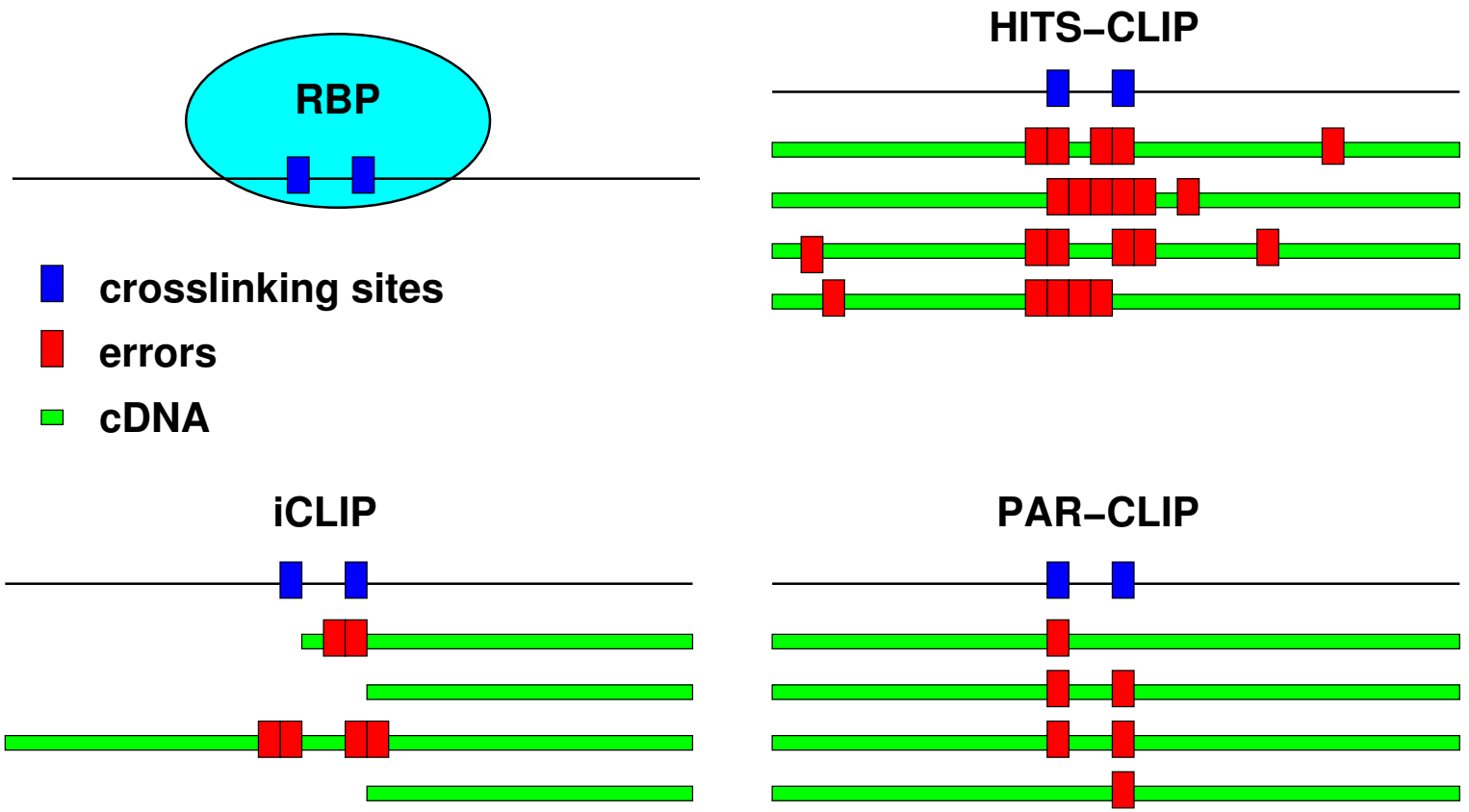
4-thiouridine (4SU)



6-thioguanosine (6SG)



CLIP-seq data – What do we get?



- HITS-CLIP-seq: nucleotide substitutions around the BS
- iCLIP: cDNA/read ends at crosslinked nucleotides
- PAR-CLIP: particular substitutions at crosslinked nucleotides

Limitations

	Main feature	Advantages	Disadvantages
HITS-CLIP	254 nm UV crosslinking	Can be performed on any tissue	30-60 nucleotide resolution Typically less accurate
PAR-CLIP	4SU incorporation into RNA followed by crosslinking with 365 nm UV light	Thymidine to cytidine and guanosine to adenosine mutations allow binding sites to be determined with high accuracy	Requires cytotoxic nucleosides Limited to *in vitro* cell cultures Uptake of the nucleoside is variable
iCLIP	Circularization of reverse transcription product	Allows for the amplification of truncated cDNA Increased accuracy for binding sites located within repetitive motifs	Circularization may introduce biases