

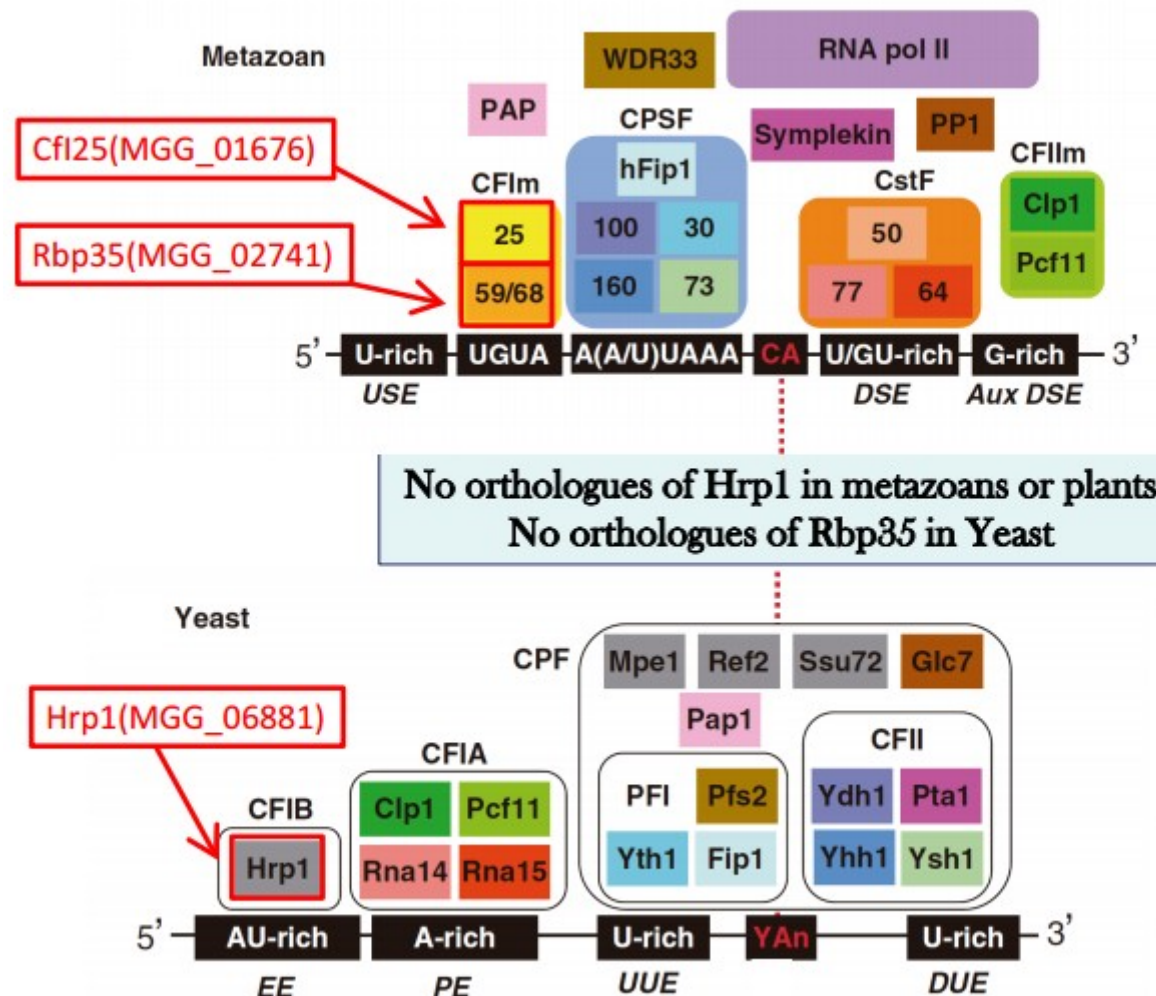
# Bioinformatic approaches to understand RNA biology in the rice blast fungus

Marco Marconi  
[marco.marconi@upm.es](mailto:marco.marconi@upm.es)

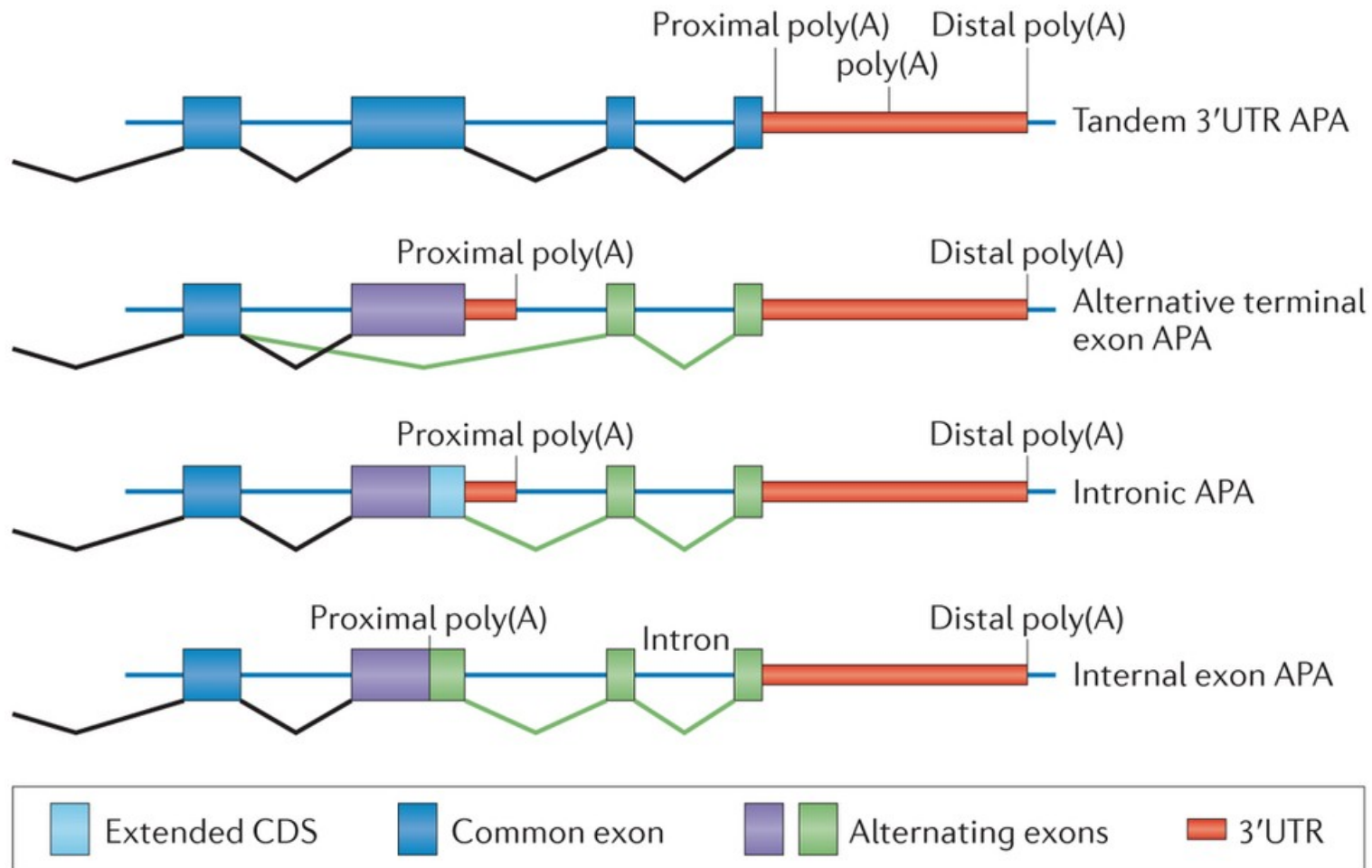
*Supervisor:*  
Mark Wilkinson

*Co-Supervisor:*  
Ane Sesma

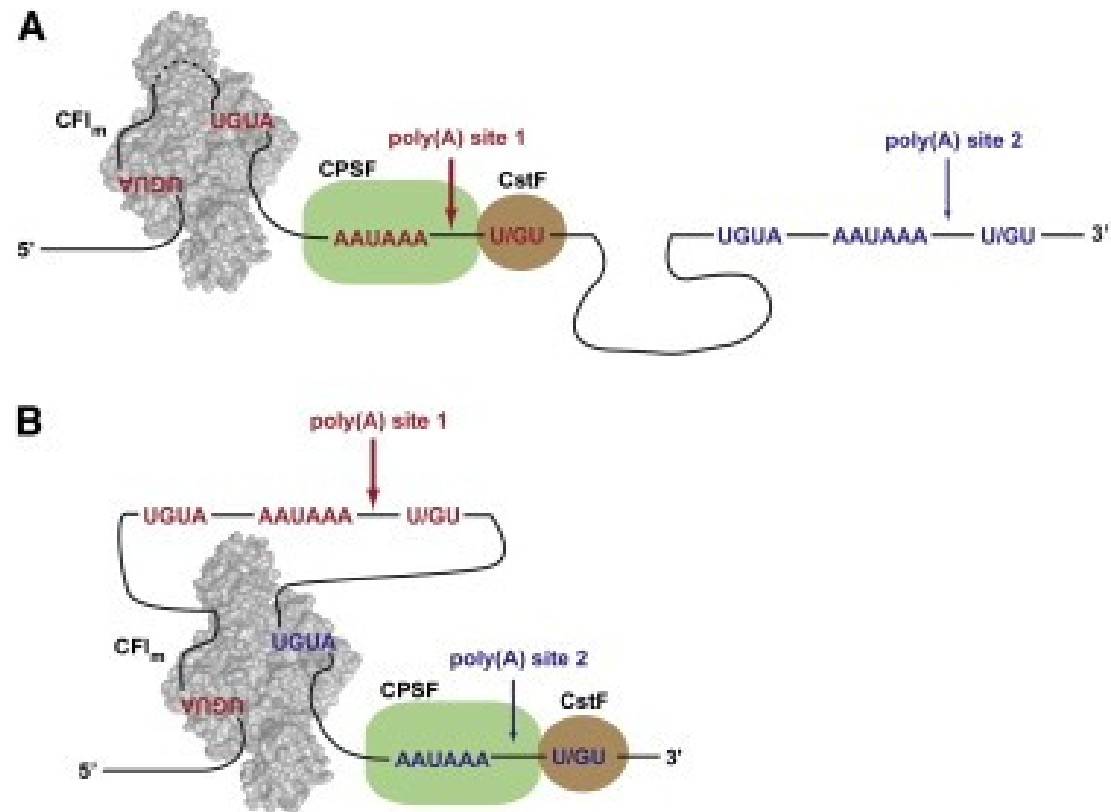
*M. oryzae* contains metazoan CFI (Rbp35/CfI25) and yeast CFI (Hrp1) protein complexes



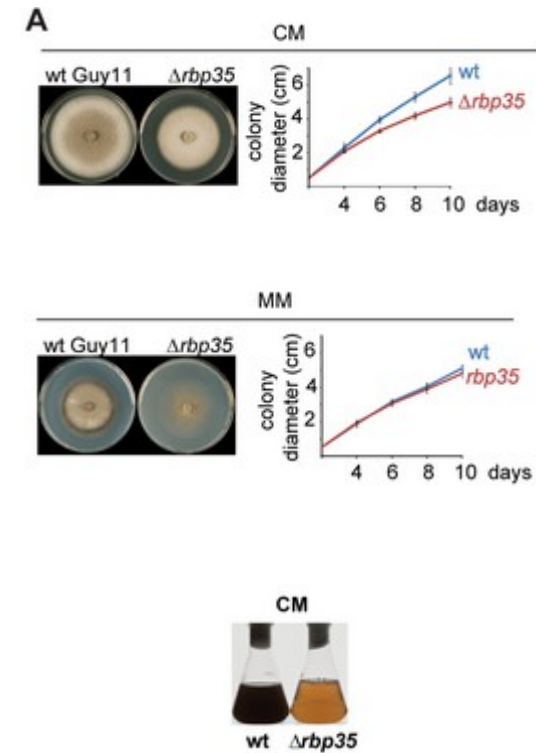
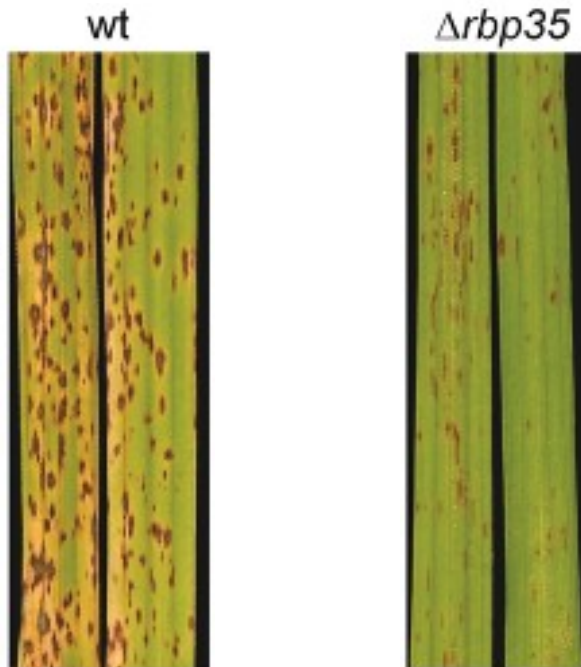
# Alternative polyadenylation is common in many organisms



# In human, RBP35 binds to UGUA motifs and affects alternative polyadenylation



# $\Delta rbp35$ is defective in pathogenicity and growth

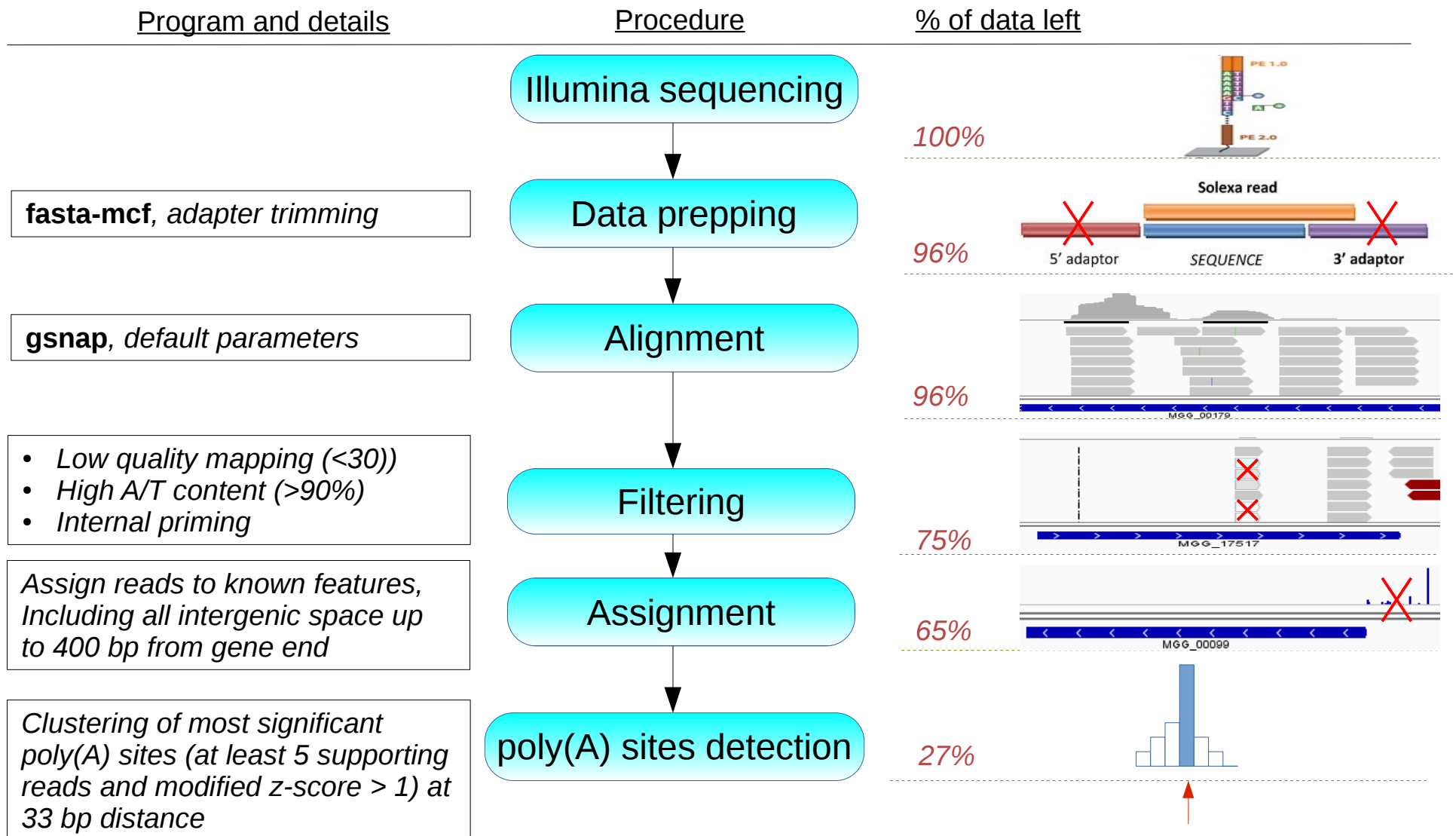


# Sequencing resume

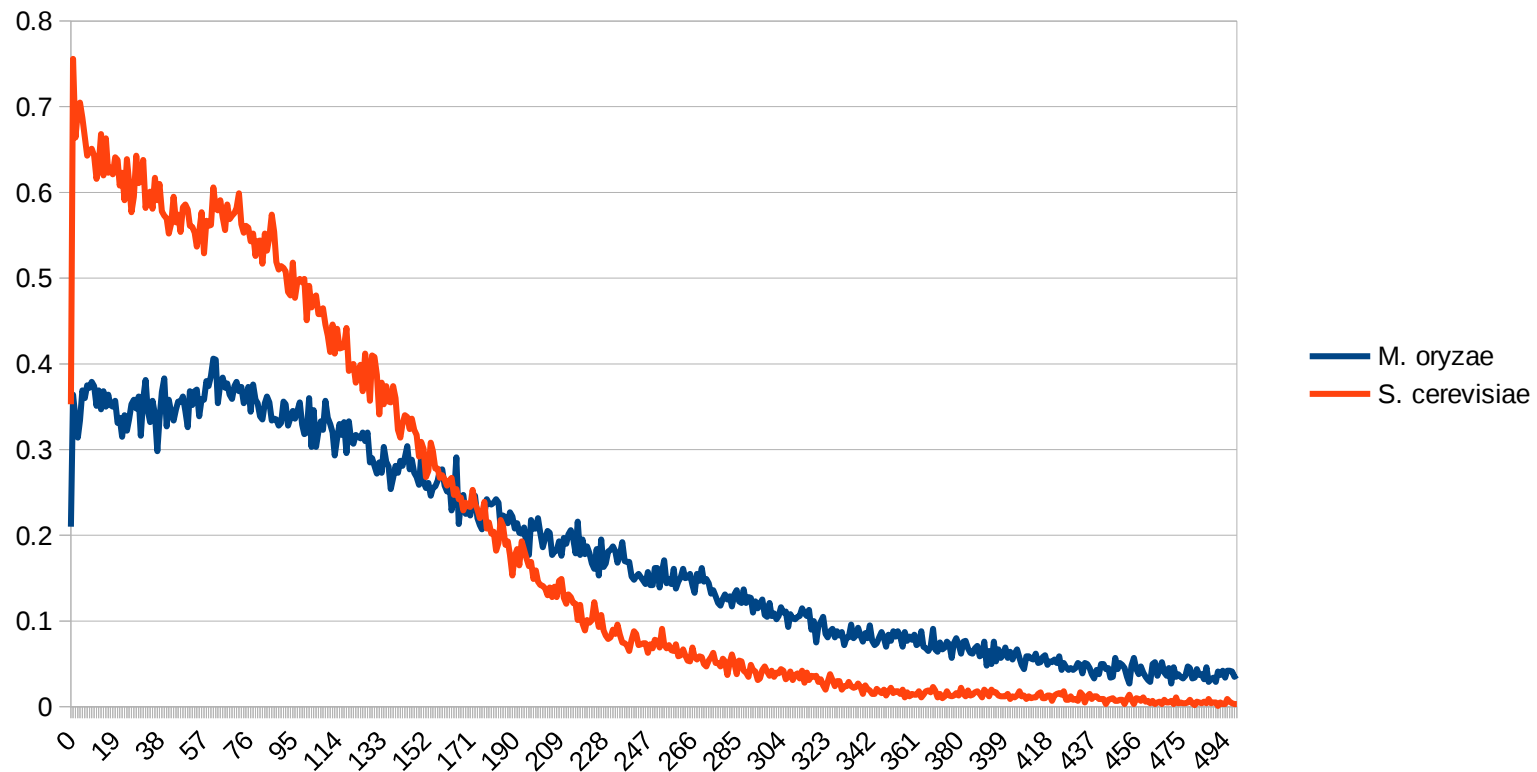
- 2 strains (*WT*,  $\Delta rbp35$ ) x 4 conditions (*CM*, *MM*, *-N*, *-C*) x 3 replicates
- HiSeq 2000 (Illumina) sequencing – poly(A) sites selection
- 4751592 – 11517077 total reads database
- ~95% replicates correlation
- ~73% successfully mapped reads
- ~400x coverage per poly(A) site\*

\* assuming an amount of 22000 mRNA molecules per cell

# Workflow



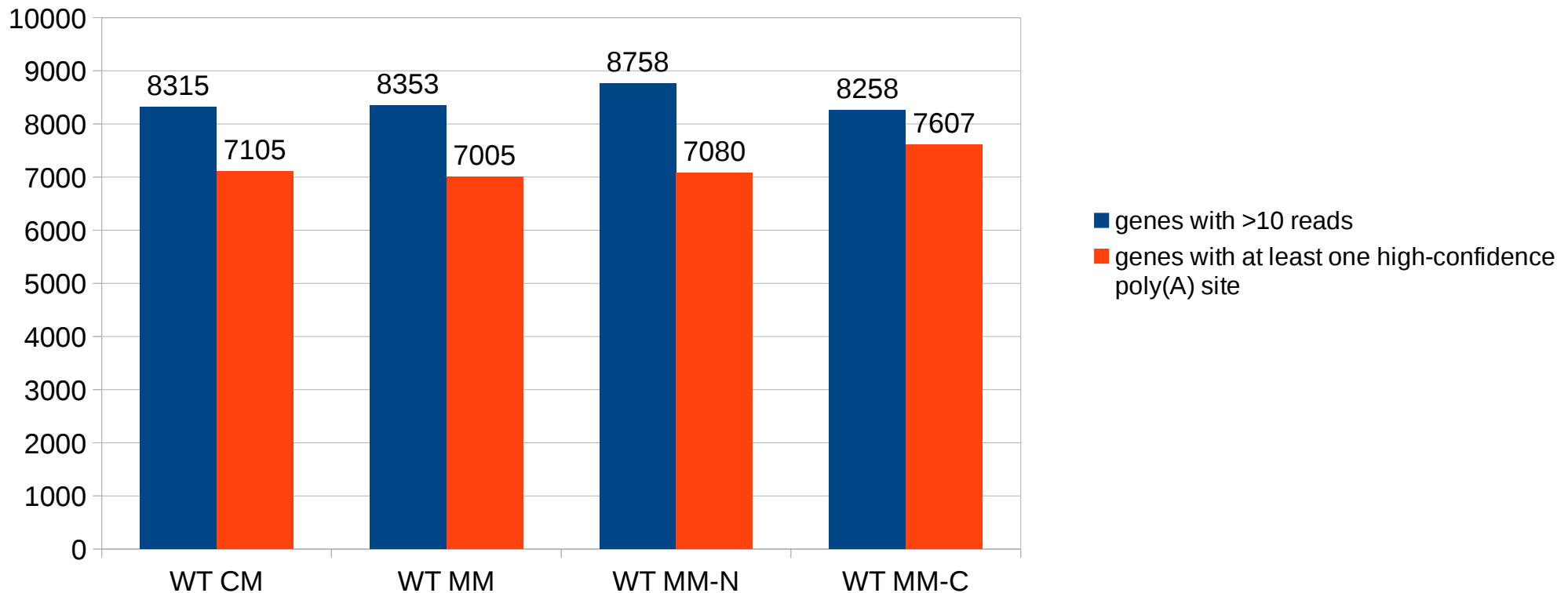
# *M. oryzae* has longer 3'UTRs than *S. cerevisiae*



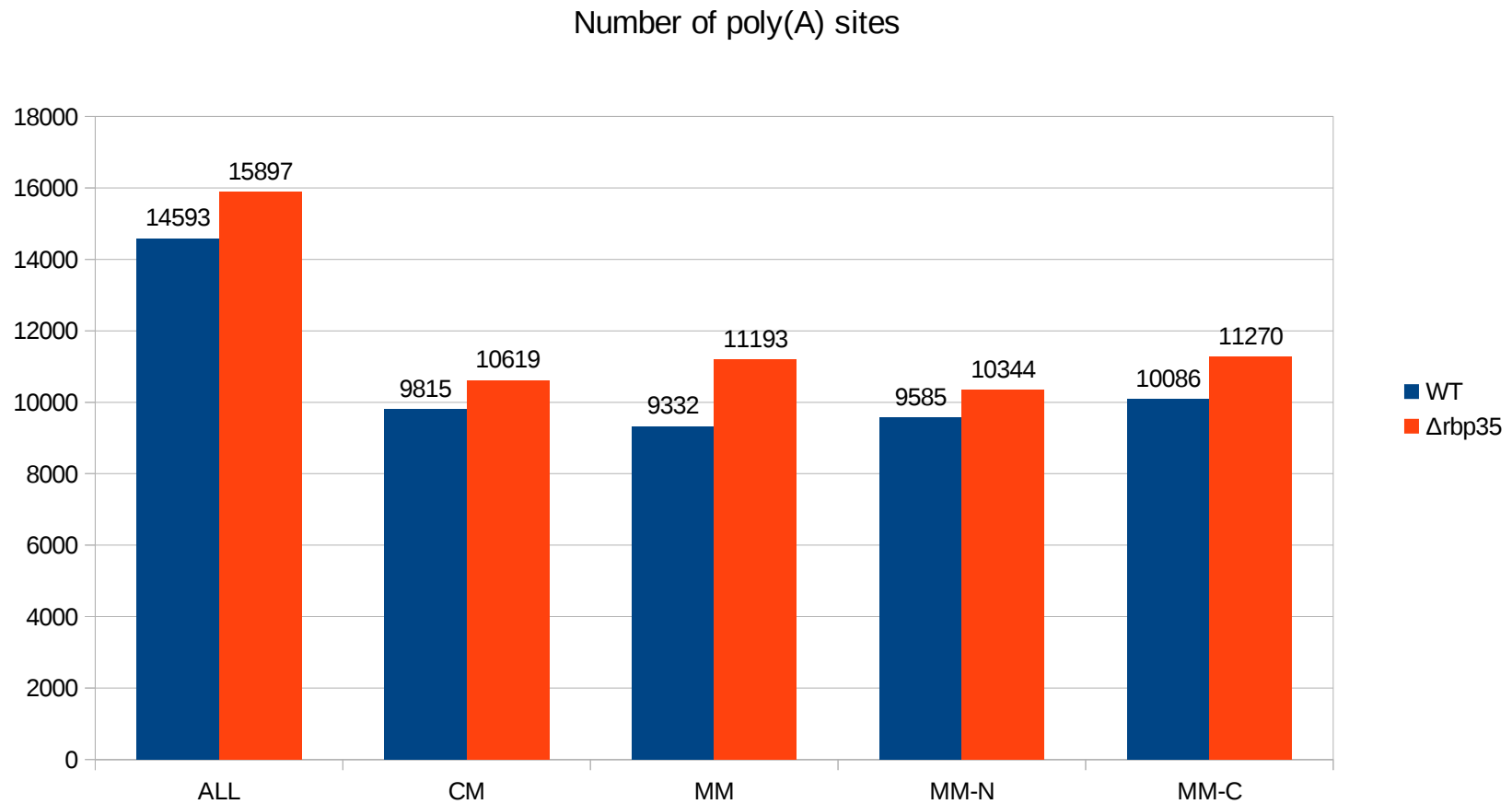


~85% of genes expressed have a recognizable poly(A) site

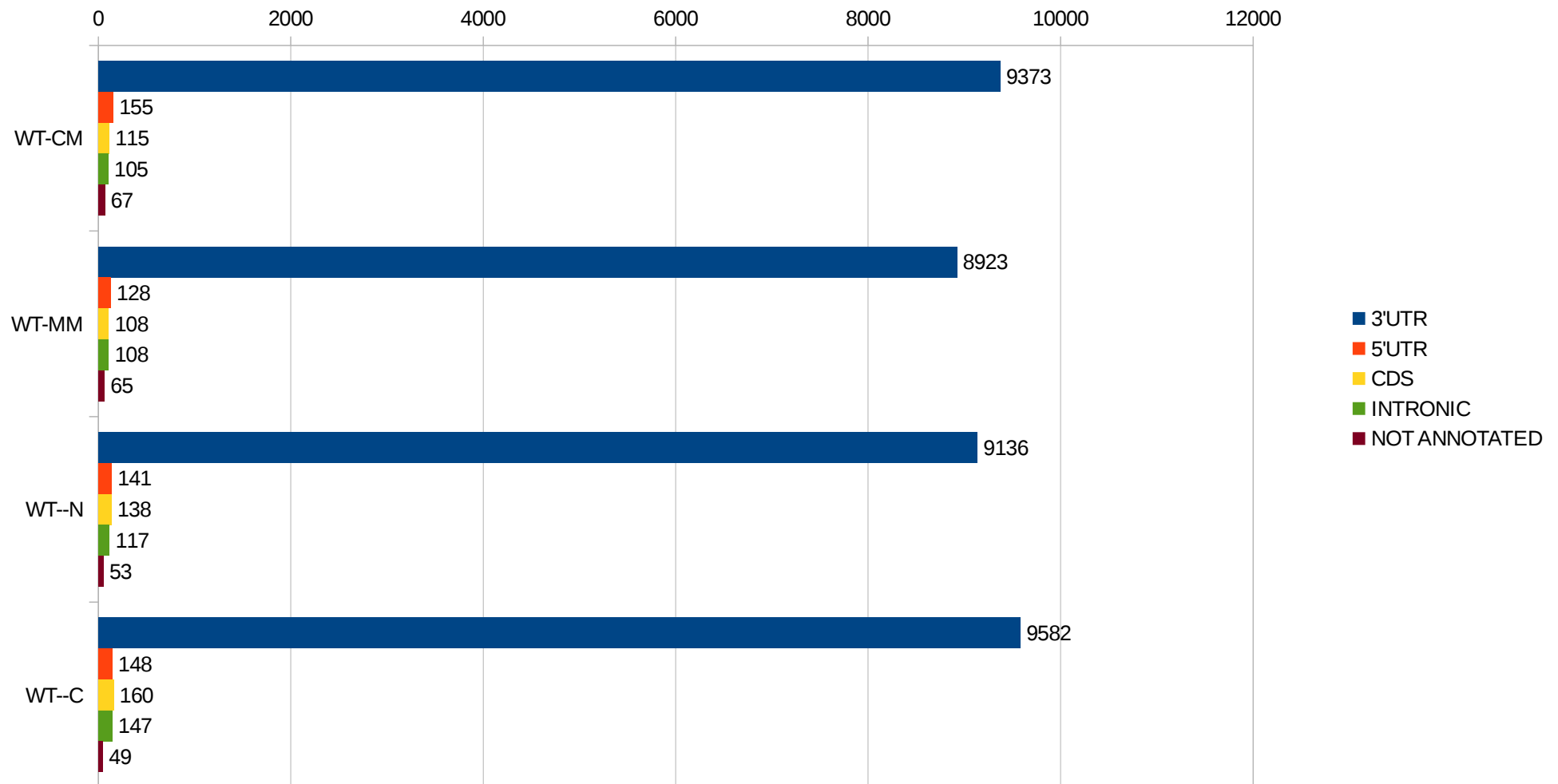
Genes with a recognizable poly-A site



~14000 poly(A) site could be assigned to annotated genes

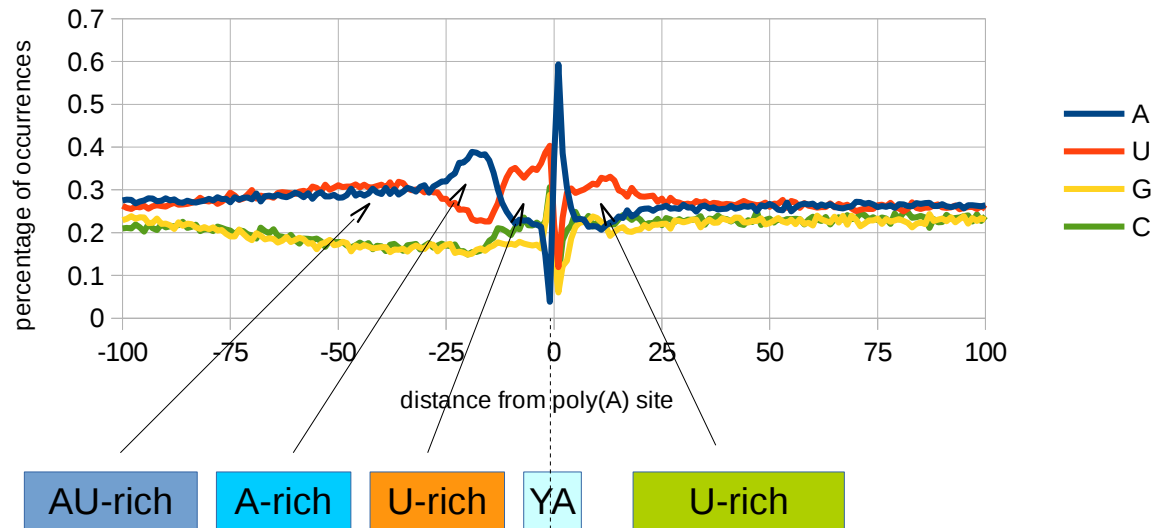


>90% of poly(A) sites are located in the 3'UTR

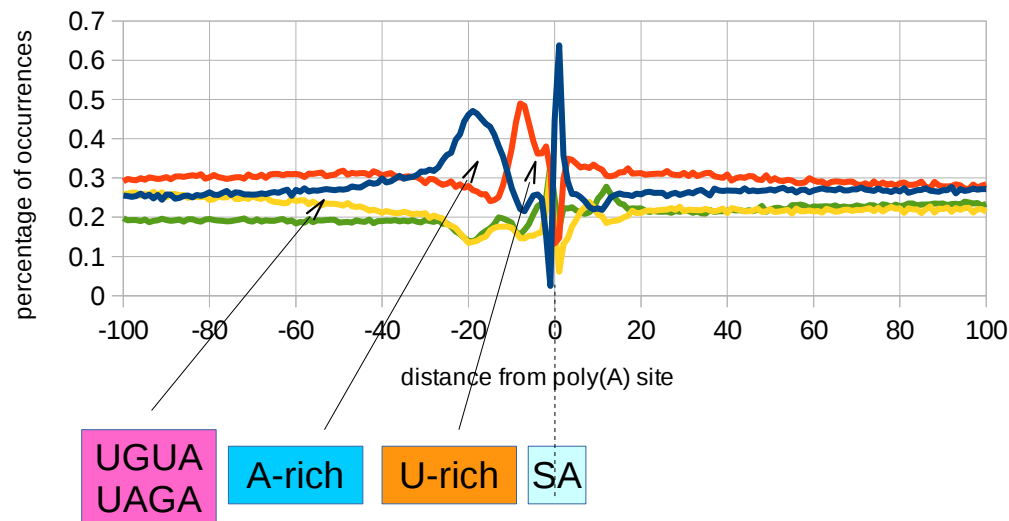


# Nucleotides profile of poly(A) sites slightly differs from *S.cerevisiae*

Poly-A site nucleotide profile - *S. cerevisiae*

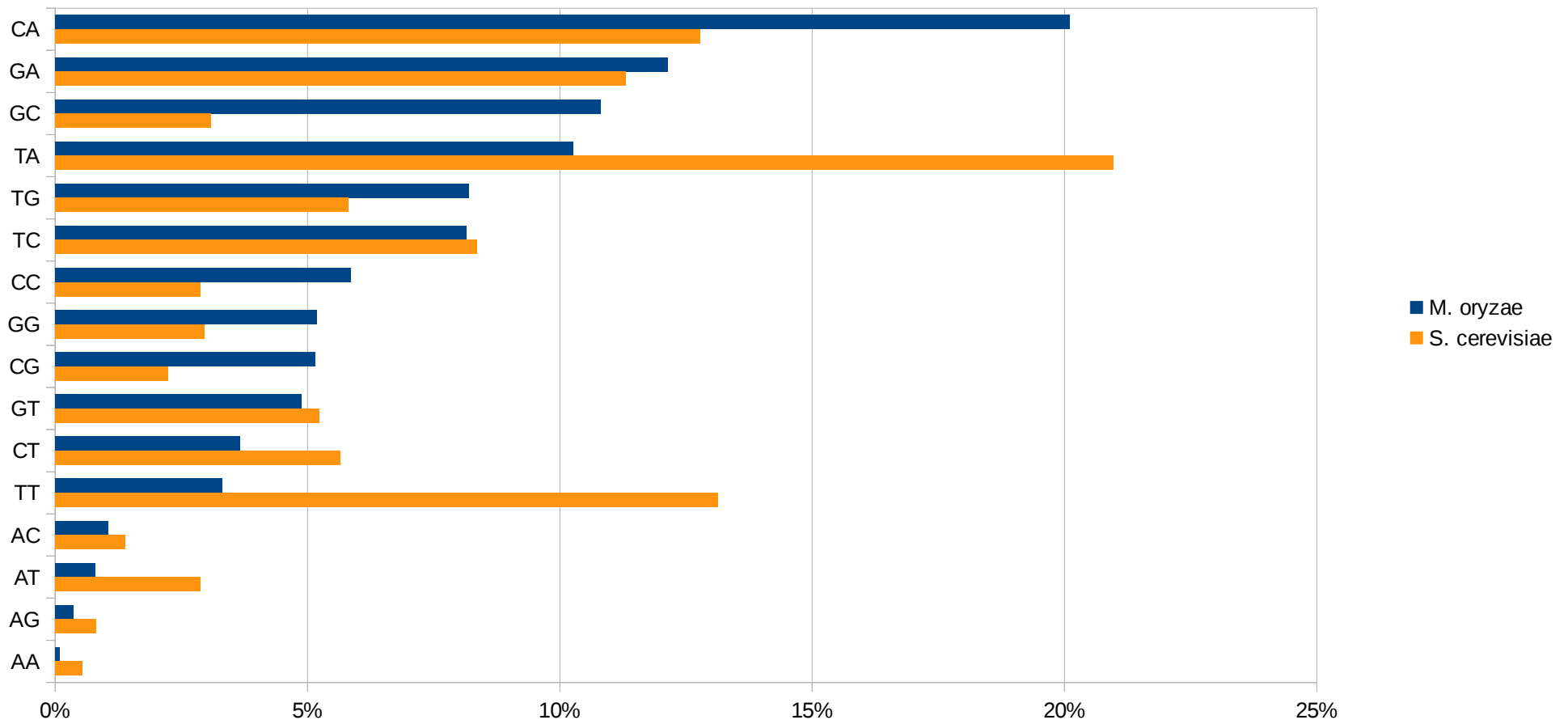


Poly(A) site nucleotide profile - *M. Oryzae*

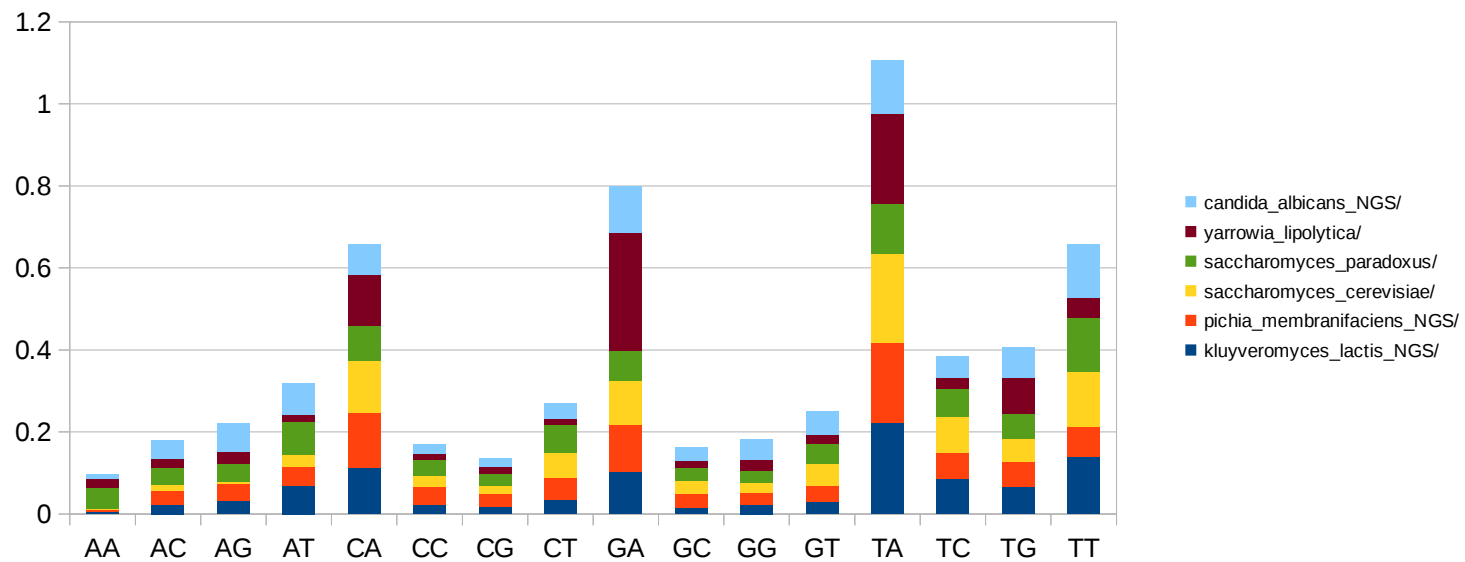
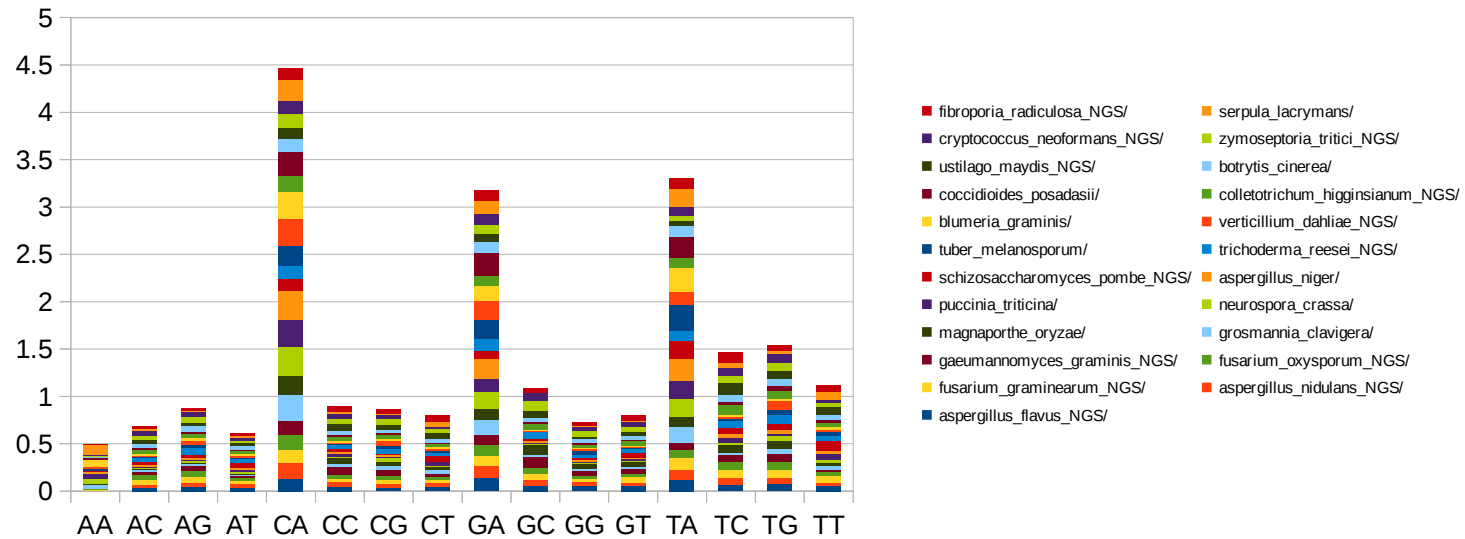


# *M. oryzae* prefers SA as cutting-site instead of YA

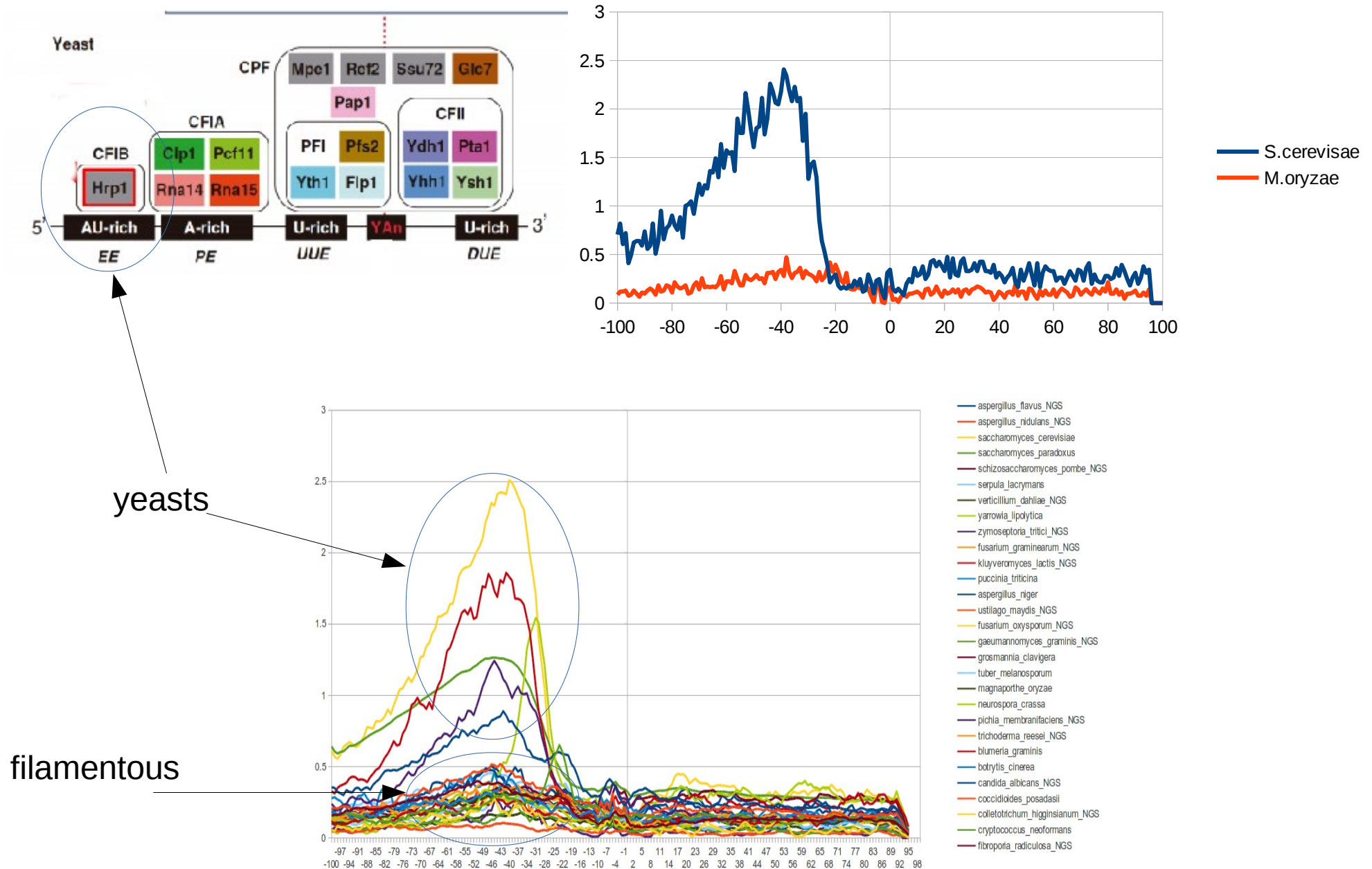
TOP CLEAVAGE SITES - *M. oryzae* vs *S. cerevisiae*



CA is the preferred cut-site in filamentous fungi, TA in yeasts

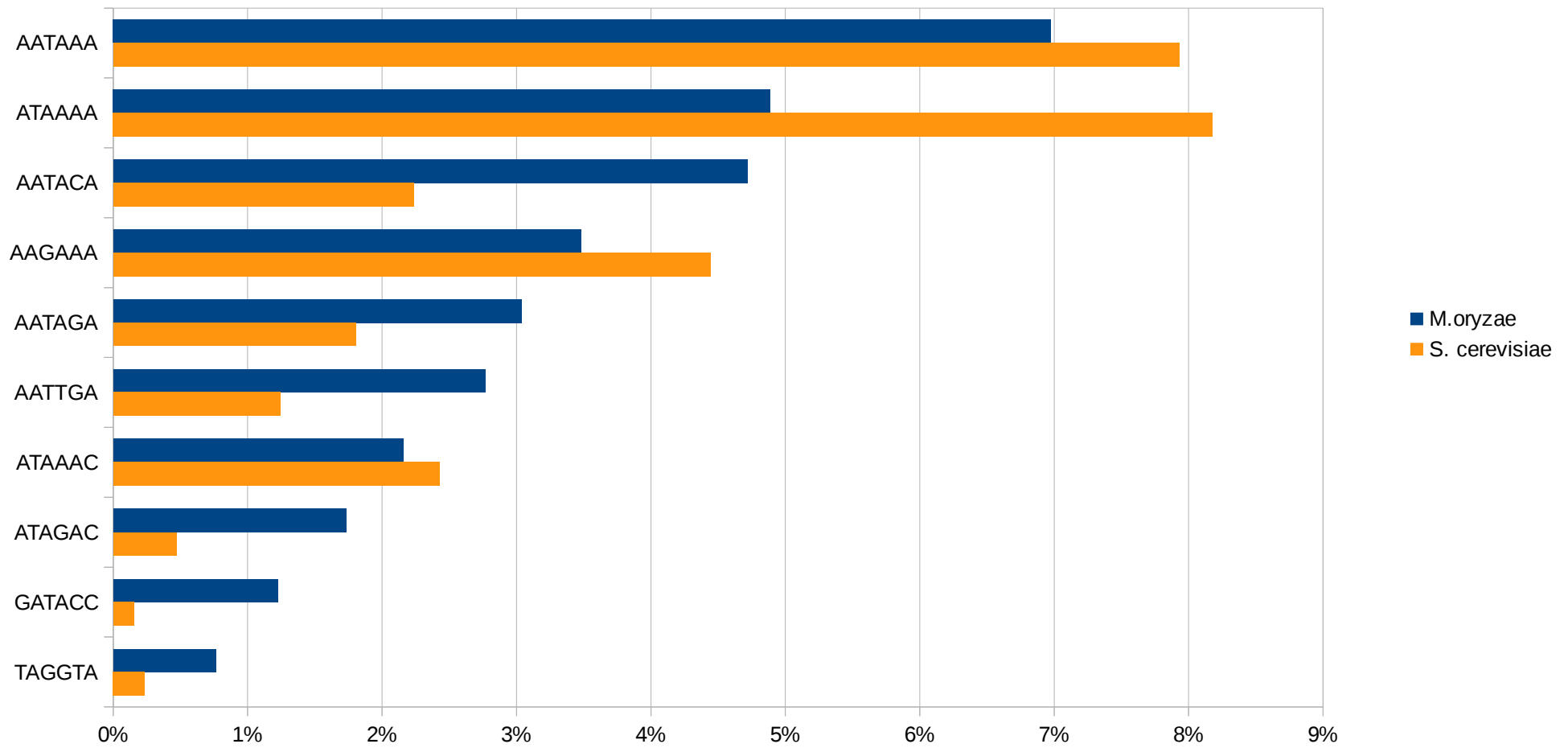


# The motif bound by HRP1 in yeast is missing in filamentous fungi



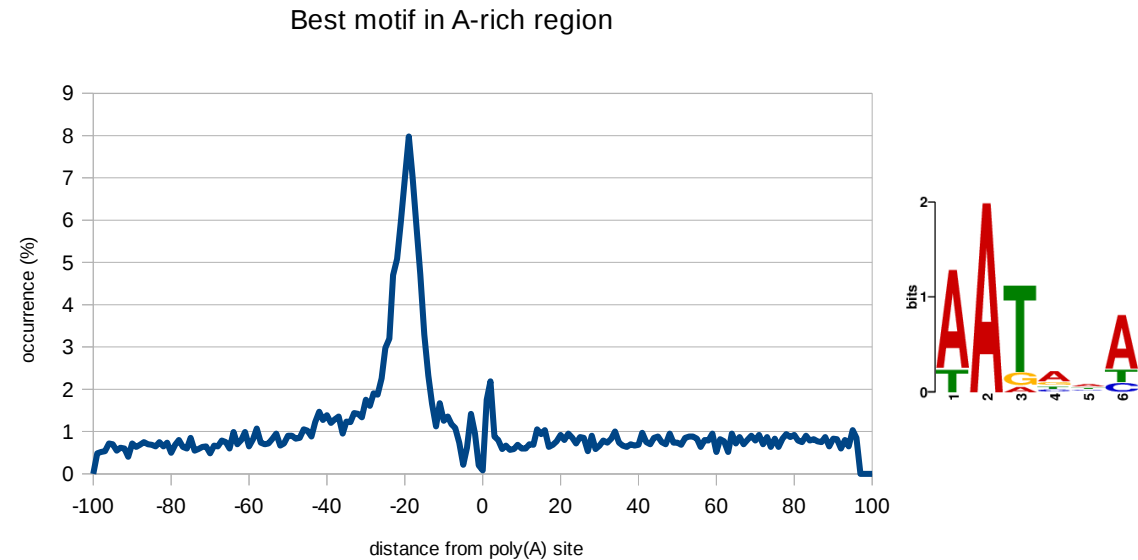
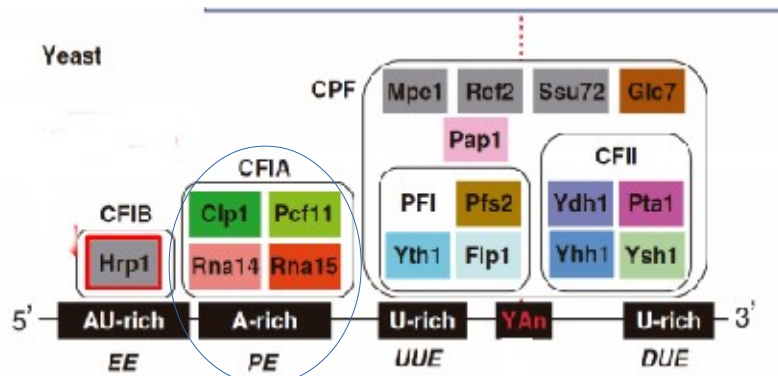
# The canonical metazoan AAUAAA motif has a frequency of only 7% in *M. oryzae*

TOP 10 MOST SIGNIFICANT HEXAMERS in A-RICH REGION



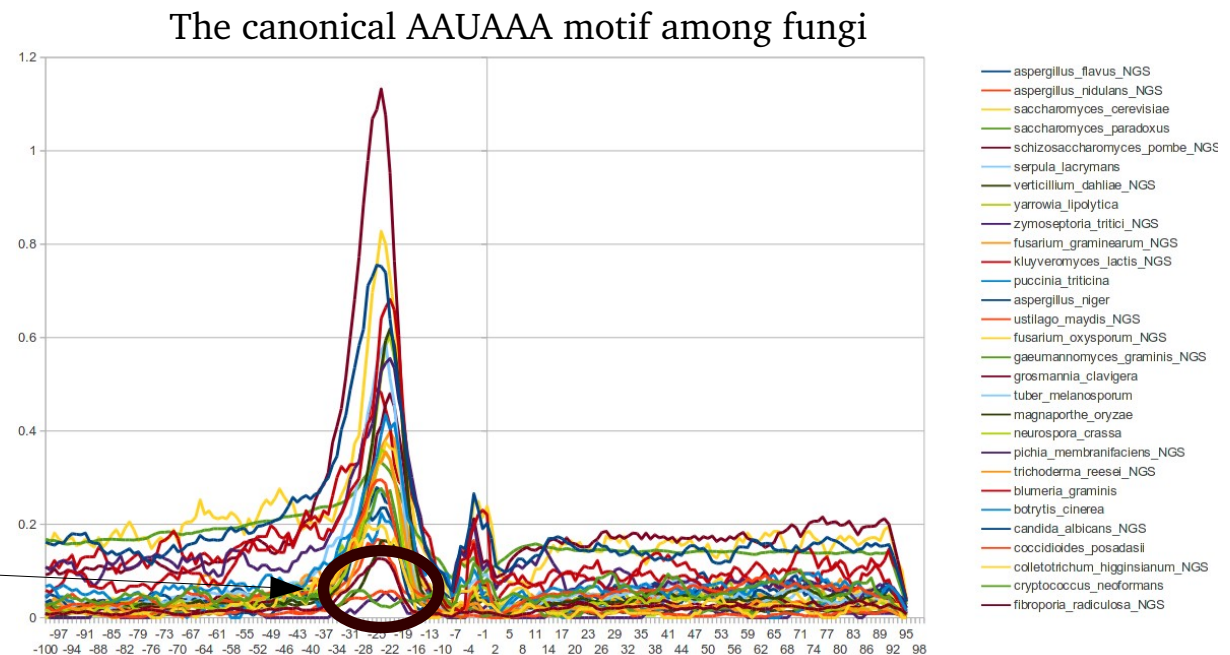


The A-RICH region is located -30 -10 bp upstream in most fungi



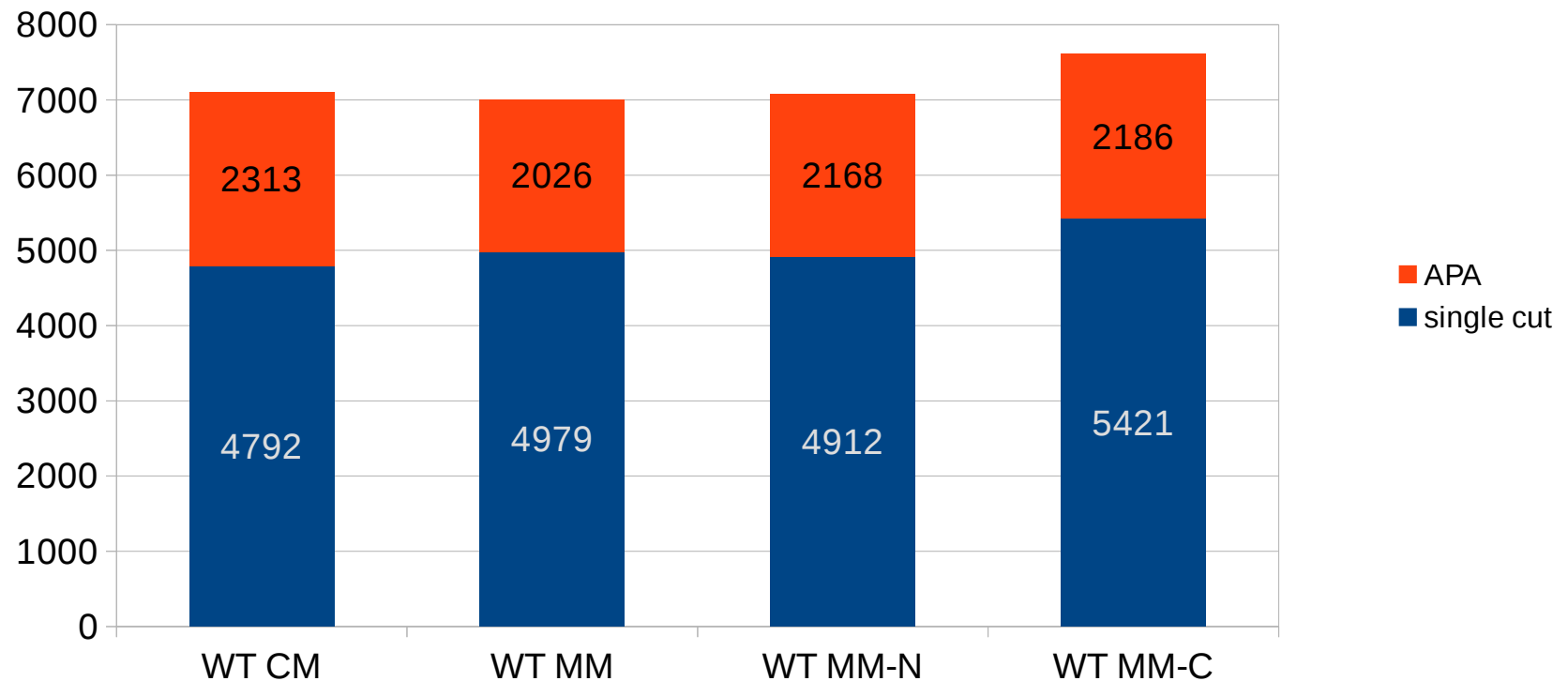
70% is AAUAAA in vertebrates

# Basidiomycota?



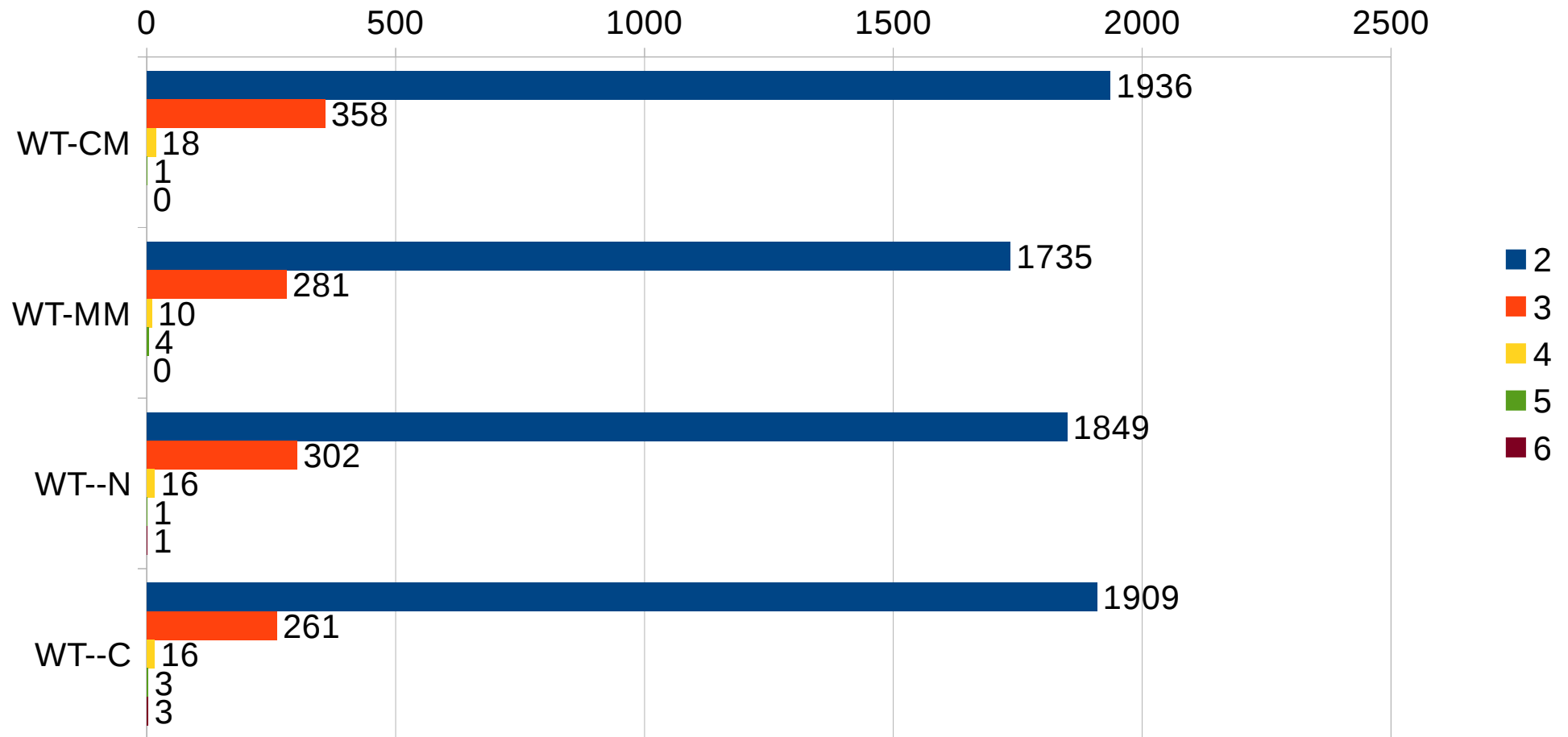
~30% of genes are alternatively polyadenylated

Number of genes with single cut or APA\*

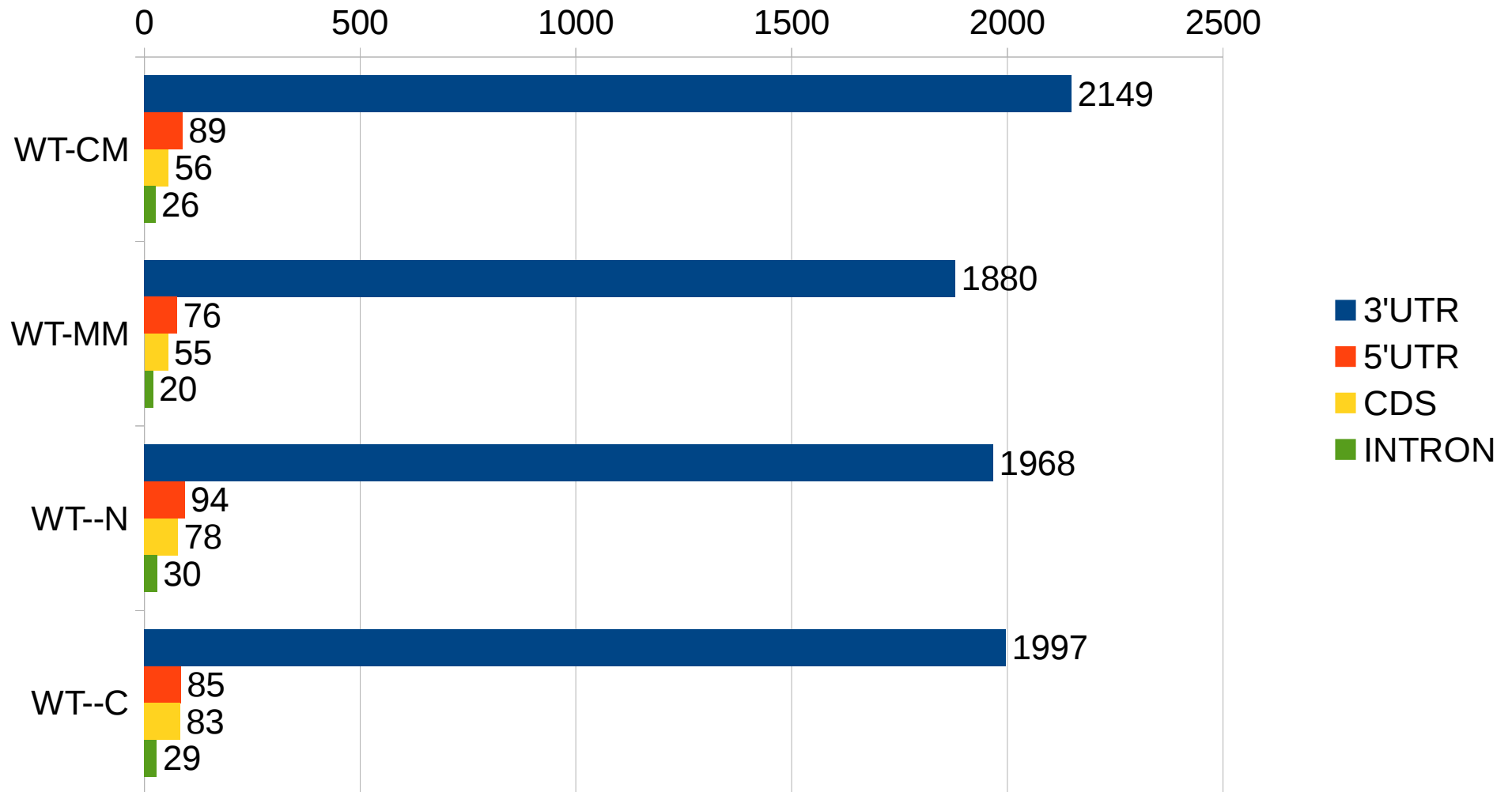


\* calculated over the global number of expressed genes with a recognizable poly(A) site

>80% of APA is composed of two cleavage sites

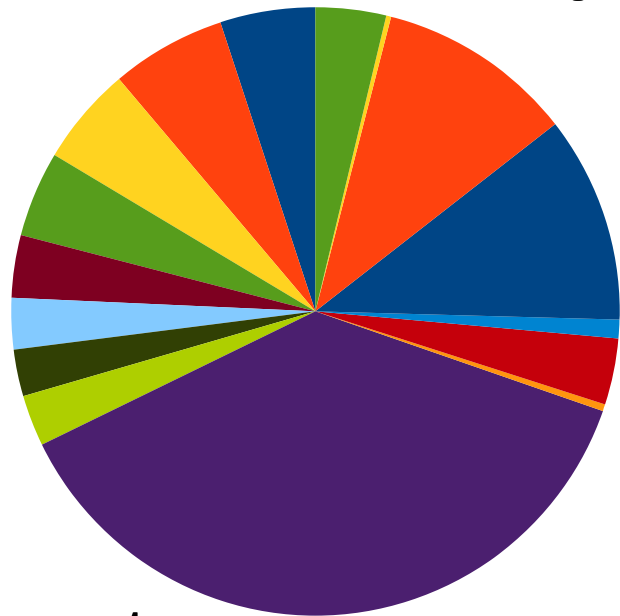


>90% of APA are tandem poly(A) sites in the 3'UTR



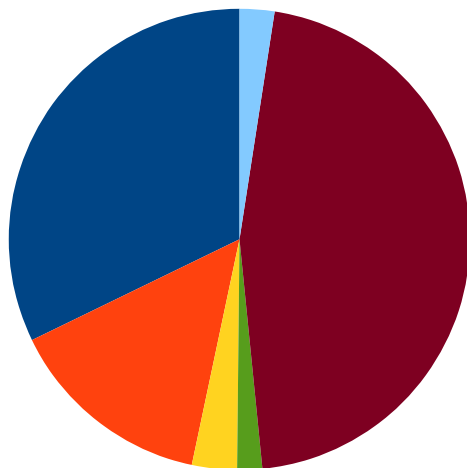
# APA is related with specific functional groups

## Biological processes



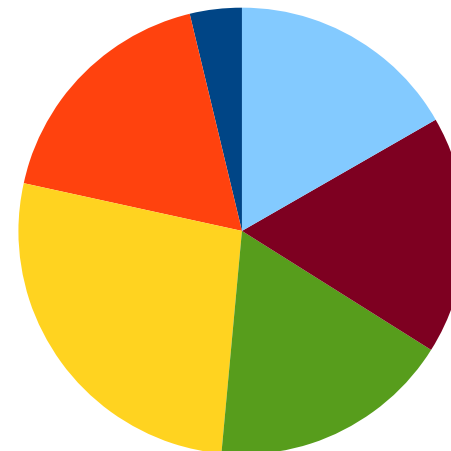
- macromolecule localization
- cellular localization
- establishment of localization in cell
- single-organism intracellular transport
- establishment of protein localization
- vesicle-mediated transport
- cell division
- signaling
- metabolic process
- positive regulation of cytoskeleton organization
- cell communication
- actin filament-based process
- macromolecule biosynthetic process
- RNA metabolic process
- positive regulation of protein polymerization
- cell cycle

## Cellular components



- membrane-bounded organelle
- protein complex
- cell periphery
- cellular bud
- intracellular
- cytoskeleton

## Molecular functions



- enzyme regulator activity
- carbohydrate derivative binding
- protein binding
- ribonucleotide binding
- purine nucleotide binding
- purine nucleoside binding

*But nothing found for saccharomyces! ;-(*

# The most conserved APA genes are related with some metabolic processes

MGG\_08078 Minor isoform of tropomyosin  
MGG\_06721 60S ribosomal protein L28  
MGG\_16201 CMD1 Calmodulin  
MGG\_08084 Cell wall biogenesis protein phosphatase Ssd1  
MGG\_04438 ADP-ribosylation factor  
MGG\_02872 40S ribosomal protein S27  
MGG\_13782 60S ribosomal protein L19  
MGG\_08006 Transmembrane nucleoporin  
MGG\_01079 Ras-like protein Rab-11A  
MGG\_16739 Eukaryotic translation initiation factor 5A  
MGG\_12122 CMGC/GSK protein kinase

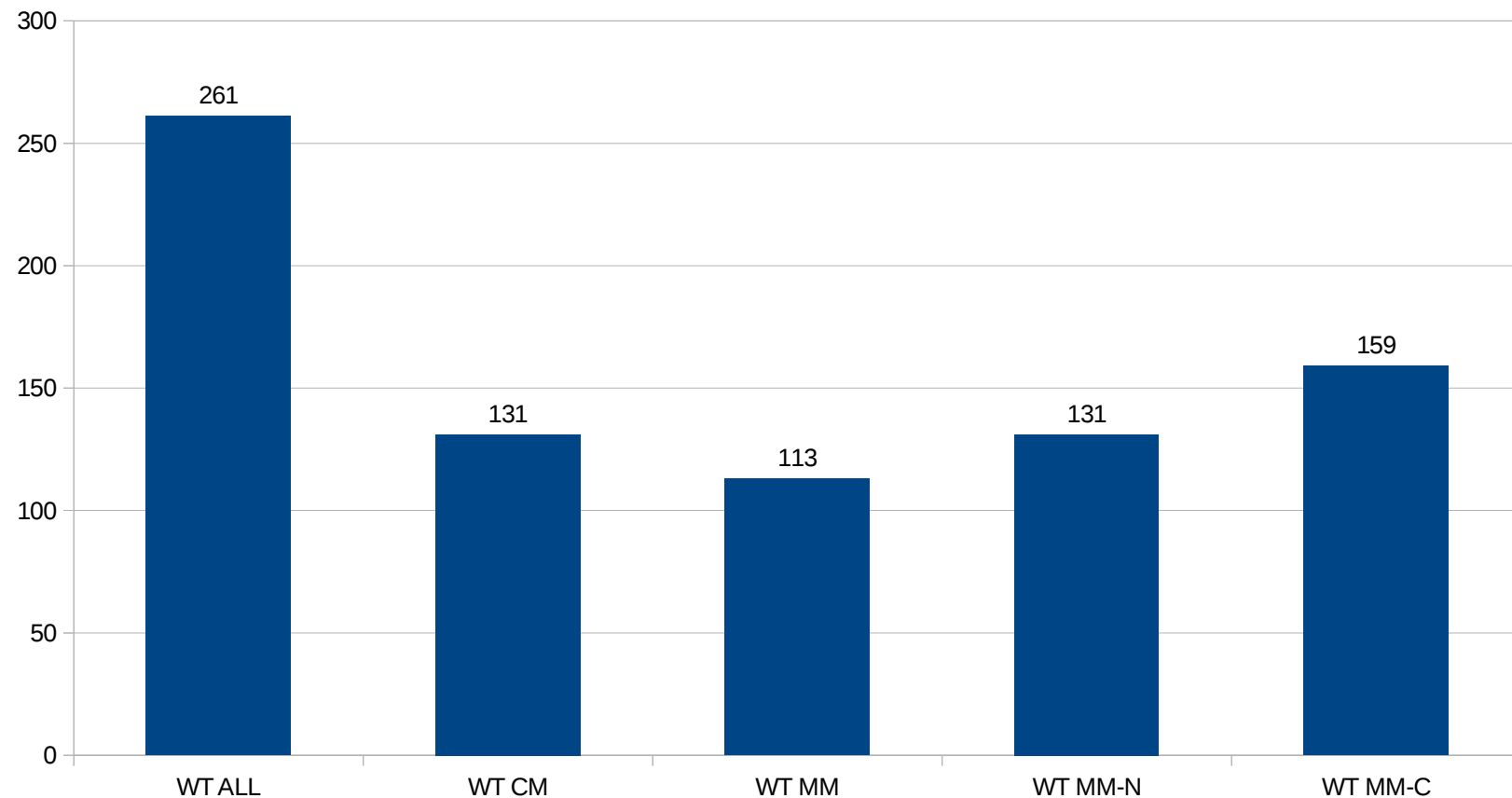
...

cytoplasm  
intracellular organelle  
macromolecular complex  
structural molecule activity  
protein metabolic process  
ribosome  
Translation

....

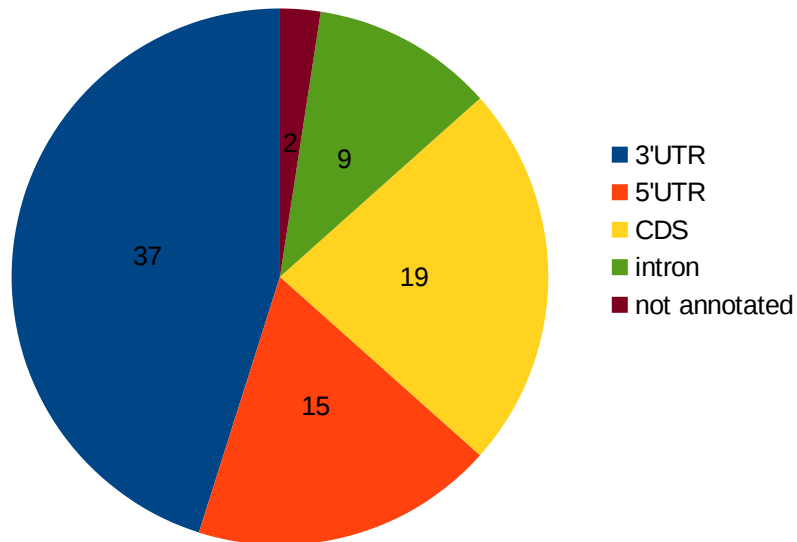
# 261 highly expressed (>100 reads) poly(A) sites could not be assigned to any annotated gene

Highly expressed poly(A) sites not mapping to any annotated gene

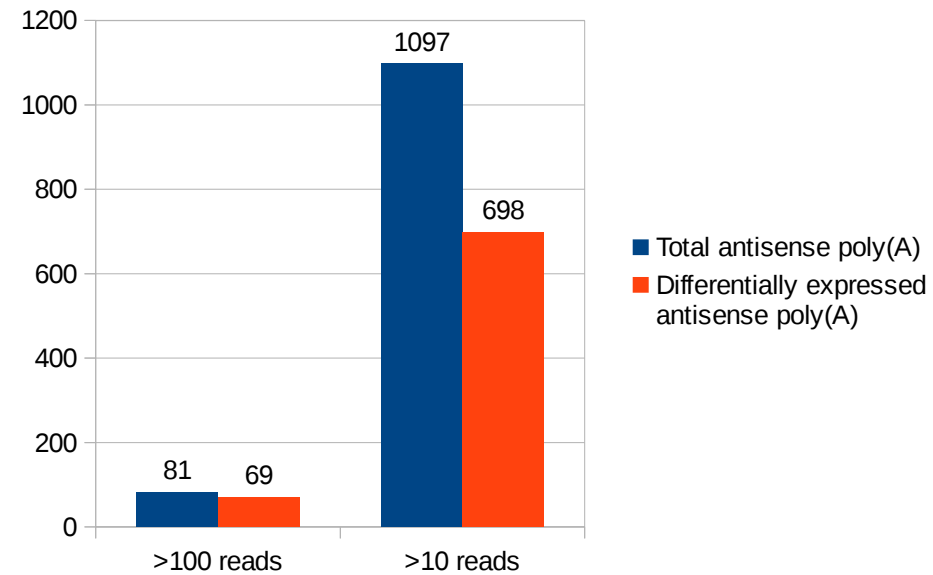
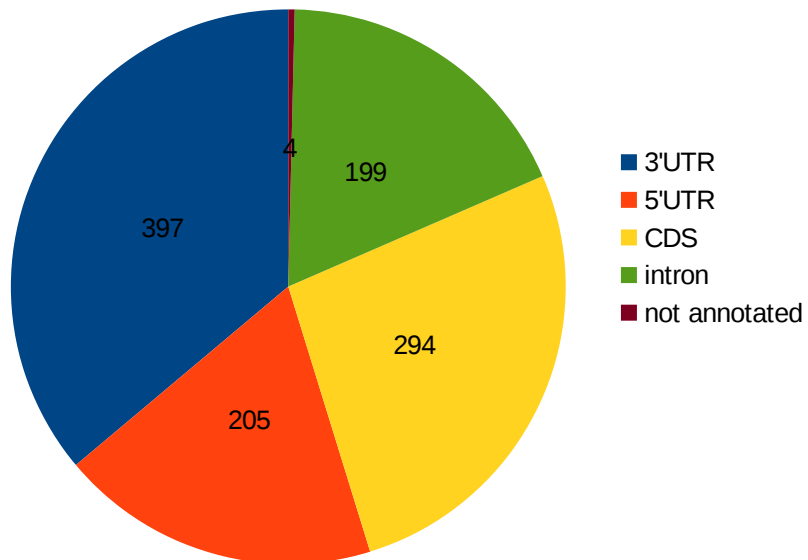


# Antisense poly(A) are usually located in the 3'UTR

Location of antisense poly(A) sites >100 reads

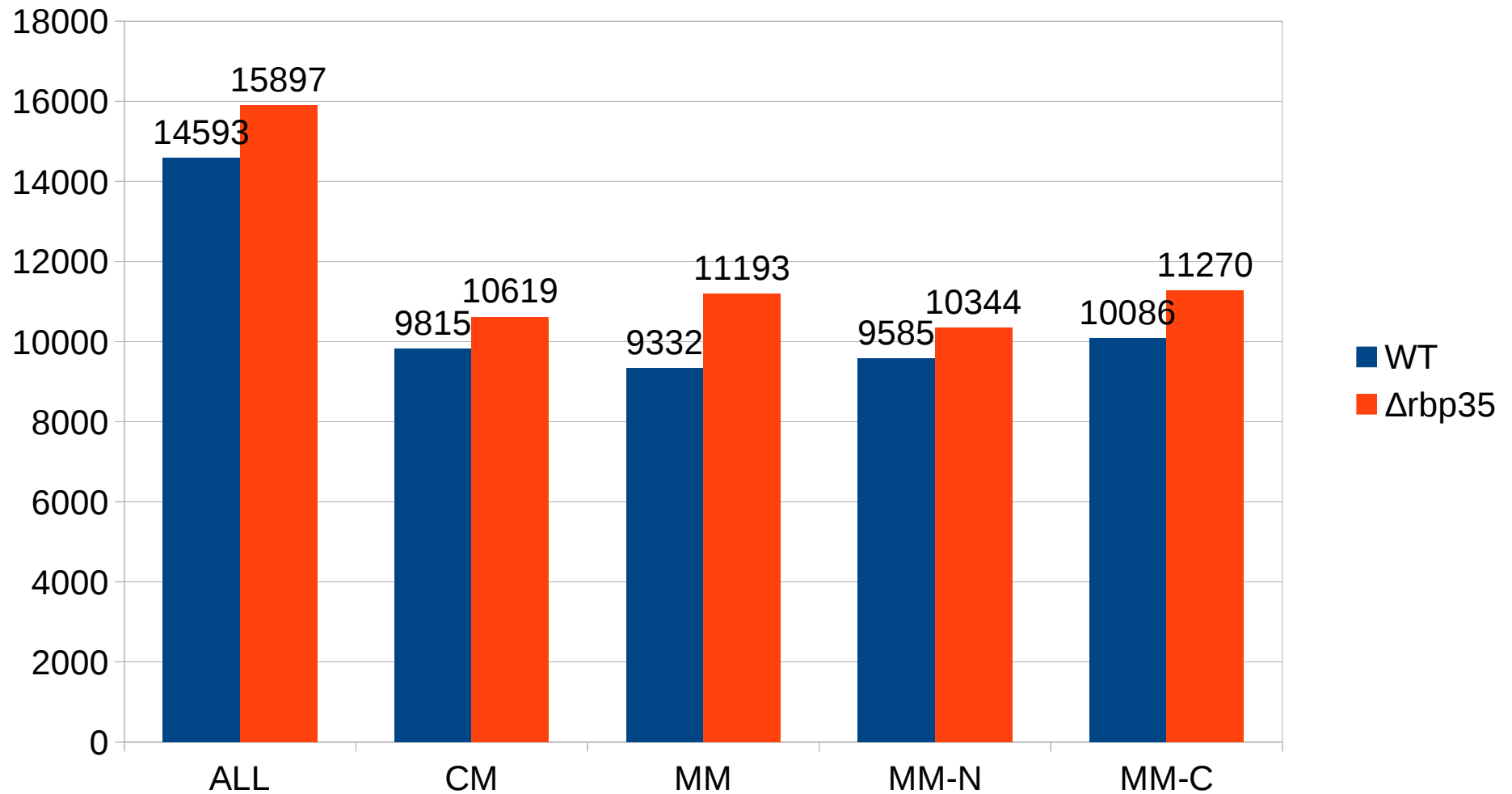


Location of poly(A) sites >10 reads

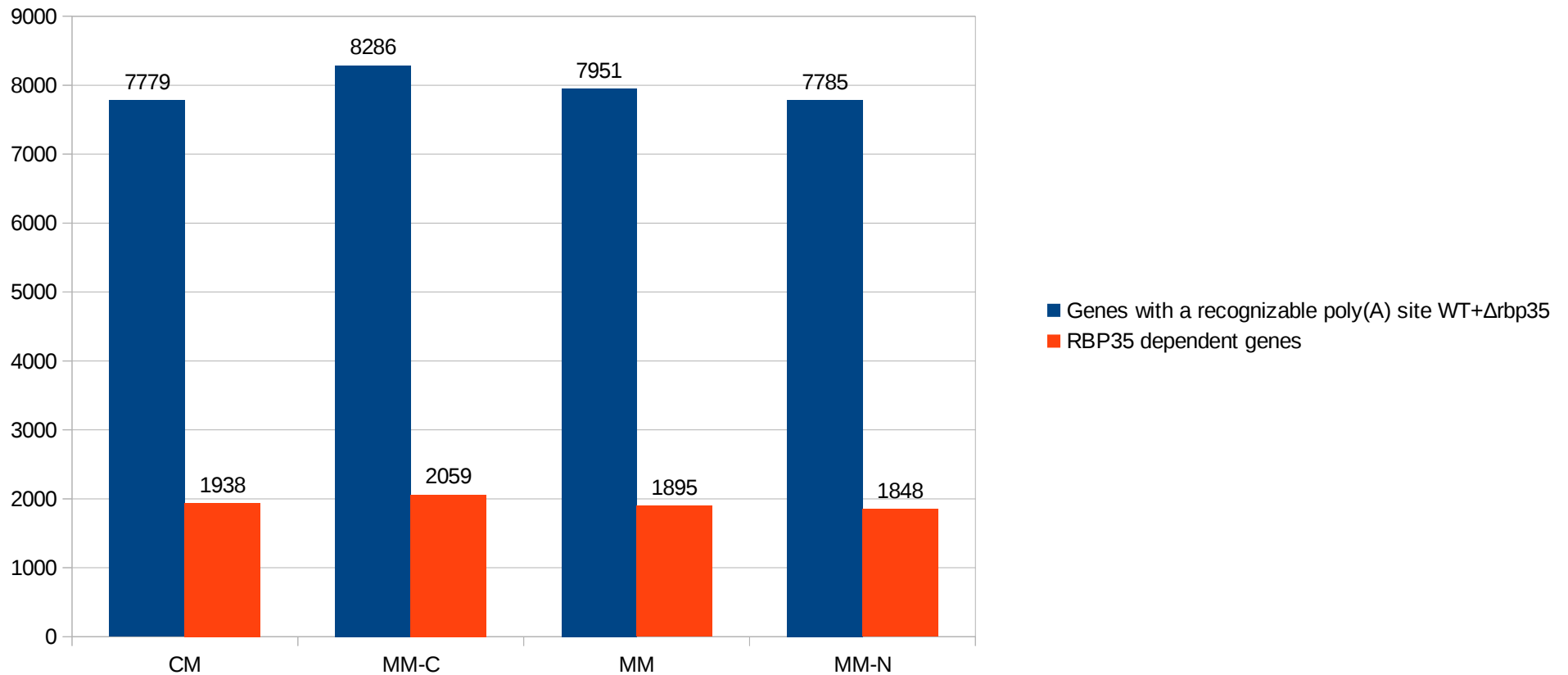




# $\Delta rbp35$ affects poly(A) sites number



# $\Delta rbp35$ affects poly(A) selection in $\sim 25\%$ of genes



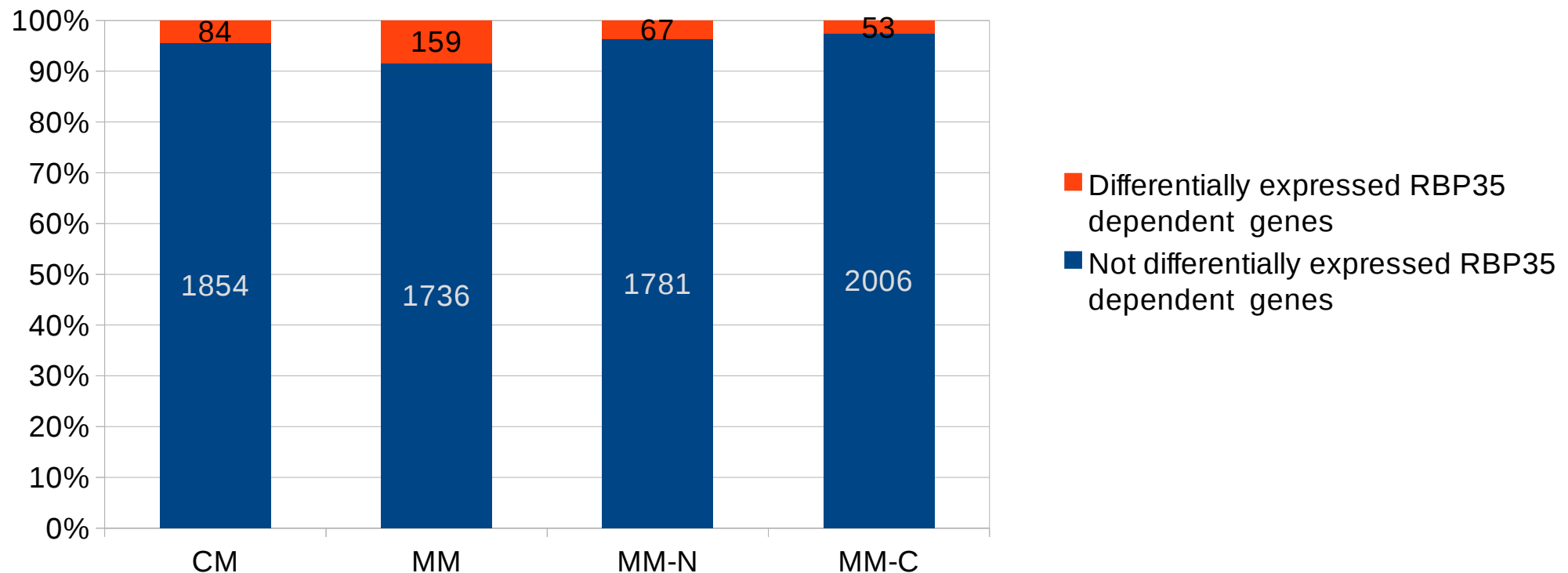
WT

$\Delta rbp35$



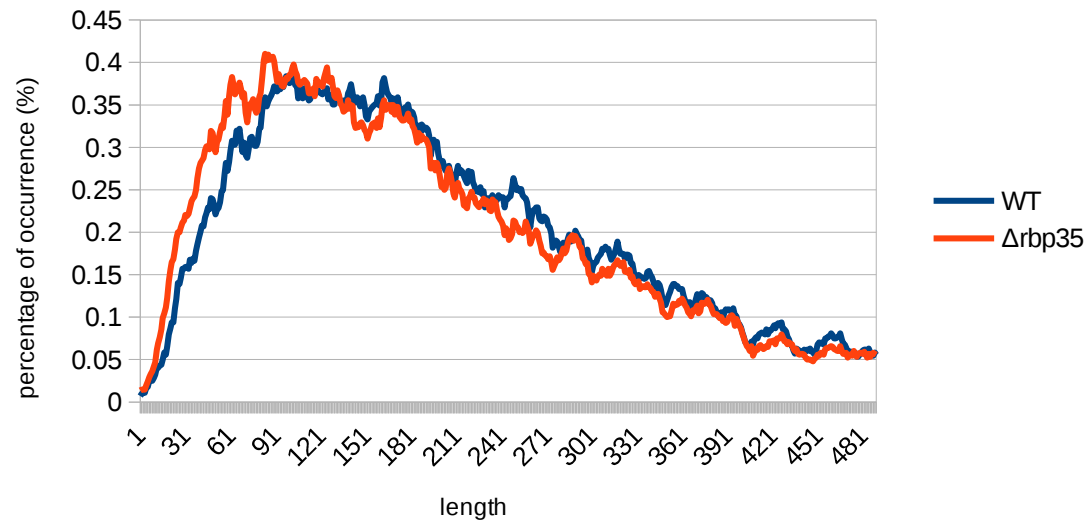
# RBP35 dependent genes are usually not differentially expressed

RBP35 dependance vs differential expression

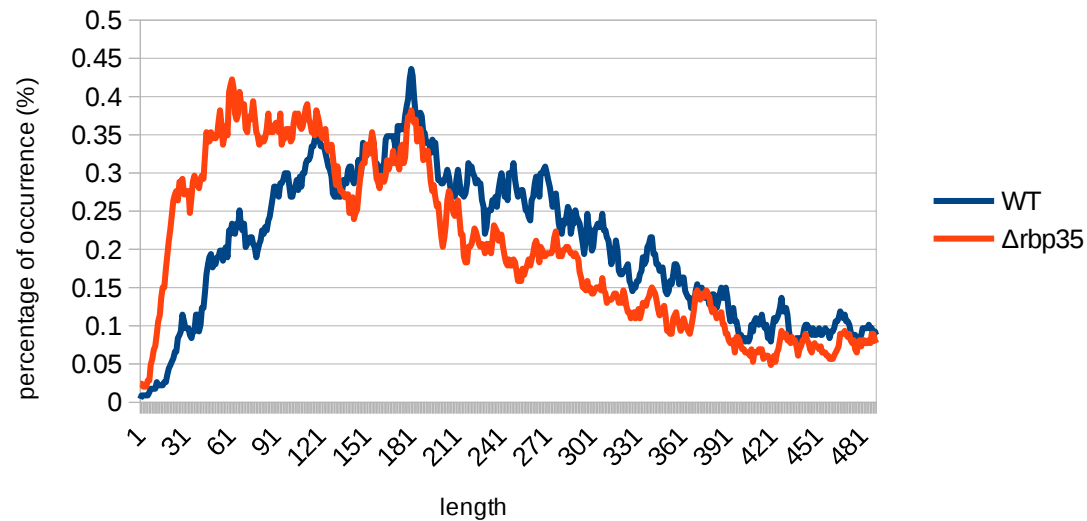


# $\Delta rbp35$ affects 3'UTR length

3'UTR length (all genes)

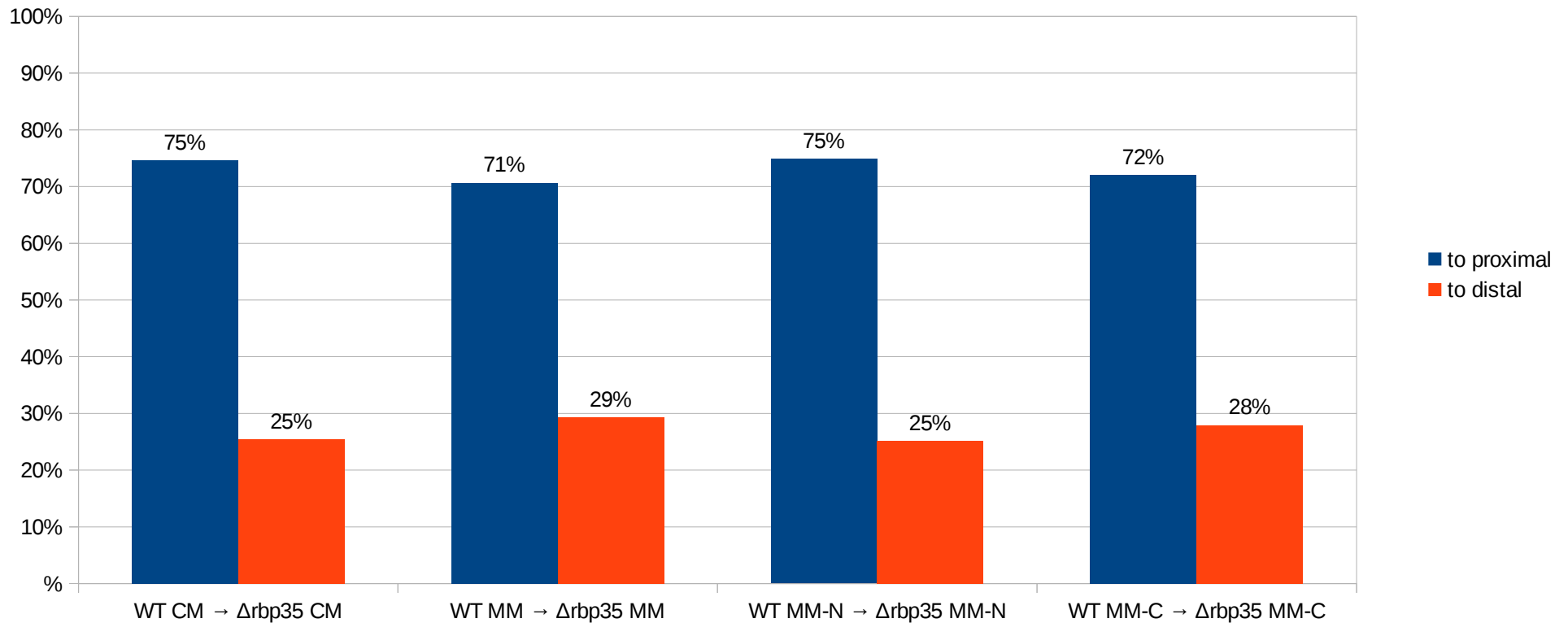


3'UTR length (RBP-dependent genes)



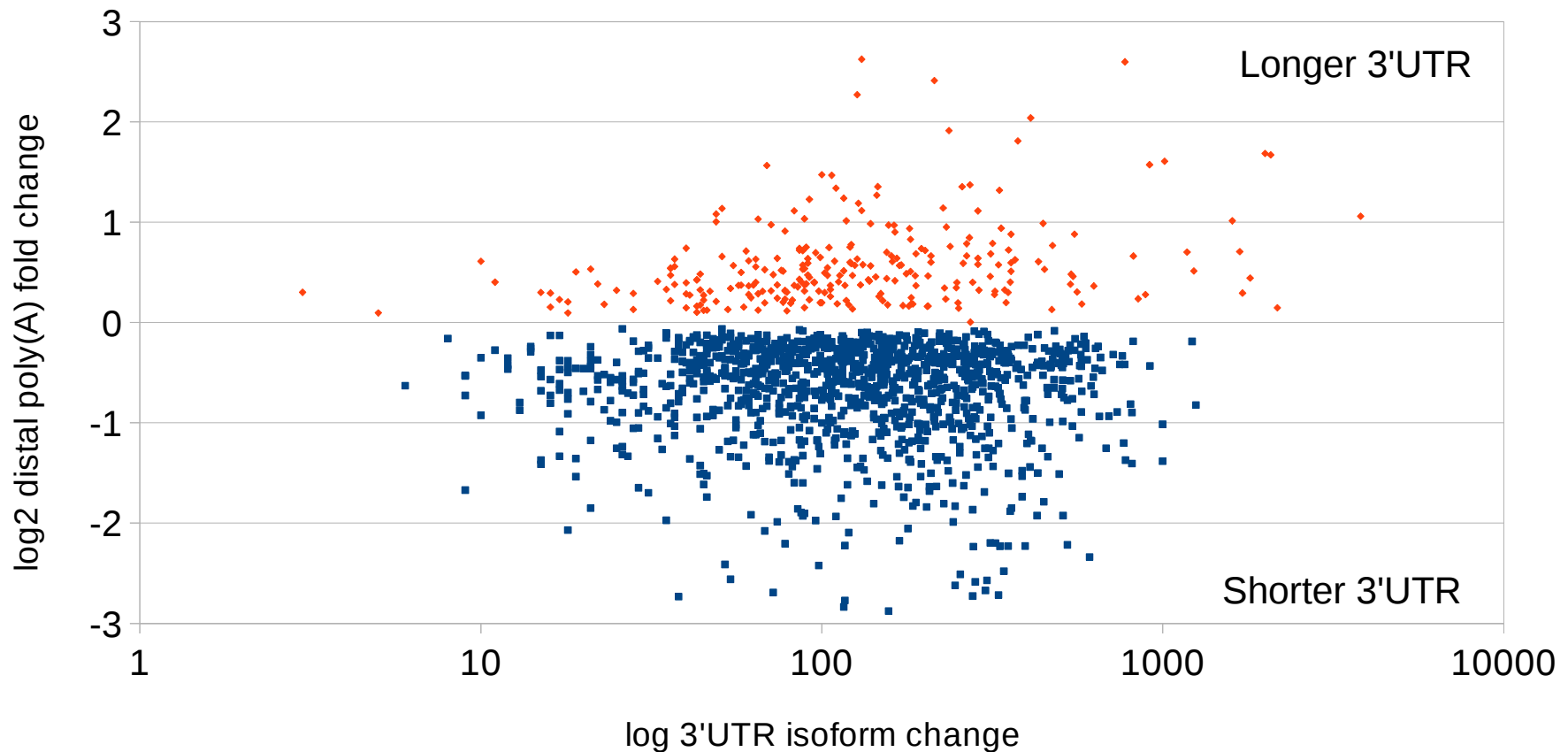
# $\Delta rbp35$ affects poly(A) sites usage, preferring proximal cuts

Poly(A) site usage change - RBP35 dependent genes

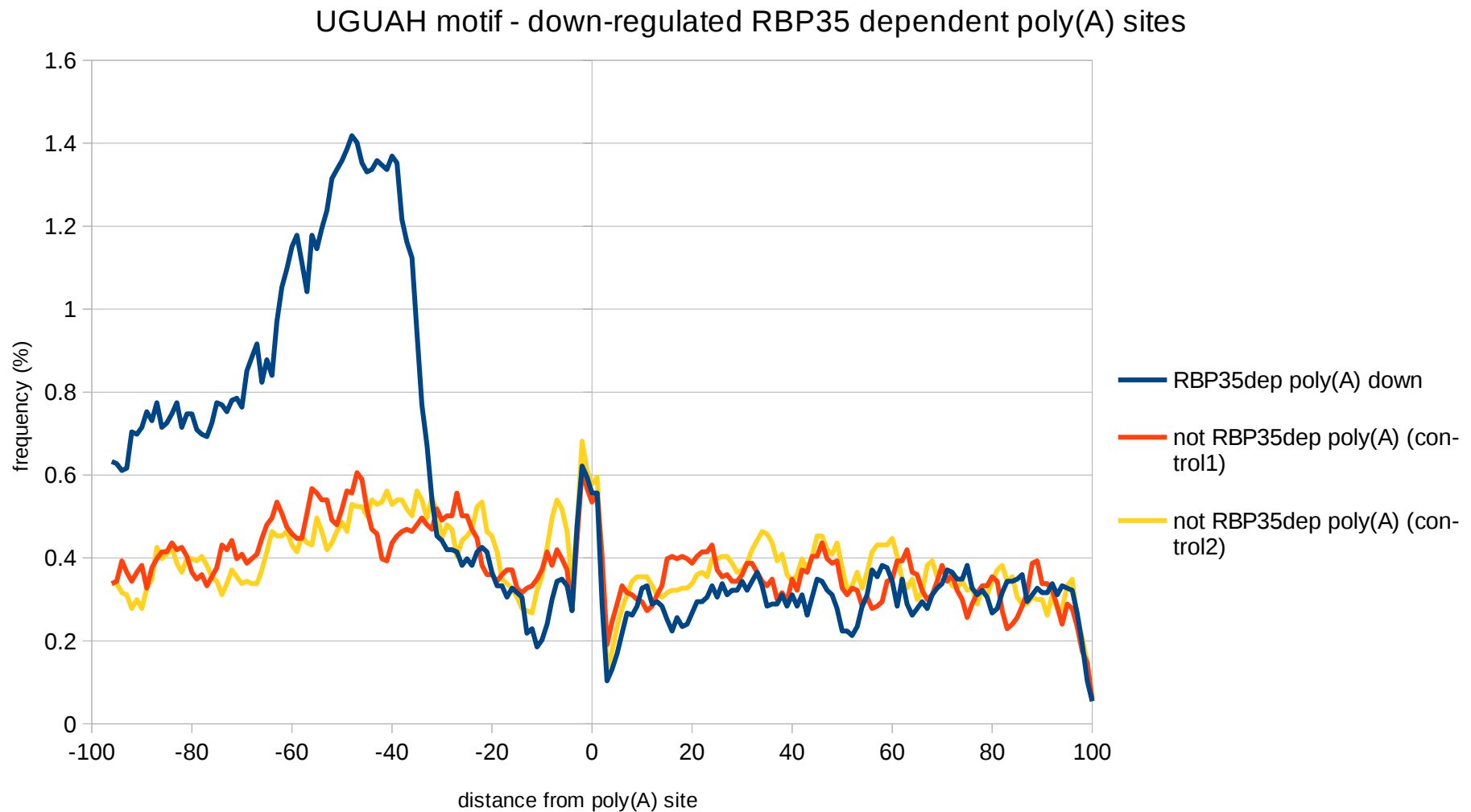


# $\Delta rbp35$ affects poly(A) sites usage, preferring proximal cuts

RBP35-dependent genes poly(A) site usage

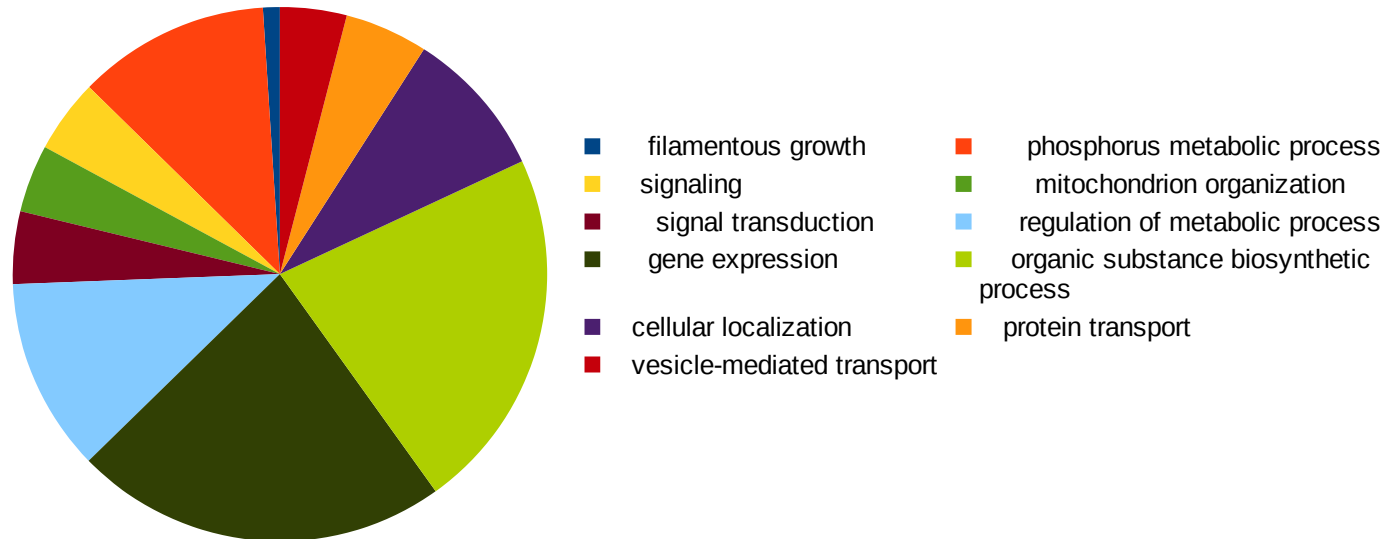


# UGUAH is enriched at -45 in poly(A) sites dependent from *RBP35*

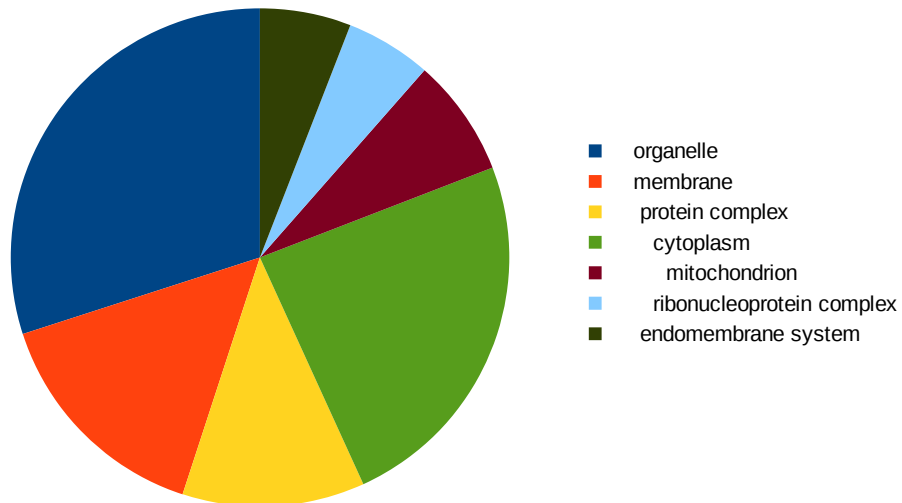


# $\Delta rbp35$ dependent genes are related with specific functional groups

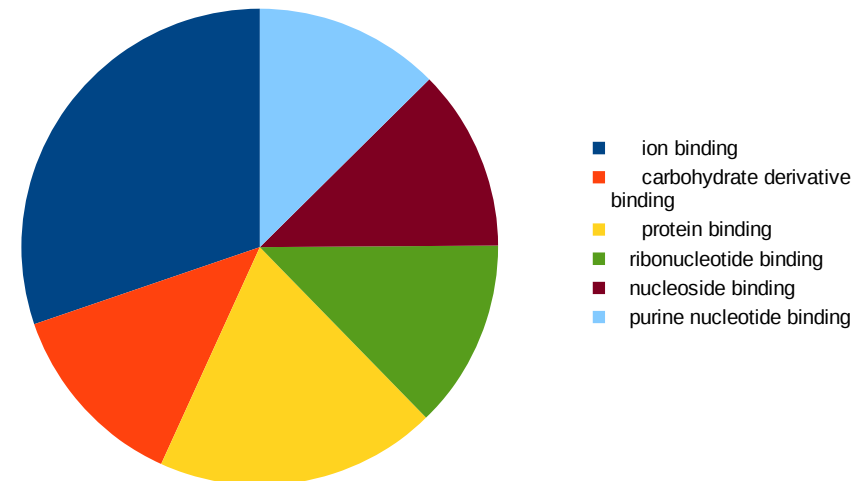
## Biological processes



## Cellular components



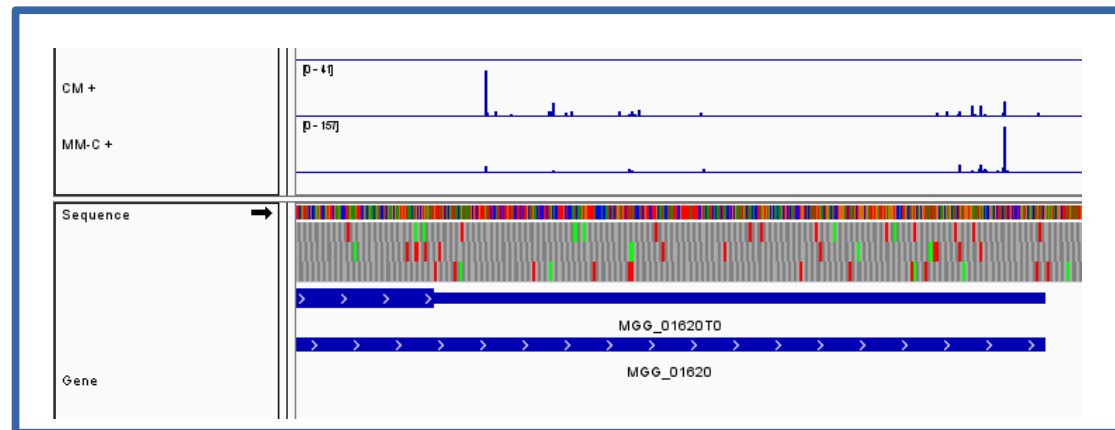
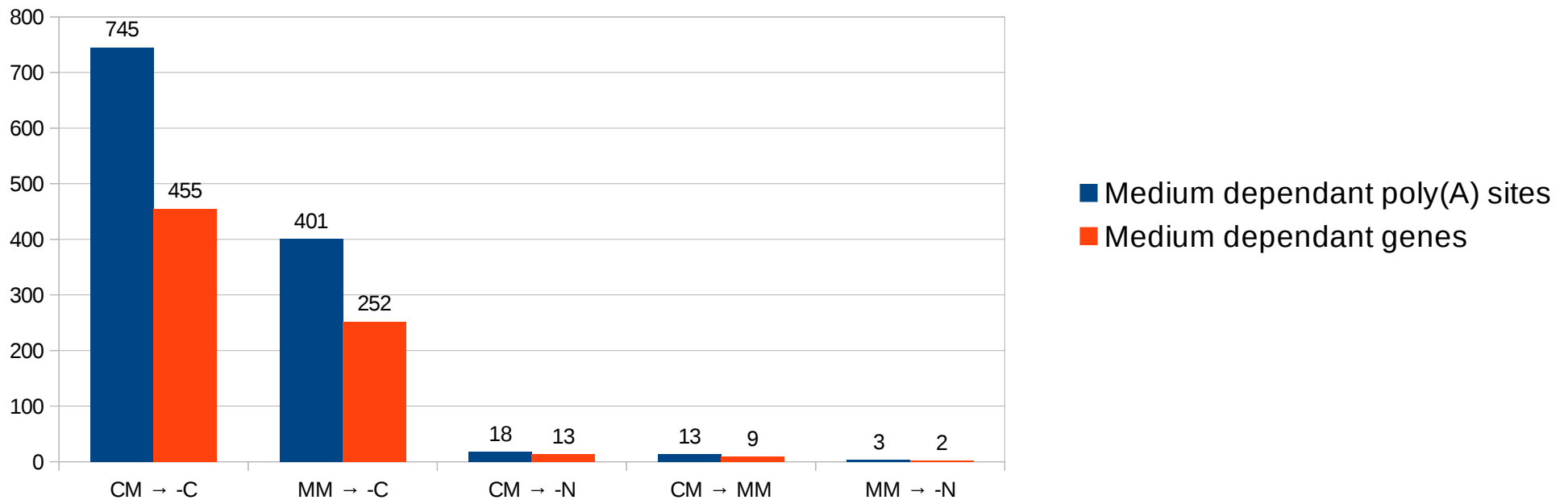
## Molecular functions





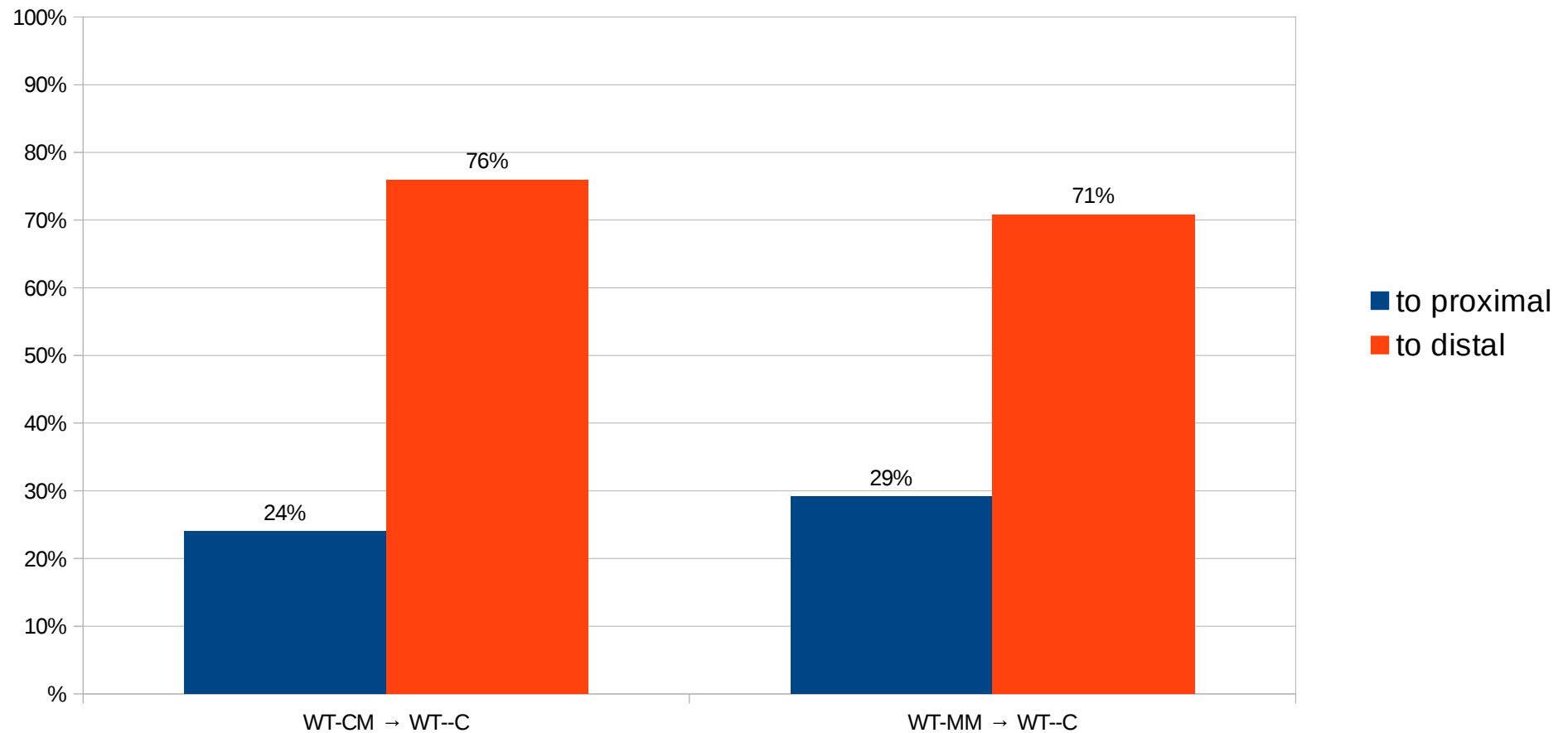
# Carbon starvation affects a great number of poly(A) sites

Medium dependent poly(A) sites and genes

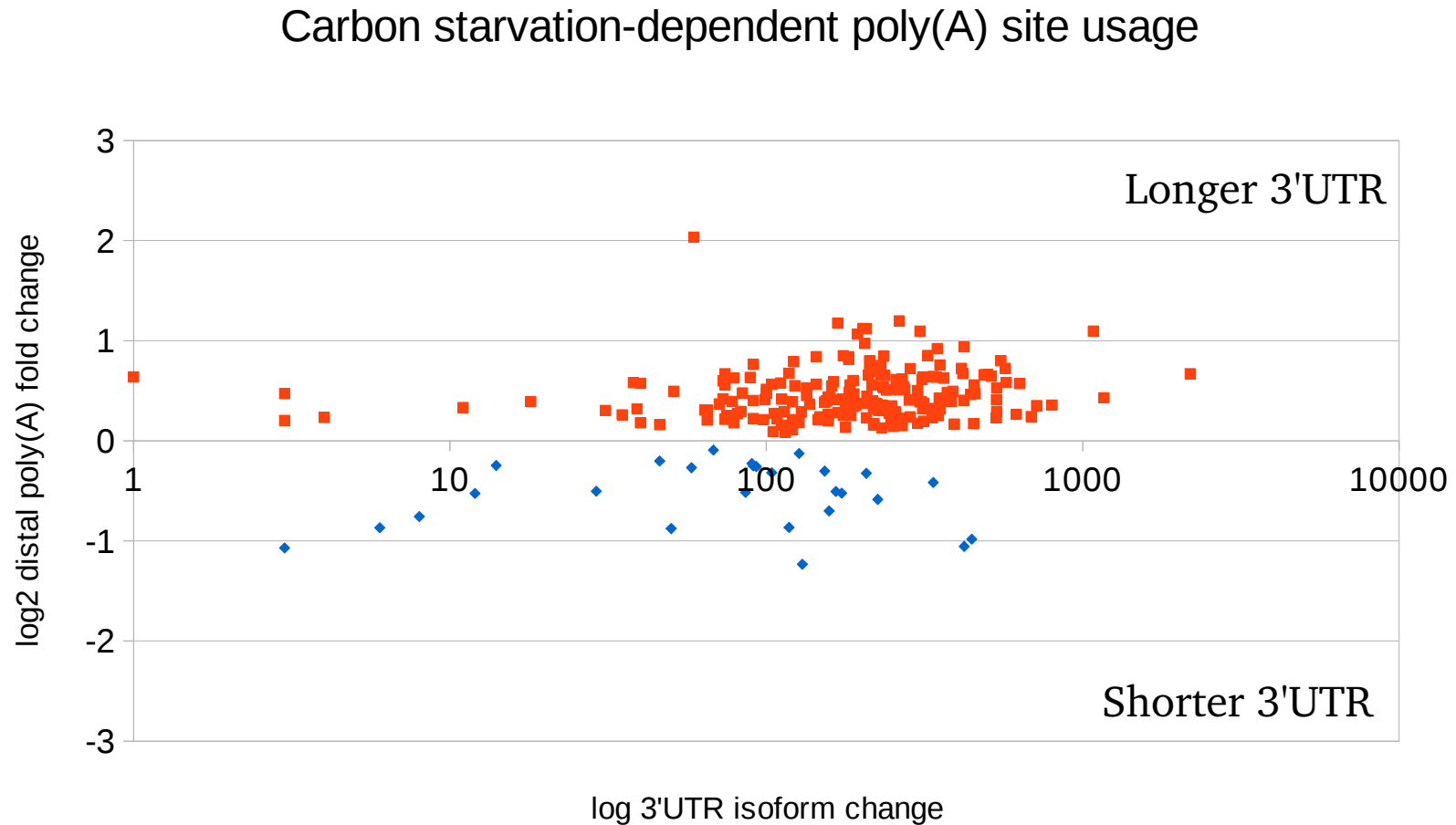


# Carbon starvation affects poly(A) sites usage, preferring distal cuts

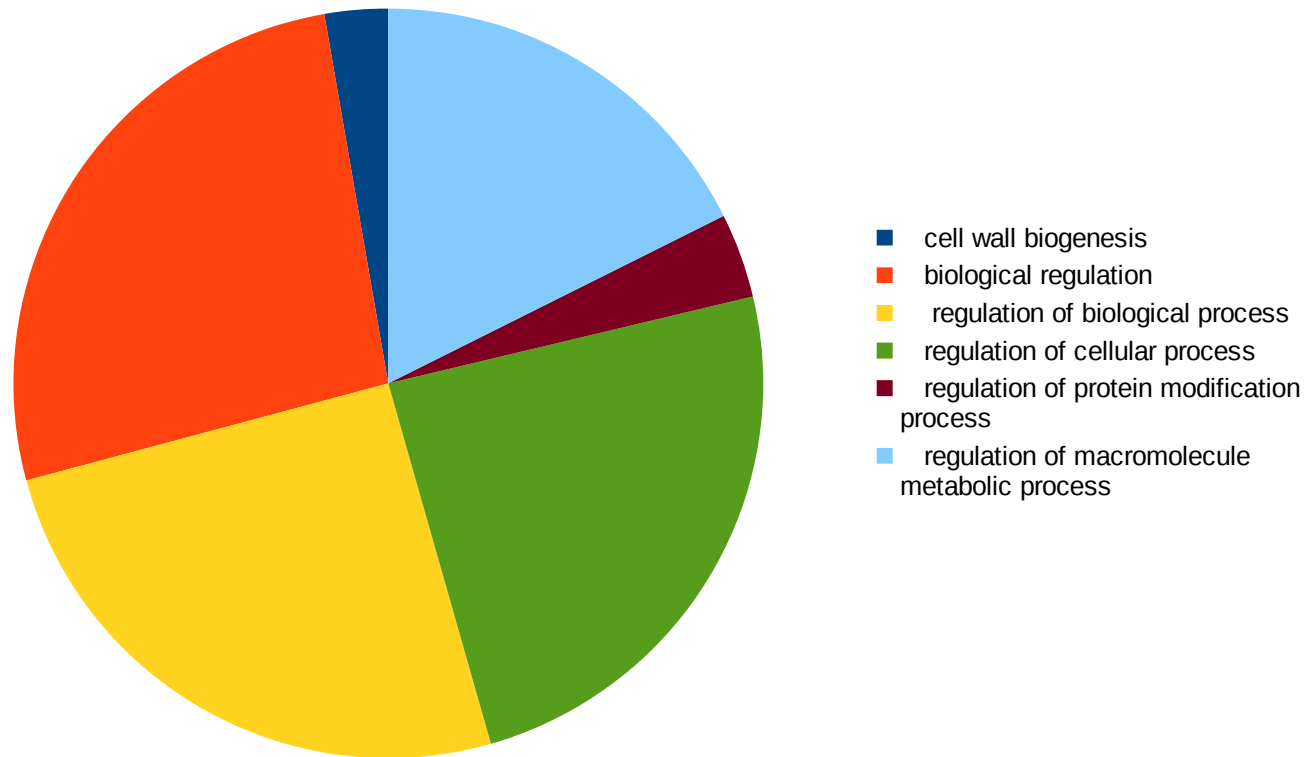
Poly(A) site usage alteration - MM-C dependent genes



# Carbon starvation affects poly(A) sites usage, preferring distal cuts

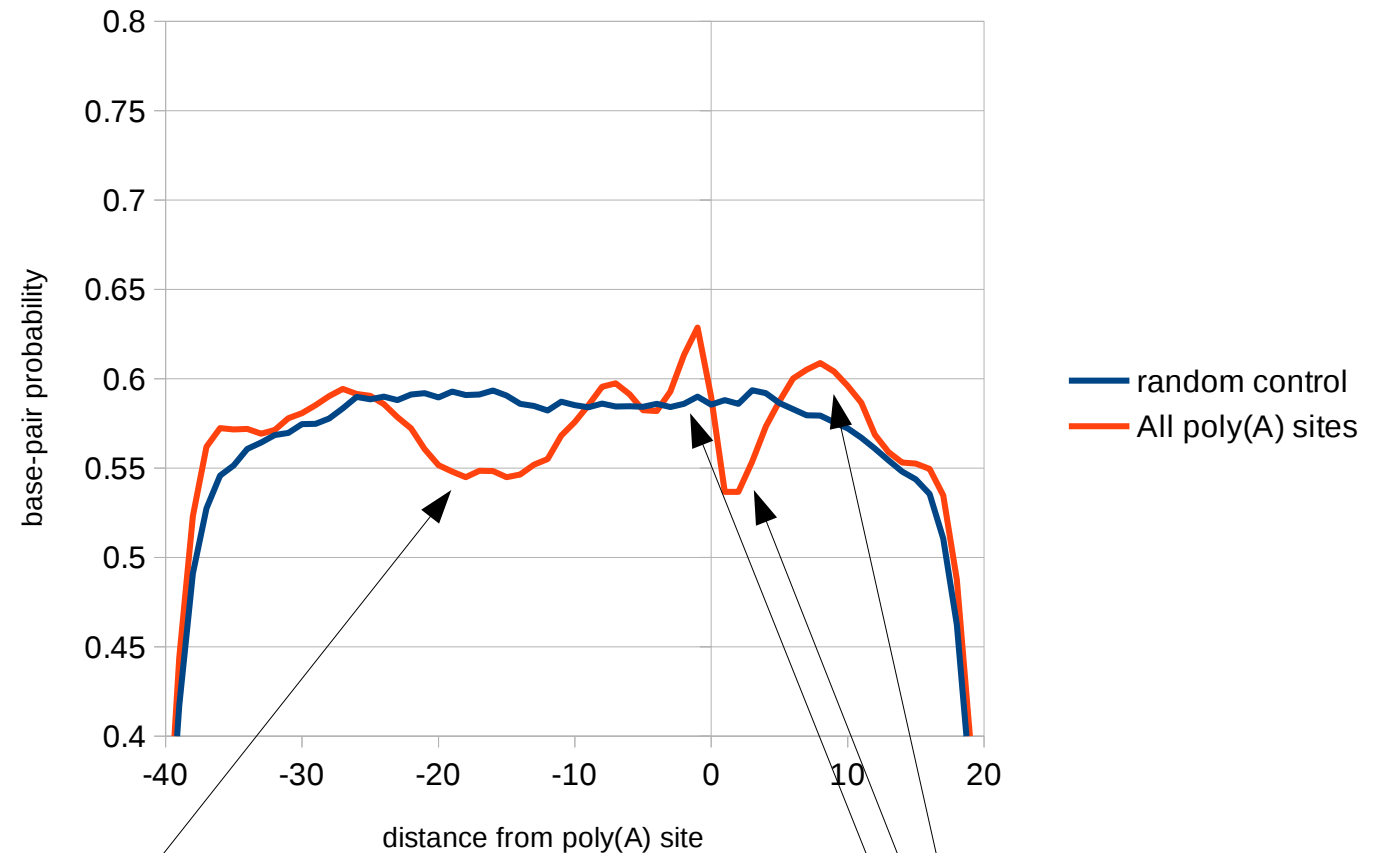


# Carbon starvation dependent genes are related with very few functional groups



# The polyadenylation site region has a defined secondary structure

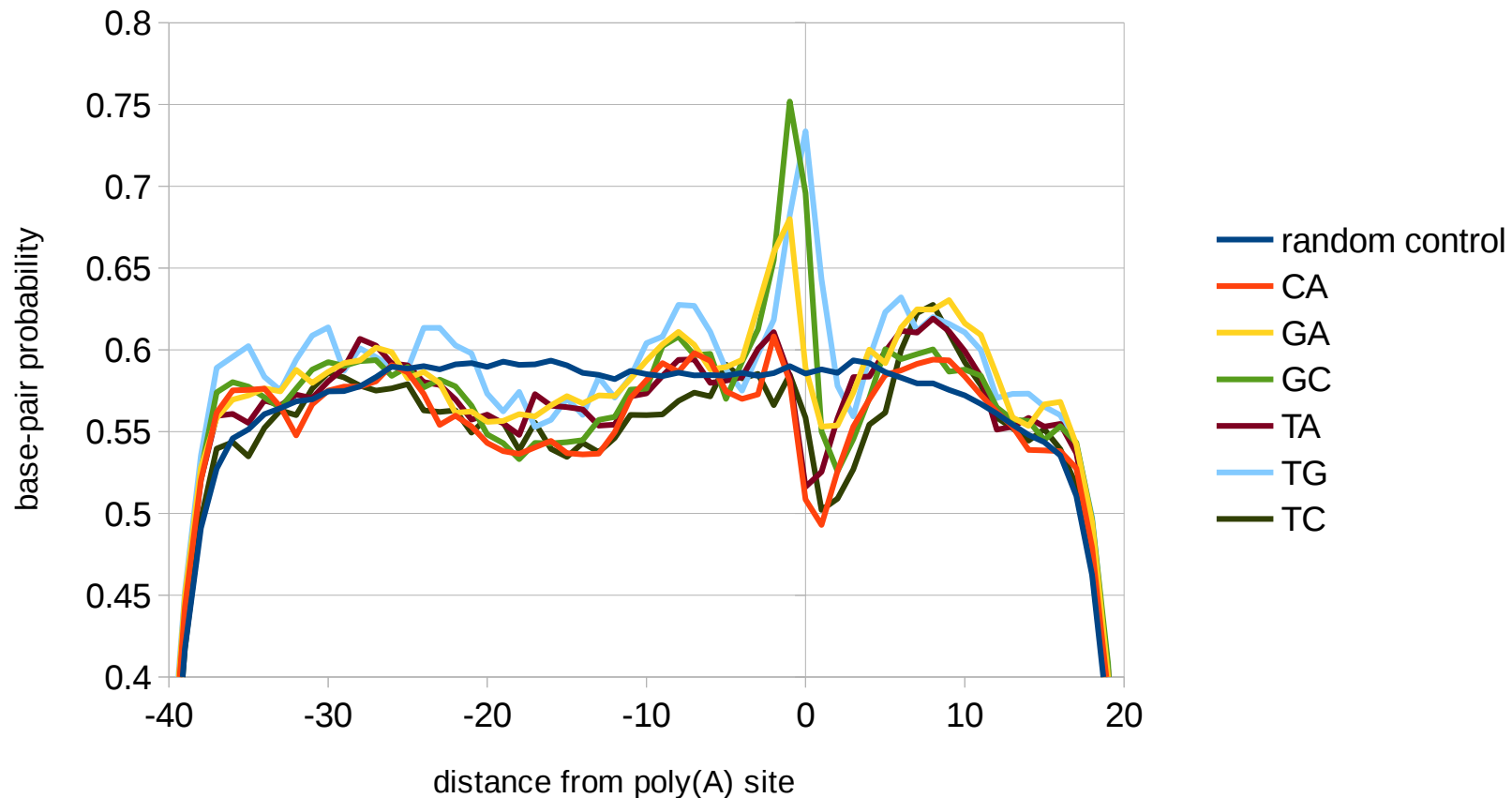
Average MFE for poly(A) sites is significantly lower (-7.48089) than random poly(A) sites ( $p$ -value =  $1.608E-015$ )



The A-rich region is usually not structured

The polyadenylation site is usually structured and located in a hairpin-loop

# The polyadenylation site region has a defined structure



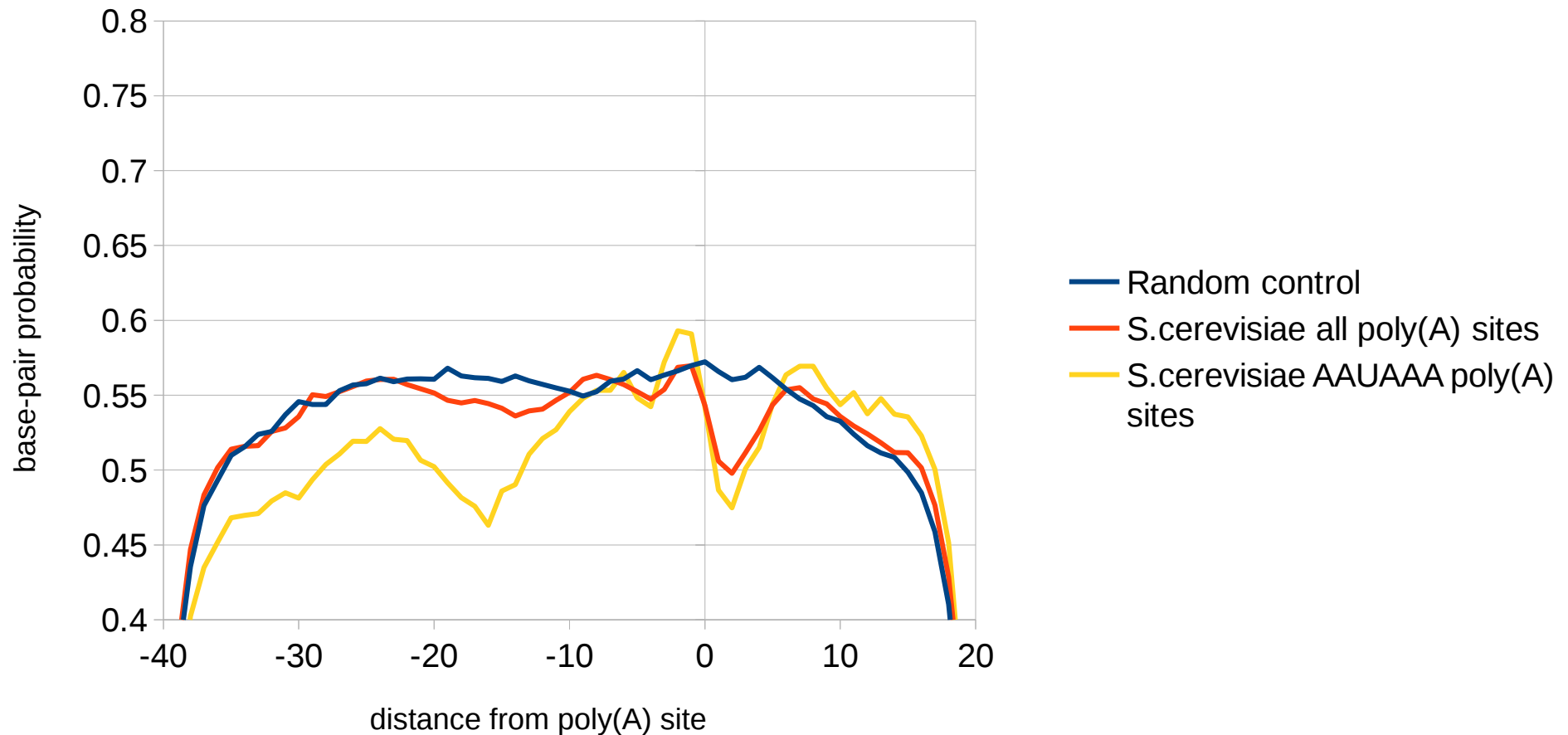
*Different cutsites have different base pairs probabilities, with TG and GC the most structured. The most common poly(A) site CA has a average conformation*

# The polyadenylation site region has a defined structure



*Different A-rich motifs results in different degrees of conformation, with AAUAAA the most unstructured*

# The polyadenylation site region has a defined structure



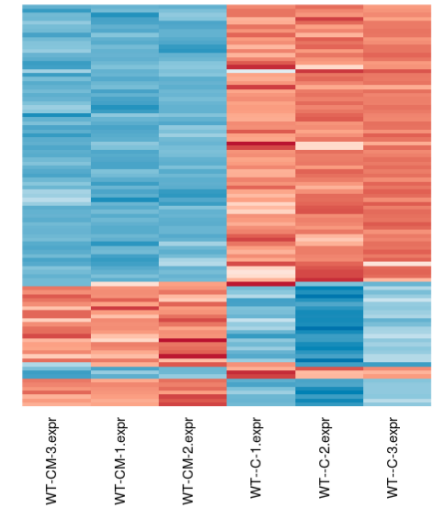
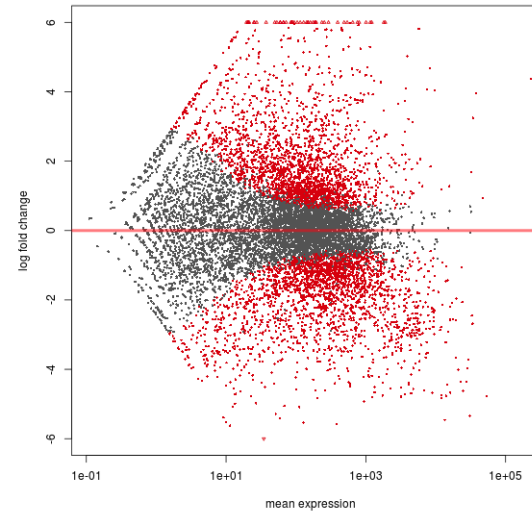
*In S.cerevisiae, the poly(A) site is not clearly structured*



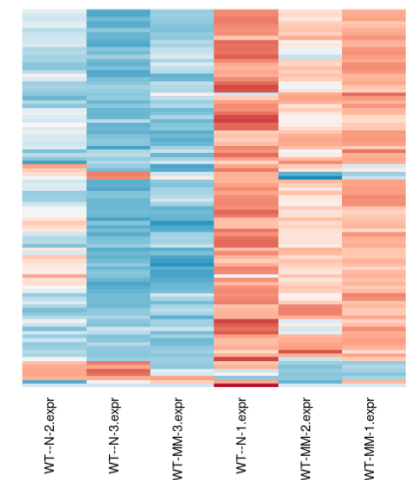
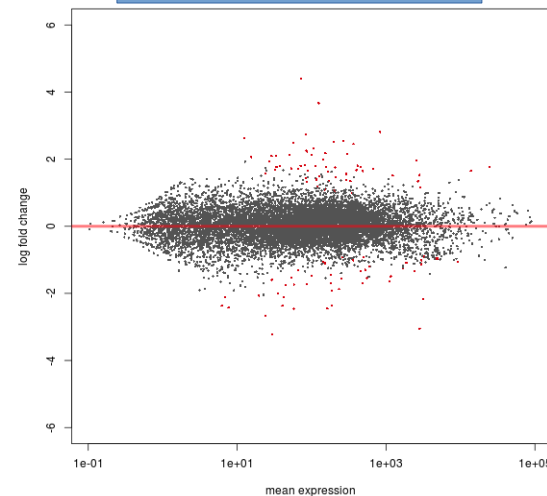
In the Wild-type The CM  $\rightarrow$  MM-C condition presents the highest number of differentially expressed genes, while MM  $\rightarrow$  MM-N the lowest

DIFFERENTIALLY EXPRESSED GENES IN THE WT			
	DOWN	UP	TOTAL
CM $\rightarrow$ MM	314	559	873
CM $\rightarrow$ MM-N	630	874	1504
CM $\rightarrow$ MM-C	2307	2342	4649
MM $\rightarrow$ MM-N	48	59	107
MM $\rightarrow$ MM-C	1882	1589	3471

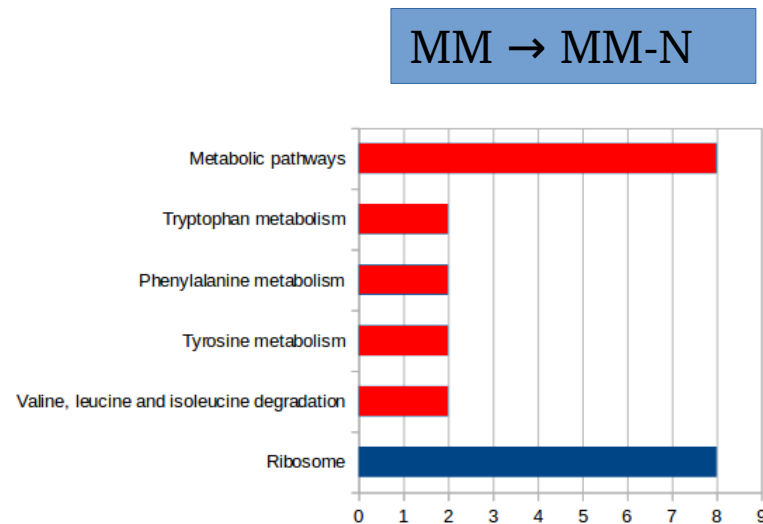
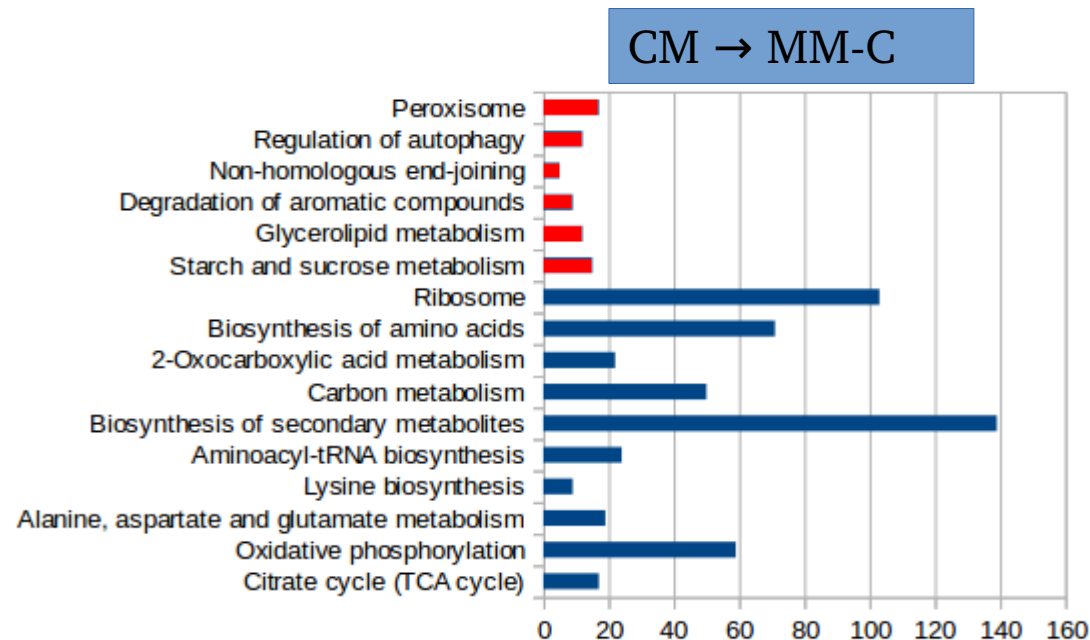
CM  $\rightarrow$  MM-C



MM  $\rightarrow$  MM-N



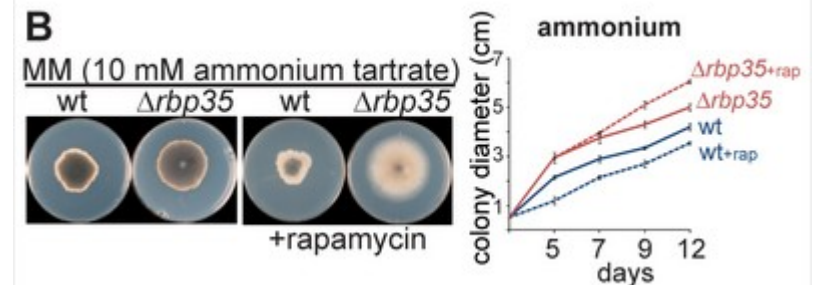
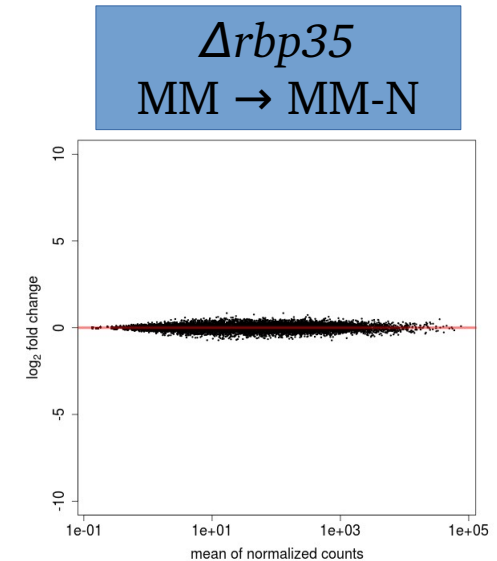
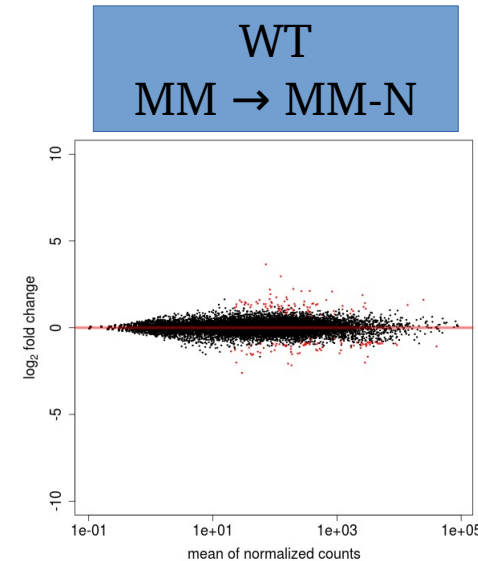
In the Wild-type The CM → MM-C condition presents the highest number of differentially expressed genes, while MM → MM-N the lowest



# $\Delta rbp35$ appears to inhibit medium recognition in MM-N

## DIFFERENTIALLY EXPRESSED GENES IN $\Delta rbp35$

	DOWN	UP	TOTAL
CM $\rightarrow$ MM	508	405	913
CM $\rightarrow$ MM-N	461	404	865
CM $\rightarrow$ MM-C	1241	1136	2377
MM $\rightarrow$ MM-N	0	0	0
MM $\rightarrow$ MM-C	475	493	968



Several genes of the *mTOR* pathway are also affected!