Methodology

* Find RBPs that were studied by different groups / methods (SELEX, RNAcompete..)
  + Develop methods for checking consistency of motifs
    - PPM1 vs PPM2:
      * matrix distance
      * similarity in result: autologous binding (next point)
      * similarity in result: rnd seq?
  + Check robustness of statistical approaches (point: **autologous binding**)
    - (benchmark statistical approaches)
* Check for RNA binding domains -> group proteins and check for similarity in motif within groups. If no result -> point **motif causality**
  + check for experiments where just domains were used

Autologous binding

(check presence in auto CDS; Arthur project)

NEW:

* New motif sets / species
  + CLIP motif set
* **New statistical approaches** (accurate mapping of PPMs to sequences; start with literature research)
* small variation: instead of coverage (on own CDS): does RBP bind own CDS at least once?
  + still normalize by length
* CDS / UTR motif presence ratio (length normalized)
* Protein binding RNAs:
  + Are mRNAs that code RBPs higher covered by motifs from all RBPs?
  + Gene-ontology on the top % mRNAs sorted by motif coverage (all motifs)

Motif causality

Was the motif-seq a cause of complementary autologous binding / important autologous feedback? (Why is the motif like it is?)

* Composition: Check for correlations between protein/CDS composition and motif-seq/comp
* Seq: Check theoretical highest alignment affinities between protein and motif
  + (energy score instead of PPM score)

Open:

Properly sliding a motif down a sequence