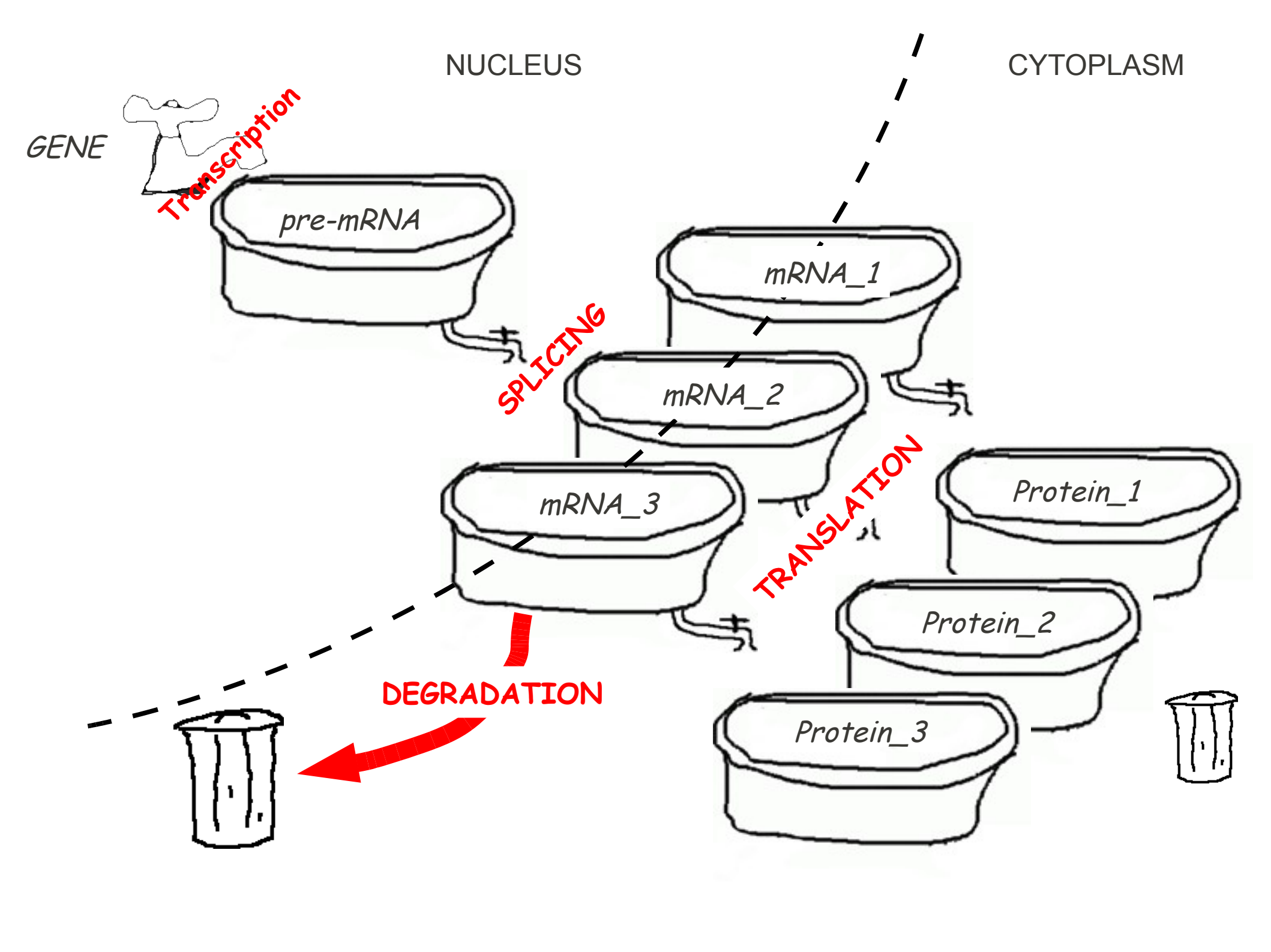
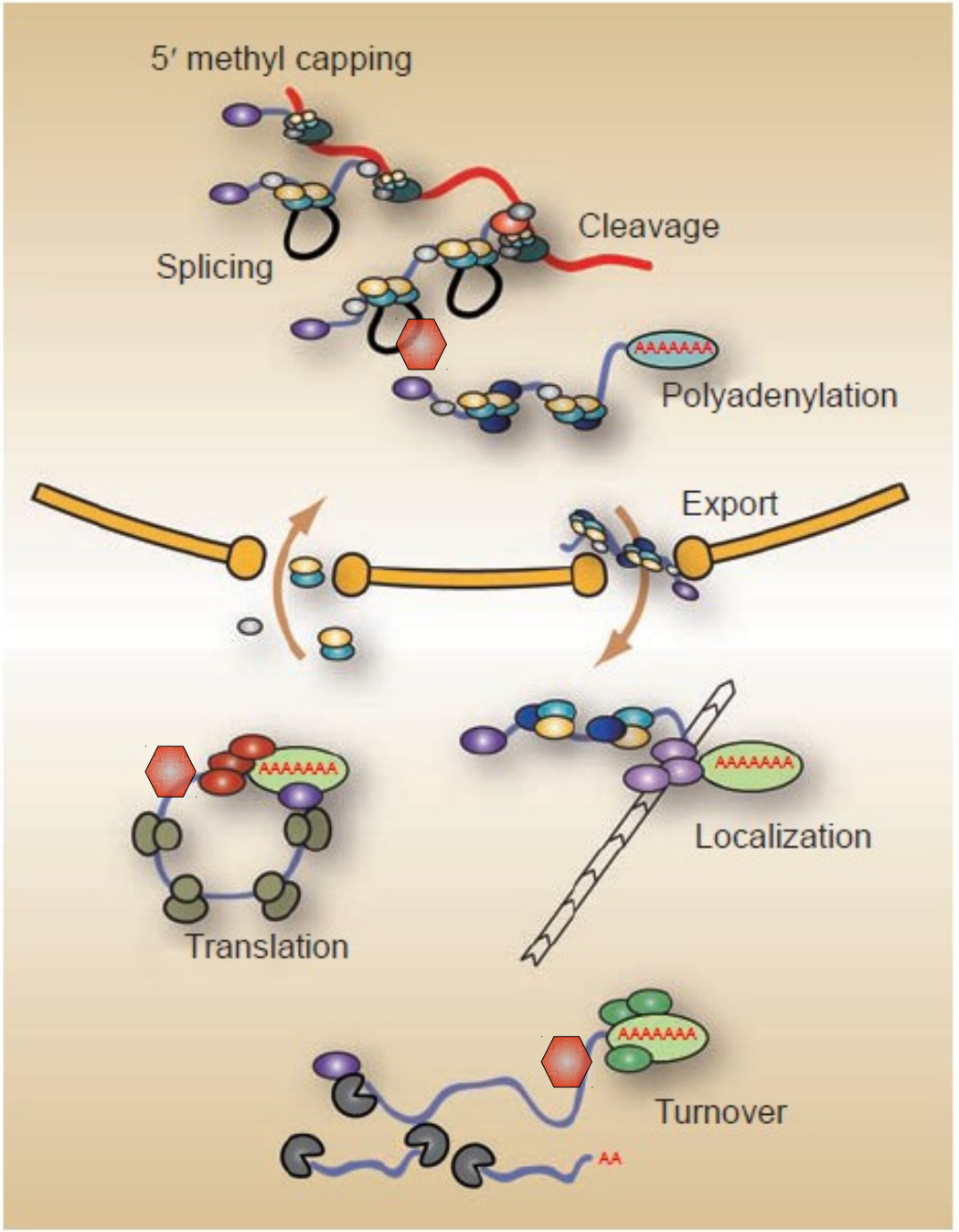
IDENTIFICATION OF RNA TARGETS FOR THE REGULATORY RNA-BINDING PROTEIN **CELF1**

by

CLIP-SoliD and RNAseq-SoliD


NUCLEUS
CYTOPLASM
GENE
Transcription
pre-mRNA
SPLICING
mRNA_1
mRNA_2
mRNA_3
TRANSLATION
Protein_1
Protein_2
Protein_3
DEGRADATION



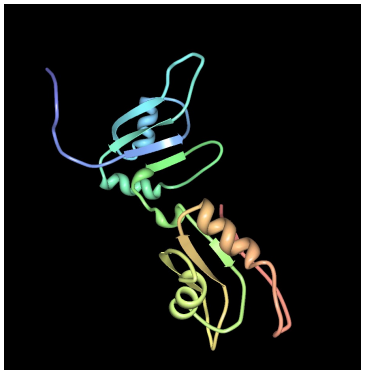
From A McKee and P Silver (2007), *Cell Research*

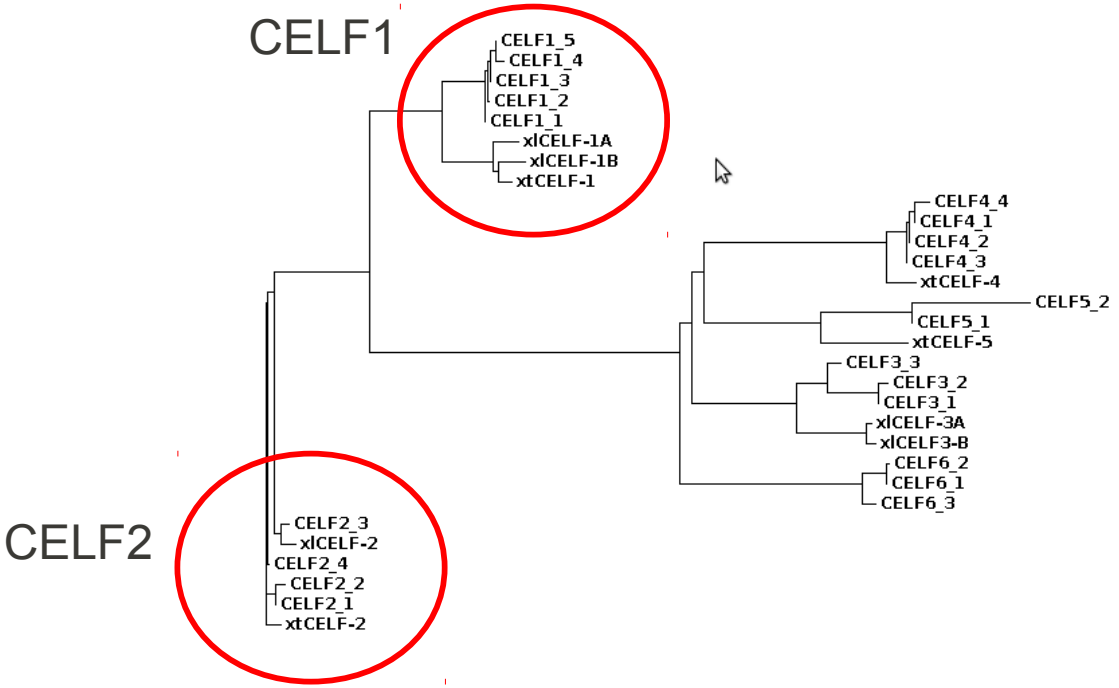
REGULATORY FACTORS :

- RNABPs (~ 600 annotated)

- small regulatory RNAs

CELF PROTEINS

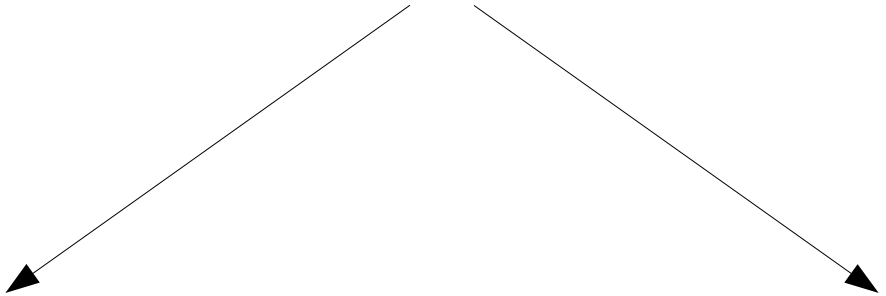




CELF1
MOLECULAR FUNCTIONS :

- alternative splicing regulation

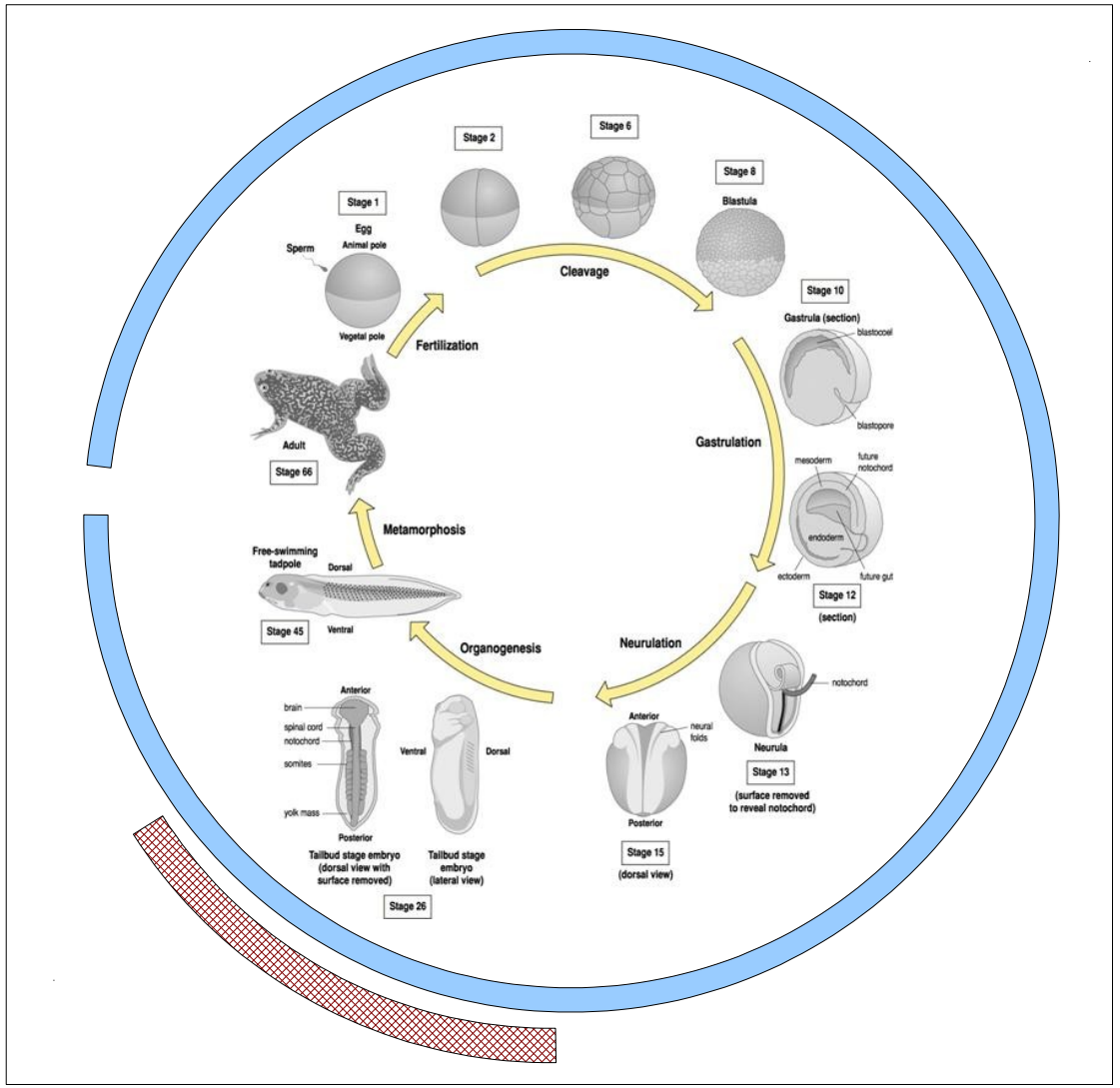
- mRNA deadenylation targeting factor

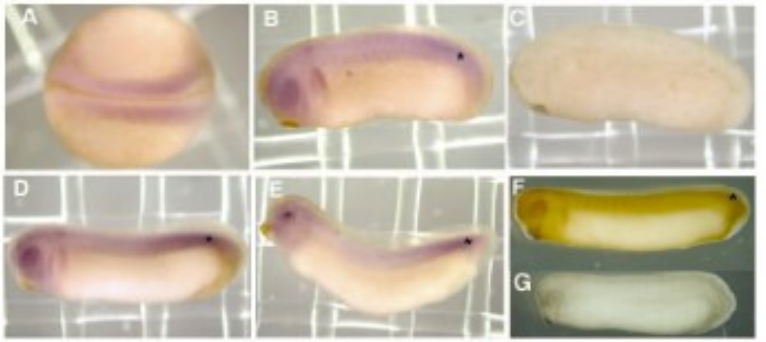


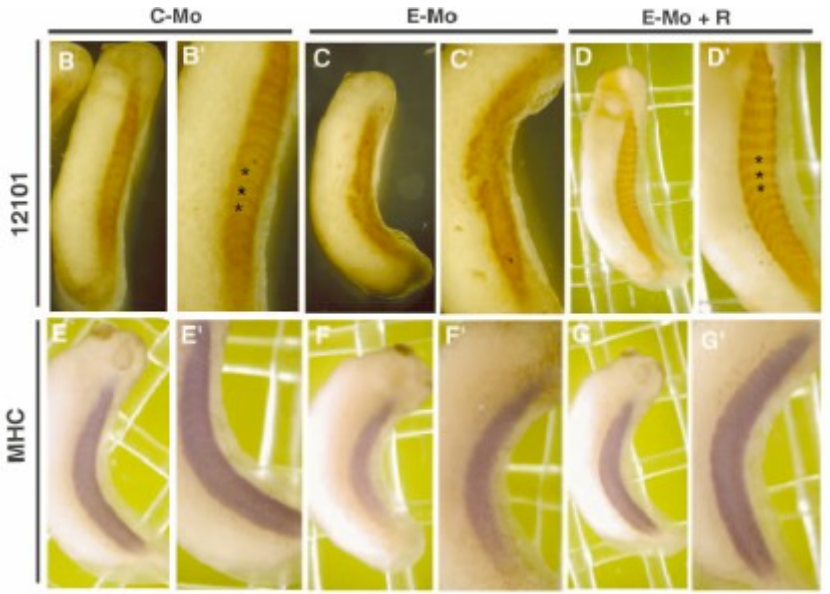
translational silencing

mRNA degradation

CELF1 FUNCTIONS IN XENOPUS



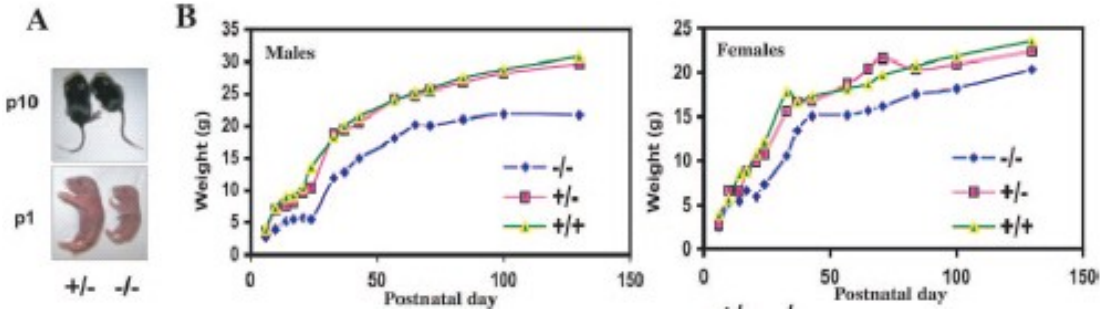




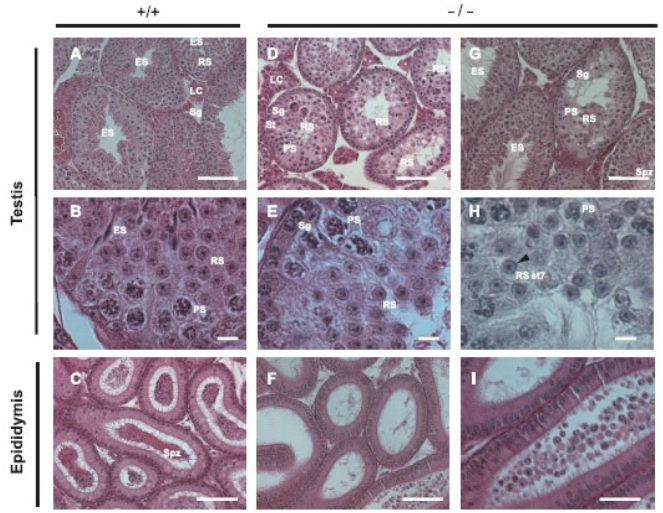
(Gautier-Courteille et al.,Dev. 2004)

CELF1 FUNCTIONS IN MICE

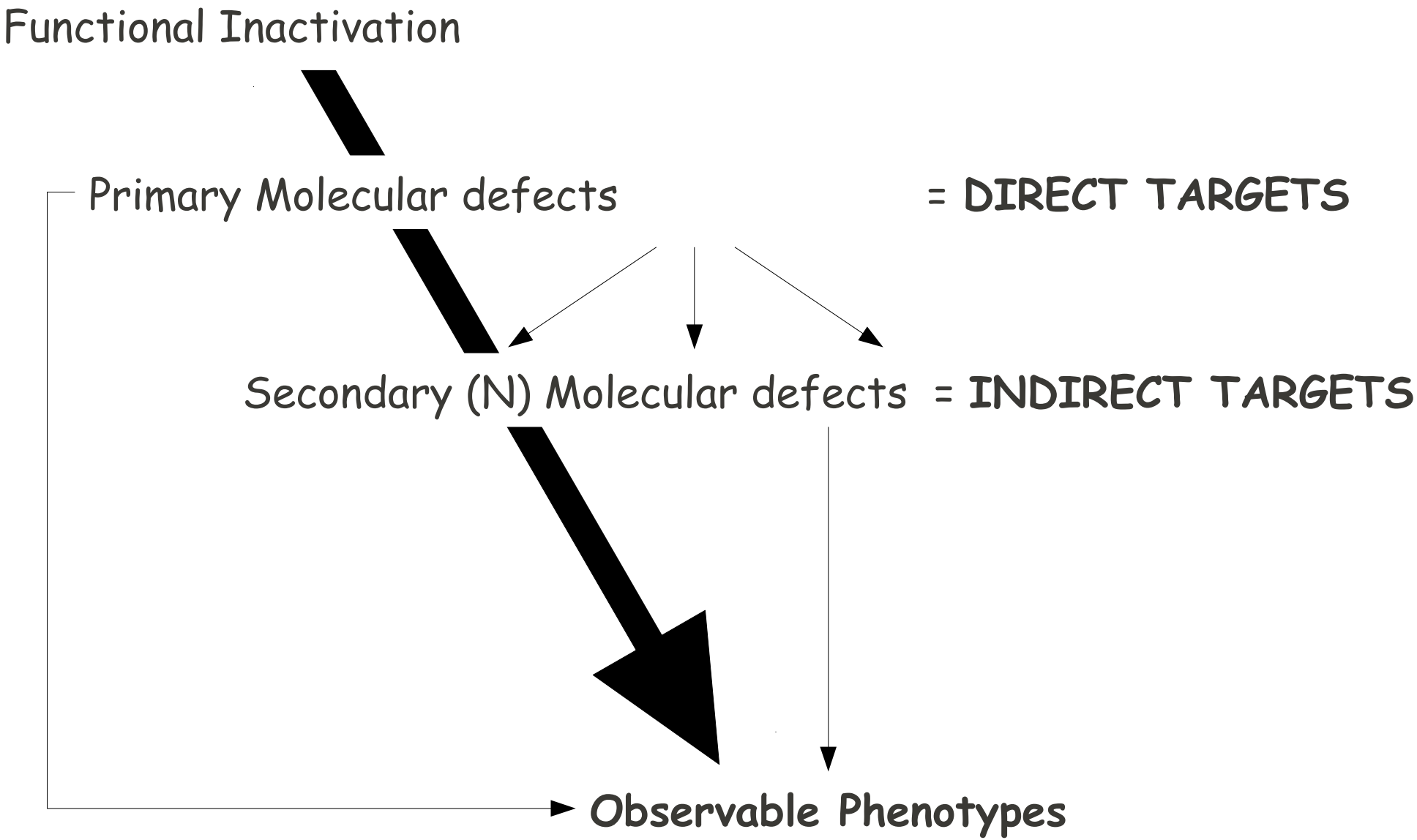
growth retardation phenotype



spermatogenesis defect



(Kress et al.,MCB. 2007)


Functional Inactivation
Primary Molecular defects
= DIRECT TARGETS
Secondary (N) Molecular defects
= INDIRECT TARGETS
Observable Phenotypes

Primary Molecular defects = **DIRECT TARGETS**

Identify RNA associated with CELF1

Secondary (N) Molecular defects = **DIRECT + INDIRECT TARGETS**

Search for GENES with altered expression
- microarray analysis on CELF1 deprived cells

Primary Molecular defects = **DIRECT TARGETS**

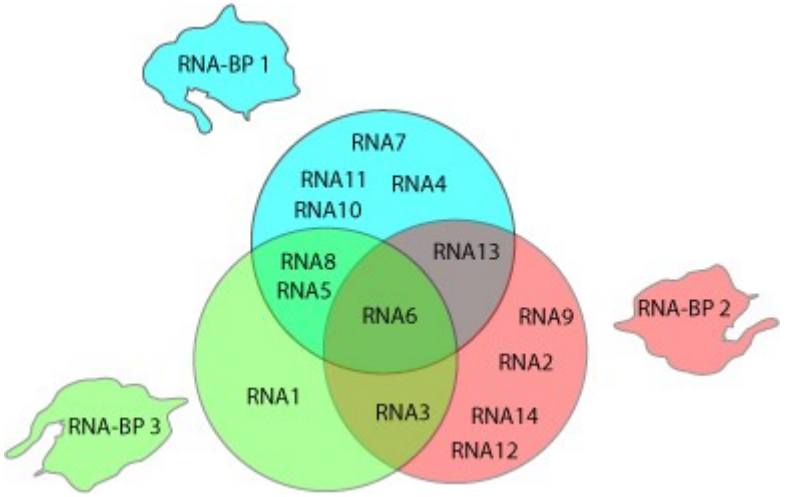
Identify RNA associated with CELF1

- define molecular landscape of CELF 1 target *IN VIVO*

- determine sequences constraint defining the specificity of CELF 1 binding

Primary Molecular defects = **DIRECT TARGETS**

Identify RNA associated with CELF1



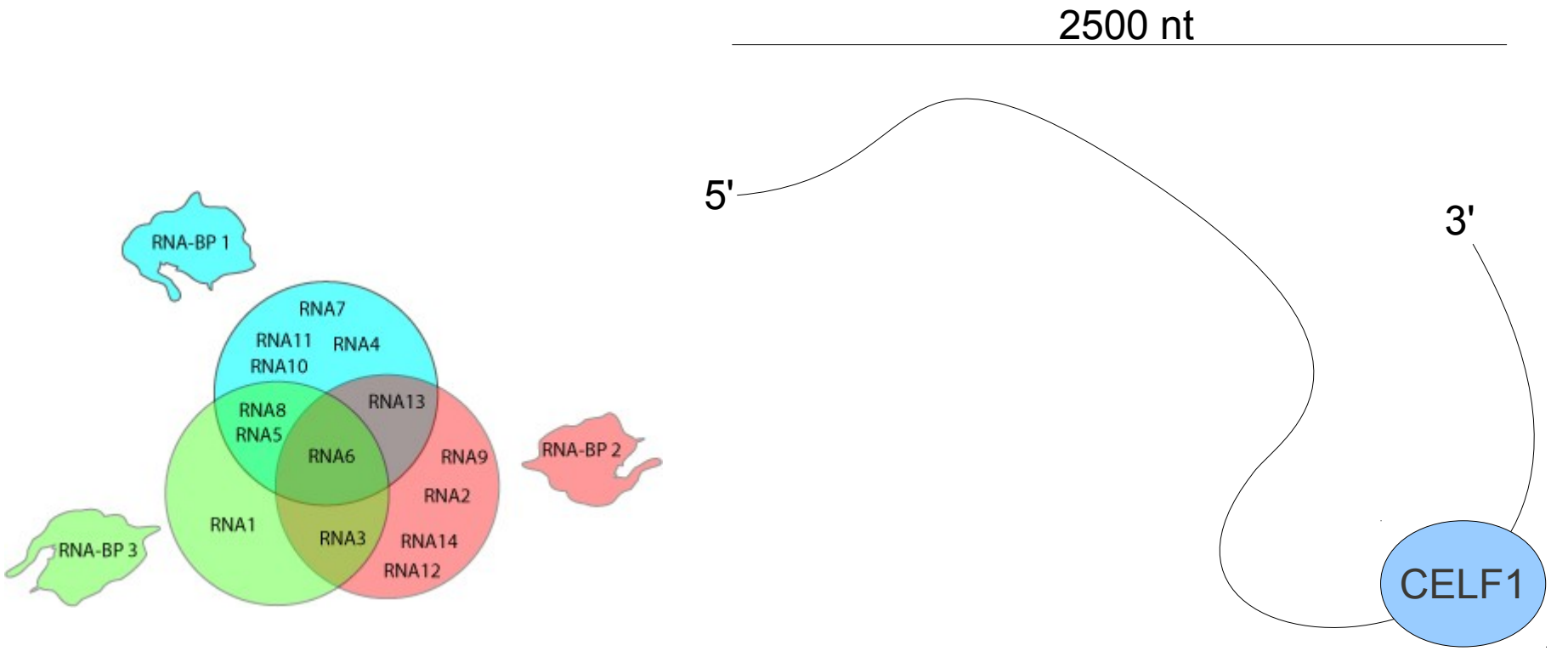
-3 hybrid screening | **GENETIC**

-SELEX
-GST pulldown | **IN VITRO**

-Immunoprecipitation
-CLIP | **IN VIVO**

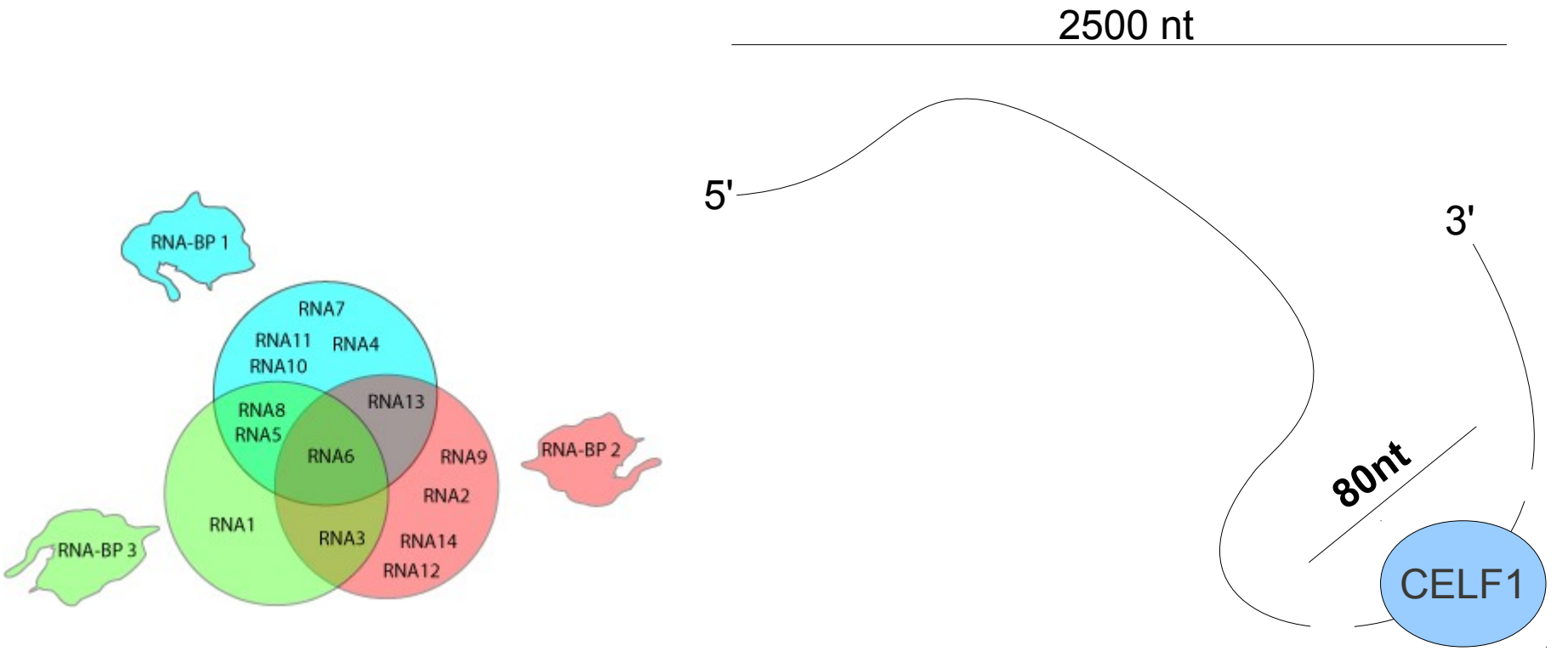
Primary Molecular defects = **DIRECT TARGETS**

Identify RNA associated with CELF1 -CLIP-SoliD

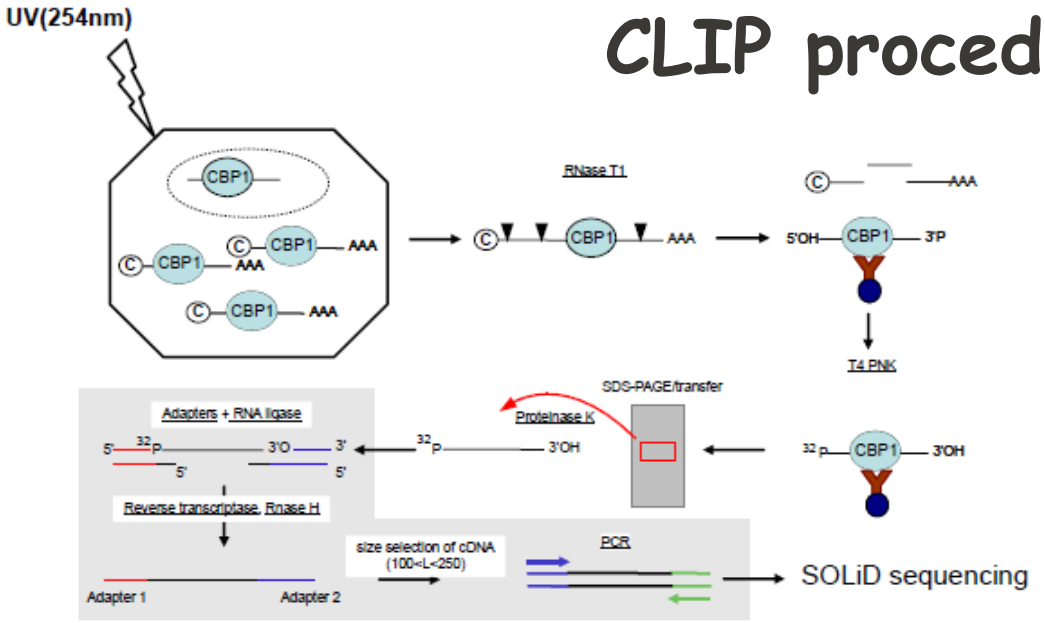


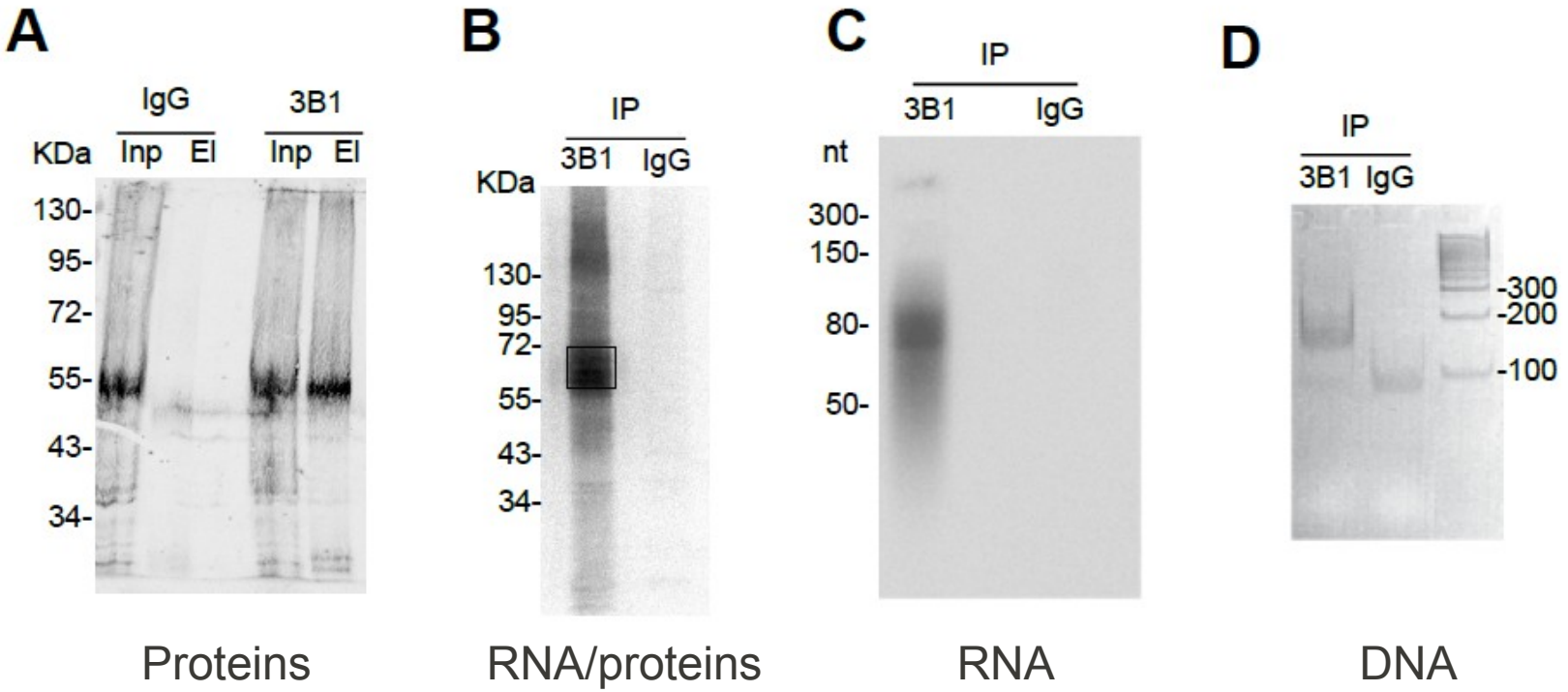
Primary Molecular defects = **DIRECT TARGETS**

Identify RNA associated with CELF1 -CLIP-SoliD



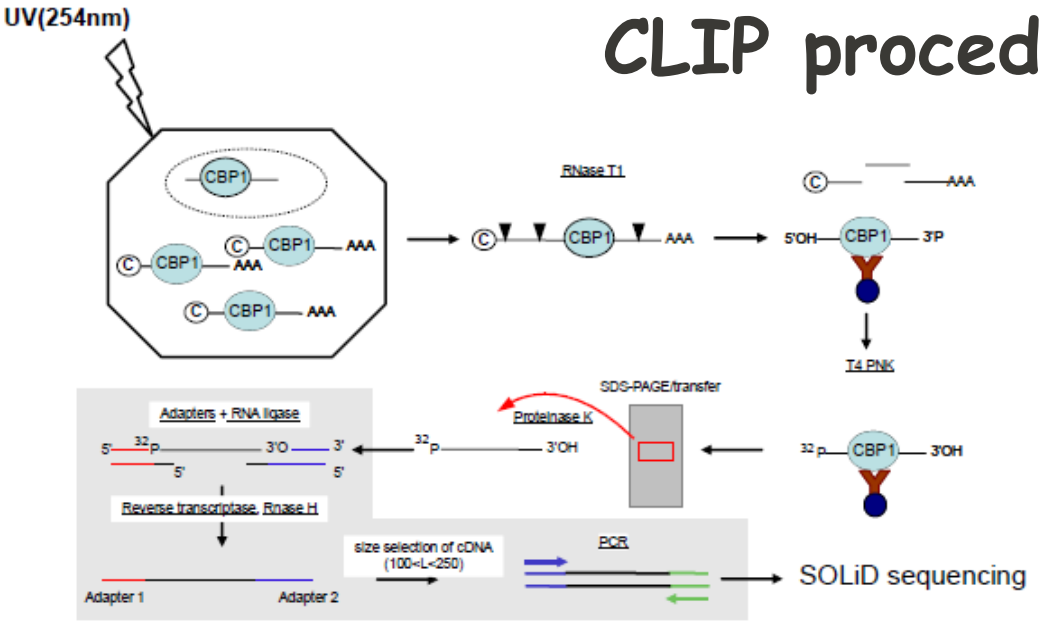
CLIP procedure

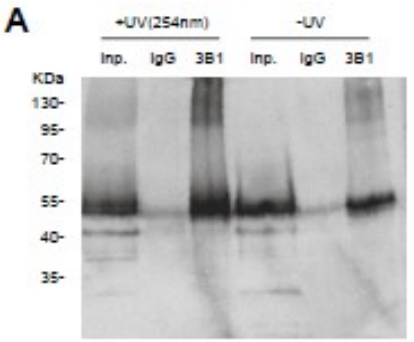


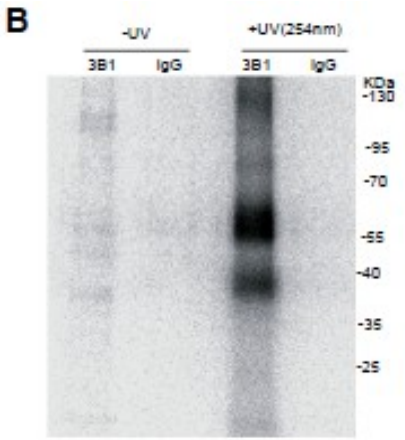


Proteins RNA/proteins RNA DNA

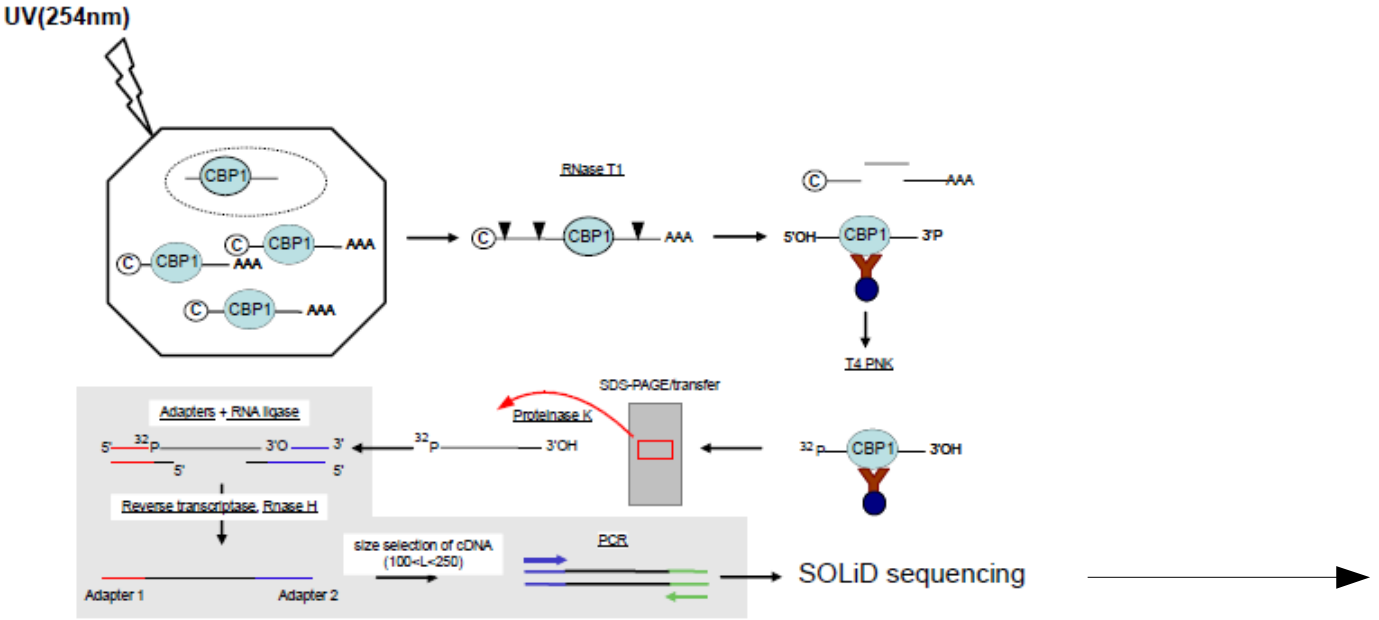
CLIP procedure



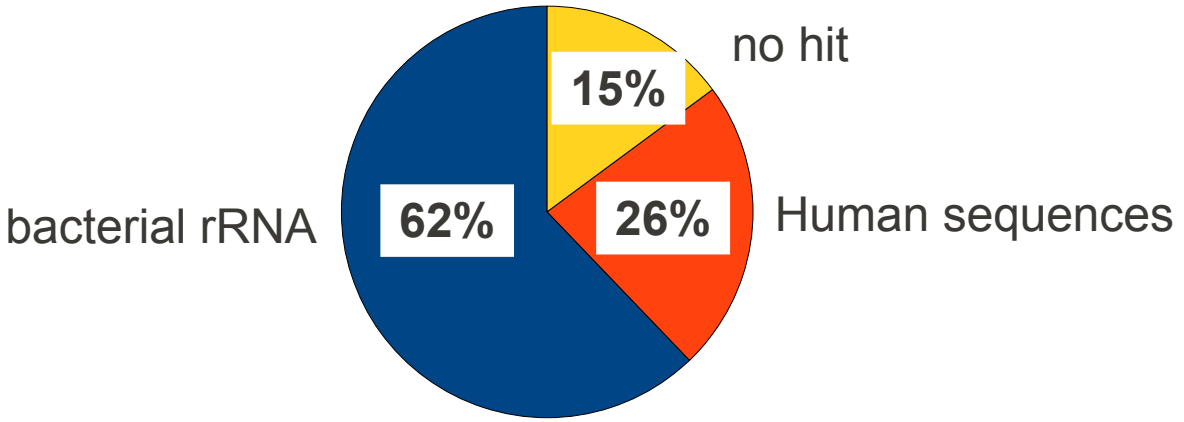




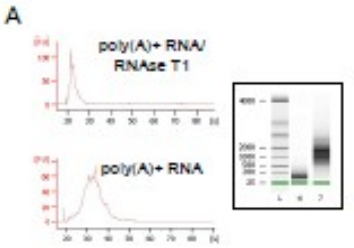
CLIP procedure



In house clonal sequencing



RNASEQ procedure



Hela Cells → TOTAL RNA → poly(A)+ RNA

Partial digest **RNAse T1**

T4 PNK

Adapters + RNA ligase

5' 32P 3'O 3'
5' 5'

Reverse transcriptase, Rnase H

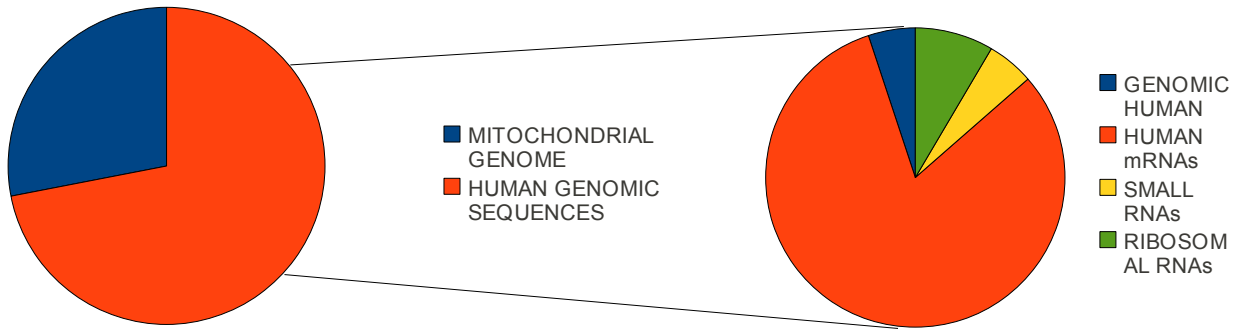
size selection of cDNA
(100<L<250)

PCR

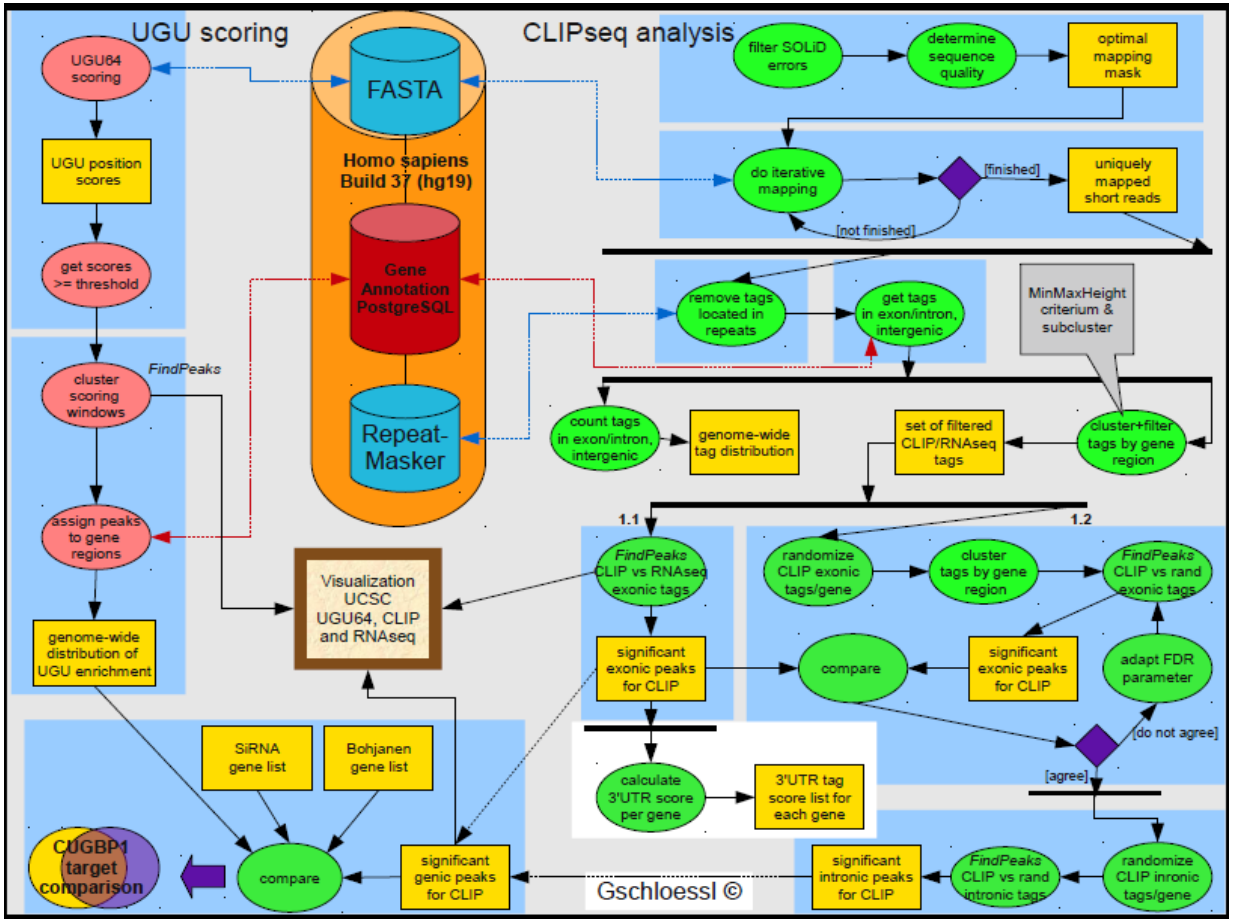
Adapter 1 Adapter 2

GENOSCOPE
Centre National de Séquençage

In house clonal sequencing



MOTIF SCORING

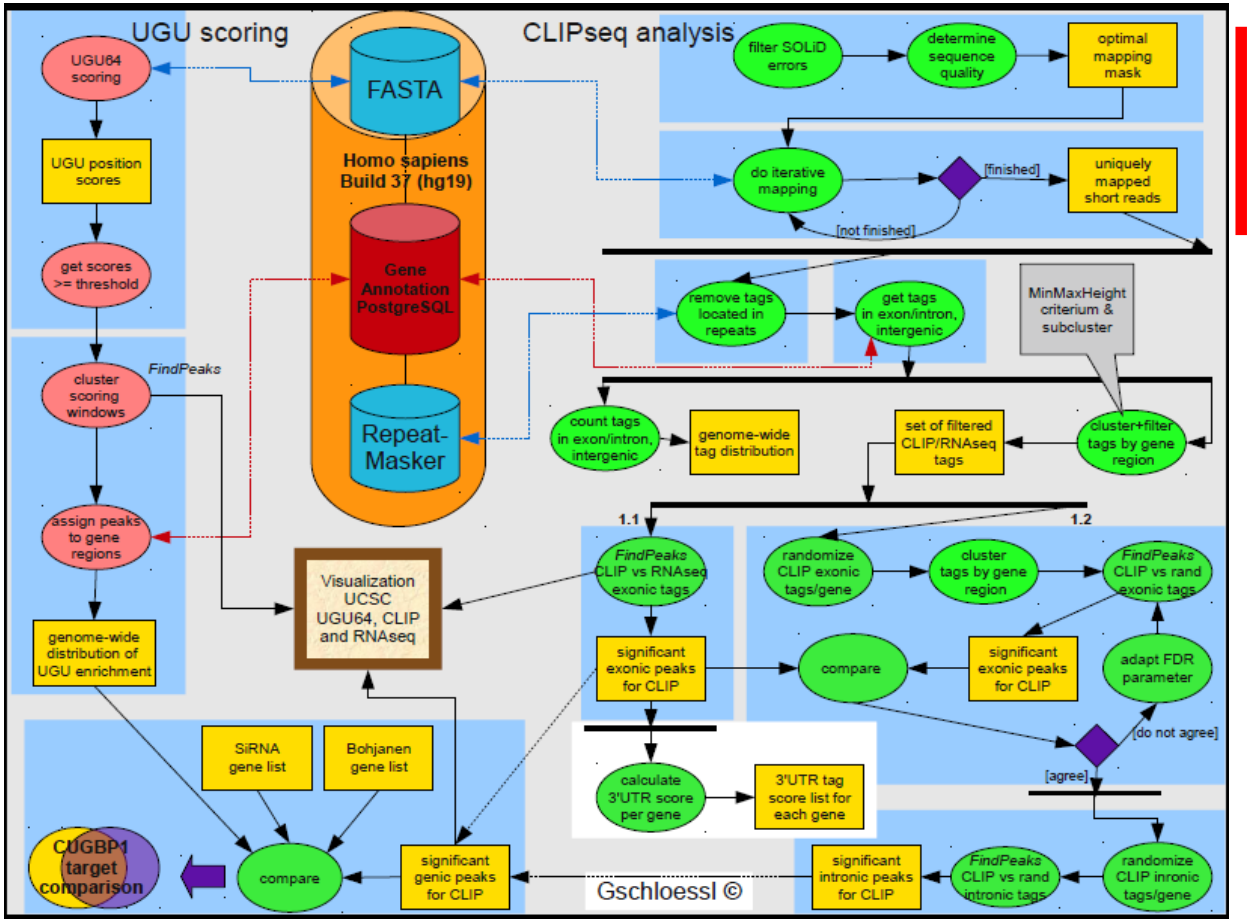


MAPPING (corona lite)

CLUSTER DEFINITION

DIFFERENTIAL ANALYSIS

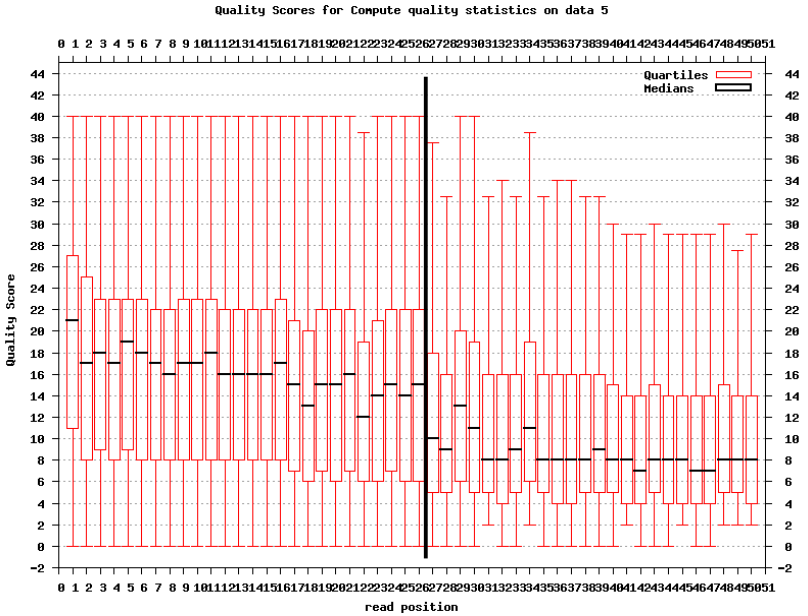
BIOLOGICAL ANALYSIS



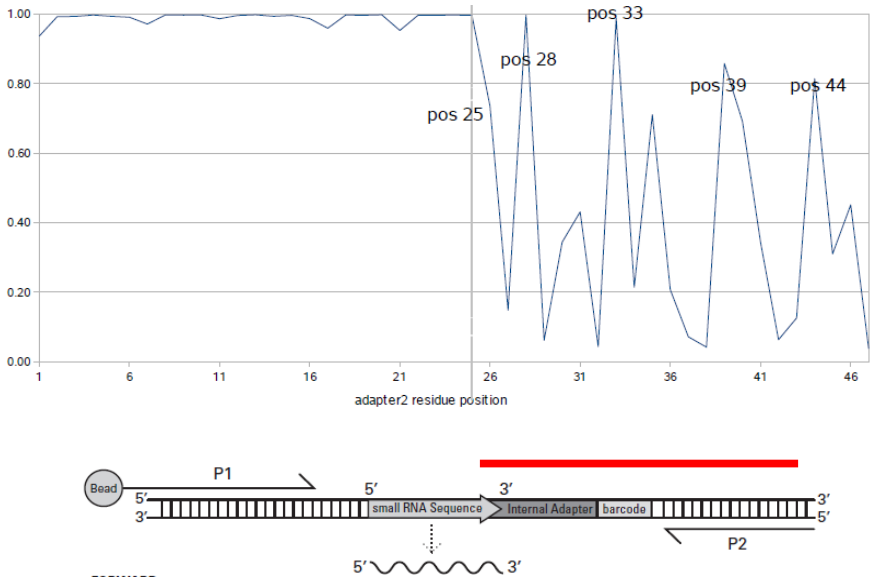
MAPPING (corona lite)

RNAs associated with CELF1 -CLIP-SoliD

6 X 1/4 runs realised → 321 968 741 sequences



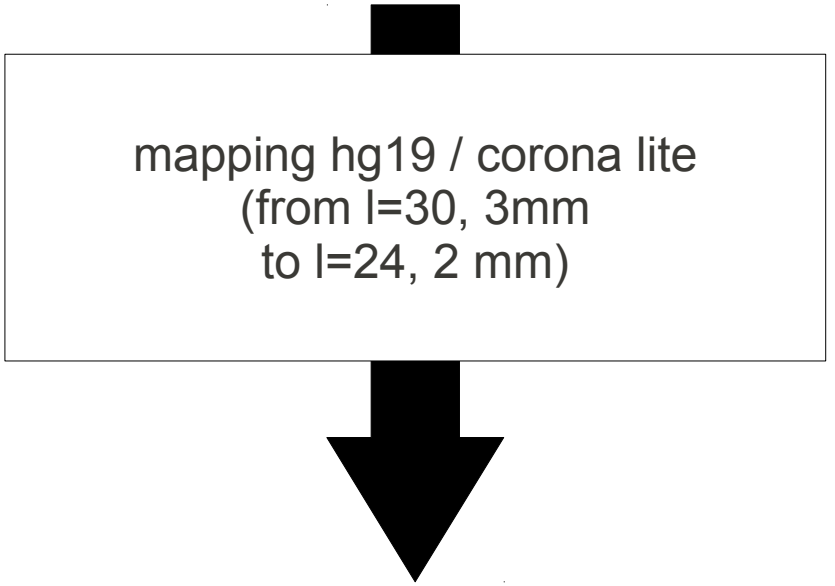
Phred quality analysis



Experimental quality analysis

RNAs associated with CELF1 -CLIP-SoliD

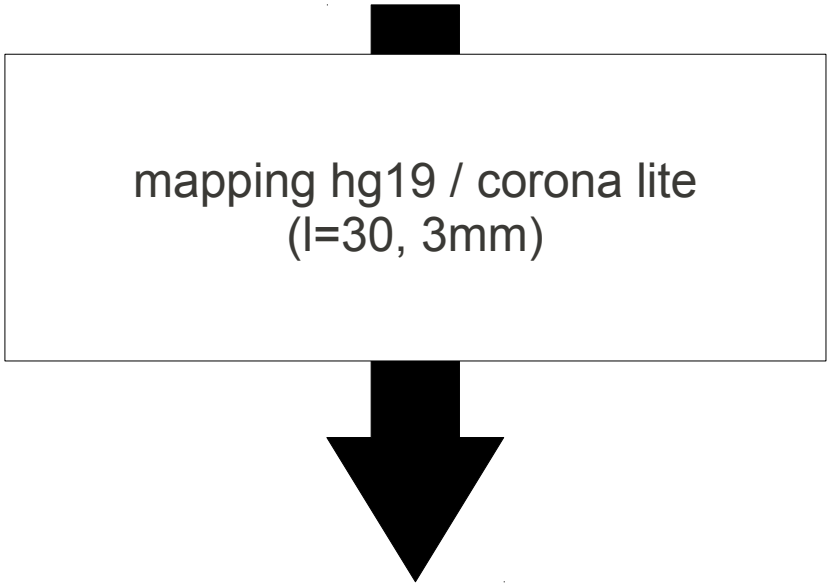
6 X 1/4 runs realised → 321 968 741 sequences
83 000 000 **expected** human sequences



20 589 217 uniquely mapped (repeat masked)

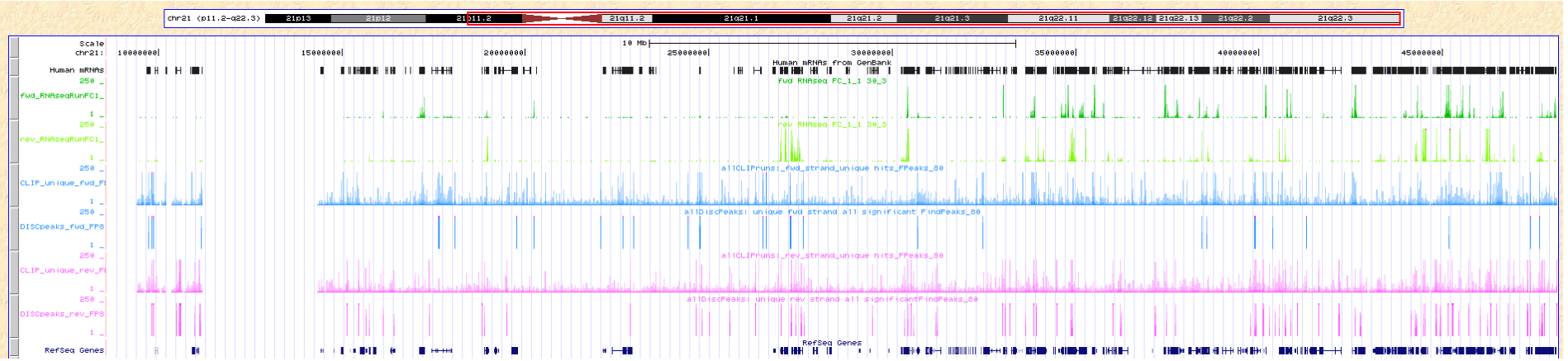
MAPPING / RNASEQ-SoliD

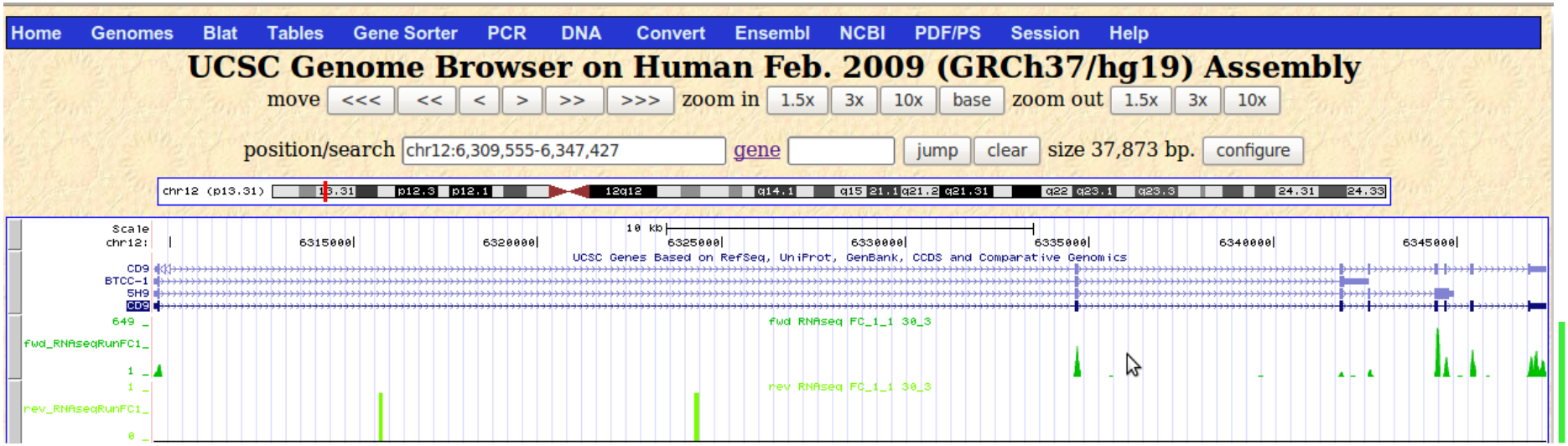
4 X 1/4 runs realised → 347 347 696 sequences



79 869 723 uniquely mapped (repeat masked)

Chr 21 Landscape



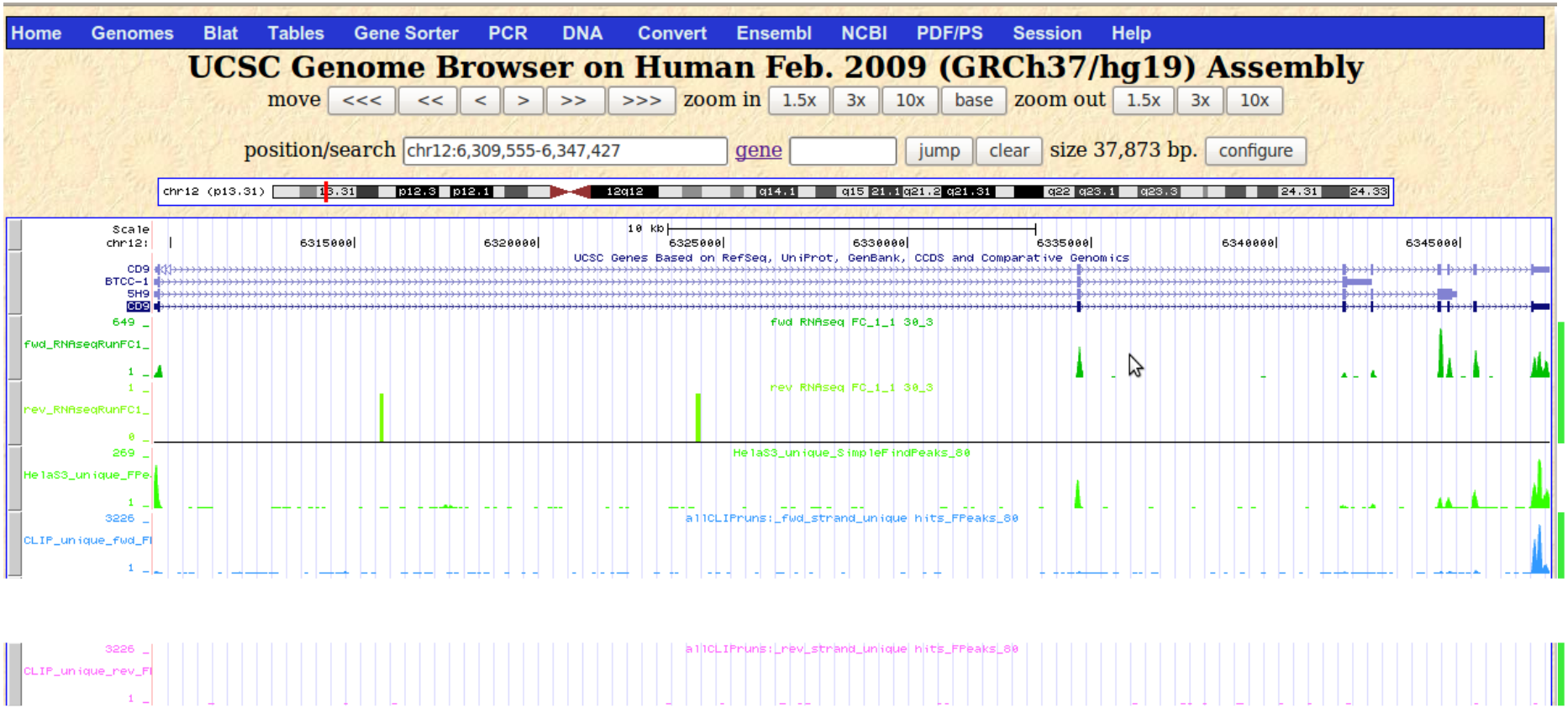


FWD

REV

The RNASEQ : - is strand specific
- cover the exons
- does not map on intronic sequences

94 % of the mapped tag are preceded by a G



fwd

rev

fwd

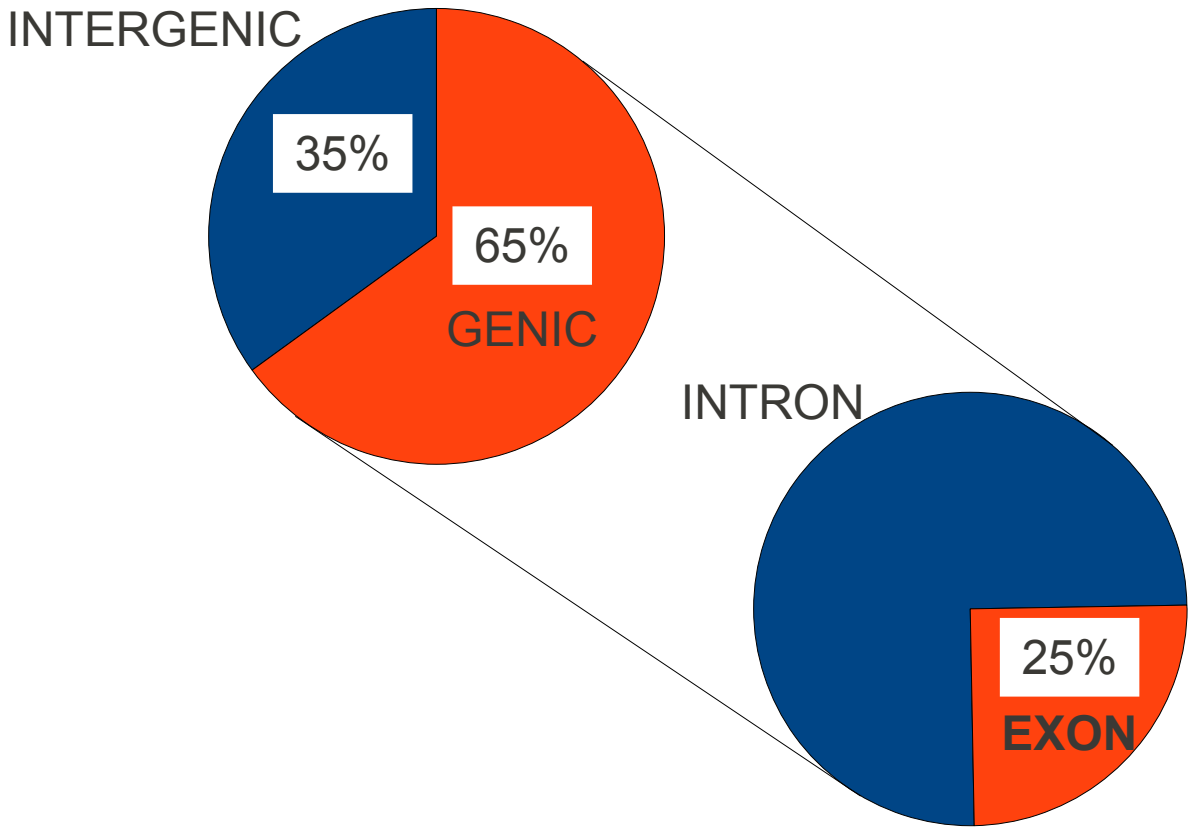
rev

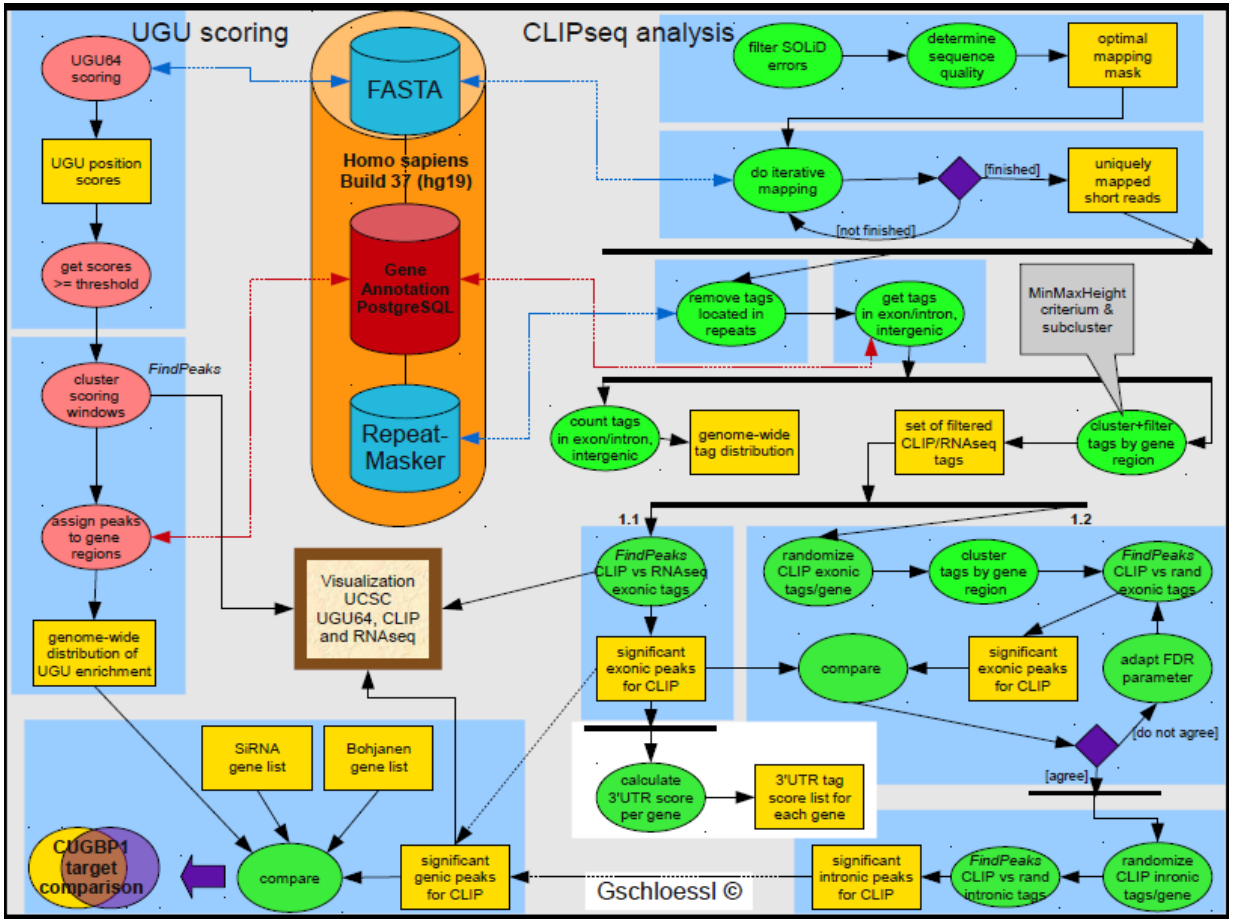
The CLIPSEQ : - is strand specific
- cover both intronic and exonic sequences
- cover specific regions of the genome

No enrichment for G residue upstream of sequence tag.

WHAT ARE THE ENRICHED REGIONS IN CLIP SEQ vs RNASEQ ?

GENOME-WIDE DISTRIBUTION OF CELF1 CLIP READS

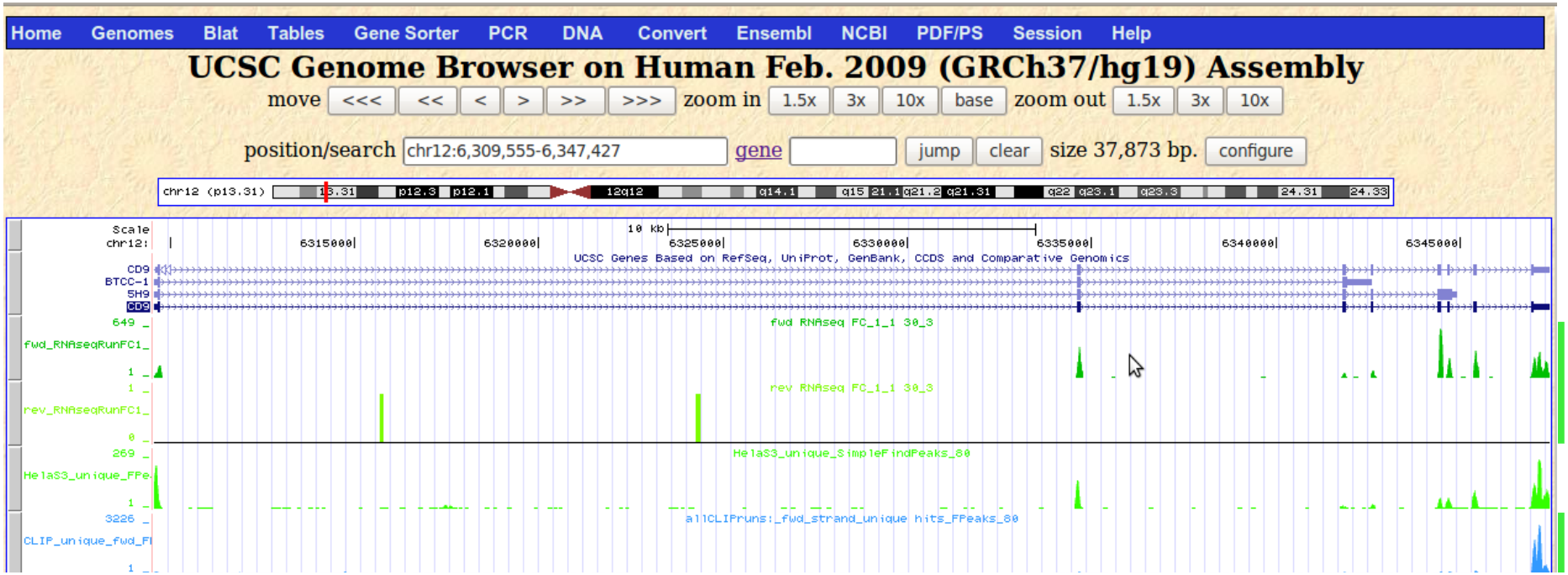



UGU scoring
UGU64 scoring
UGU position scores
get scores >= threshold
cluster scoring windows
FindPeaks
assign peaks to gene regions
genome-wide distribution of UGU enrichment
FASTA
Homo sapiens Build 37 (hg19)
Gene Annotation PostgreSQL
Repeat-Masker
Visualization UCSC UGU64, CLIP and RNAseq
SiRNA gene list
Bohjanen gene list
CUGBP1 target comparison
compare
significant genic peaks for CLIP
CLIPseq analysis
filter SOLiD errors
determine sequence quality
optimal mapping mask
do iterative mapping
[finished]
[not finished]
uniquely mapped short reads
remove tags located in repeats
get tags in exon/intron, intergenic
MinMaxHeight criterium & subcluster
count tags in exon/intron, intergenic
genome-wide tag distribution
set of filtered CLIP/RNAseq tags
cluster+filter tags by gene region
1.1
1.2
FindPeaks CLIP vs RNAseq exonic tags
significant exonic peaks for CLIP
calculate 3'UTR score per gene
3'UTR tag score list for each gene
randomize CLIP exonic tags/gene
cluster tags by gene region
FindPeaks CLIP vs rand exonic tags
compare
significant exonic peaks for CLIP
adapt FDR parameter
[do not agree]
[agree]
significant intronic peaks for CLIP
FindPeaks CLIP vs rand intronic tags
randomize CLIP inronic tags/gene
Gschloessl ©
CLUSTER DEFINITION
DIFFERENTIAL ANALYSIS

Genome analysis

FindPeaks 3.1: a tool for identifying areas of enrichment from massively parallel short-read sequencing technology

Anthony P. Fejes[1,*], Gordon Robertson[1], Mikhail Bilenky[1], Richard Varhol[1], Matthew Bainbridge[2] and Steven J. M. Jones[1,*]

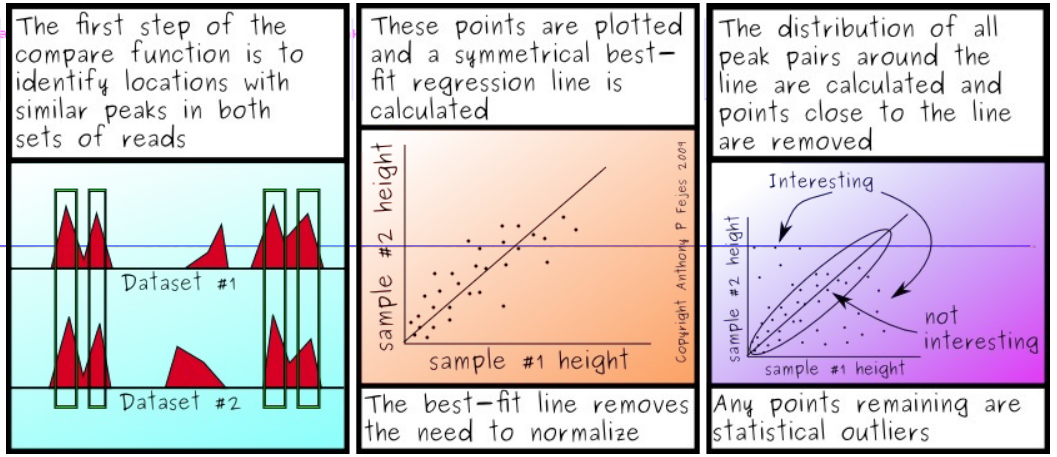


fwd

rev

fwd

rev

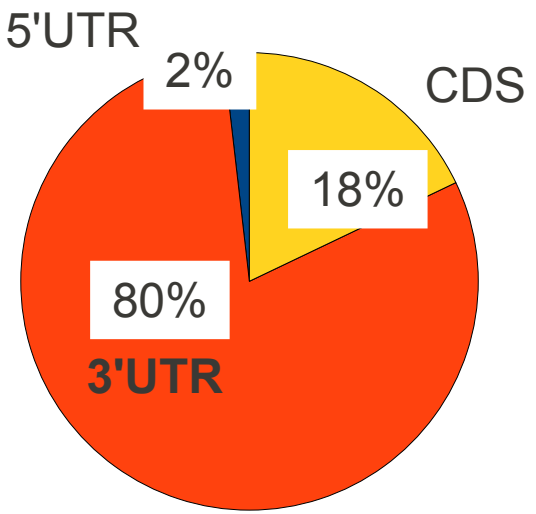


TRANSCRIPT-WIDE DISTRIBUTION OF CELF1 CLUSTERS

CELF1 CLIP vs RNA-SEQ

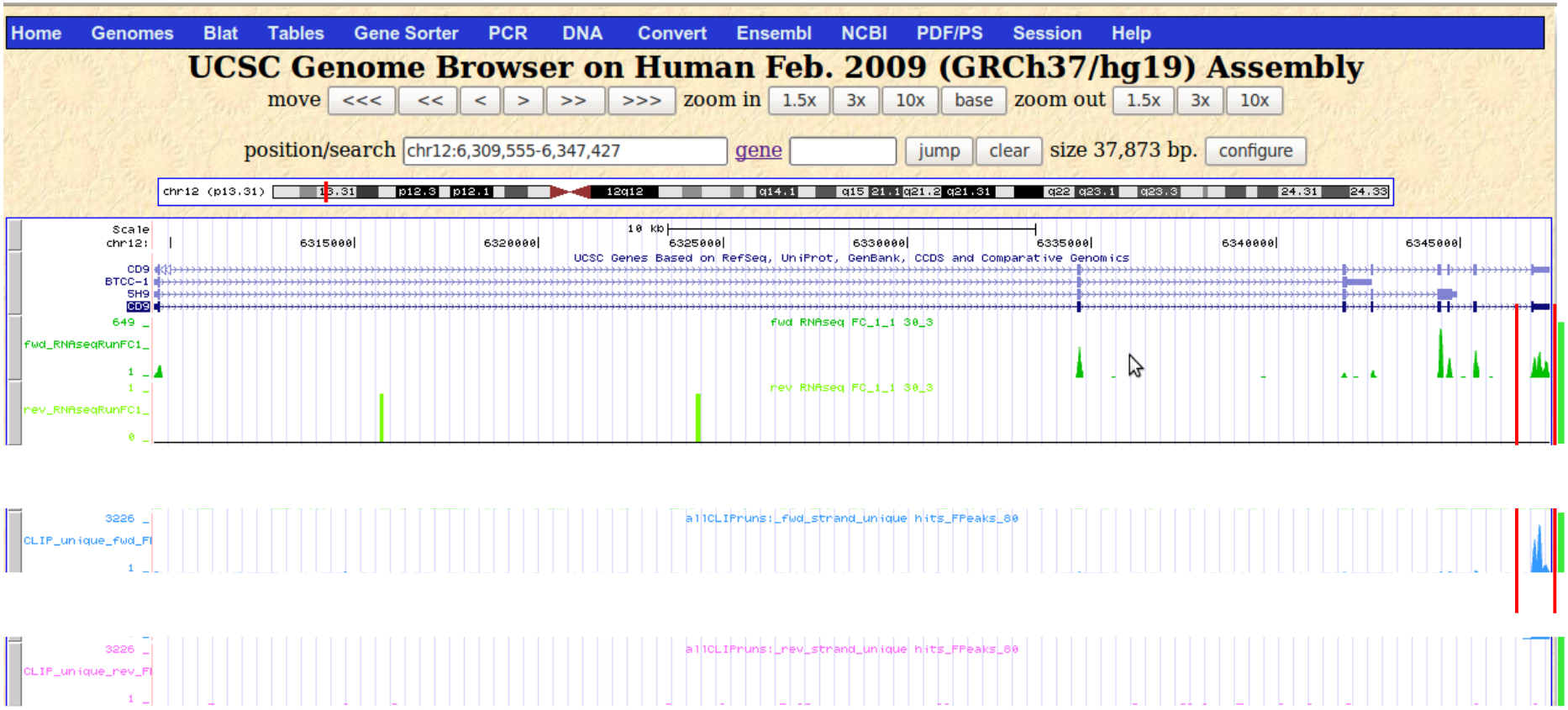
FINDPEAKS on exonic reads

2788 enriched clusters → **1277 GENES**



In mRNAs 3'UTR are the major binding sites for CELF1

DO PREVIOUSLY IDENTIFIED TARGET CONFIRM THE VALIDITY OF THE CELF1 CLIP ?

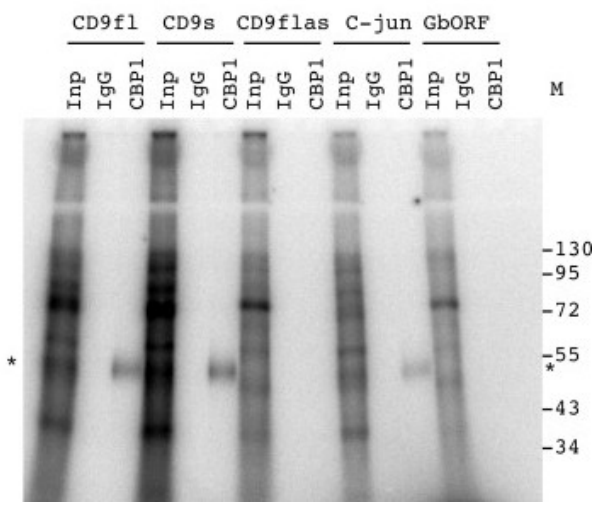


fwd

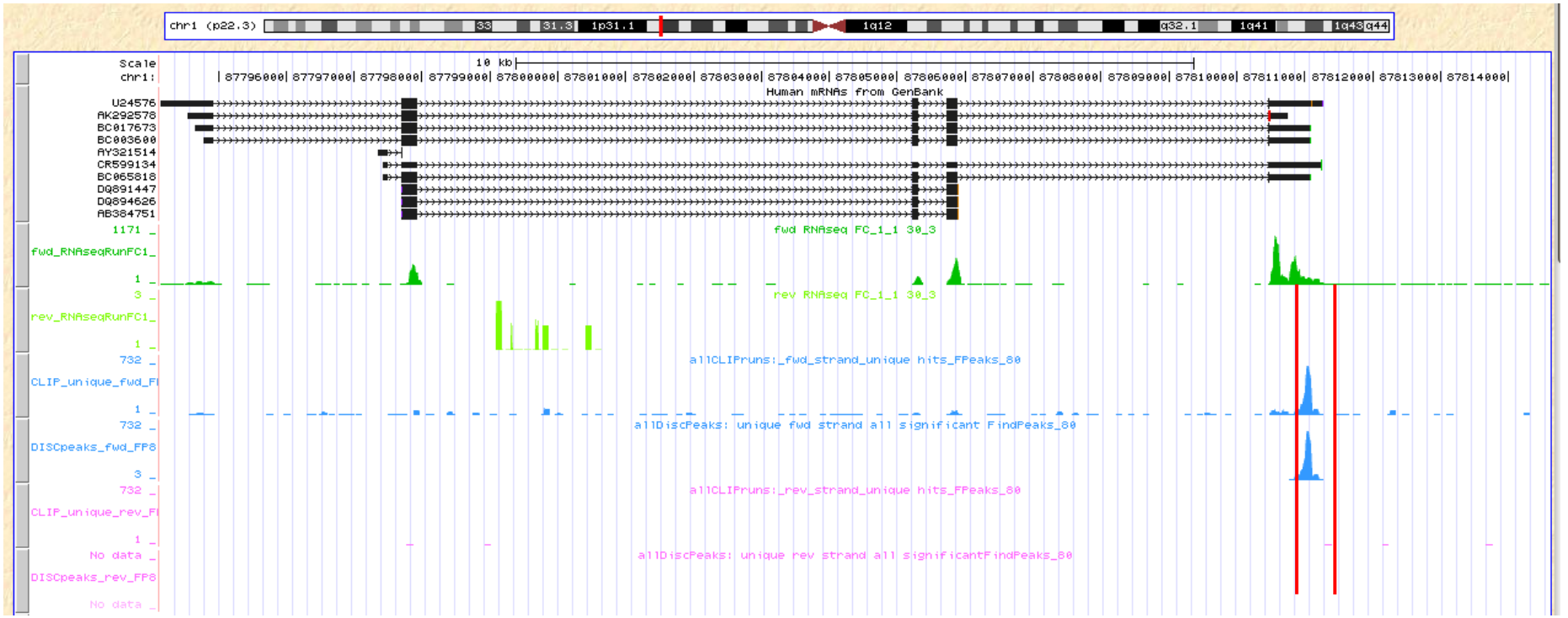
rev

fwd

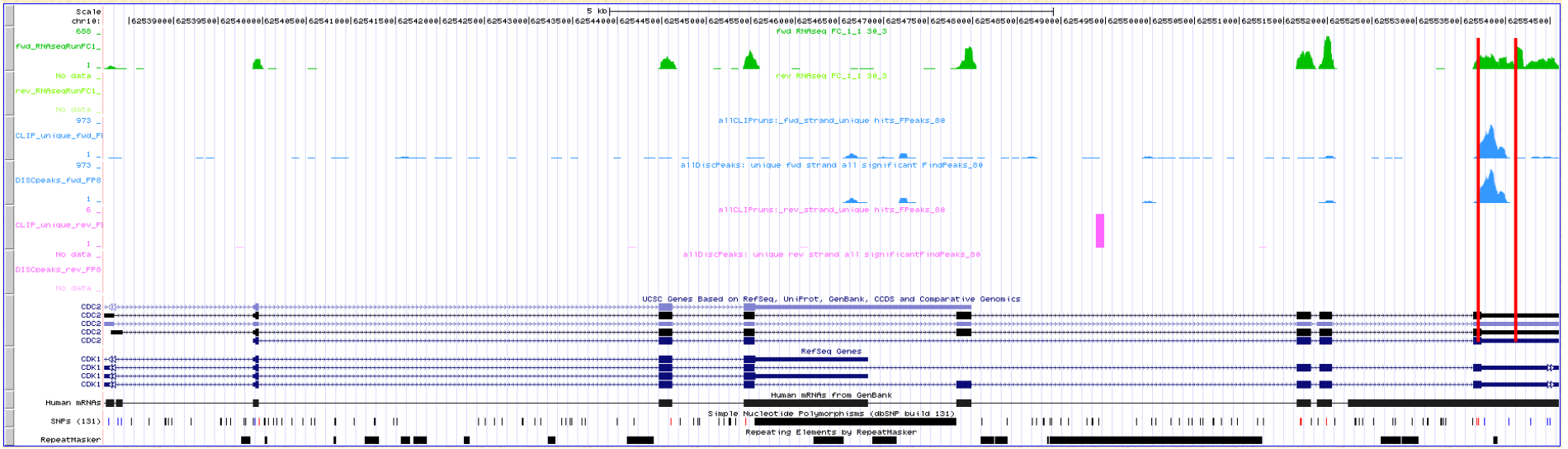
rev



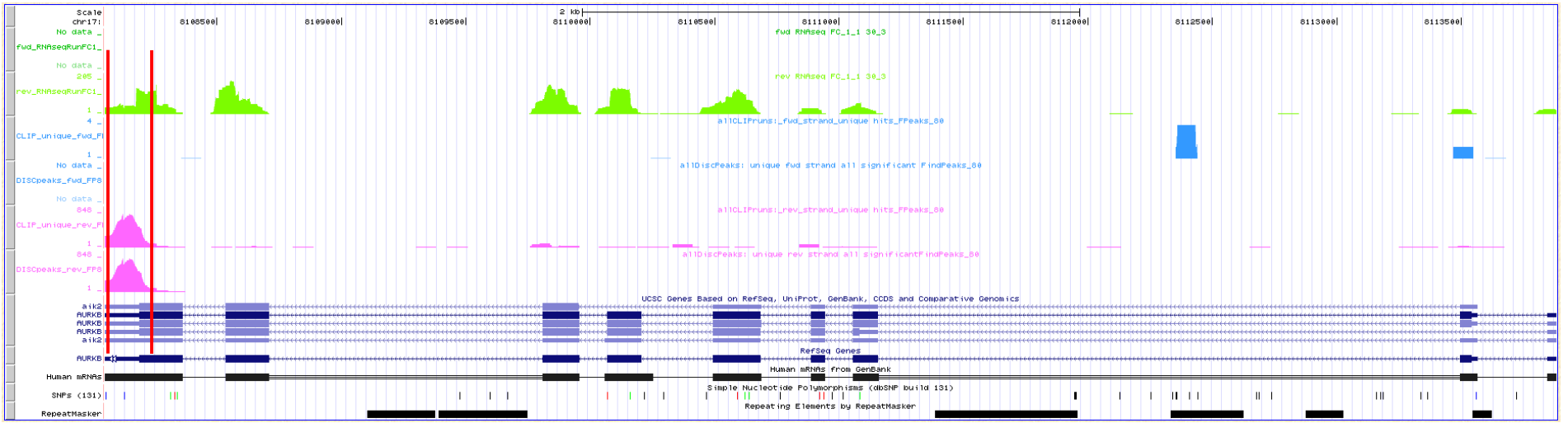
LMO4


chr1 (p22.3)
Scale
chr1:
10 kb
Human mRNAs from GenBank
U24576
AK292578
BC017673
BC003600
AY321514
CR599134
BC065818
DQ891447
DQ894626
AB384751
fwd_RNAseqRunFC1_
fwd RNAseq FC_1_1 30_3
rev_RNAseqRunFC1_
rev RNAseq FC_1_1 30_3
CLIP_unique_fwd_FI
allCLIPruns:_fwd_strand_unique hits_FPeaks_80
DISCpeaks_fwd_FP8
allDiscPeaks: unique fwd strand all significant FindPeaks_80
CLIP_unique_rev_FI
allCLIPruns:_rev_strand_unique hits_FPeaks_80
No data
DISCpeaks_rev_FP8
allDiscPeaks: unique rev strand all significantFindPeaks_80

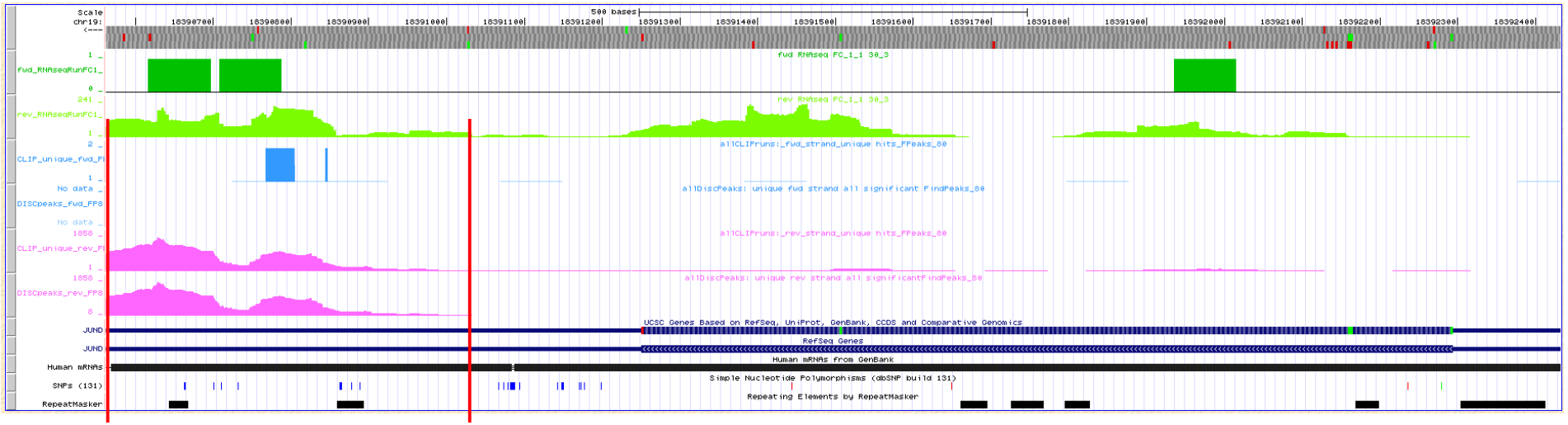
CDK1


Scale
chr10:
5 kb
fwd RNAseq FC_1_1 30_3
rev RNAseq FC_1_1 30_3
allCLIPruns:_fwd_strand_unique hits_FPeaks_80
allDiscPeaks: unique fwd strand all significant FindPeaks_80
allCLIPruns:_rev_strand_unique hits_FPeaks_80
allDiscPeaks: unique rev strand all significantFindPeaks_80
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics
CDC2
RefSeq Genes
CDK1
Human mRNAs from GenBank
Human mRNAs
Simple Nucleotide Polymorphisms (dbSNP build 131)
SNPs (131)
Repeating Elements by RepeatMasker
RepeatMasker

AURKB


Scale
chr17:
2 kb
fwd RNAseq FC_1_1 30_3
rev RNAseq FC_1_1 30_3
allCLIPruns:_fwd_strand_unique hits_FPeaks_80
allDiscPeaks: unique fwd strand all significant FindPeaks_80
allCLIPruns:_rev_strand_unique hits_FPeaks_80
allDiscPeaks: unique rev strand all significantFindPeaks_80
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics
RefSeq Genes
Human mRNAs from GenBank
Simple Nucleotide Polymorphisms (dbSNP build 131)
Repeating Elements by RepeatMasker

JUND


Scale
chr19:
500 bases
fwd RNAseq FC_1_1 30_3
rev RNAseq FC_1_1 30_3
allCLIPruns:_fwd_strand_unique hits_FPeaks_80
allDiscPeaks: unique fwd strand all significant FindPeaks_80
allCLIPruns:_rev_strand_unique hits_FPeaks_80
allDiscPeaks: unique rev strand all significantFindPeaks_80
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics
RefSeq Genes
Human mRNAs from GenBank
Simple Nucleotide Polymorphisms (dbSNP build 131)
Repeating Elements by RepeatMasker

ARE CELF1 BINDING TARGETS CONSERVED ACROSS SPECIES ?

Identification of CUG-BP1/EDEN-BP target mRNAs in *Xenopus tropicalis*

Antoine Graindorge[1], Olivier Le Tonquèze[1], Raphaël Thuret[2], Nicolas Pollet[2], H. Beverley Osborne[1] and Yann Audic[1,*]

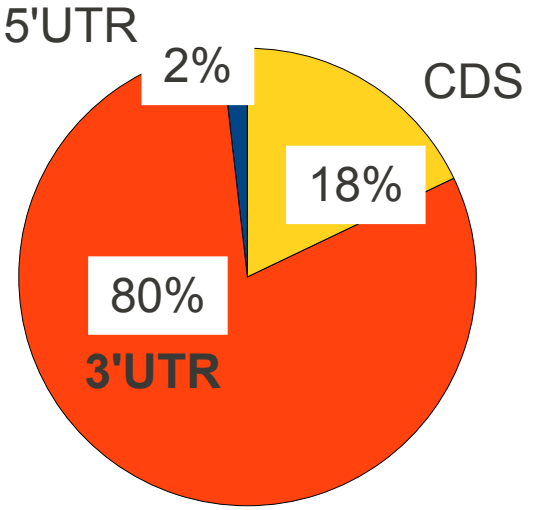
137 mRNAs (/ 3000 tested) with human orthologs are specifically enriched in CELF1 IP in XENOPUS → **43 are identified as CELF1 targets in Hela cells**

TRANSCRIPT-WIDE DISTRIBUTION OF CELF1 CLUSTERS

CELF1 CLIP vs RNA-SEQ

FINDPEAKS on exonic reads

2788 enriched clusters



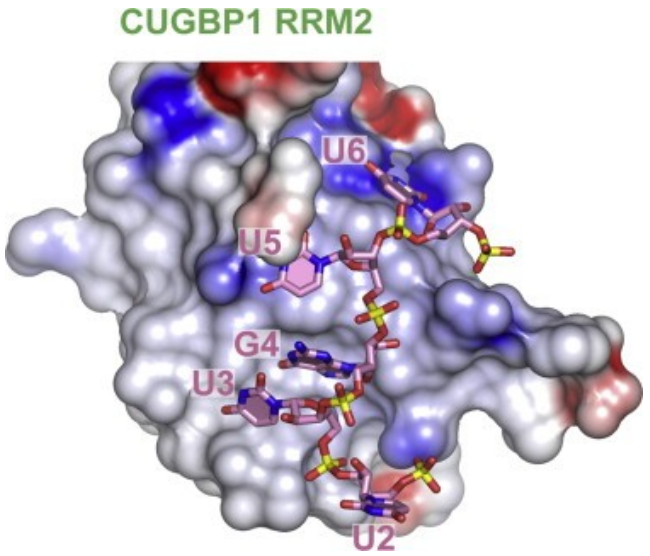
In mRNAs 3'UTR are the major binding sites for CELF1

ARE THE CELF 1 CLUSTERS ENRICHED IN
POTENTIAL BINDING SITES FOR CELF1 ?

CUG-BP1/CELF1 requires UGU-rich sequences for high-affinity binding

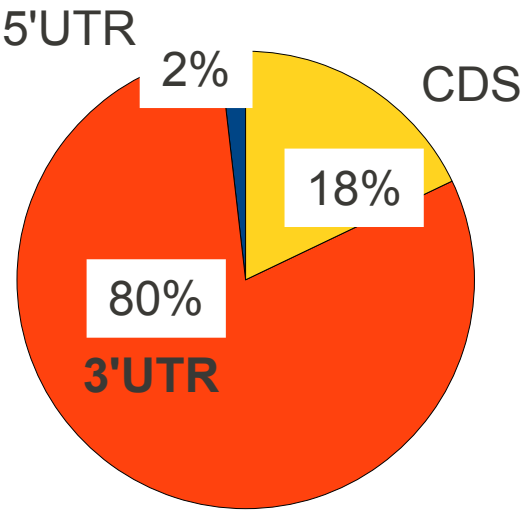
Julien MARQUIS*[1], Luc PAILLARD†, Yann AUDIC†, Bertrand COSSON†[2], Olivier DANOS*, Christine LE BEC* and H. Beverley OSBORNE†[3]

*Généthon, CNRS UMR 8115, 1 bis rue de l'Internationale 91002 Evry cedex 2, France, and †CNRS UMR 6061, Génétique et Développement, IFR 140 GFAS, Université de Rennes 1, Faculté de Médecine, 2 Avenue Léon Bernard, CS 34317, 35043 Rennes Cedex, France



Structural Insights into RNA Recognition by the Alternate-Splicing Regulator CUG-Binding Protein 1

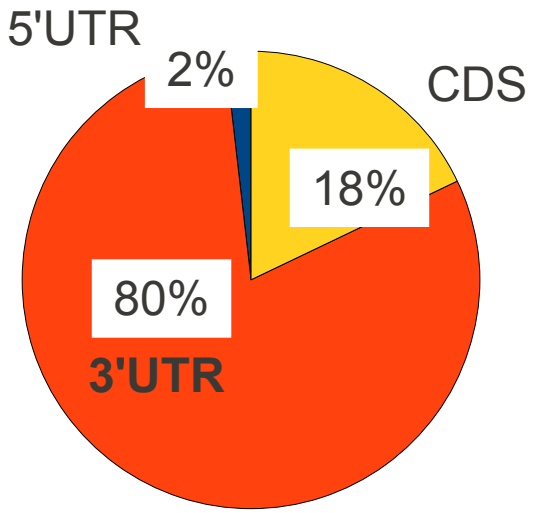
Marianna Teplova,[1] Jikui Song,[1] Hai Yan Gaw,[1] Alexei Teplov,[1] and Dinshaw J. Patel[1,*]



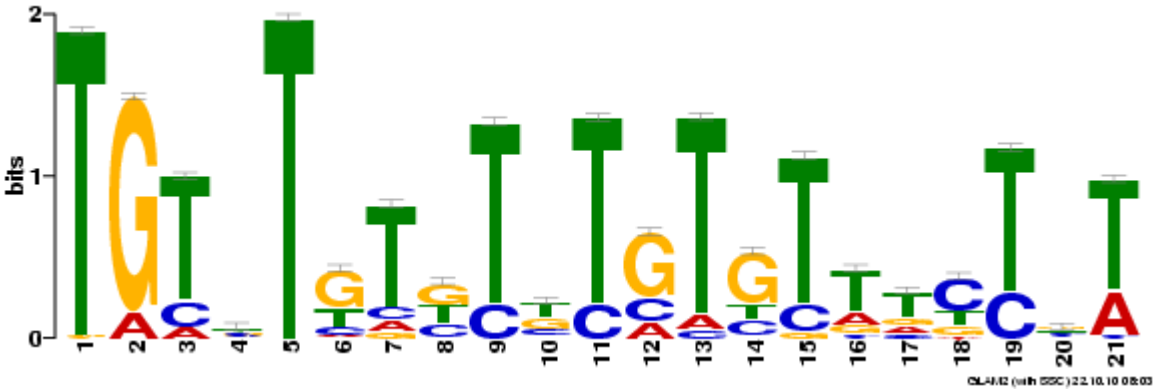
R'MES

Hoebeke, M. and Schbath, S. (2006). R'MES: Finding Exceptional Motifs,

word	count	expect	sigma2	score	rank
gtgtgt	1472	633.09	635.12	33.29	4096
tgtgtg	1674	850.57	798.05	29.15	4095
tttgtt	1429	878.34	765.65	19.9	4094
ttgttt	1402	878.34	765.65	18.92	4093
tatata	536	245.36	250.25	18.37	4092
cctccc	633	323.98	300.65	17.82	4091



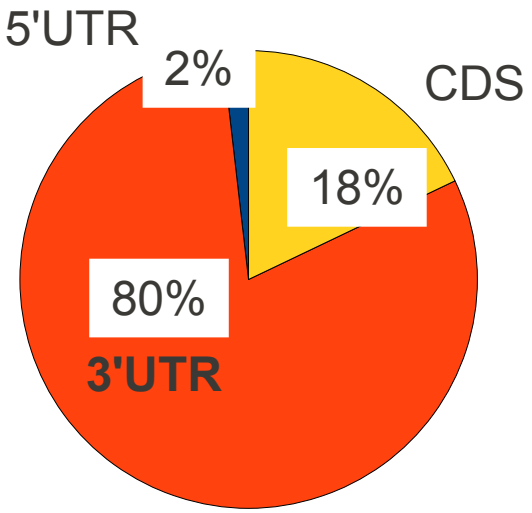
Gapped Local Alignment of Motifs
on TOP 100 CLUSTERS



TRANSCRIPT-WIDE DISTRIBUTION OF CELF1 CLUSTERS

CELF1 CLIP vs RNA-SEQ

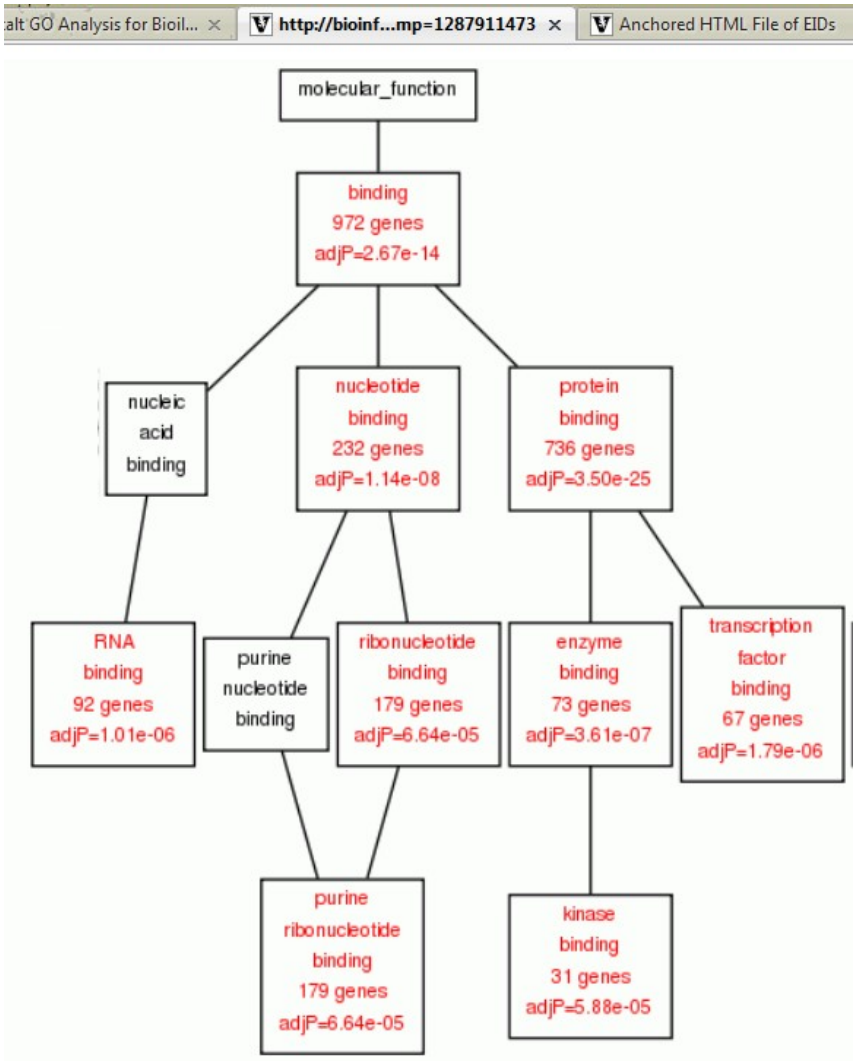
FINDPEAKS on exonic reads



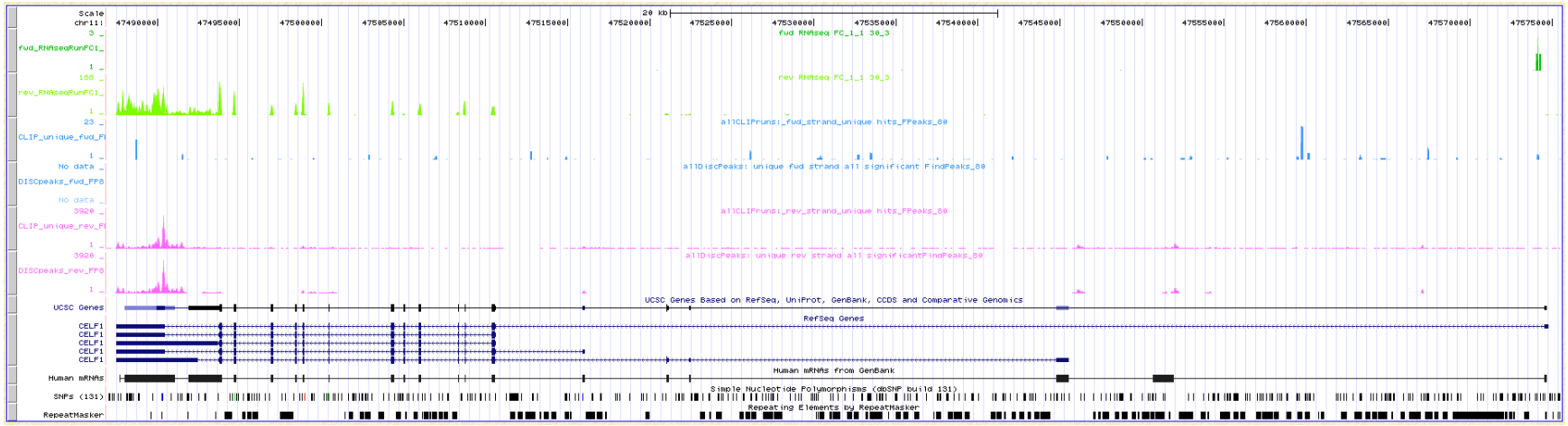
IS THERE ANY ENRICHMENT IS SOME BIOLOGICAL FUNCTIONS ASSOCIATED WITH CELF1 TARGETS ?

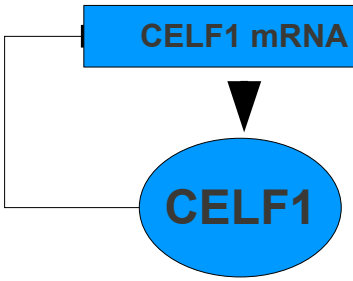
1205 GENES

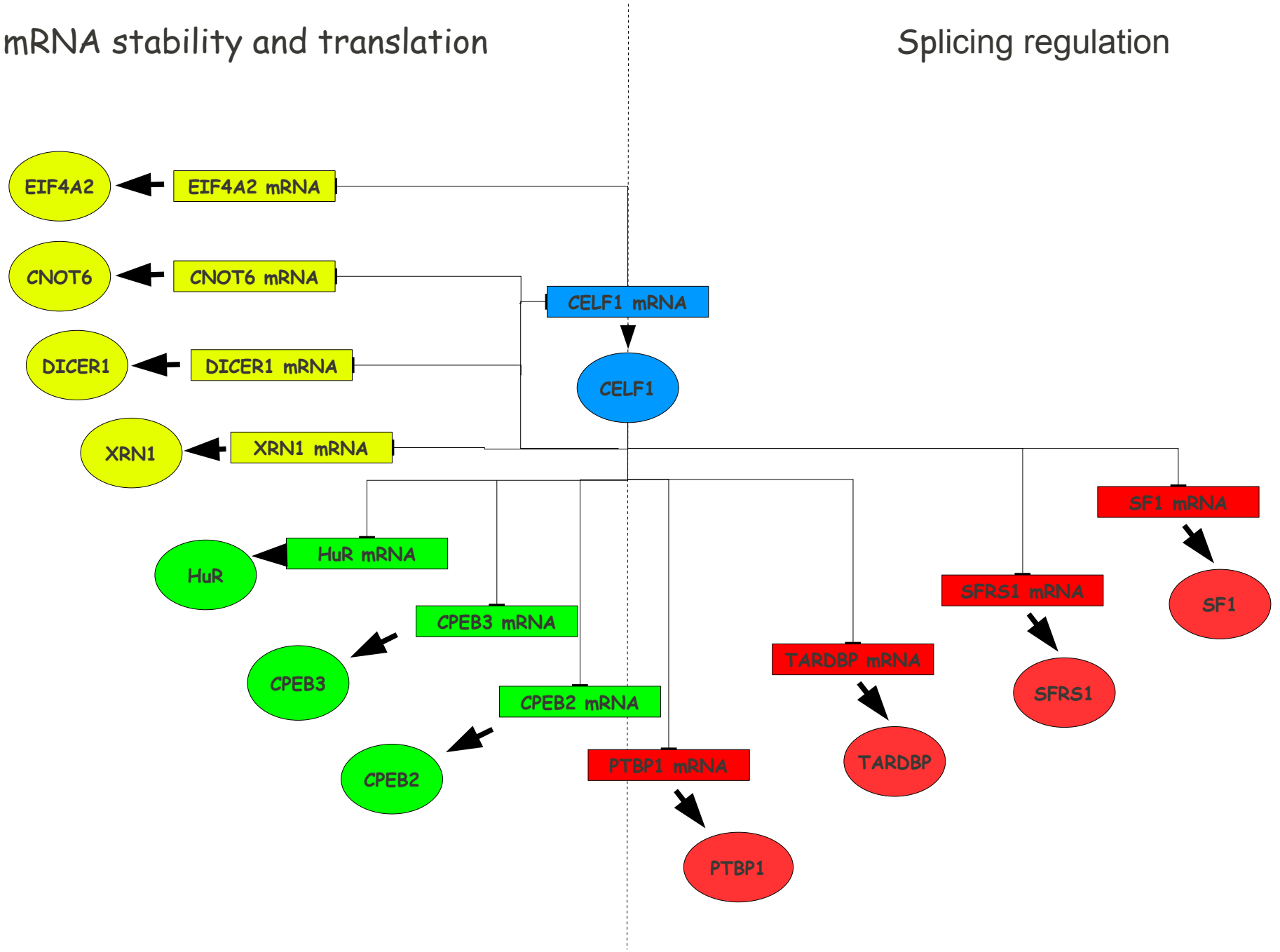
WebGestalt



CELF1


Scale
chr11:
20 kb
fwd RNAseq FC_1_1 30_3
rev RNAseq FC_1_1 30_3
allCLIPruns:_fwd_strand_unique hits_FPeaks_80
allDiscPeaks: unique fwd strand all significant FindPeaks_80
allCLIPruns:_rev_strand_unique hits_FPeaks_80
allDiscPeaks: unique rev strand all significantFindPeaks_80
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics
RefSeq Genes
CELF1
Human mRNAs from GenBank
Simple Nucleotide Polymorphisms (dbSNP build 131)
Repeating Elements by RepeatMasker


CELF1 mRNA
CELF1


mRNA stability and translation
Splicing regulation
EIF4A2
EIF4A2 mRNA
CNOT6
CNOT6 mRNA
DICER1
DICER1 mRNA
XRN1
XRN1 mRNA
CELF1 mRNA
CELF1
HuR
HuR mRNA
CPEB3 mRNA
CPEB3
CPEB2 mRNA
CPEB2
PTBP1 mRNA
PTBP1
TARDBP mRNA
TARDBP
SFRS1 mRNA
SFRS1
SF1 mRNA
SF1

CONCLUSIONS

- Genome wide landscape for CELF1 mRNAs targets

- Widespread association of CELF1 with mRNAs

- Enrichement for potential CELF1 binding motif in the target sequences

- CELF 1 is probably self regulatory

- CELF 1 binds (controls ?) a large number of RNA BP involved in cytoplasmic and/or nuclear regulation of gene expressions

PERSPECTIVES

WHAT ARE THE FUNCTIONAL CONSEQUENCES OF CELF 1 BINDING ?

- MOLECULAR CONSEQUENCES :
  - TRANSCRIPTOME
  - POLYSOMAL ANALYSIS
  - SPLICING

-

CNRS-UMR6061 Genetic and Development

Gene Expression and Development

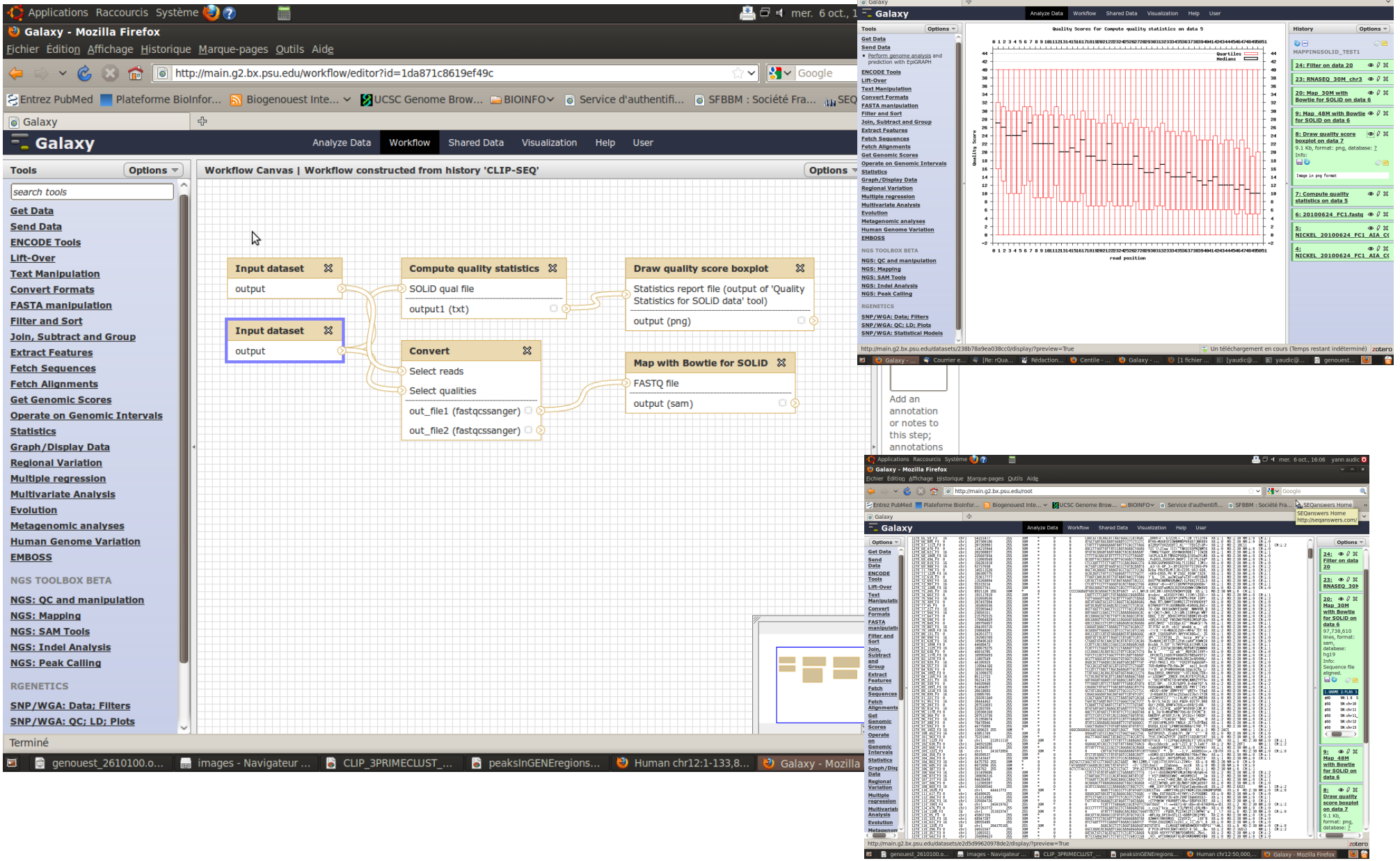
Carole Gautier-Courteille
Serge Hardy
Vincent Legagneux
Olivier Le Tonquèze
Hubert Lerivray
Stephanie Mottier
Agnès Méreau
Marie Cibois
Luc Paillard
Maud Noiret
Gaëlla Boulanger
Yann Audic
Bernhard Gschloessl

former lab member

Antoine Graindorge

Corinne Cruaud
Corine Da Silva

ET MAINTENANT ?
MAPPING -> GALAXY FRAMEWORK
QUANTIFICATION RNA SEQ -> GALAXY FRAMEWORK (RQUANT /RDIFF)



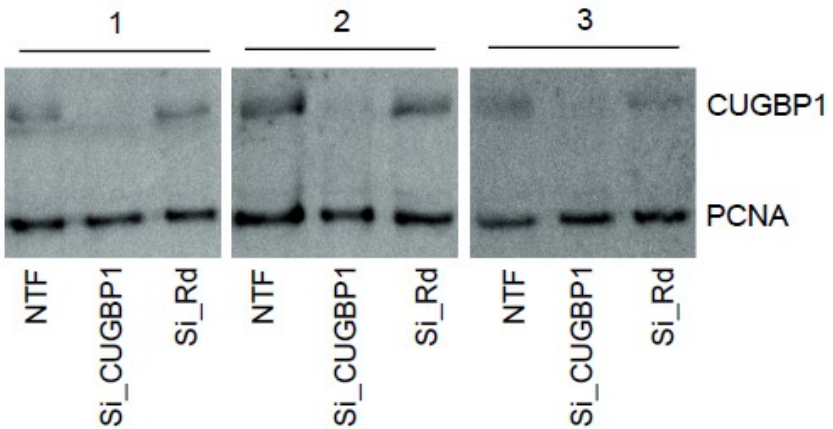
Secondary (N) Molecular defects = **INDIRECT TARGETS**

Search for GENES with altered expression
- microarray analysis on CELF1 deprived cells

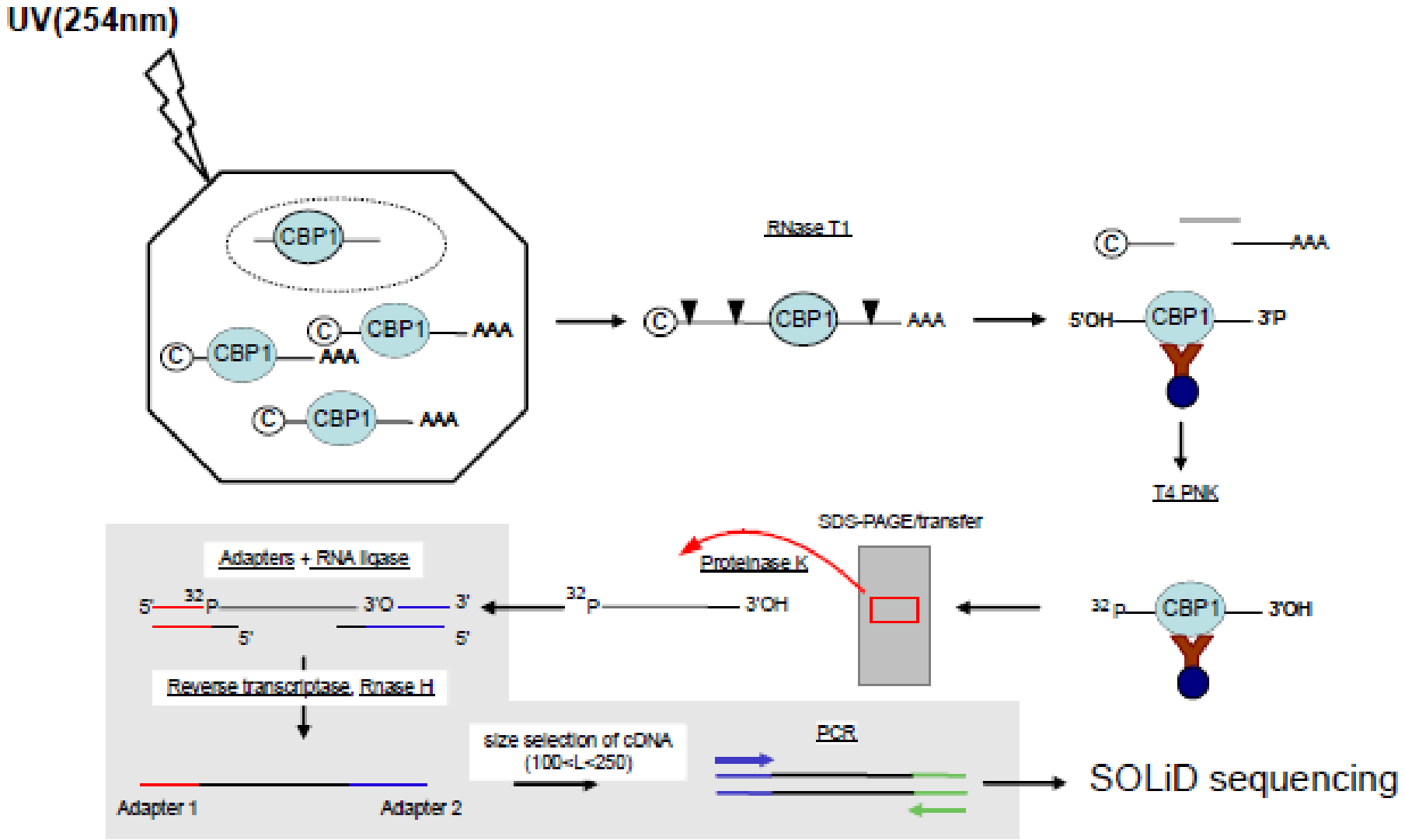
Hela Cells → SiRNA treatment → TOTAL RNA → ILLUMINA HT12V4 (DNAVISION)

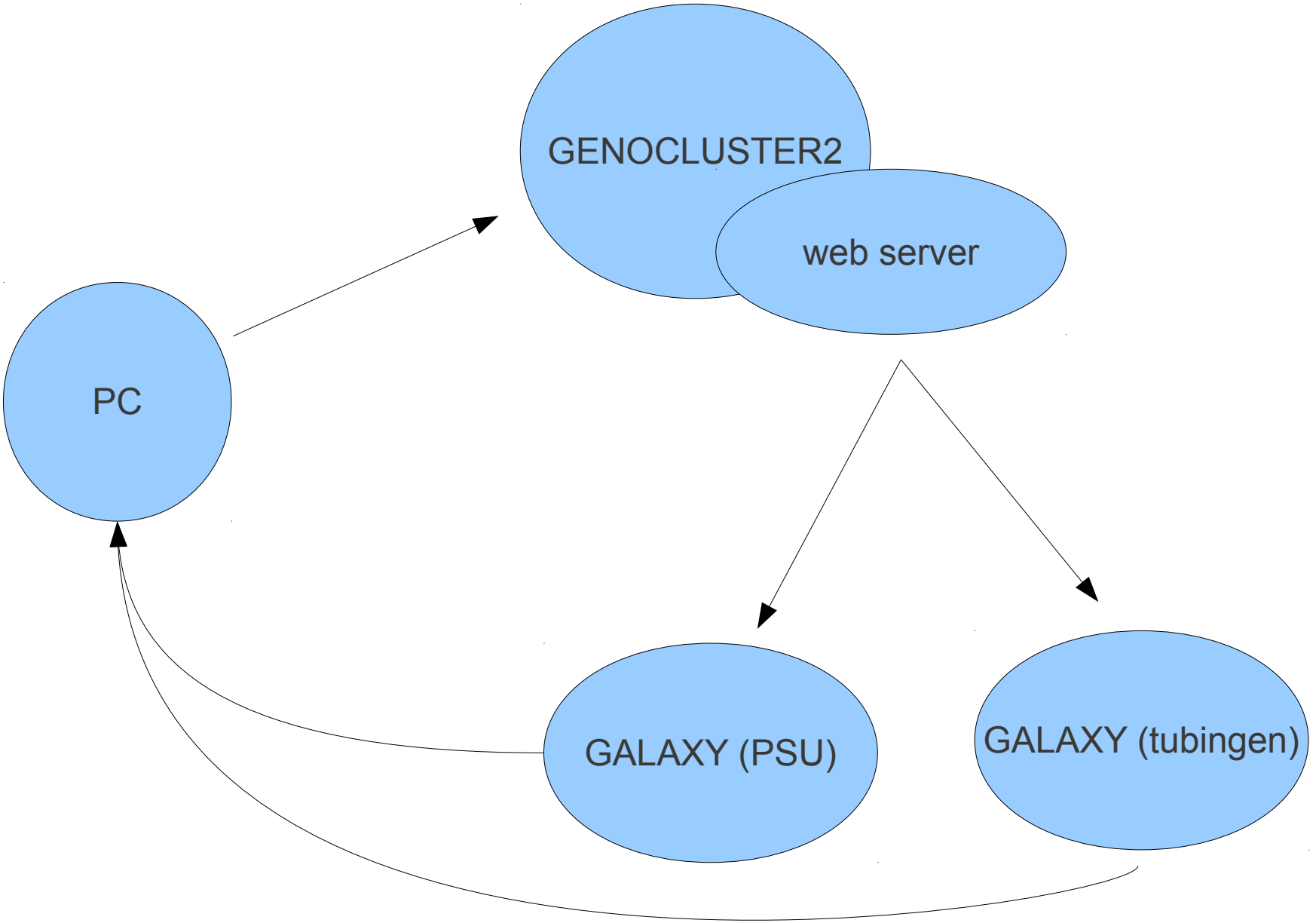
Significance **A**nalysis of **M**icroarrays

PROTEINS



CLIP procedure




GENOCLUSTER2
web server
PC
GALAXY (PSU)
GALAXY (tubingen)