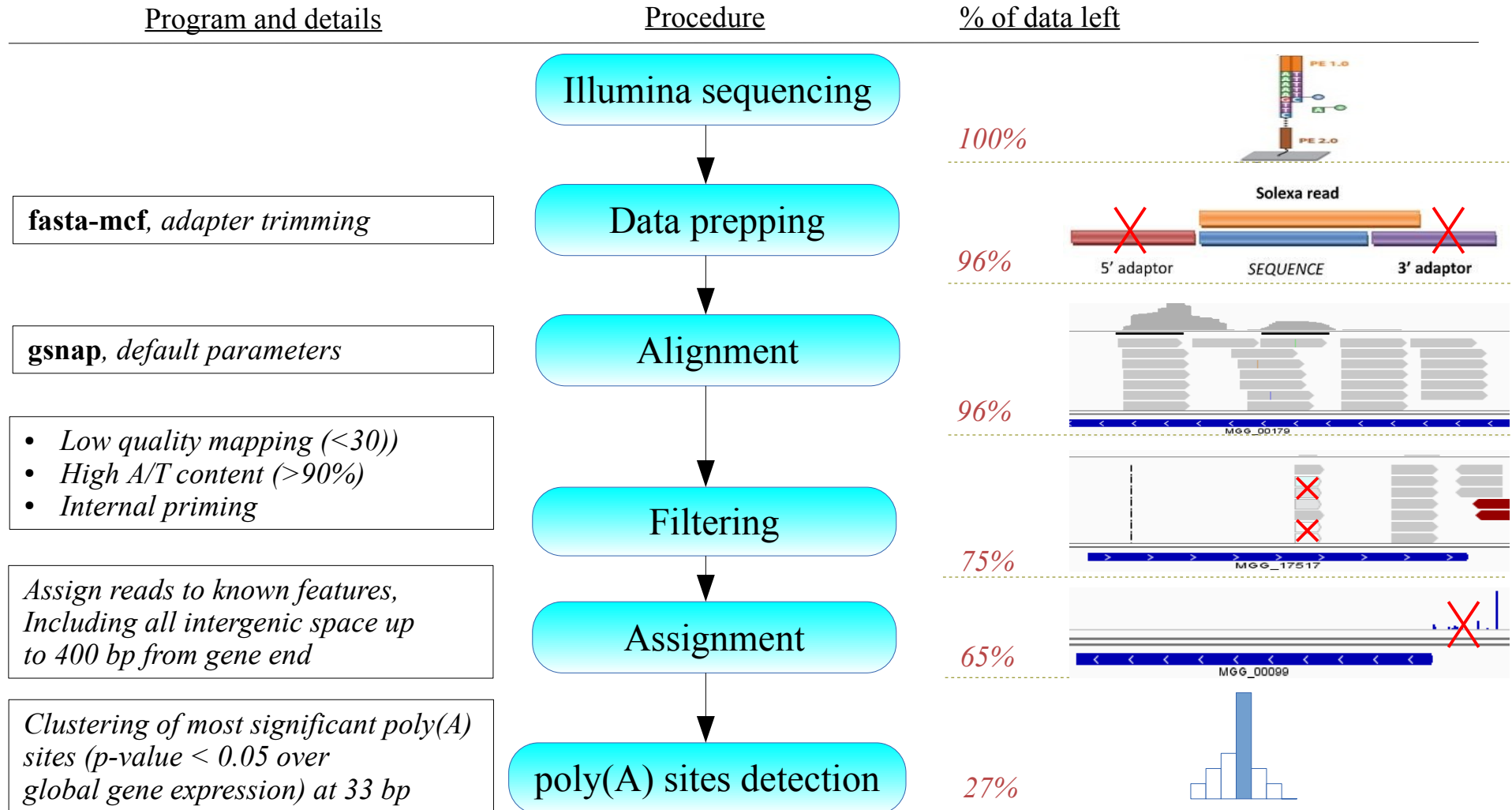


Sequencing resume

- 2 strains (*WT*, *2D4*) x 4 conditions (*CM*, *MM*, *-N*, *-C*) x 3 replicates
- 4751592 – 11517077 total reads database
- ~63% - ~83% successfully mapped reads
- 43 bp mean read length
- ~92% - ~98% replicates correlation
- ~100bp mean pair ended distance
- ~400x coverage per poly(A) site*

* assuming an amount of 22000 mRNA molecules per cell

Workflow

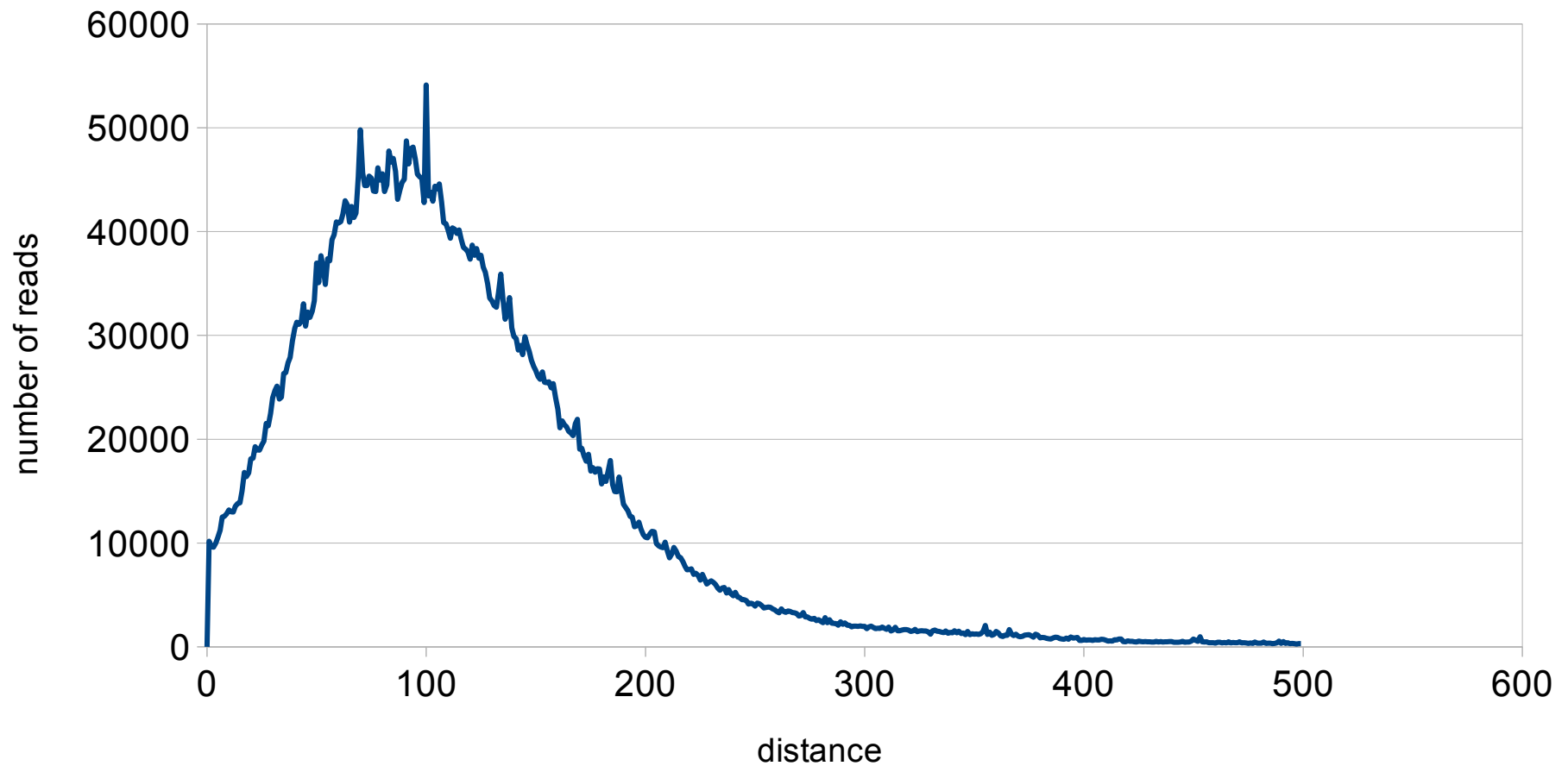


Example: Actin, single cut poly(A):

- whole gene expression ~7000 reads
- poly(A) site expression ~3000 reads

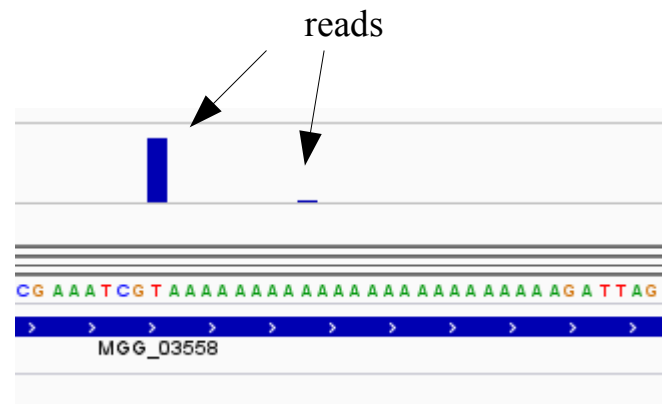
Pair ended reads distance

Pair ended distance (WT-CM-1)

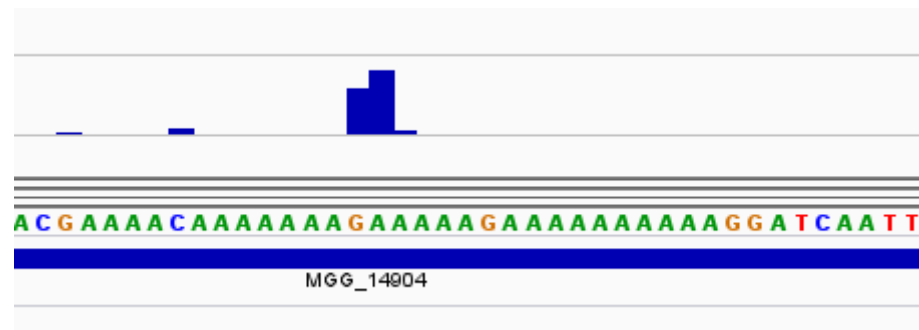


~2.5% of poly(A) sites are internal priming

- Some poly(A) sites are just a side effect of poly(A) *genomic* regions

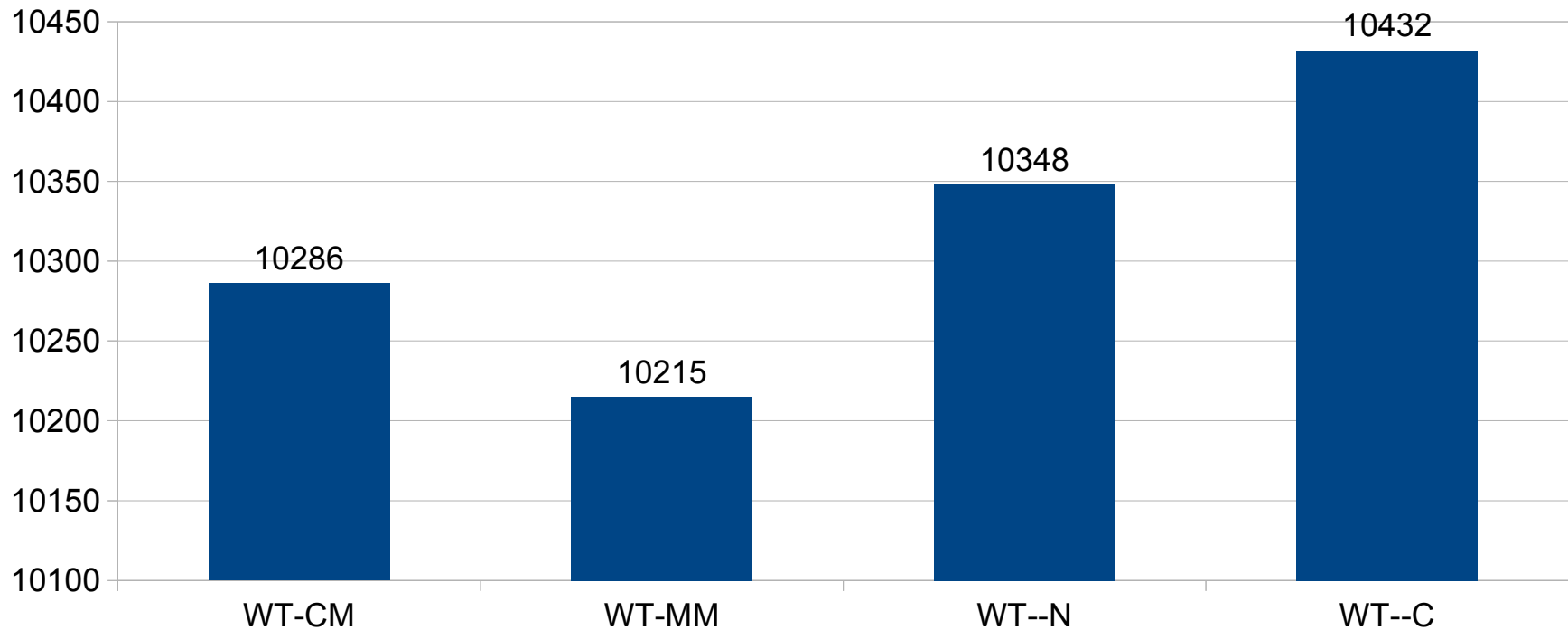


stretches of As



~10000 genes are expressed, out of a total of 13218 annotated genes (WT)

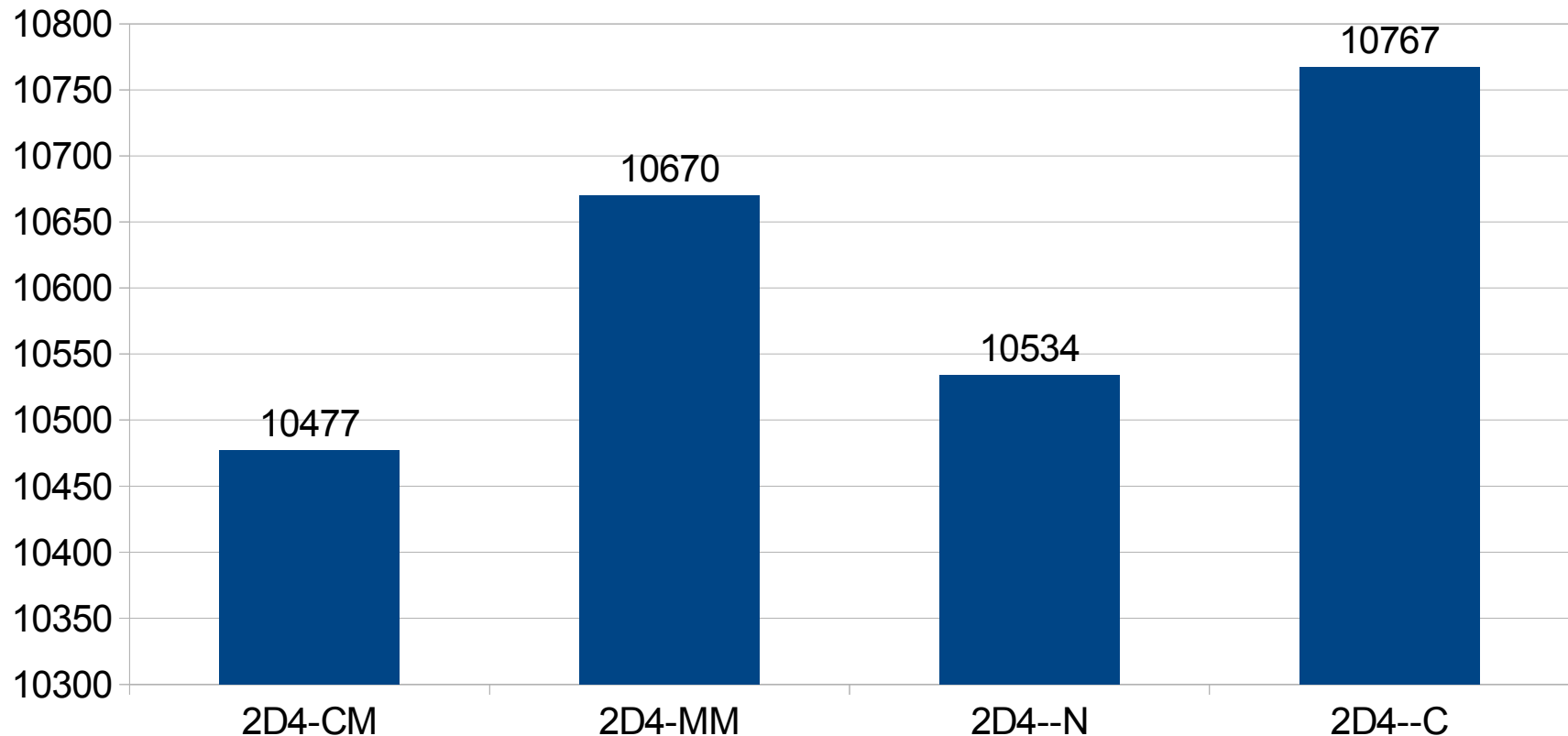
Number of expressed genes



- 9546 genes are expressed in every condition (WT only)
- 2166 genes are never expressed (WT only)

~10000 genes are expressed, out of a total of 13218 annotated genes (2D4)

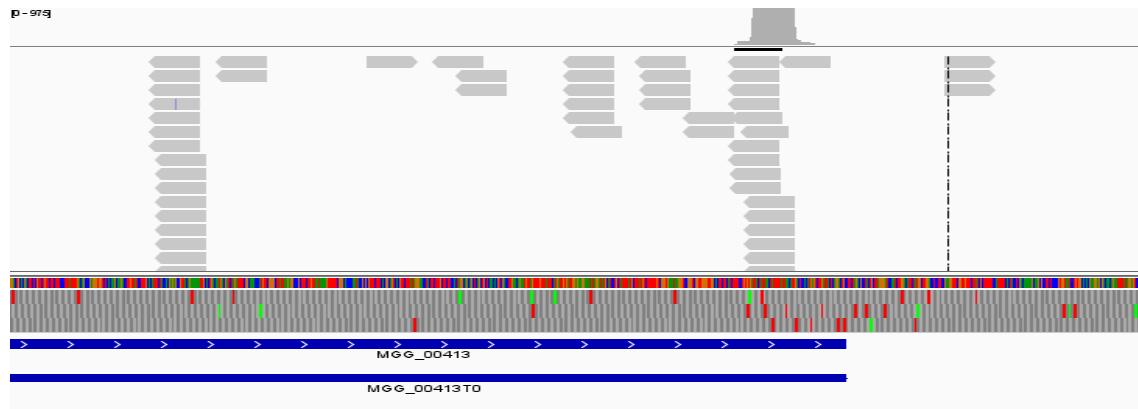
Number of expressed genes



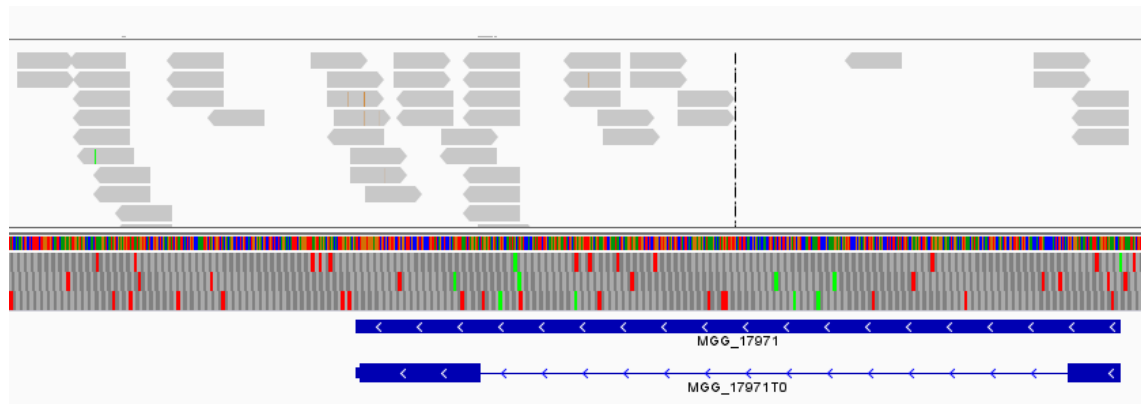
- 9918 genes are expressed in every condition (2D4 only)
- 1964 genes are never expressed (2D4 only)

Not every expressed gene has a recognizable poly(A) site

Expressed gene with a recognizable poly(A) site:

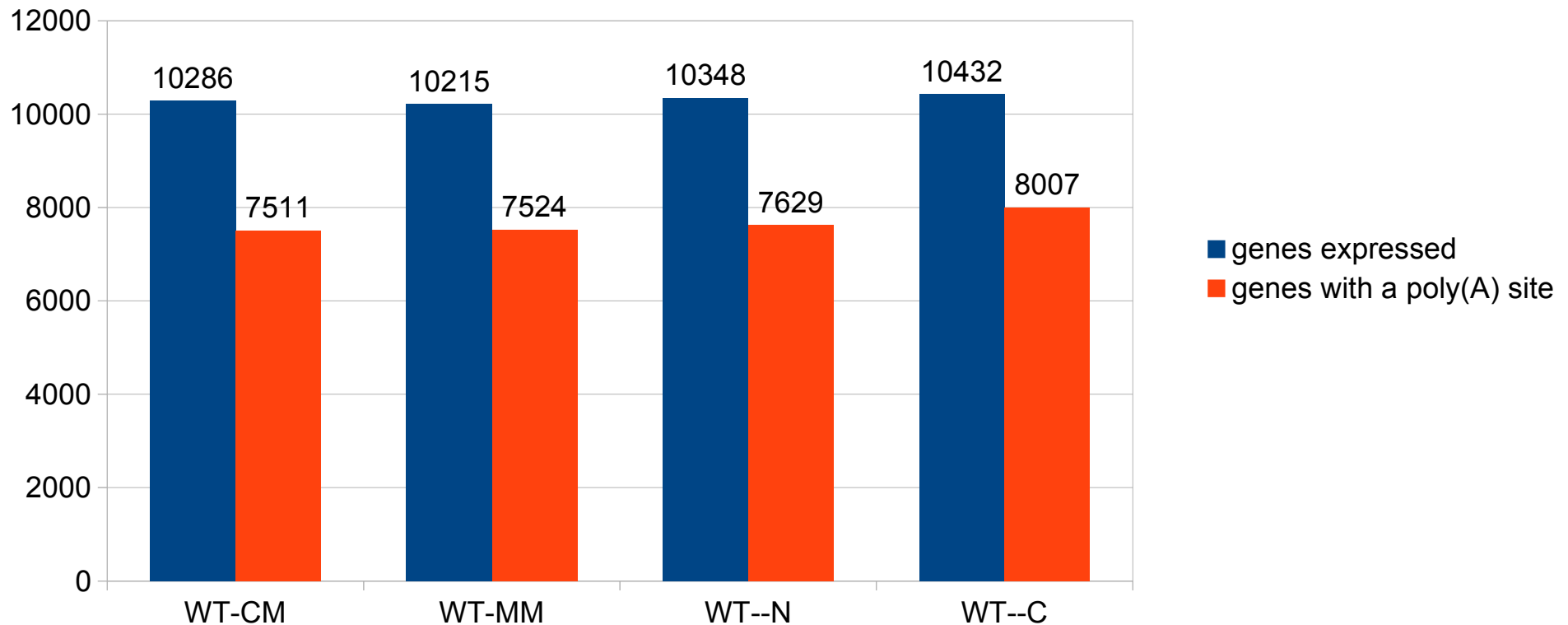


Expressed gene without a recognizable poly(A) site:



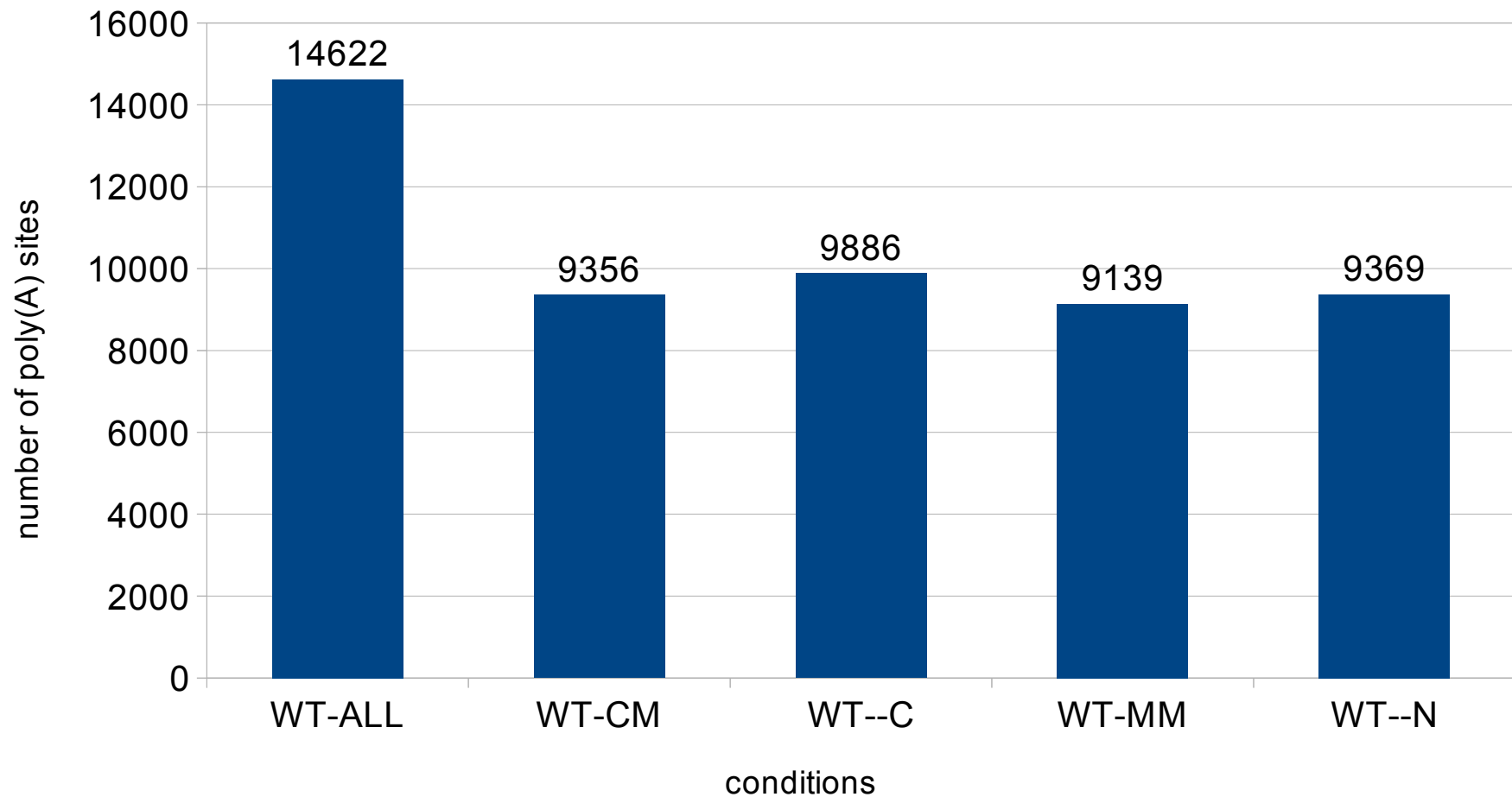
Only ~75% of genes expressed have a recognizable poly(A) site

Genes with a recognizable poly-A site



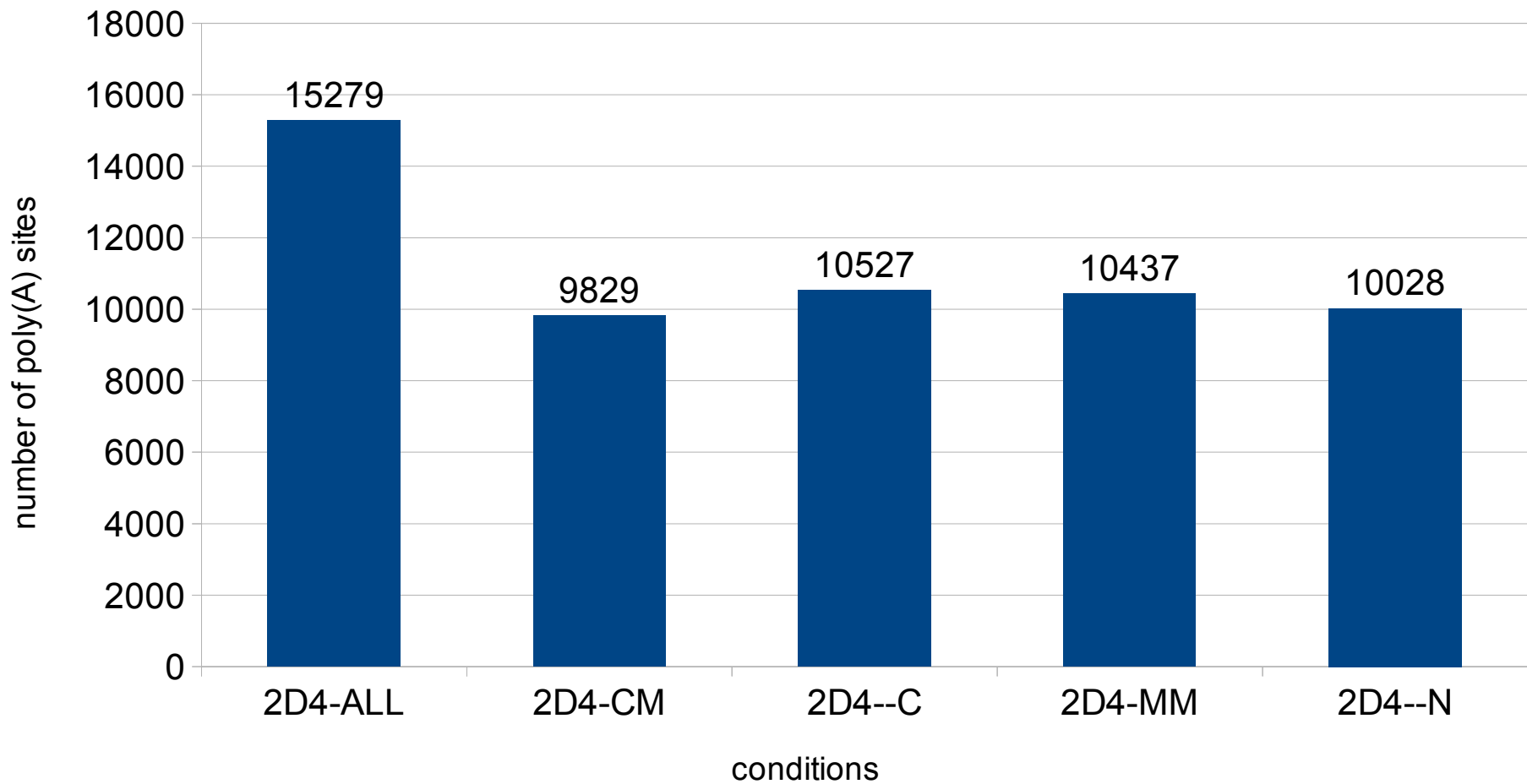
~15000 poly(A) site could be assigned to
annotated genes

Number of poly(A) sites (WT)



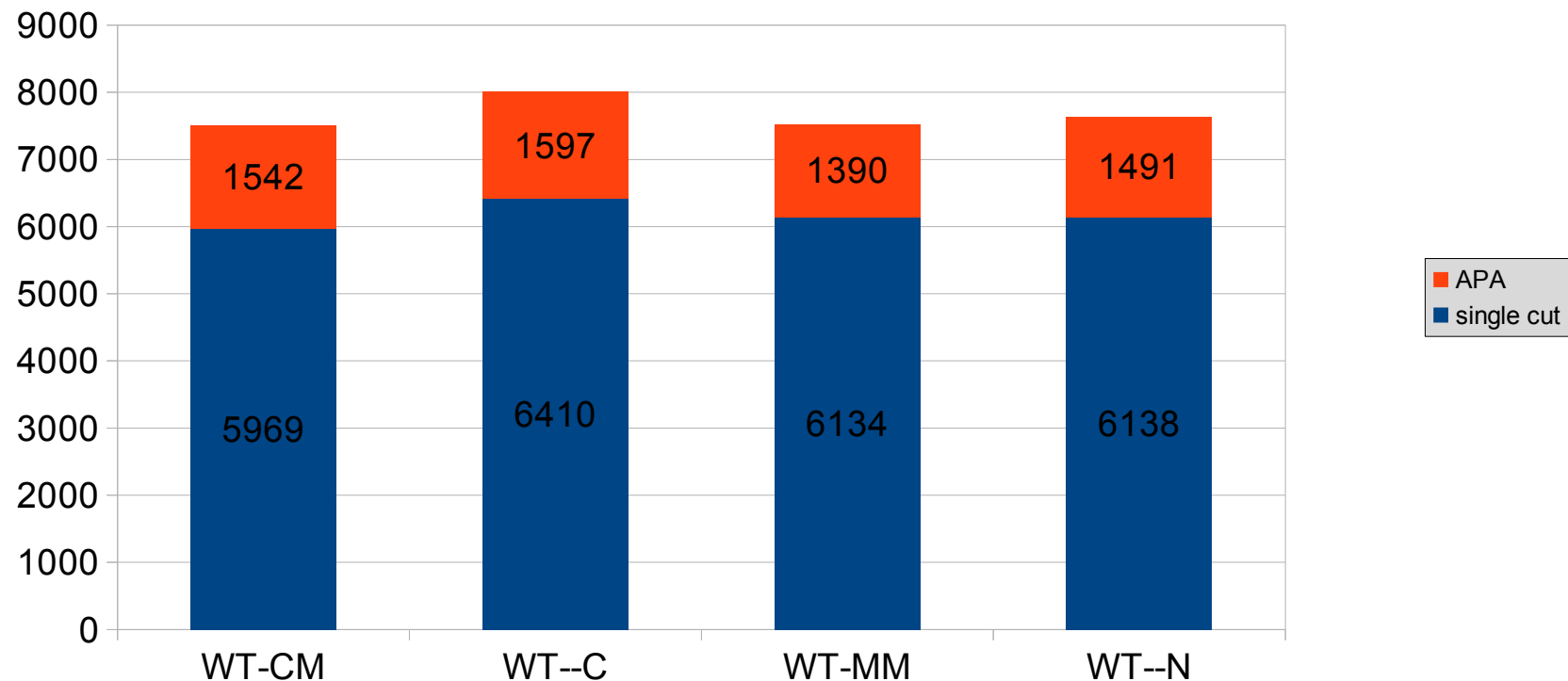
~15000 poly(A) site could be assigned to
annotated genes (2D4)

Number of poly(A) sites (2D4)



~20% of genes are alternatively polyadenilated

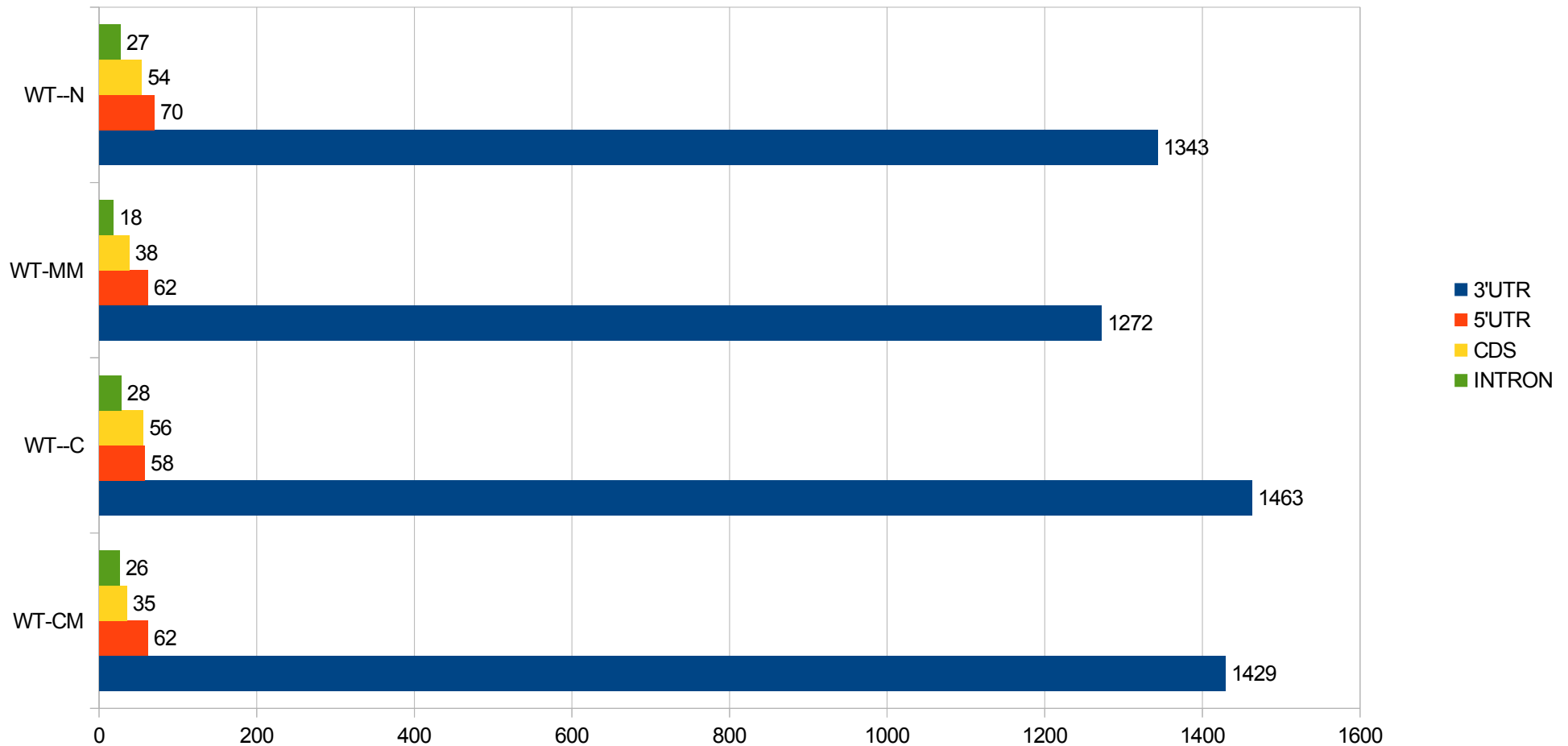
Number of genes with single cut or APA*



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

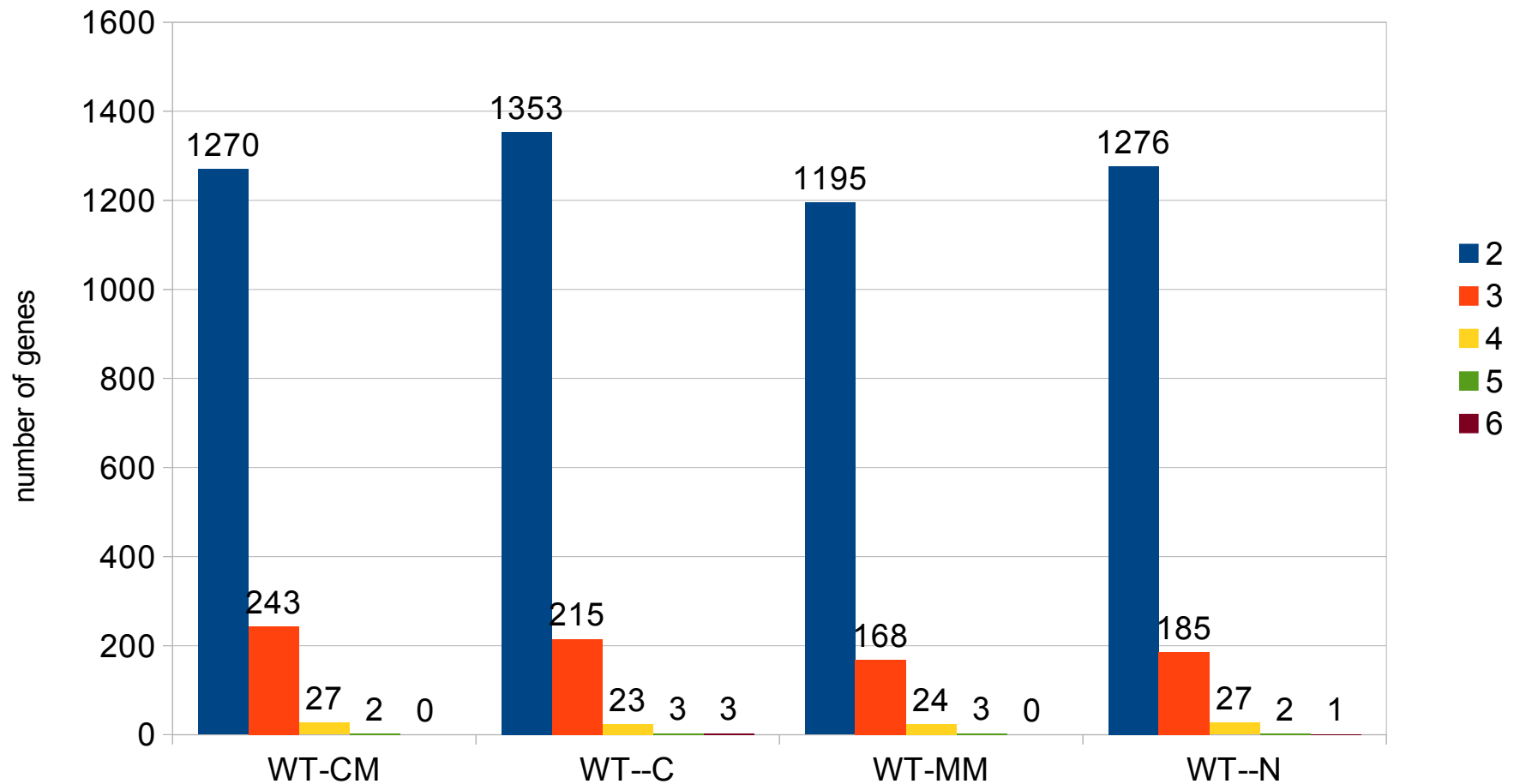
>90% of APA is located in the 3'UTR

Distribution of APA

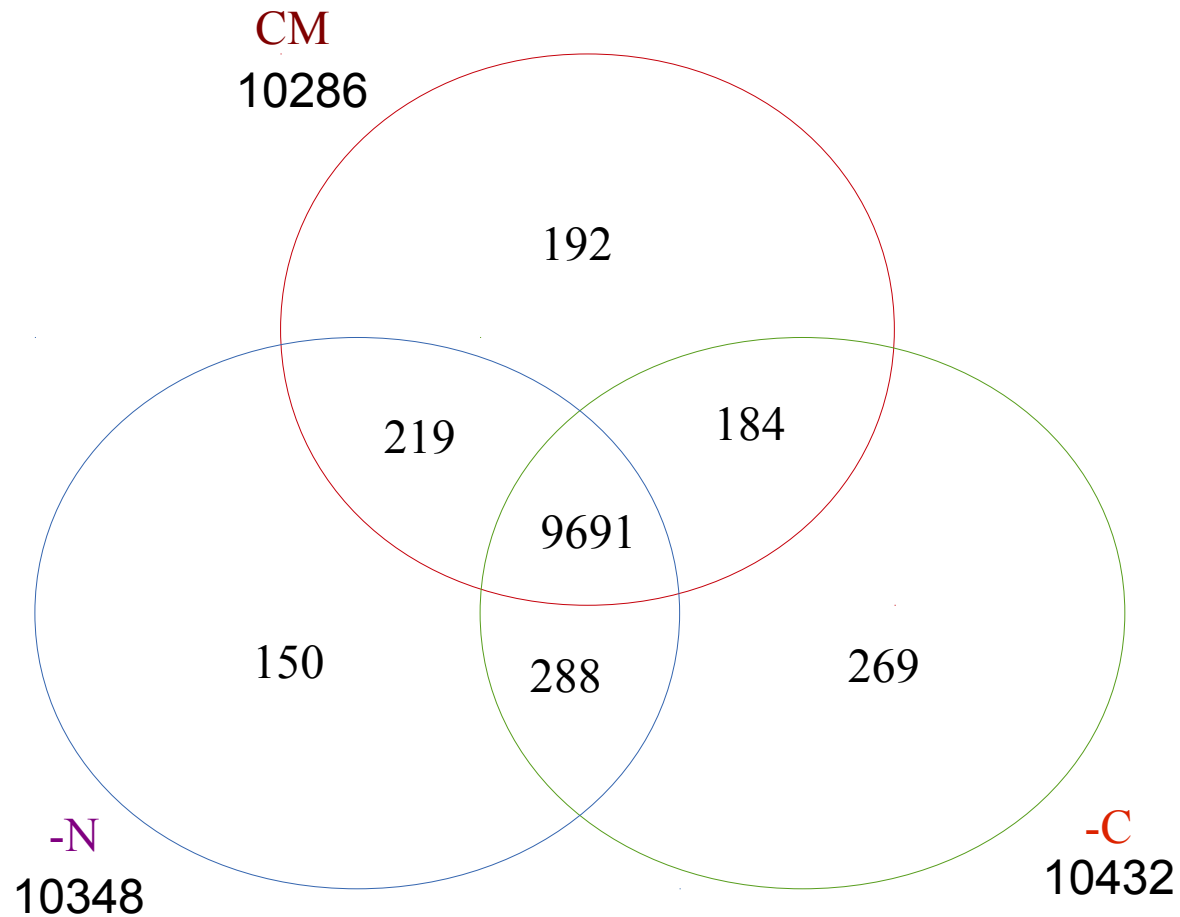


>80% of APA is composed of two cleavage sites

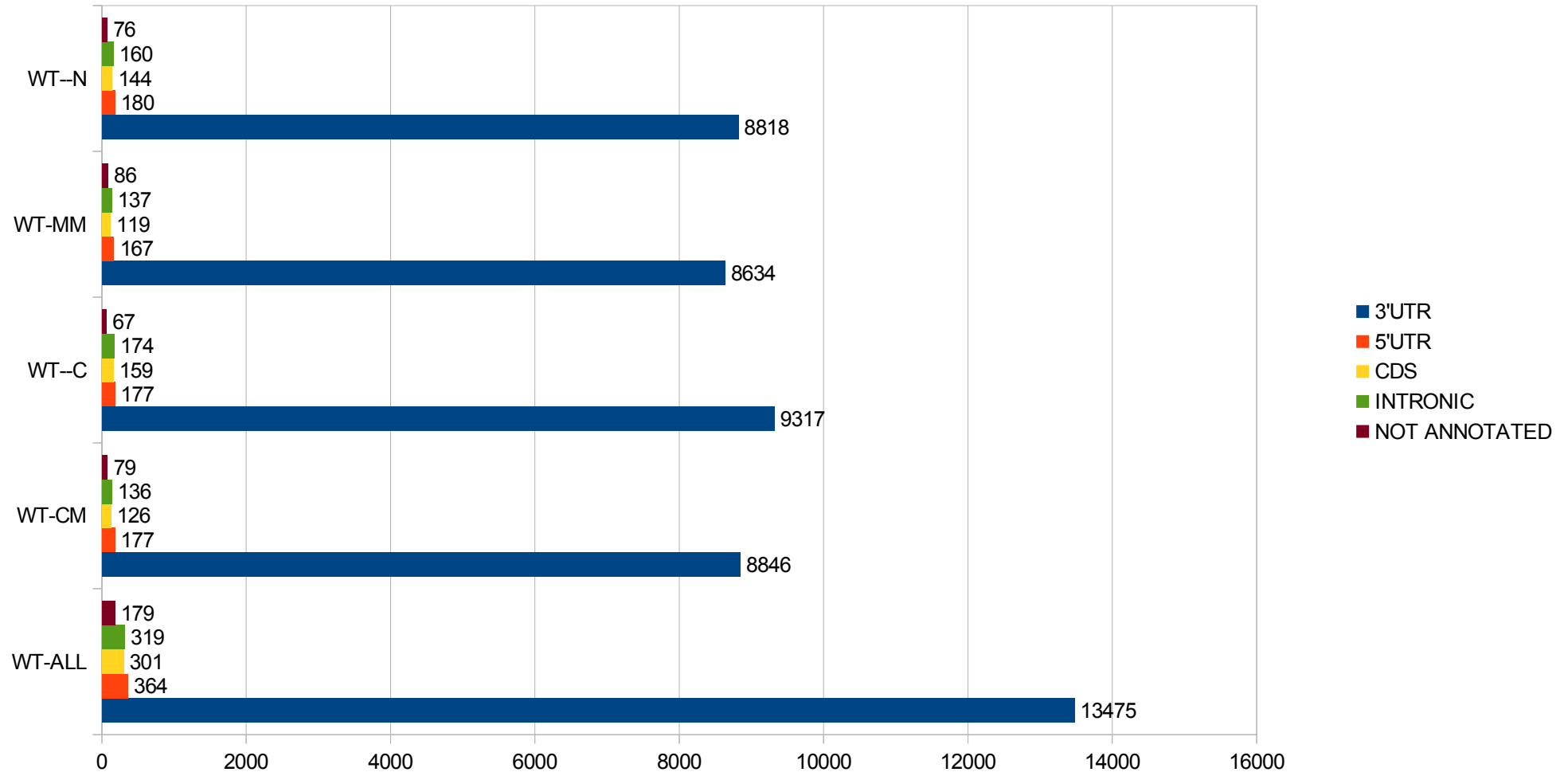
APA - number of cuts per gene



Gene expression between WT conditions

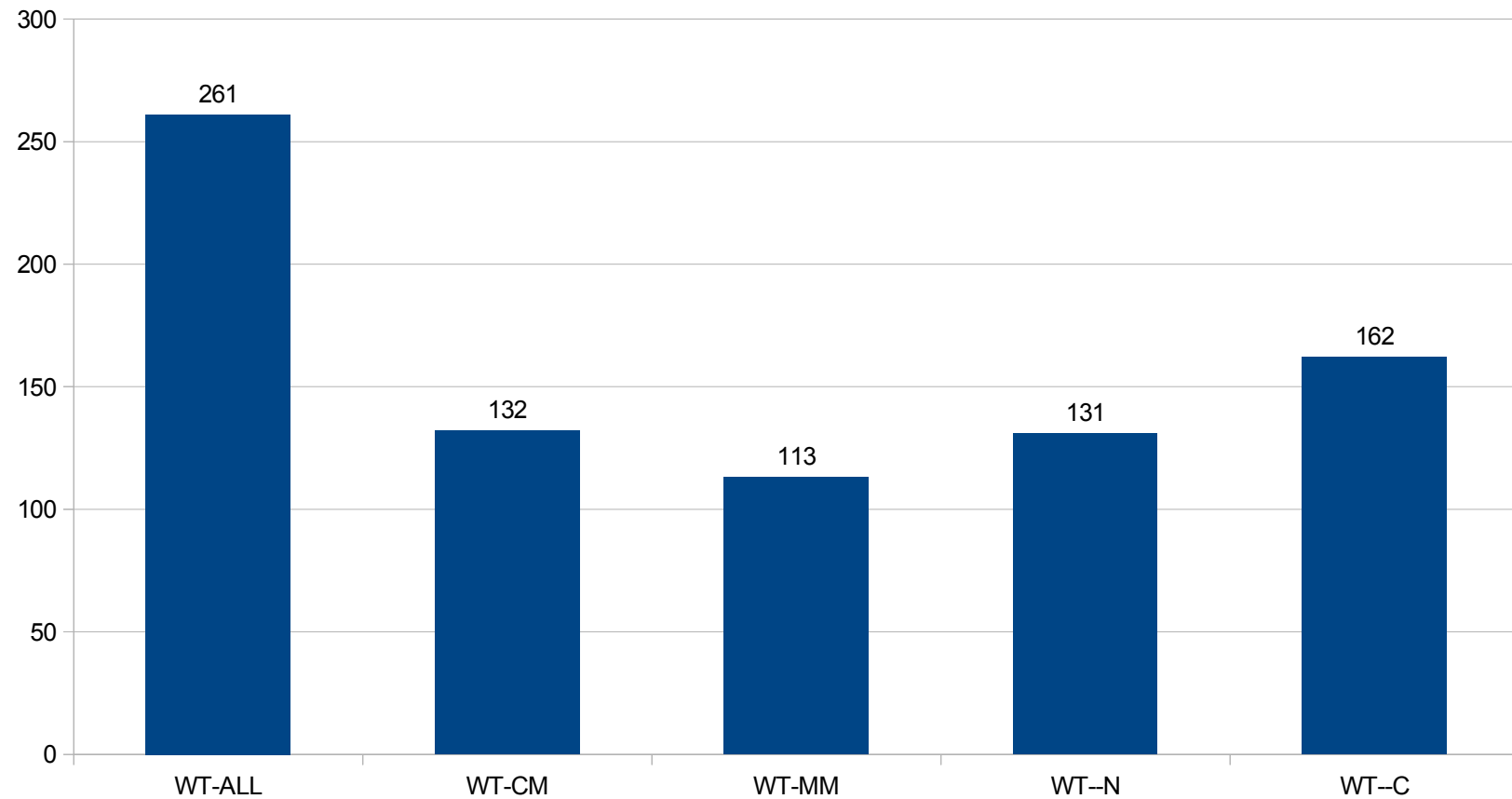


>90% of poly(A) sites are located in the 3'UTR



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



261 orphan poly(A) sites highly expressed in WT (>100 reads)

- 14 hits against other gene copies in *M.oryzae*
- 44 hits against Uniprot nt/nr database
- 4 hits against Rfam(ncRNA) database
- 81 overlapping annotated genes antisense
- 63 matching CPA-sRNA sequences
- 16 matching retrotransposons
- 7 matching avirulence genes

3165 orphan poly(A) sites expressed in WT (>10 reads)

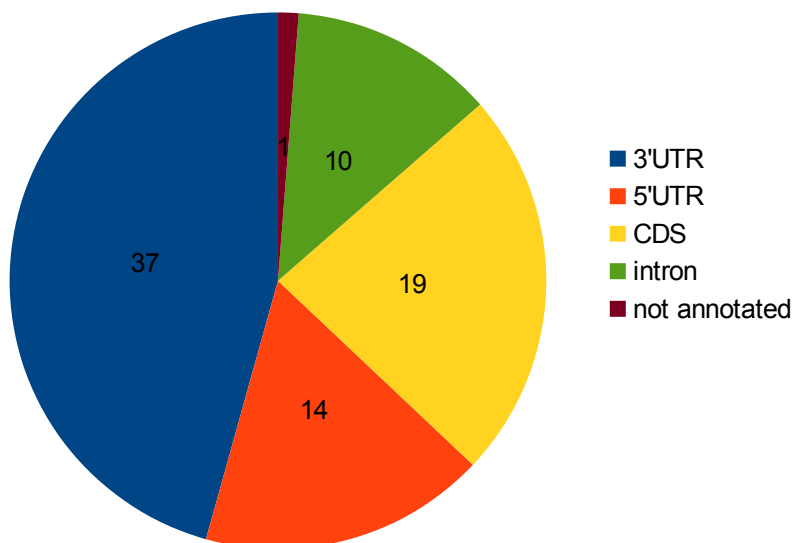
- 102 hits against other gene copies in *M.oryzae*
- 438 hits against Uniprot nt/nr database
- 10 hits against Rfam(ncRNA) database
- 1098 overlapping annotated genes antisense
- 253 matching CPA-sRNA sequences
- 129 matching retrotransposons
- 57 matching avirulence genes

Orphans differentially expressed

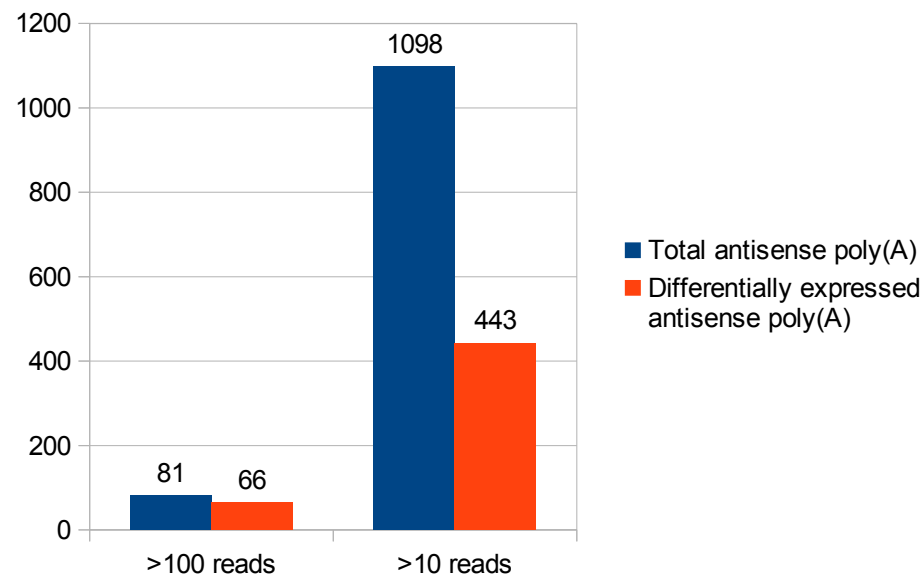
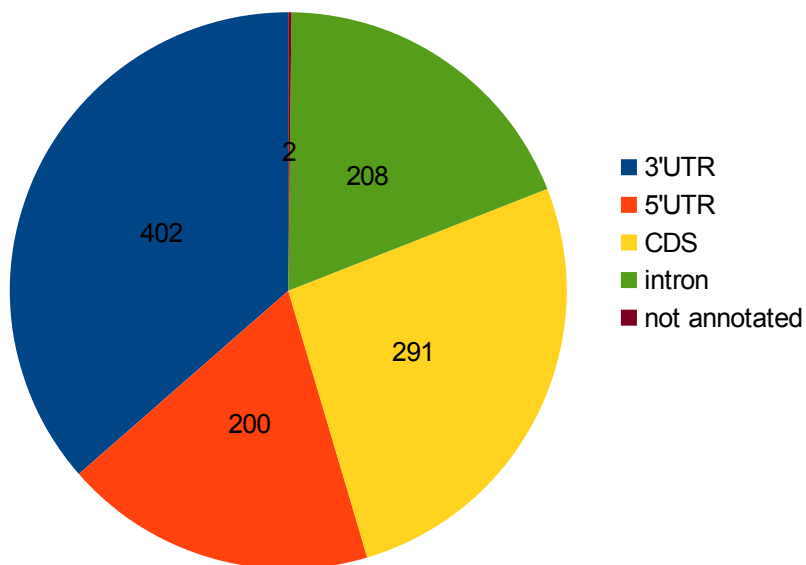
| | up | down | total |
|---------------|-----|------|-------|
| (>100 reads) | | | |
| WT-CM → WT--C | 64 | 73 | 137 |
| WT-CM → WT-MM | 4 | 2 | 6 |
| WT-CM → WT--N | 11 | 13 | 24 |
| WT-MM → WT--C | 37 | 54 | 91 |
| WT-MM → WT--N | 0 | 0 | 0 |
| WT--N → WT--C | 24 | 31 | 55 |
| (>10 reads) | | | |
| WT-CM → WT--C | 344 | 109 | 453 |
| WT-CM → WT-MM | 8 | 1 | 9 |
| WT-CM → WT--N | 8 | 7 | 15 |
| WT-MM → WT--C | 181 | 62 | 243 |
| WT-MM → WT--N | 0 | 0 | 0 |
| WT--N → WT--C | 115 | 23 | 138 |

Antisense poly(A) are usually located in the 3'UTR, most of antisense poly(A) are differentially expressed in any condition

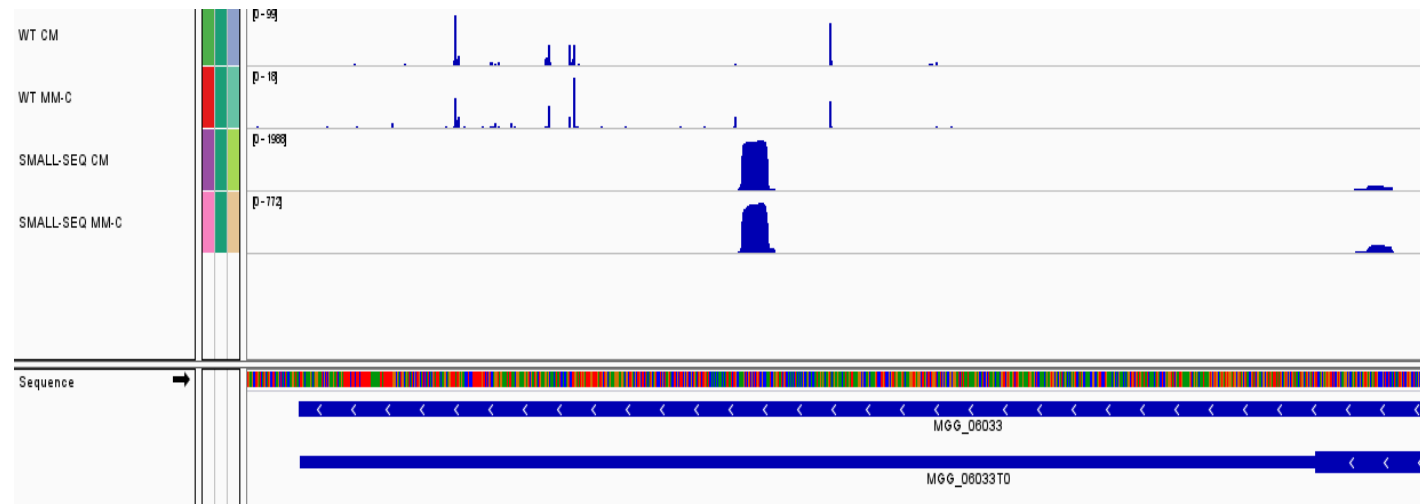
Location of antisense poly(A) sites >100 reads



Location of poly(A) sites >10 reads

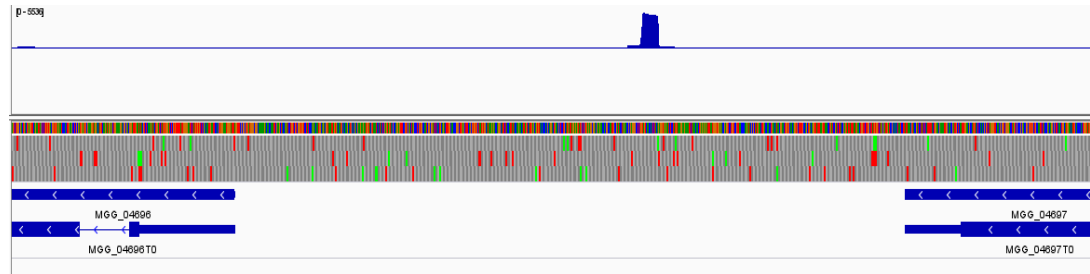


Small RNA expression and alternative polyadenylation

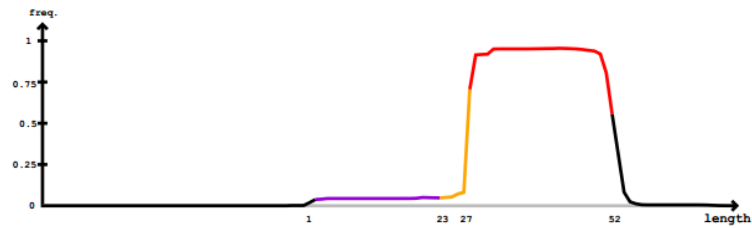
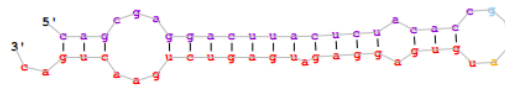


- 72 gene in CM with peaks of small RNA expression in intra-APA regions
- 36 gene in -C with peaks of small RNA expression in intra-APA regions

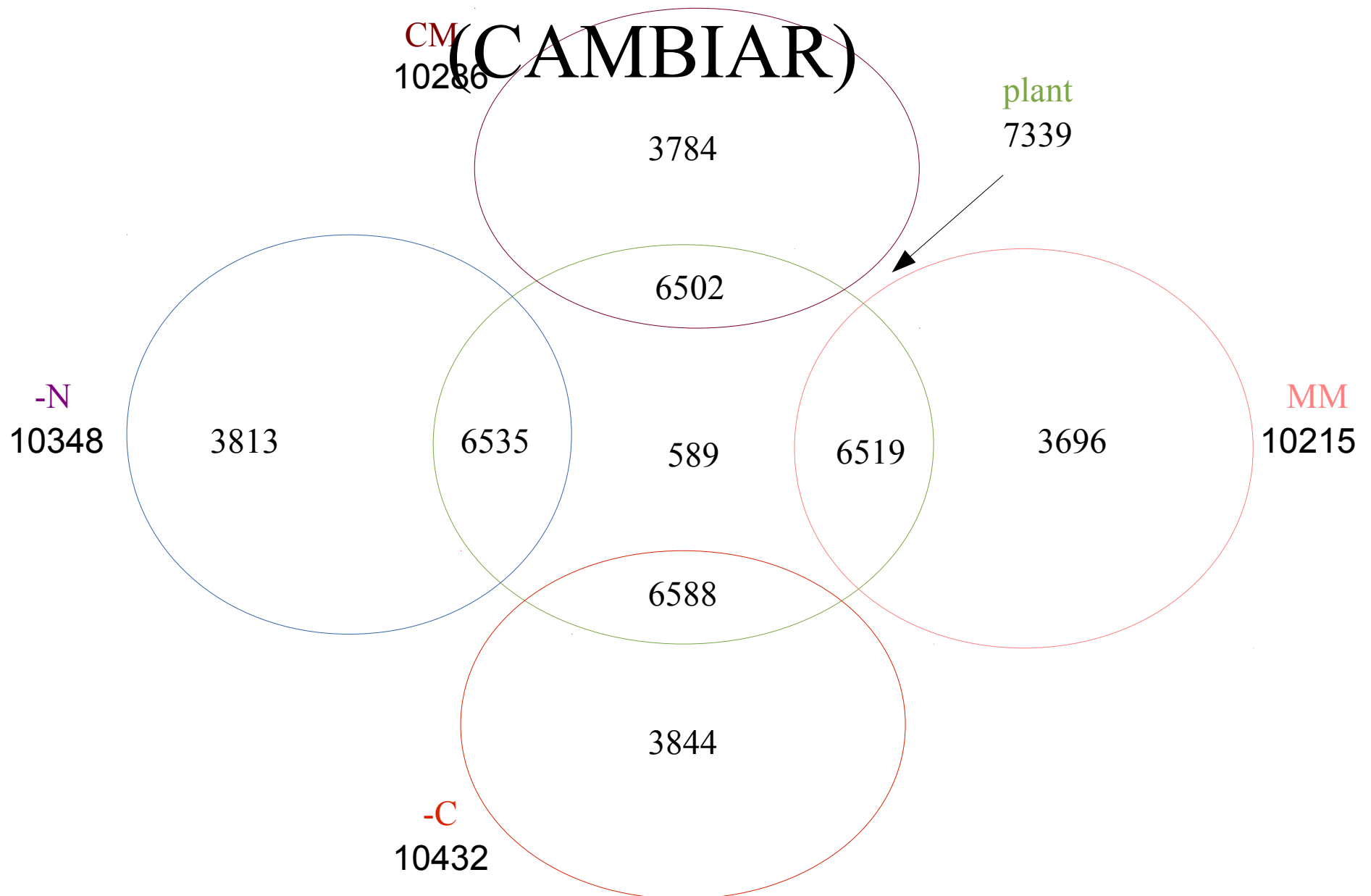
Putative microRNAs?



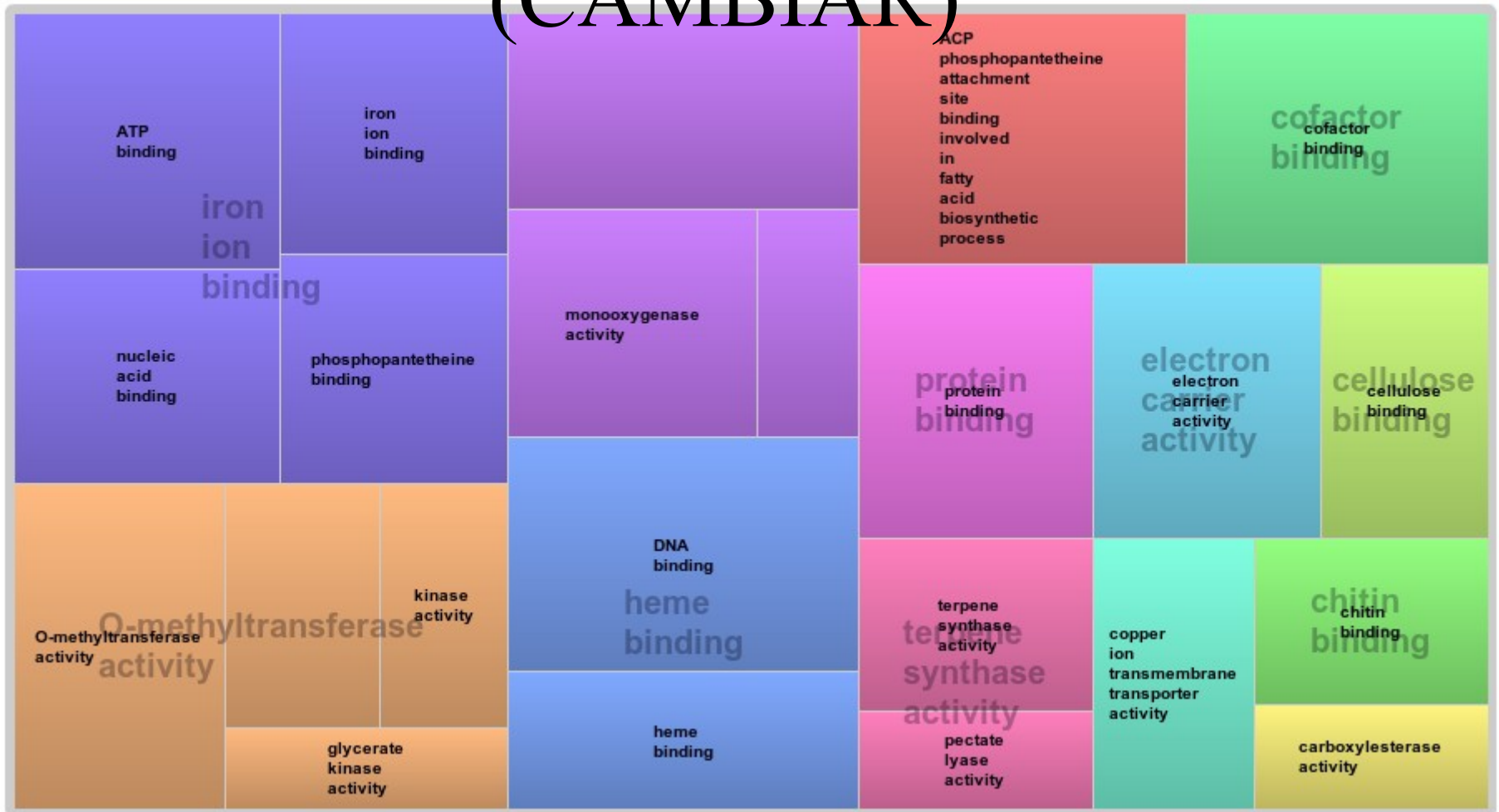
| | |
|------------------------|----------------------|
| Provisional ID | : Chromosome_2_36780 |
| Score total | : 2887.6 |
| Score for star read(s) | : 3.9 |
| Score for read counts | : 2882.2 |
| Score for mfe | : 0.7 |
| Score for randfold | : -2.2 |
| Score for cons. seed | : 3 |
| Total read count | : 5665 |
| Mature read count | : 5415 |
| Loop read count | : 0 |
| Star read count | : 250 |

[illegible]

589 genes expressed in plant are never expressed in vitro



Plant specific genes functional analysis (CAMBIAR)

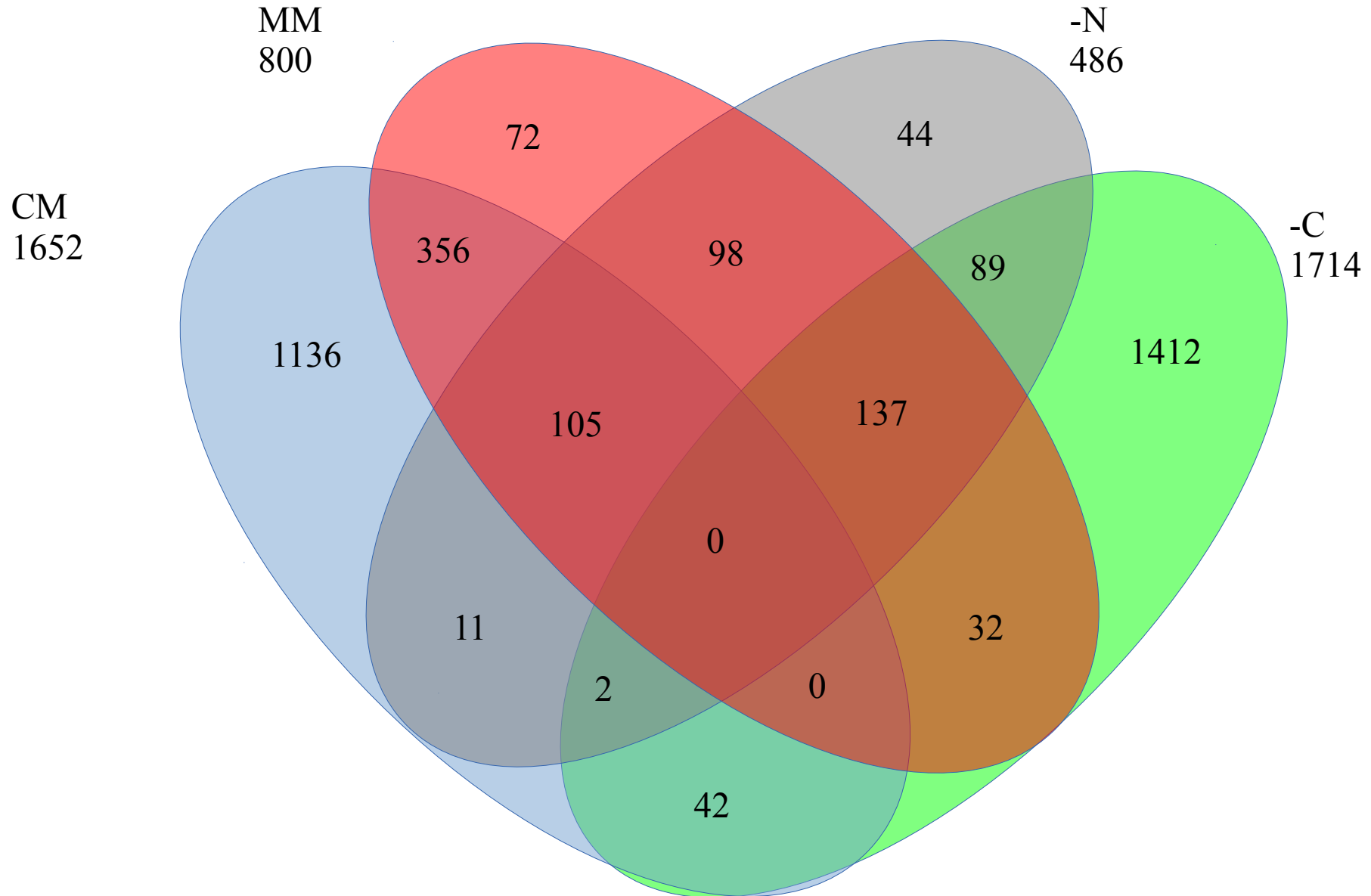


$p\text{-value} = 0.05$

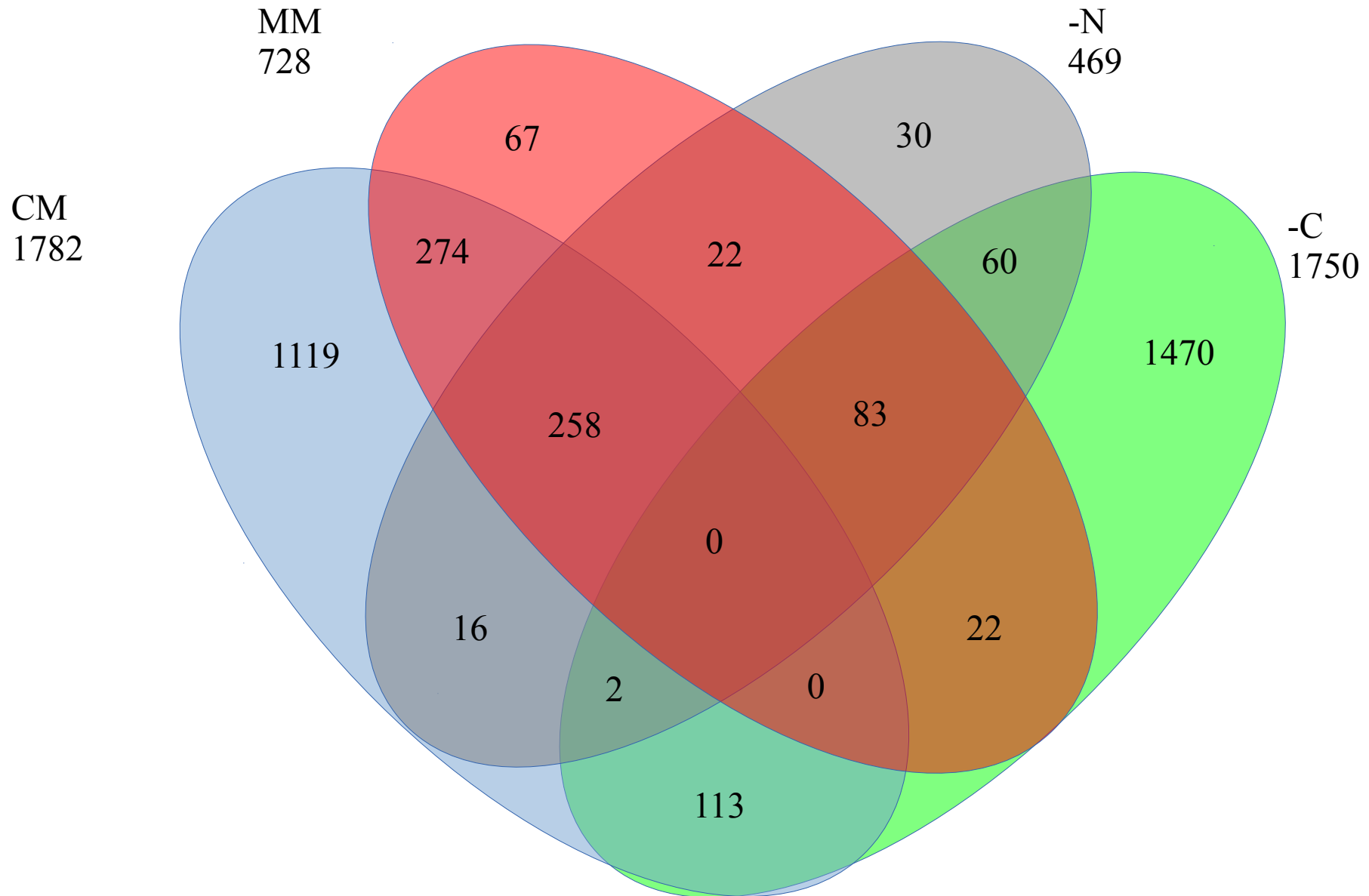
The CM \rightarrow -C condition presents the highest number of differentially expressed genes, while MM \rightarrow -N the lowest

| DIFFERENTIALLY EXPRESSED GENES IN THE WT | | | |
|--|------|------|-------|
| | DOWN | UP | TOTAL |
| CM \rightarrow MM-C | 1586 | 1617 | 3203 |
| CM \rightarrow MM | 151 | 285 | 436 |
| CM \rightarrow MM-N | 177 | 347 | 524 |
| MM \rightarrow MM-C | 601 | 608 | 1209 |
| MM \rightarrow MM-N | 1 | 5 | 6 |
| MM-N \rightarrow MM-C | 216 | 313 | 529 |

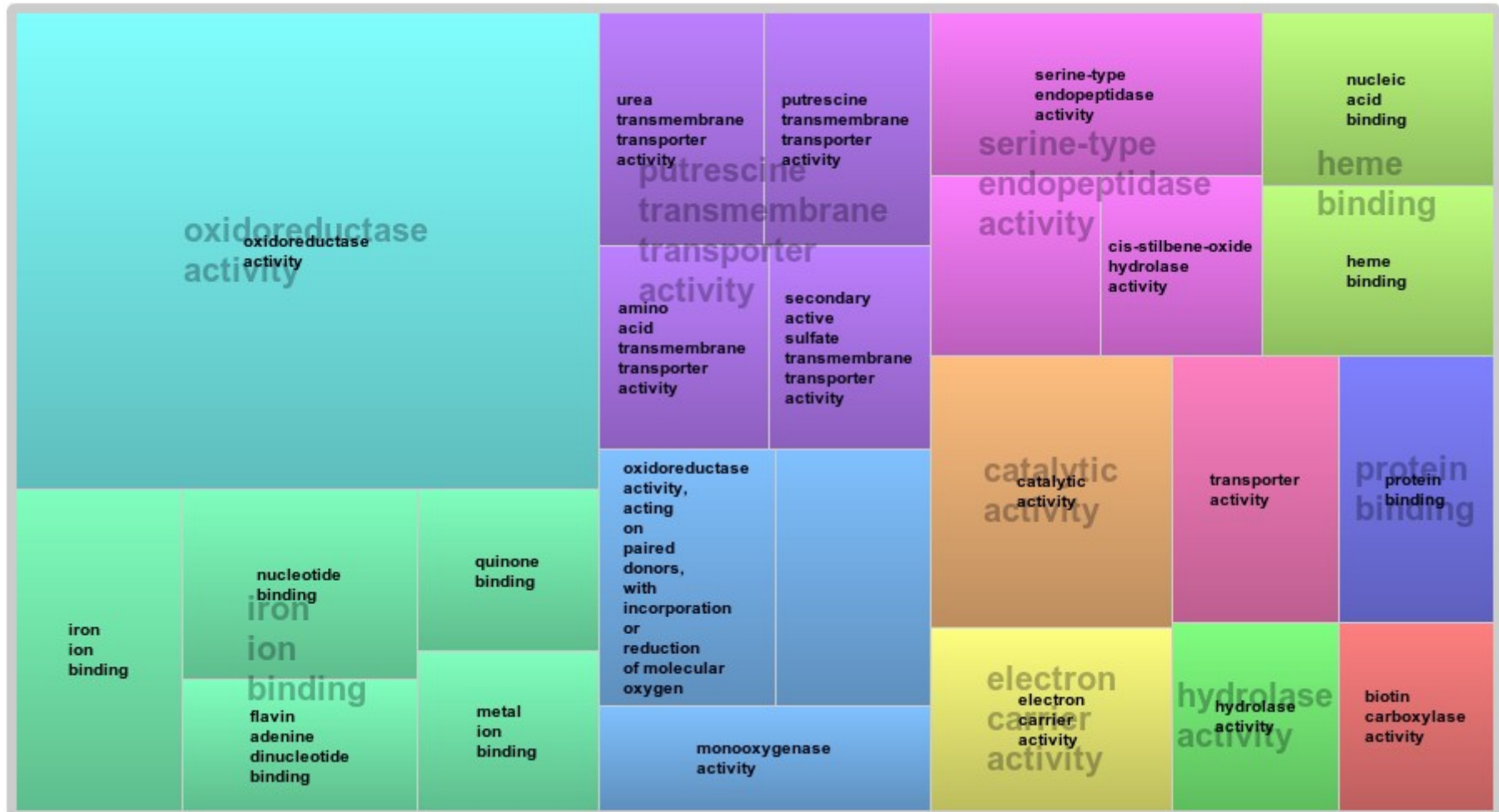
Up-regulated genes between conditions



Down-regulated genes between conditions

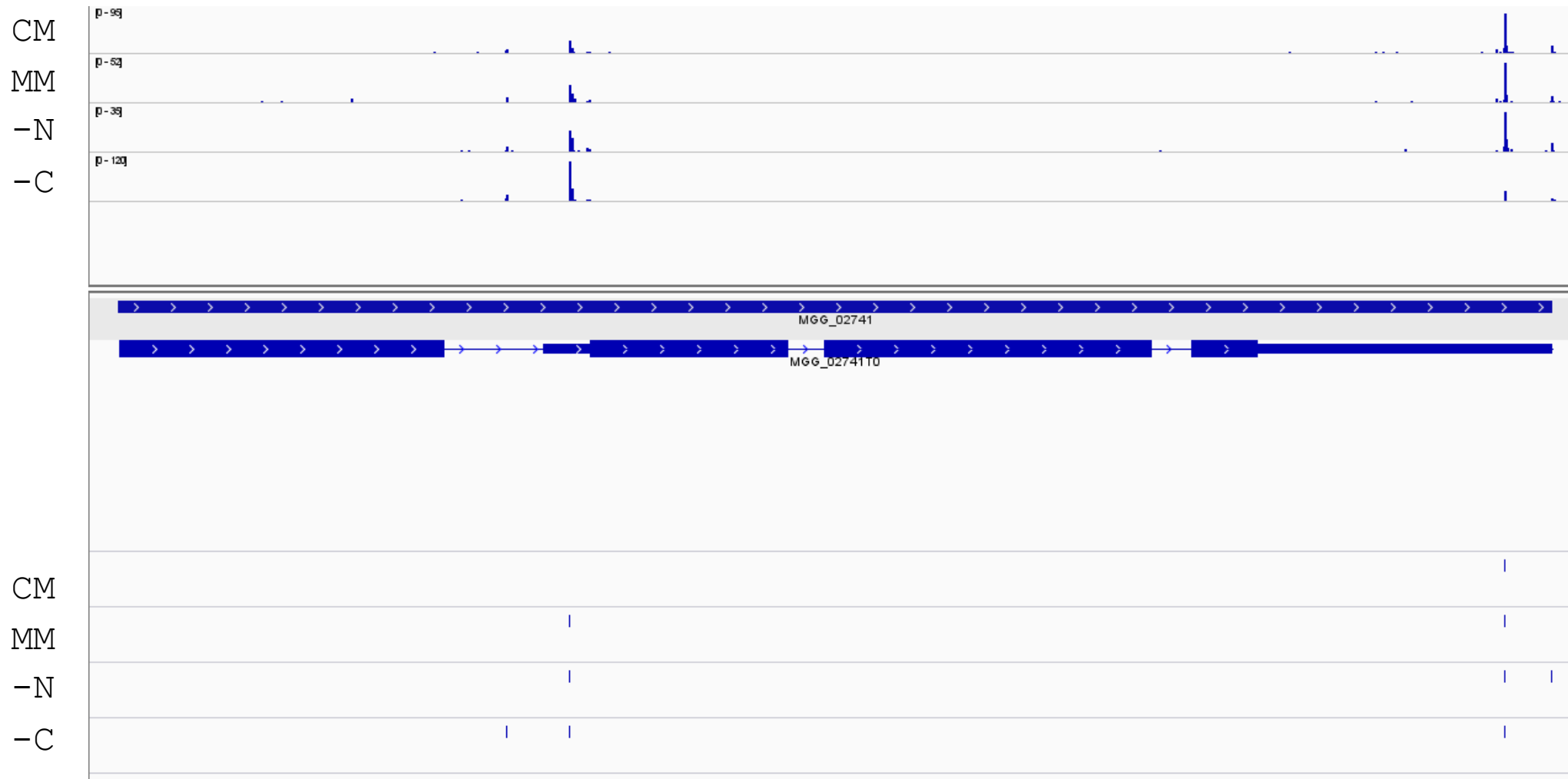


CM → MM functional analysis

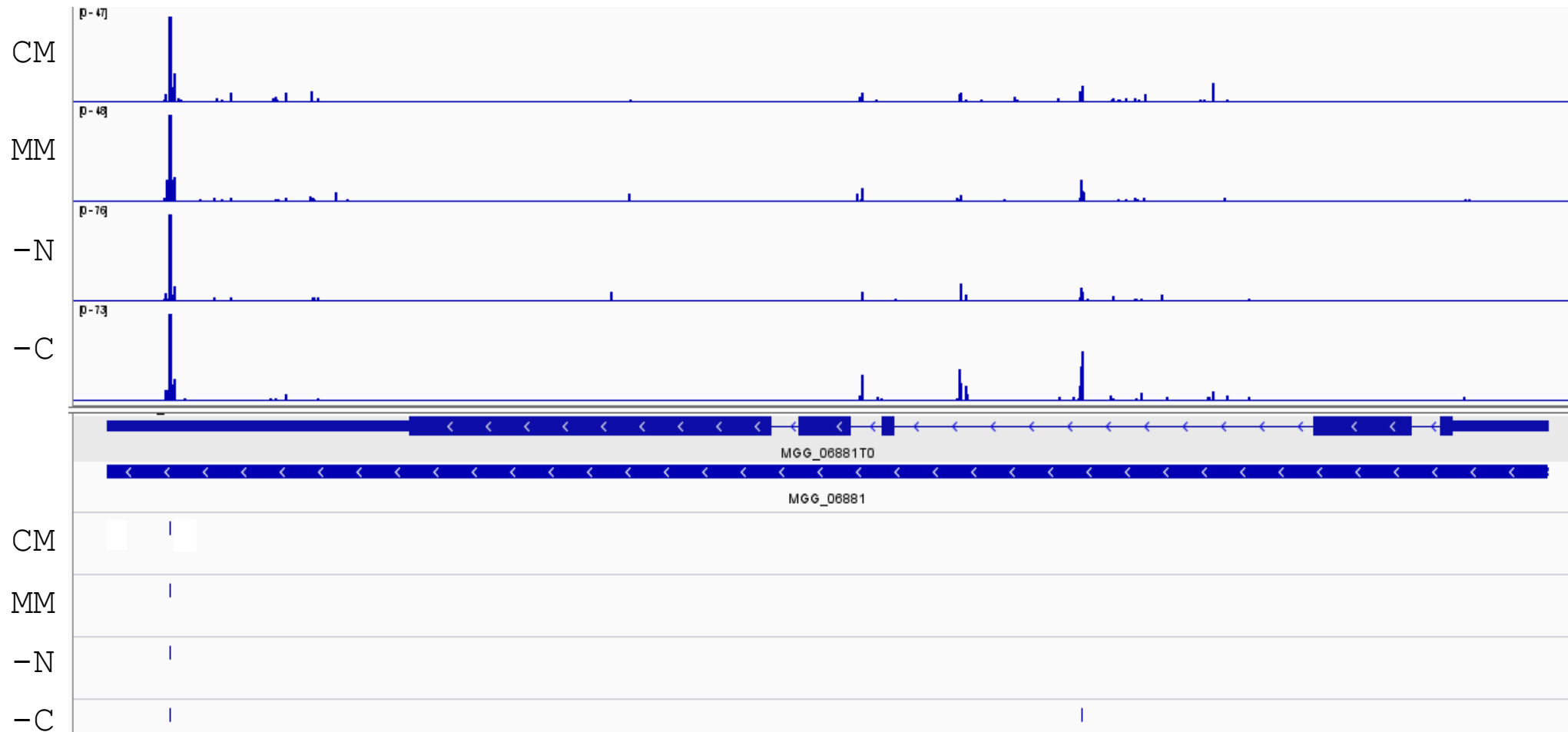


$p\text{-value} = 0.1$

rbp35 shows different polyadenylation in each medium, with strong differences in MM-C

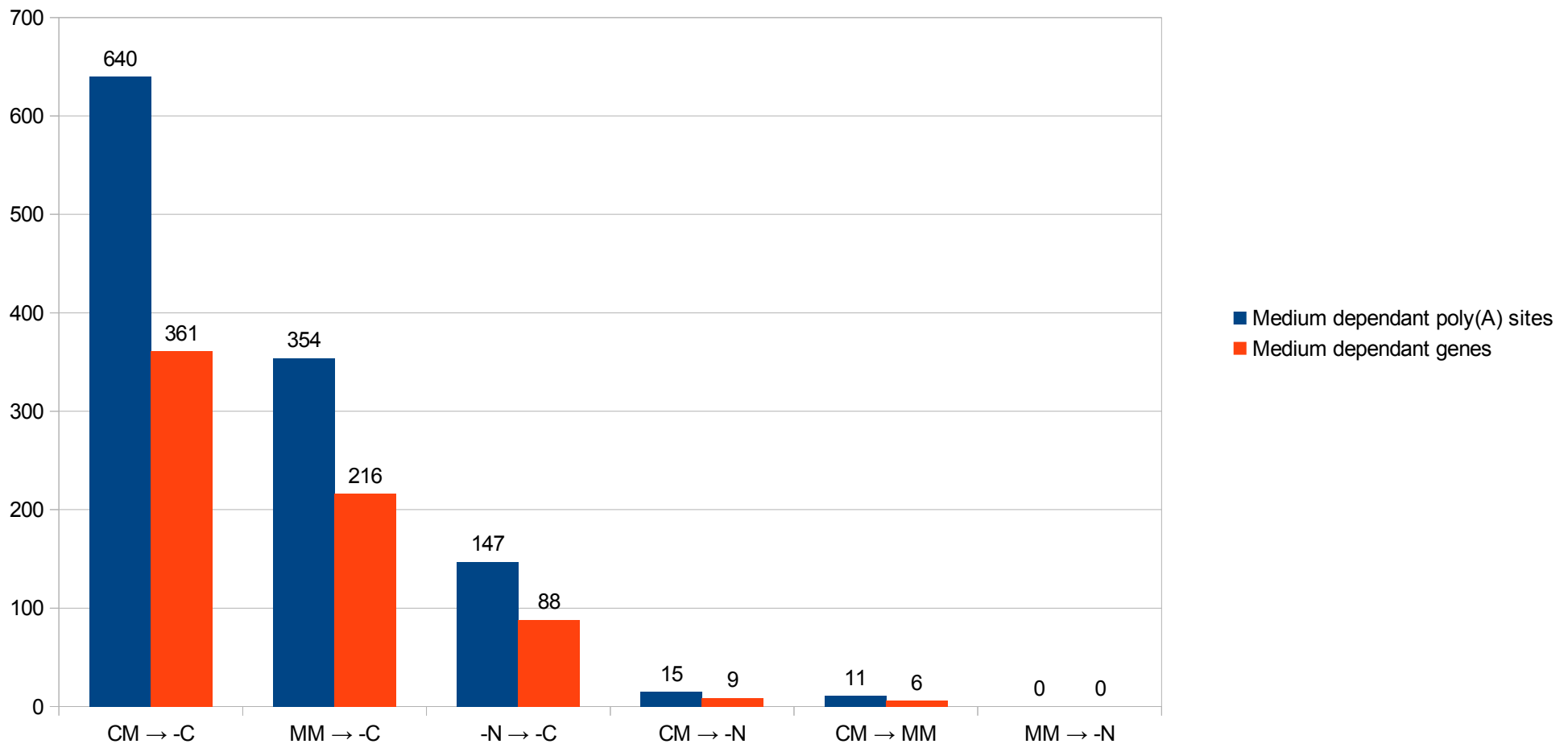


hrp1 shows an up-regulated intronic poly(A) site in MM-C



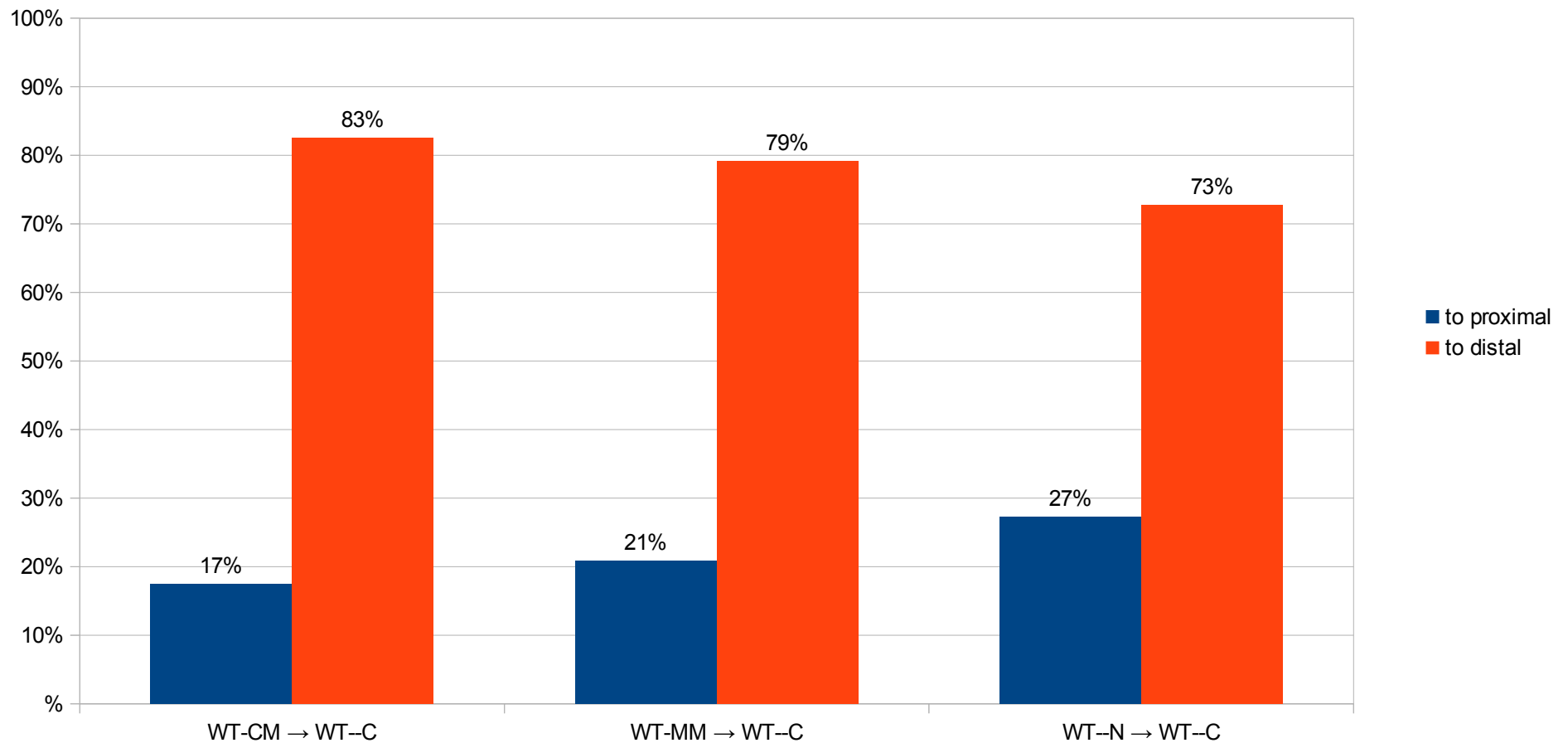
MM-C affects a great number of poly(A) sites

Medium dependant poly(A) sites and genes



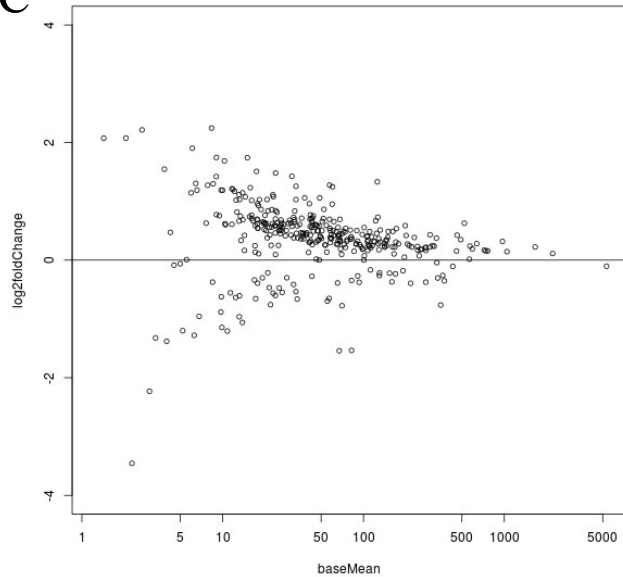
MM-C affects poly(A) sites usage, preferring distal cuts - percentages

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

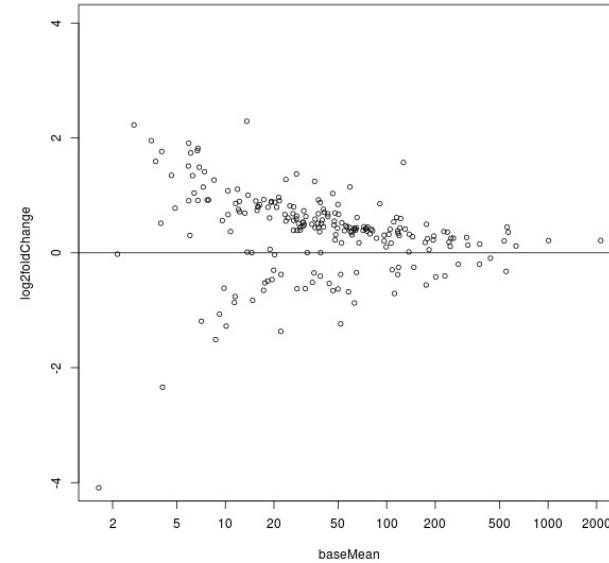


MM-C affects poly(A) sites usage, preferring distal cuts - foldChange

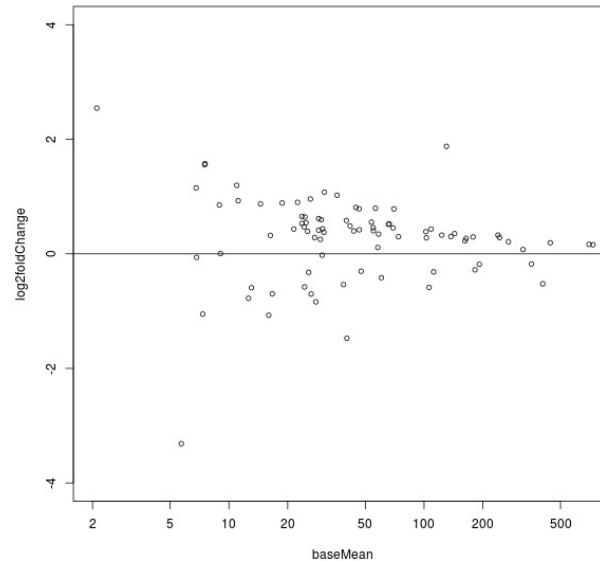
CM \rightarrow -C



MM \rightarrow -C

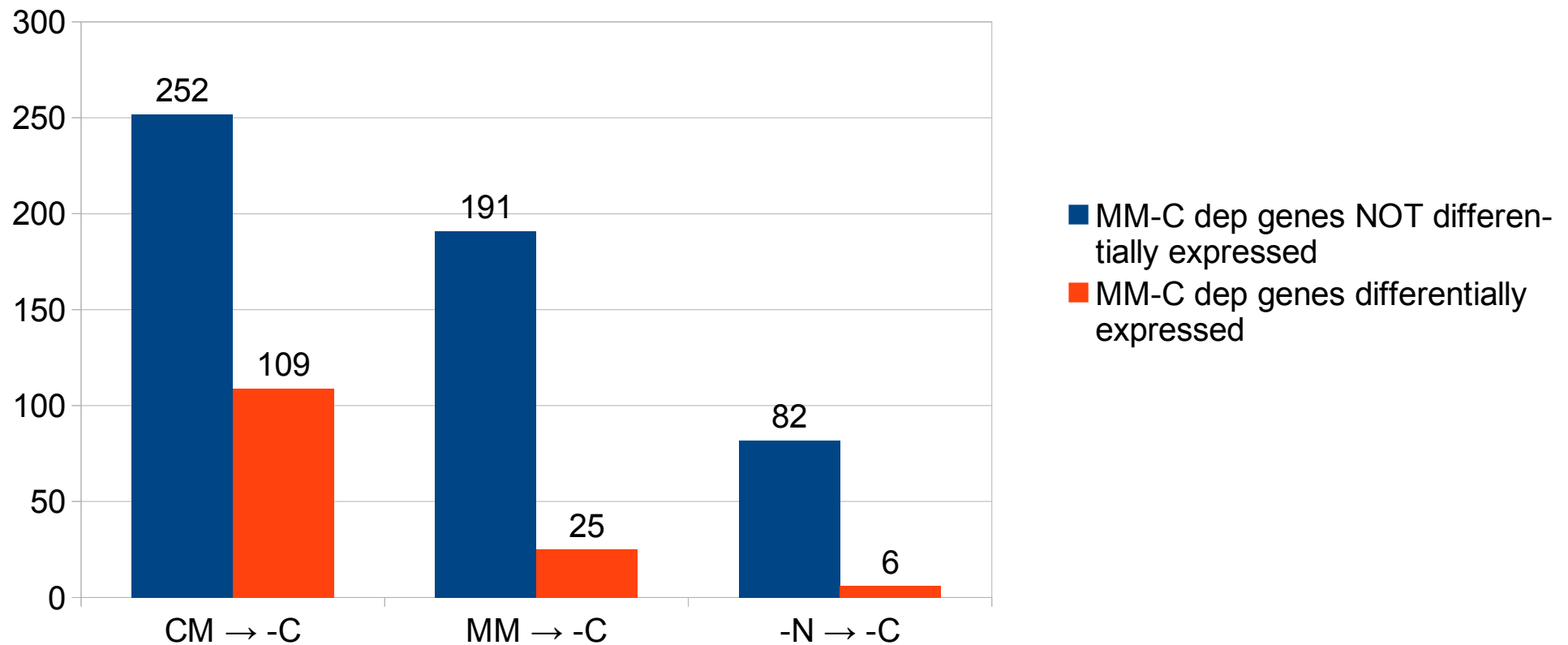


-N \rightarrow -C



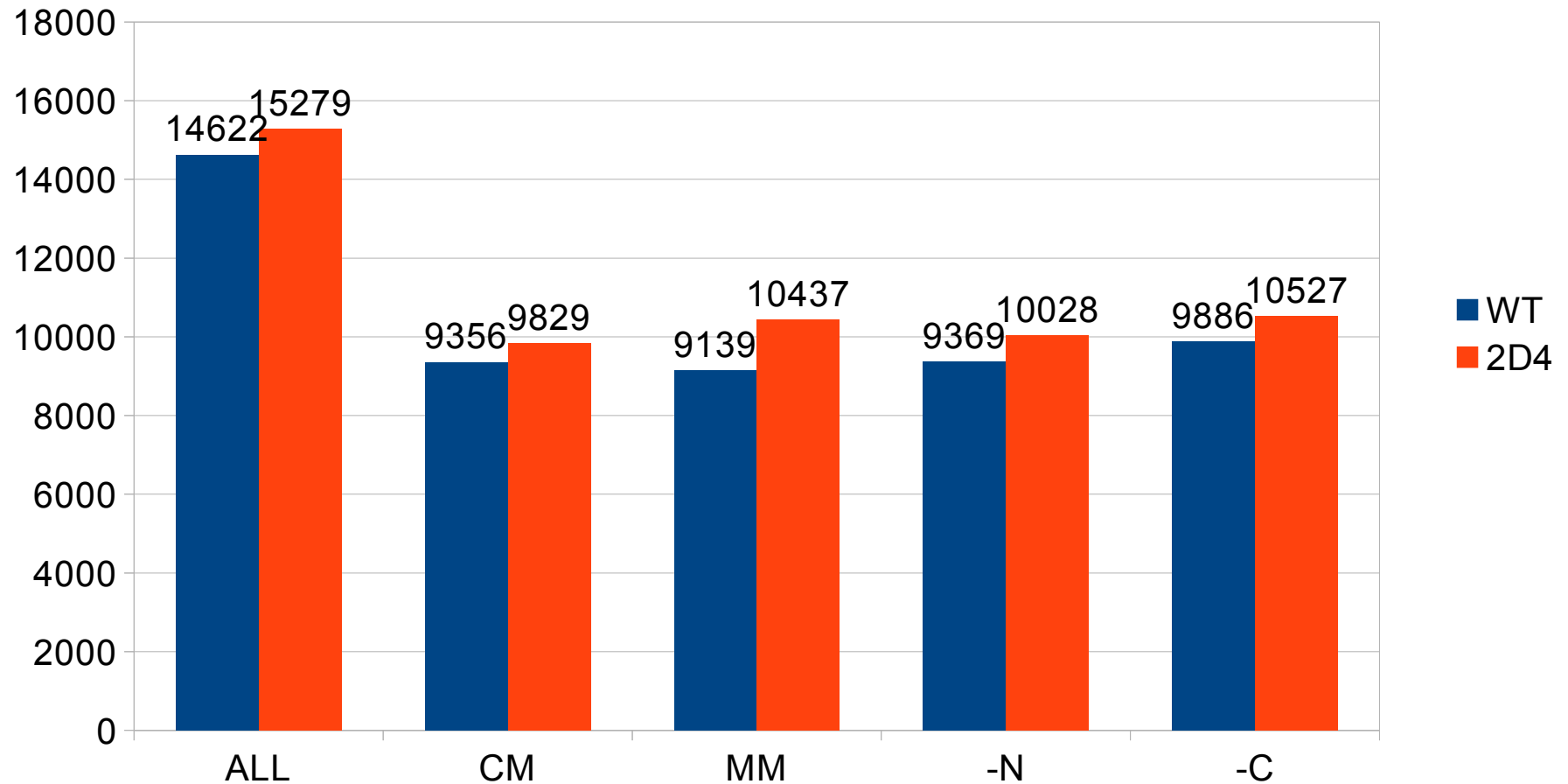
MM-C dependent genes are not always differentially expressed

MM-C dependance vs differential expression



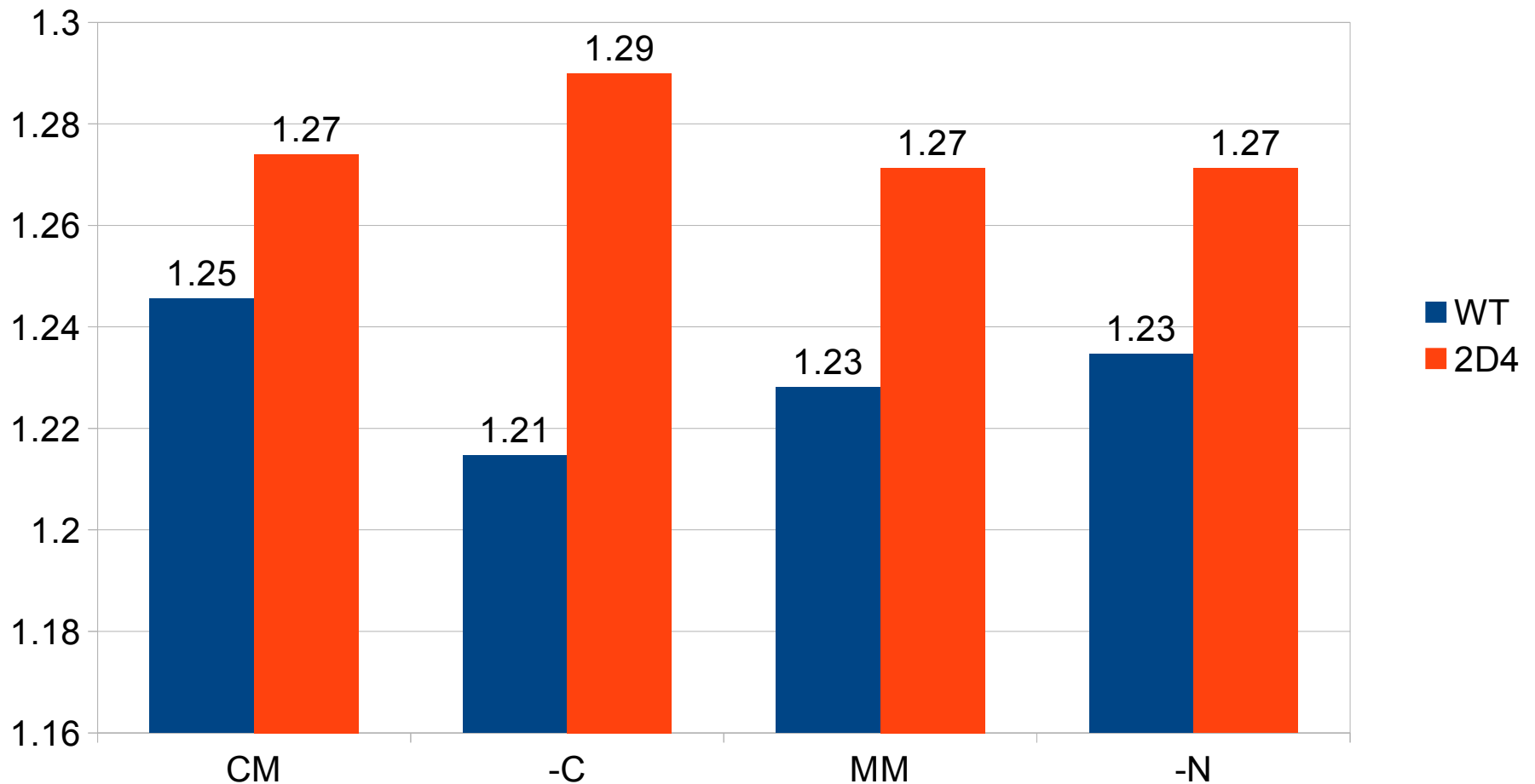
Δrbp35 affects poly(A) sites number

Poly-A sites number WT vs 2D4

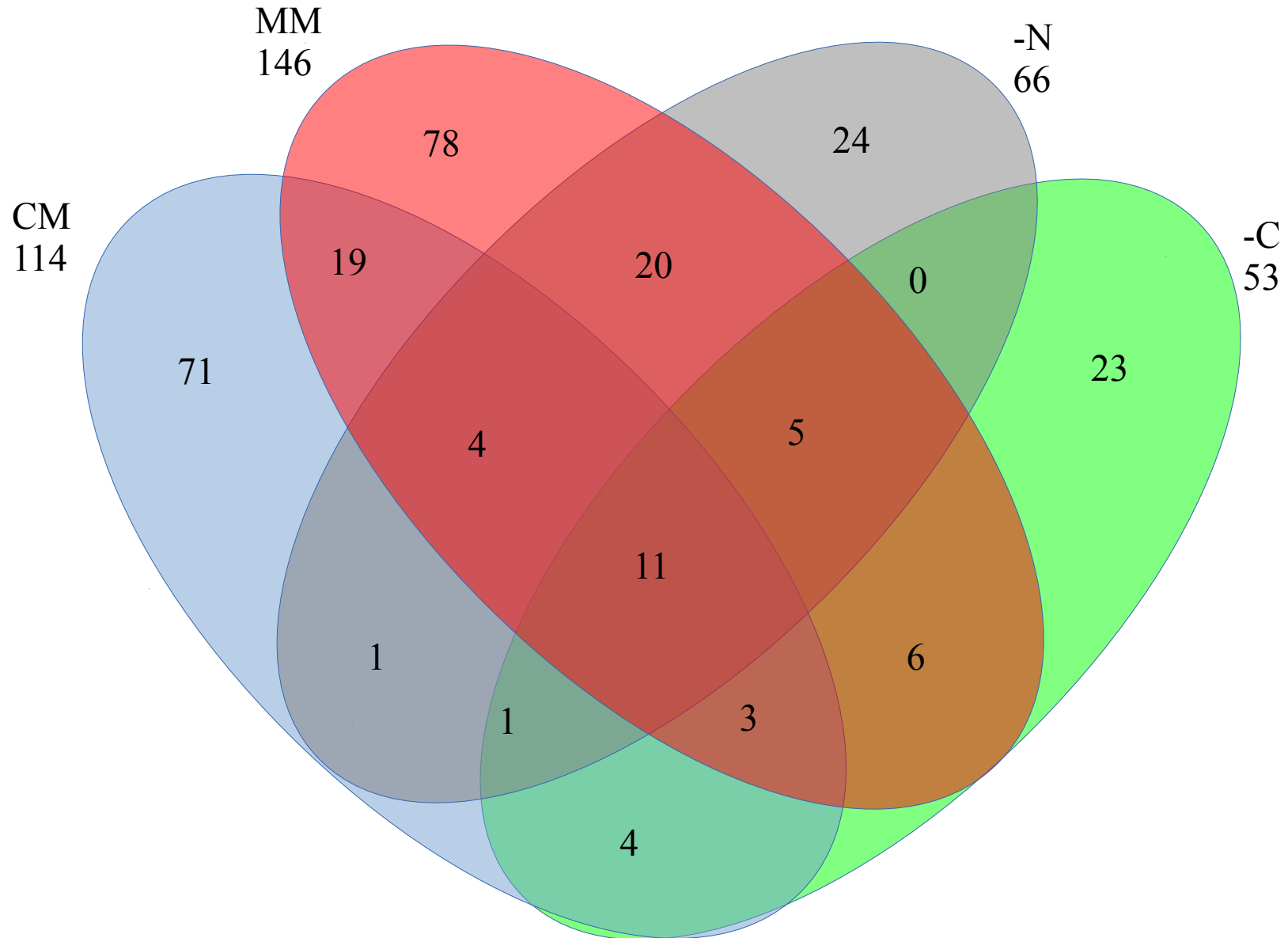


Δrbp35 affects number of cut sites per gene

Number of cut sites per gene



Differentially expressed genes WT \rightarrow $\Delta rbp35$



WT → 2D4 functional analysis



$p\text{-value} = 0.05$

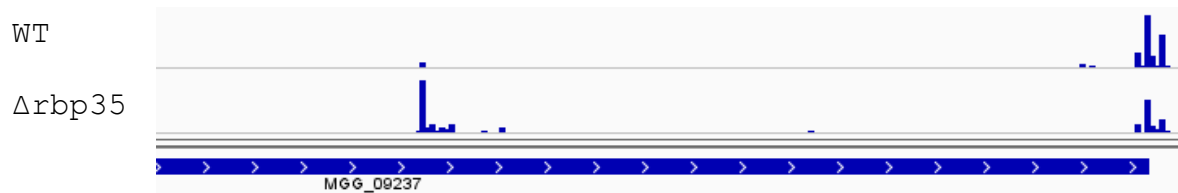


Terminology

- **pRBP35dep**: poly(A) sites that show a differential expression between wild-type and Δ rbp35. We call it “RBP35 dependent poly(A) sites”
- **pRBP35dep_down**: a down-regulated RBP35 dependent poly(A):

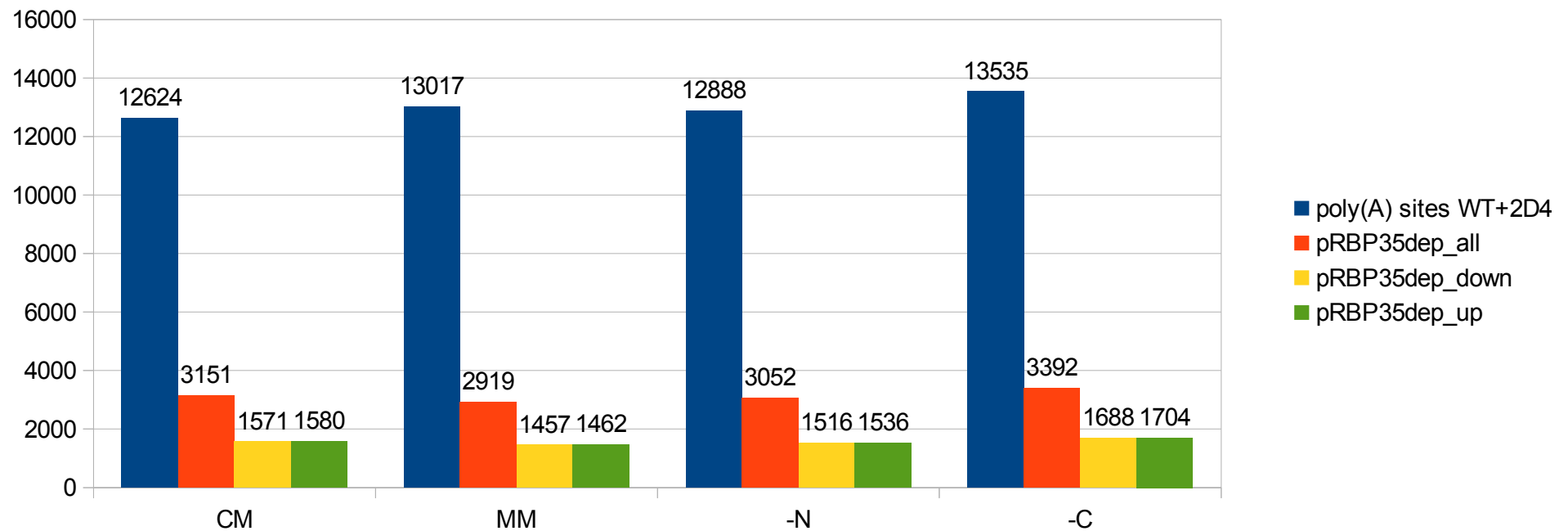


- **pRBP35dep_up**: an up-regulated RBP35 dependent poly(A):

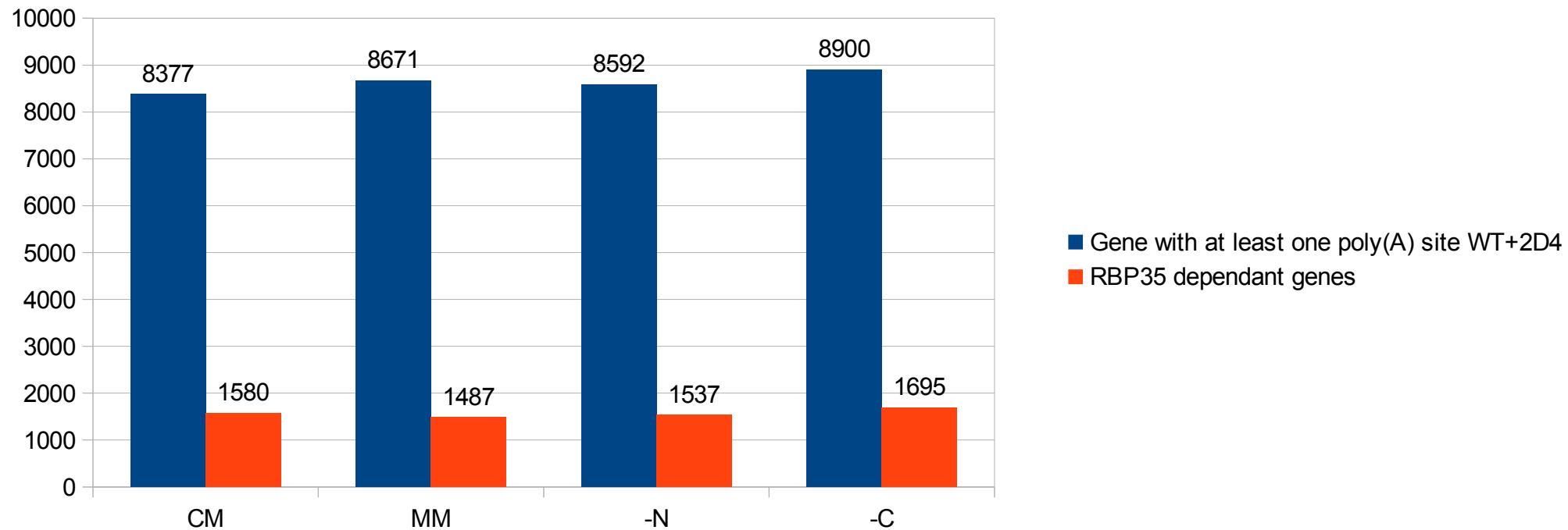


- A gene is defined “RBP35 dependent gene” (or simply **RBP35dep**) when one or more of its poly(A) belong to the previous groups

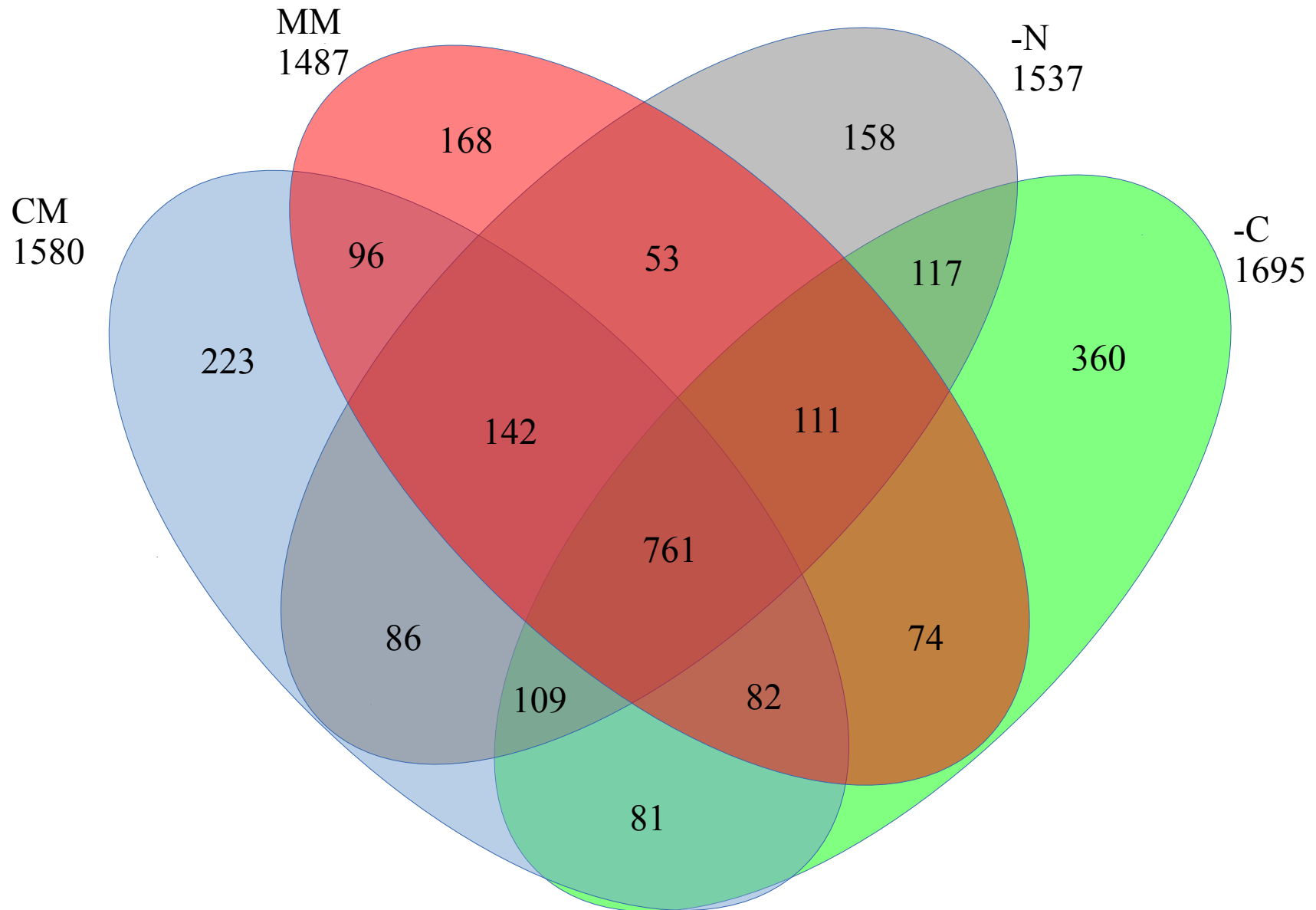
22%-25% of poly(A) sites are dependent from RBP35 in all media



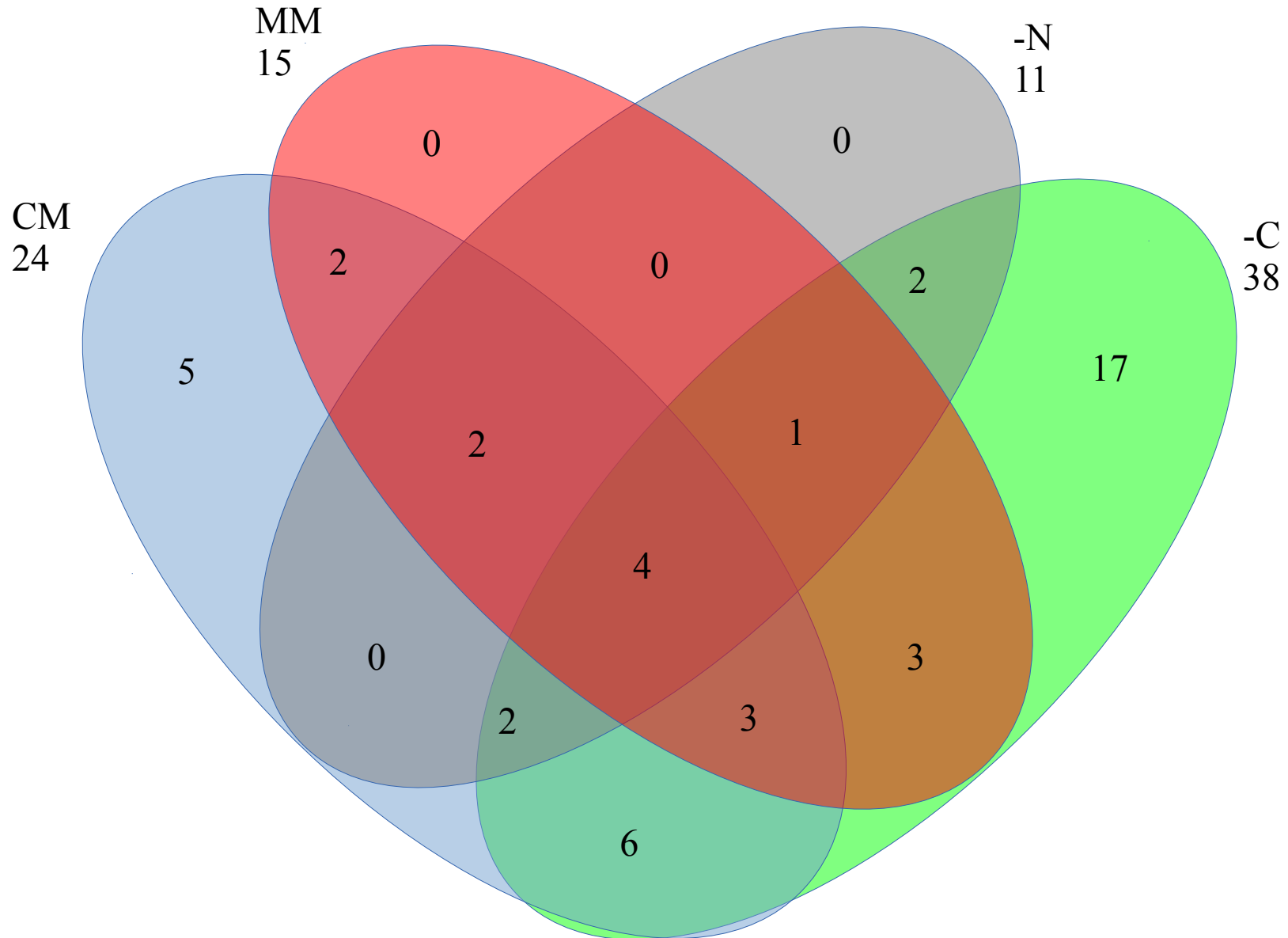
17%-19% of genes are dependent from RBP35 in all media



RBP35 dependant genes

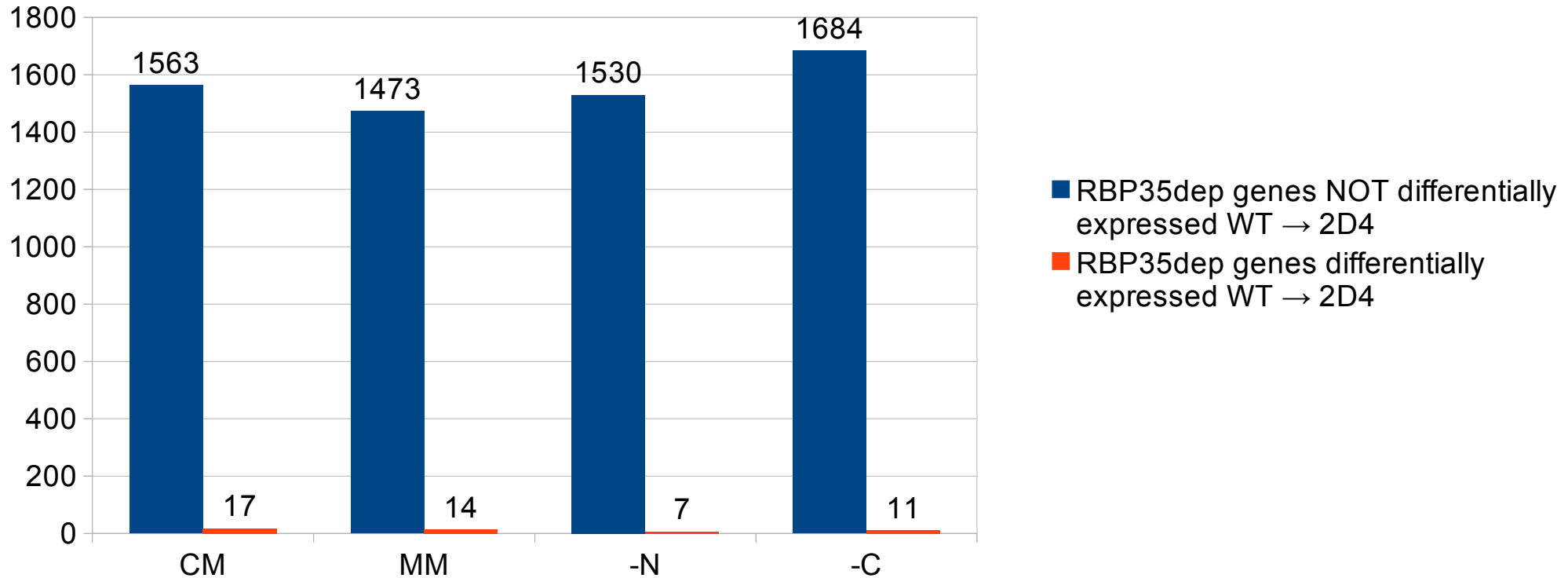


RBP35 dependant **Pathogenic** genes (CAMBIAR)



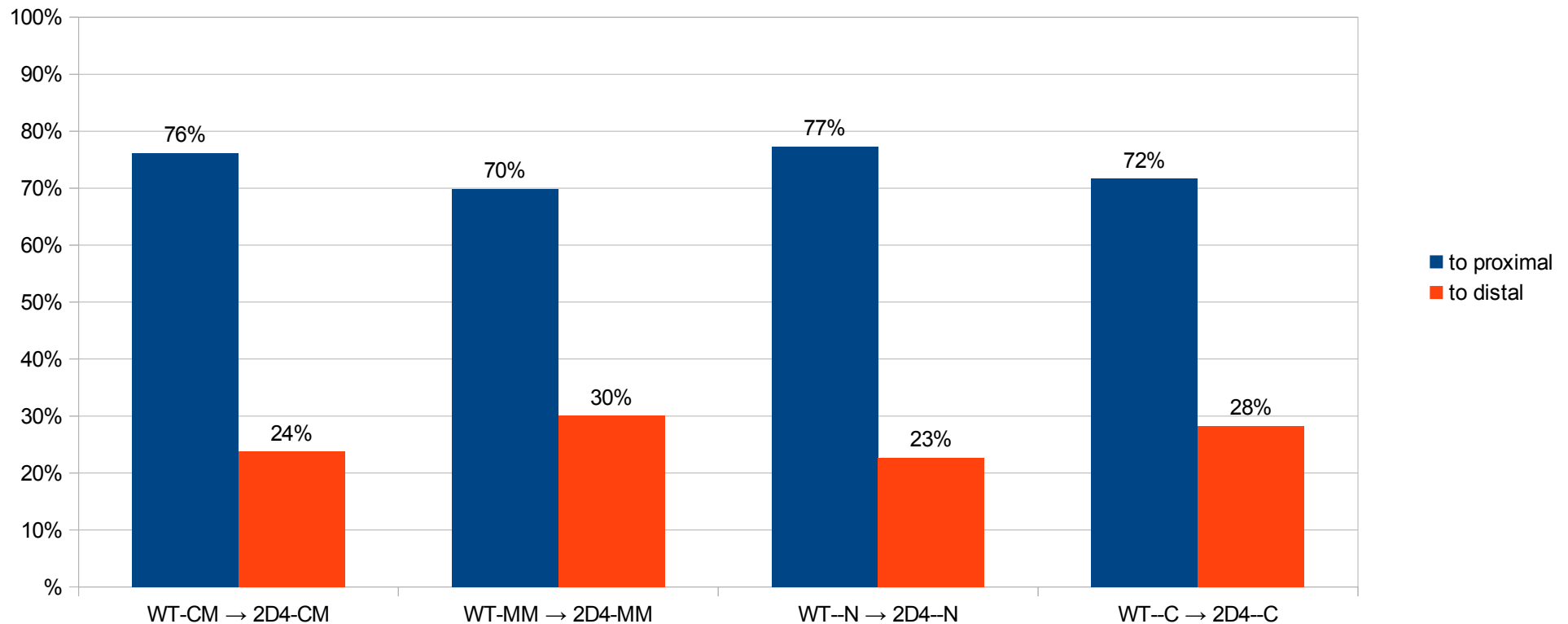
There is no correlation between RBP35 dependance and differential expression

RBP35 dependance vs differentially expression



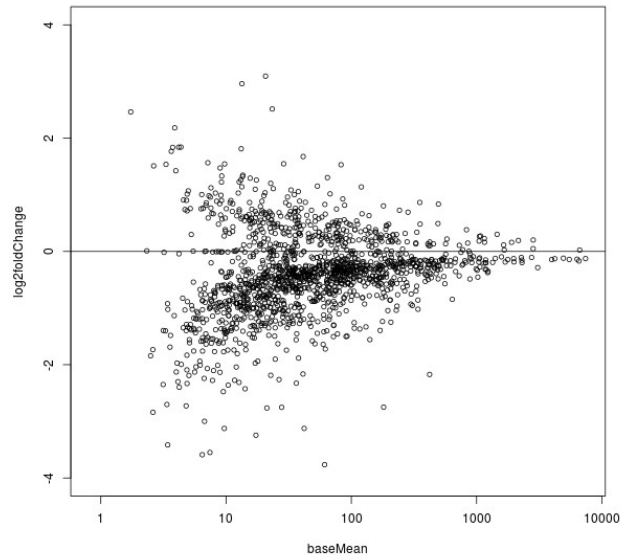
Δrbp35 affects poly(A) sites usage, preferring proximal cuts - percentages

Poly(A) site usage change - RBP35 dependent genes

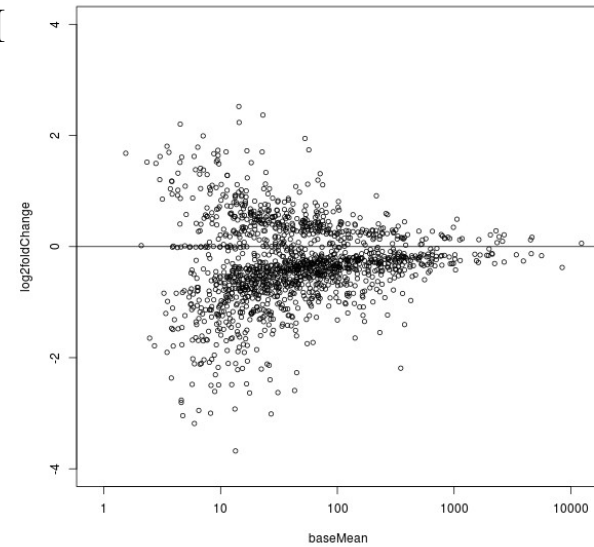


$\Delta rbp35$ affects poly(A) sites usage, preferring proximal cuts - foldChange

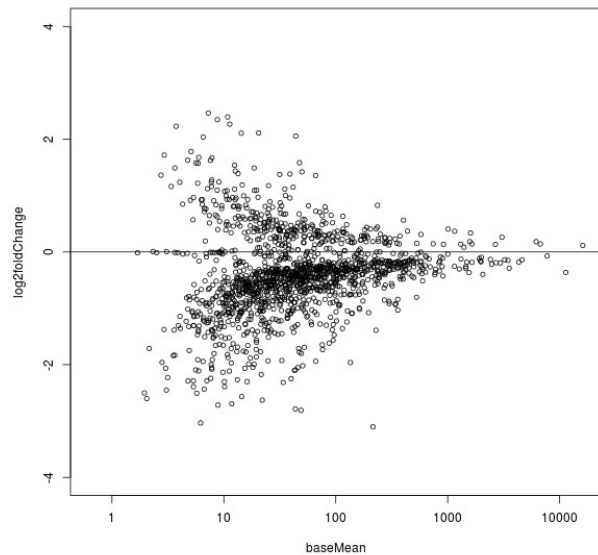
CM



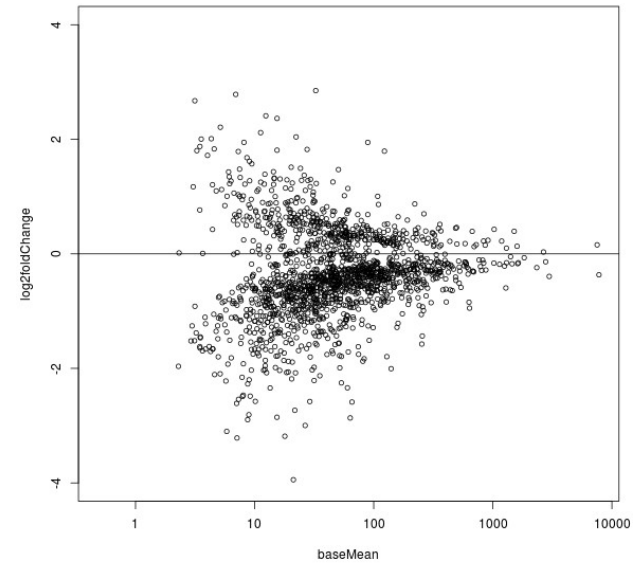
MM



-N

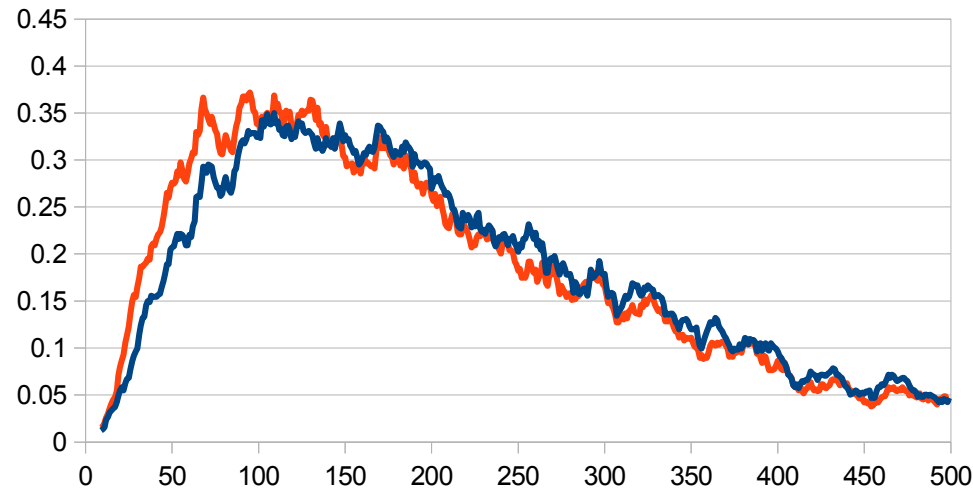


-C

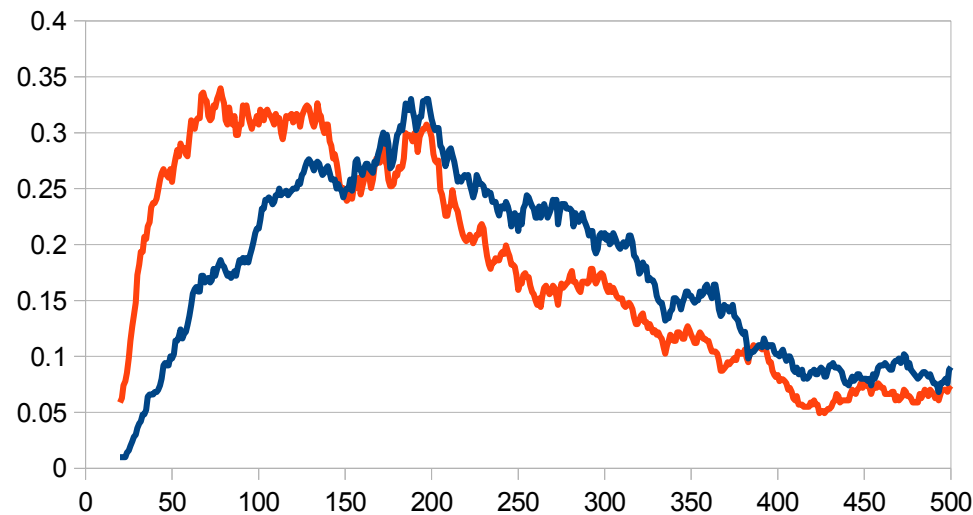


$\Delta rbp35$ affects 3'UTR length

3'UTR length (bp) CM

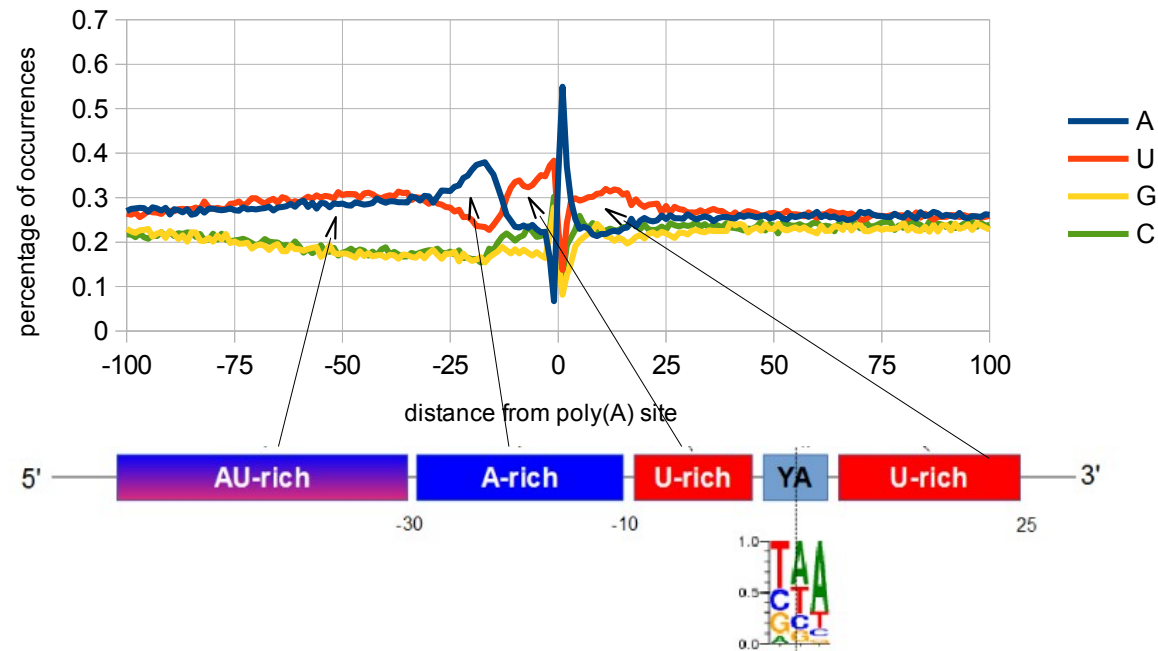


3'UTR length (bp) RBP35dep genes only CM

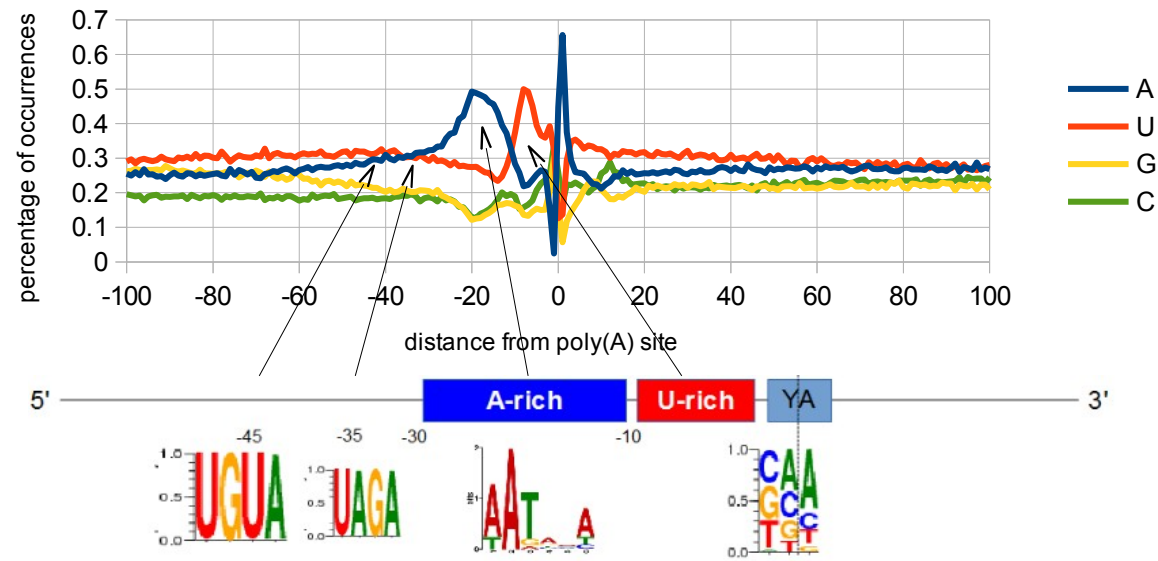


Nucleotides profile of poly(A) sites slightly differs from *S.cerevisiae*

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

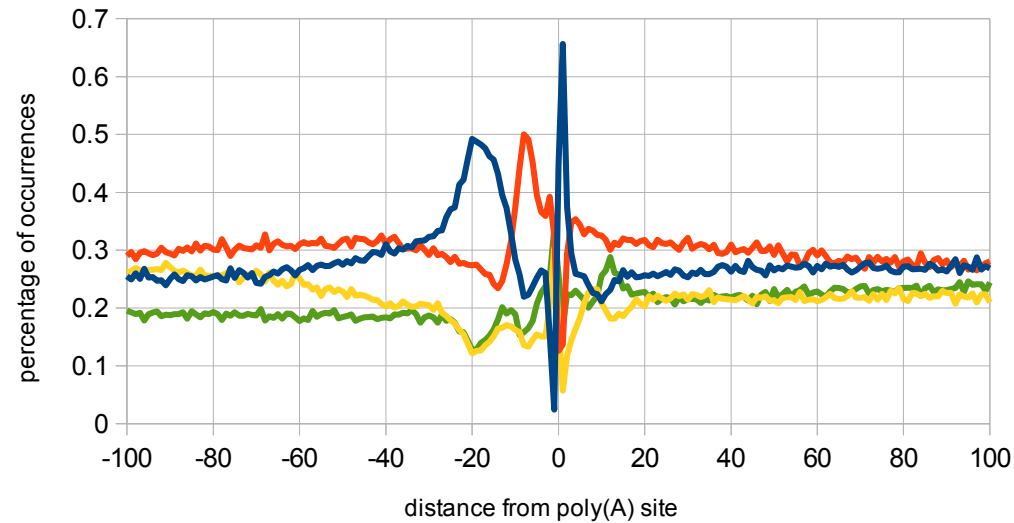


Poly-A site nucleotide profile - *M. Oryzae*

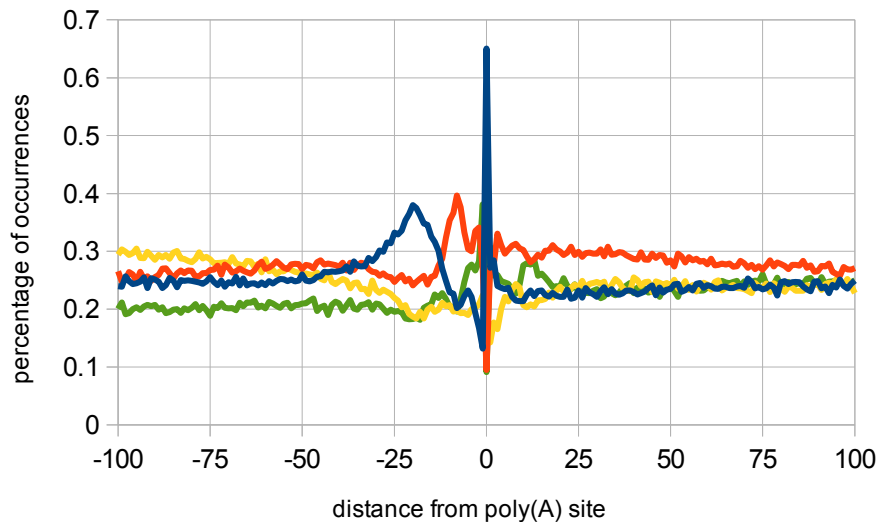


Nucleotides profile of poly(A) sites resembles *N. crassa*

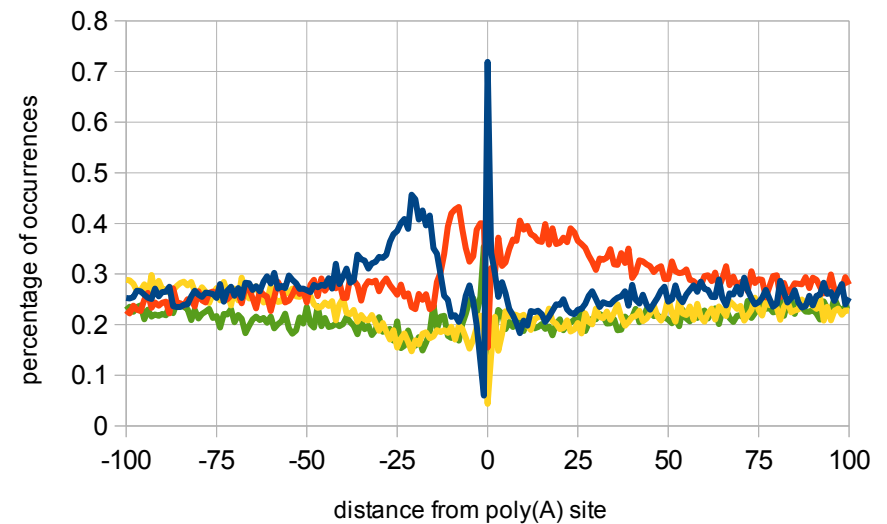
Poly-A site nucleotide profile - *M. Oryzae*



Poly-A site nucleotide profile - *N. Crassa*

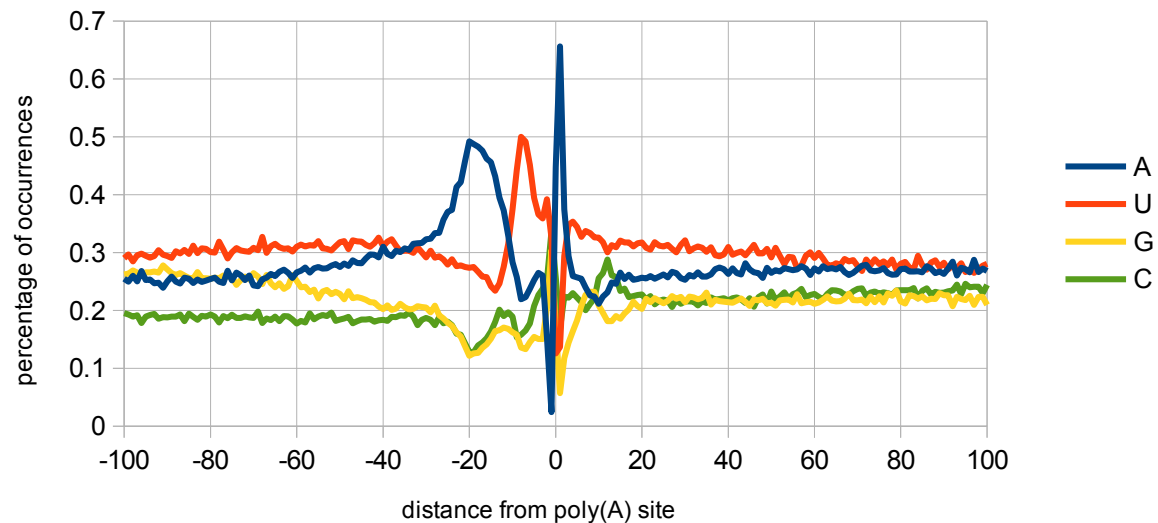


Poly-A site nucleotide profile - *P. Infestans*

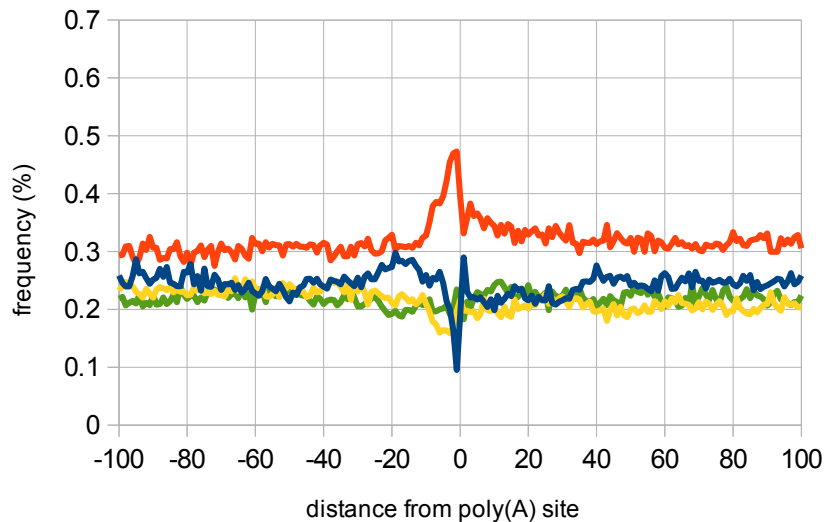


Nucleotides profile of poly(A) of ncRNA and CDS poly(A) is different

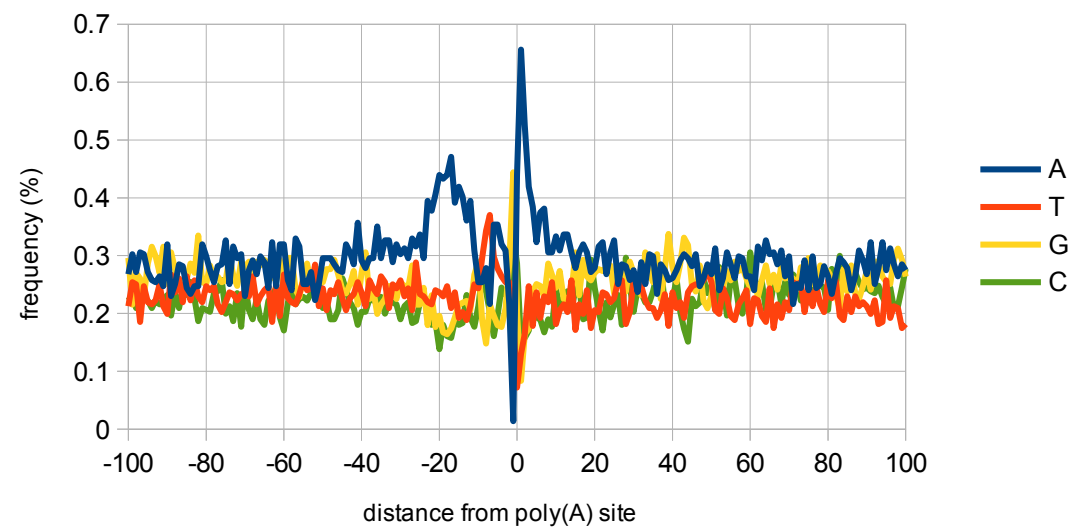
Poly-A site nucleotide profile - *M. Oryzae*



ncRNA poly(A) nucleotide profile

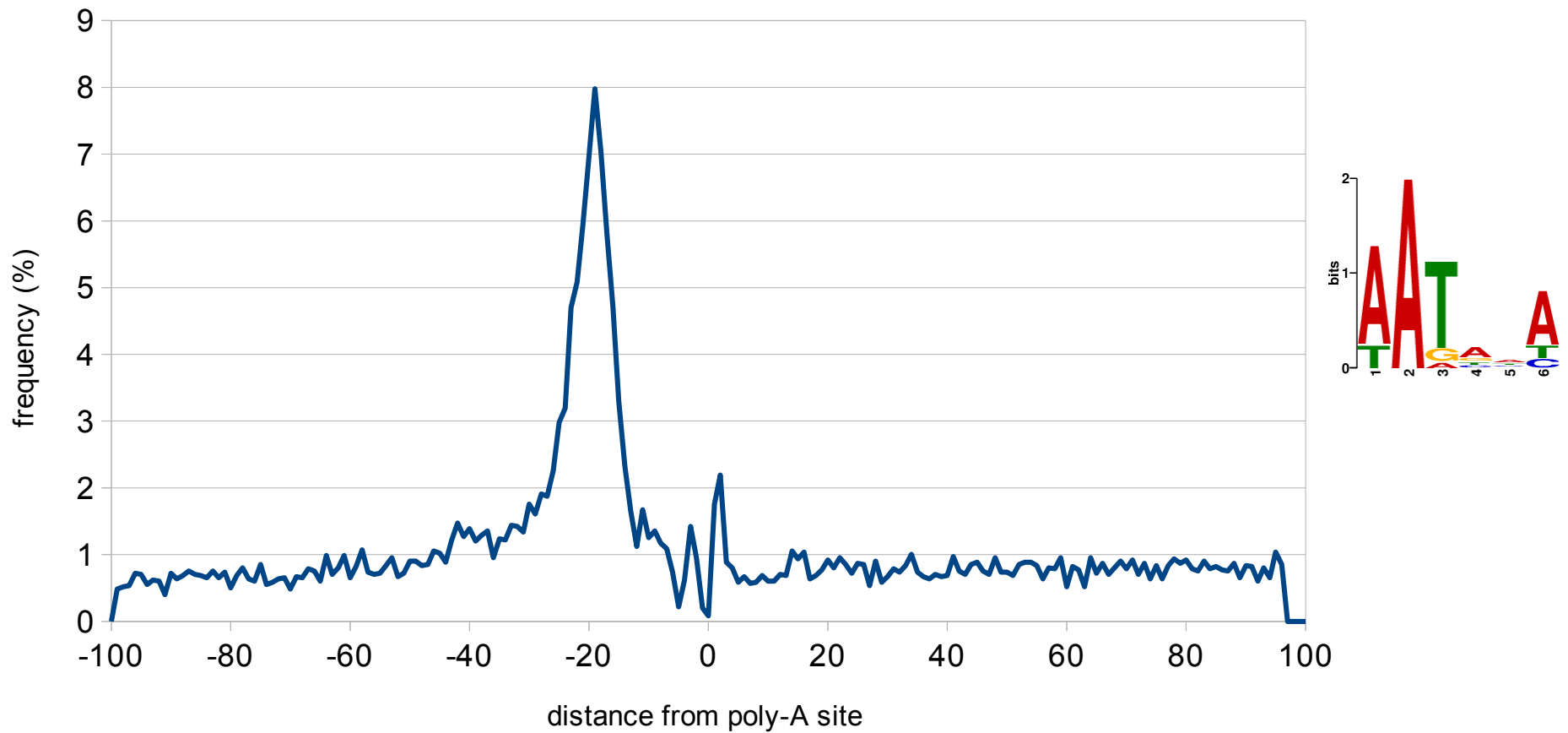


CDS poly(A) sites nucleotide profile



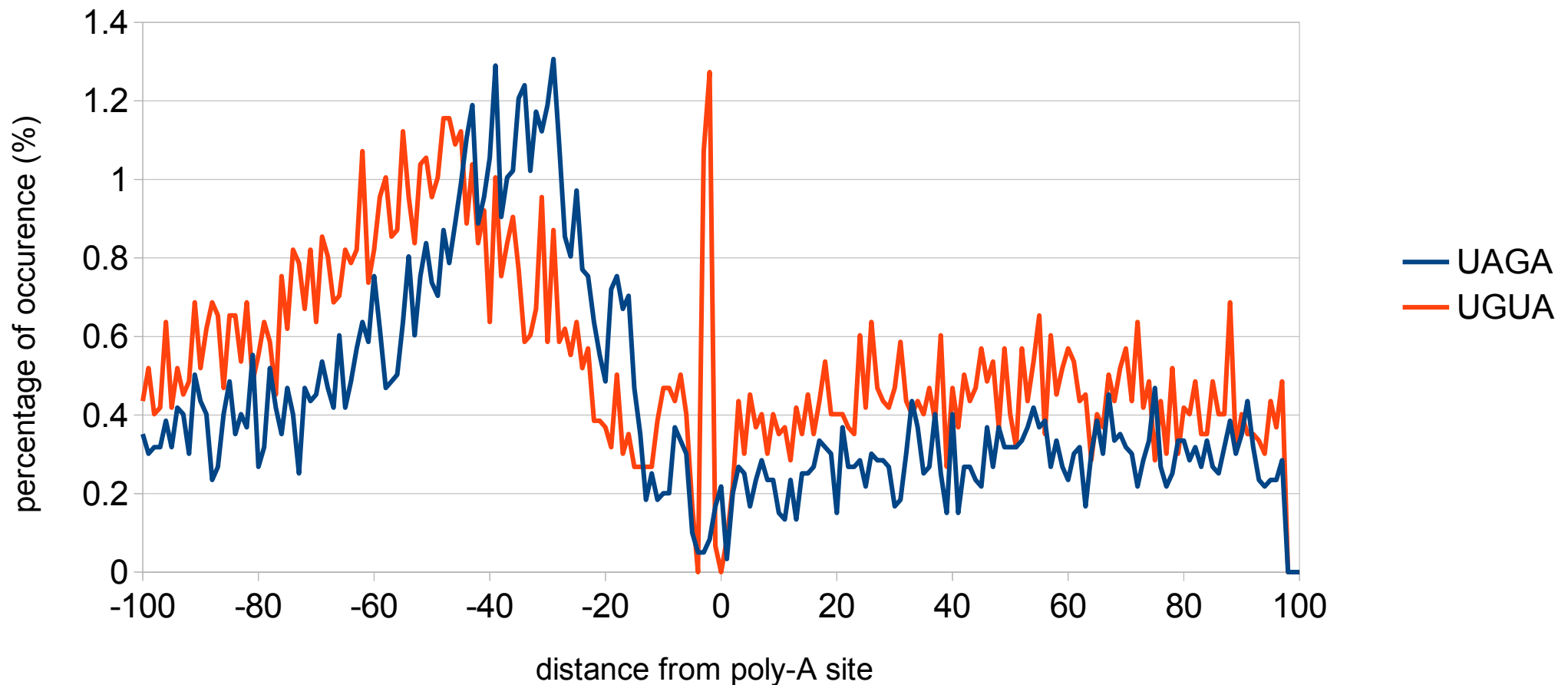
The A-RICH region is located -30 -10 bp upstream

Best motif in A-rich region



UAGA & UGUA motifs

UGUA & UAGA motifs - all genes single cut



Polyadenylation signals in common genes

MPG1

...GG**UAGA**GAAGUCUCUUCUCGUUCCACUCAUUU**AAUAAA**ACCCCUUCCAGACC**UA**...

PMK1

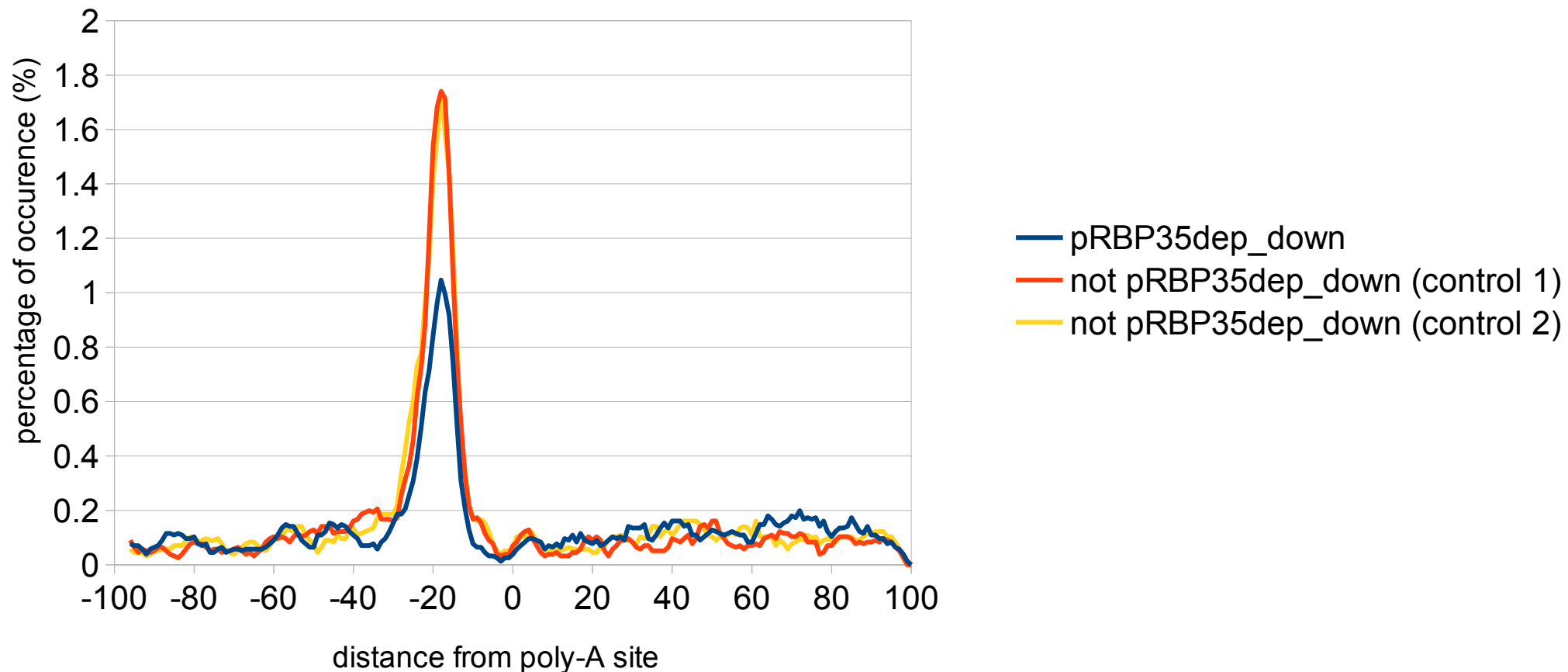
...CGUU**UAGA**AUGUGCAGGAGACACGAGUGGGAAAAUG**AAUACA**UGGAUGCCAG**CA**...

MST12

...CAGUGGCAUAAAAUCACAAAAUCUU**UAGA**AAGAUCAC**AGAAAA**CCUUUUGUC**CA**...

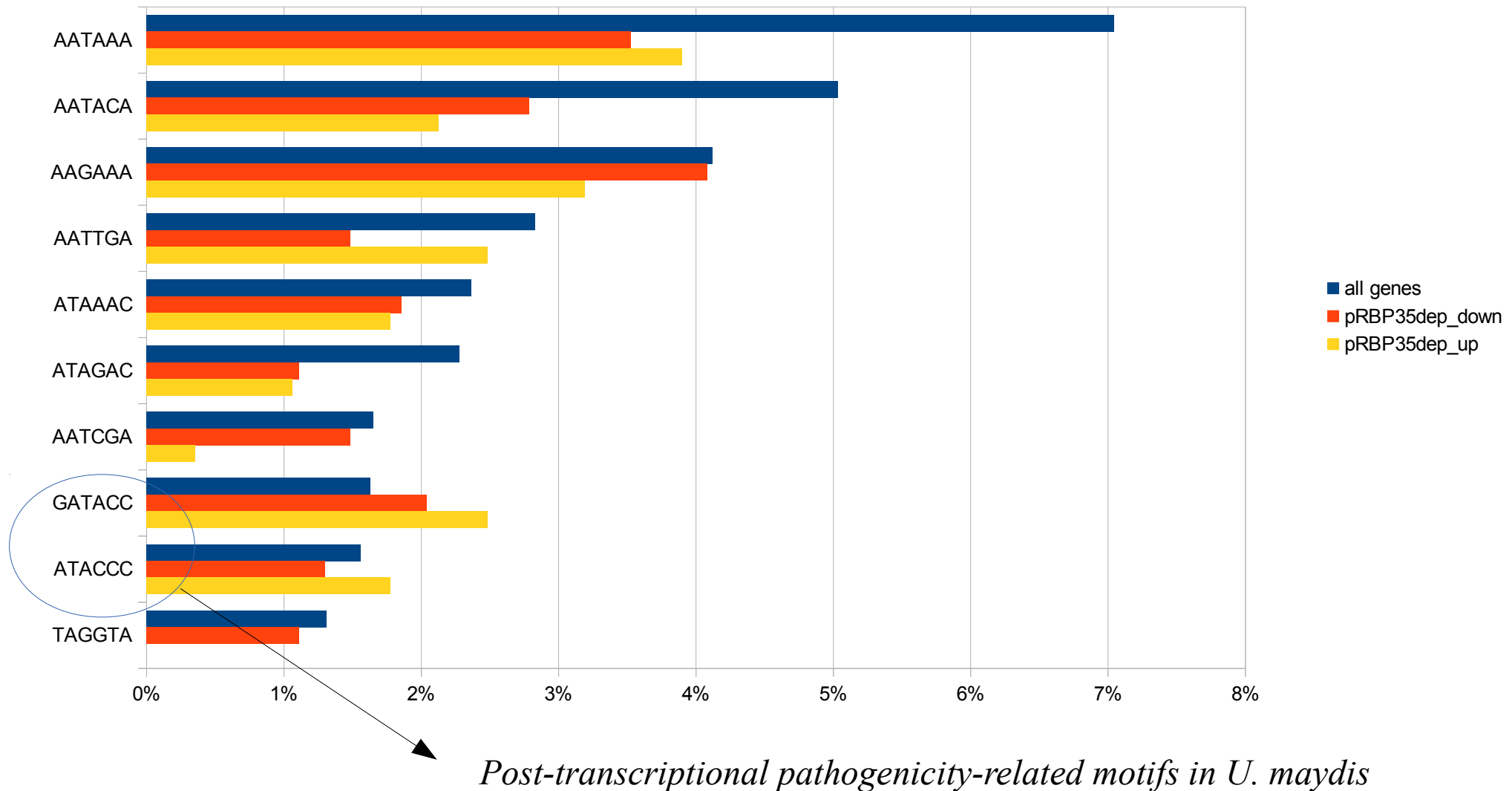
poly(A) sites dependent from *RBP35* show a less pronounced A-rich region

A-rich motif - pRBP35dep vs not pRBP35dep



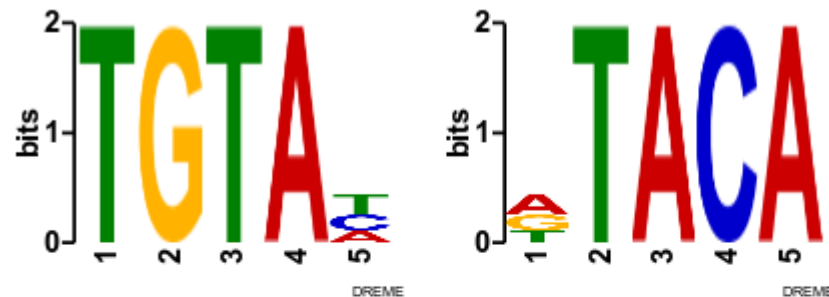
poly(A) sites dependent from *RBP35* are low in the canonical AATAAA signal

Top 10 hexamers in A-rich region (CM)



UGUAH motif is enriched in poly(A) sites dependent from *RBP35* down-regulated in $\Delta rbp35$, in the region -100 -30

1. TGTAAH



Details

| Positives ? | Negatives ? | P-value ? | E-value ? | Unersased E-value ? |
|-------------|-------------|-----------|-----------|---------------------|
| 1608/3115 | 815/3115 | 9.6e-96 | 7.6e-91 | 7.6e-91 |

Enriched Matching Words

| Word ? | Positives ? | Negatives ? | P-value ? | ▼E-value ? |
|--------|-------------|-------------|-----------|------------|
| TGTAT | 802/3115 | 363/3115 | 3.6e-47 | 2.9e-42 |
| TGTAC | 616/3115 | 297/3115 | 7.0e-31 | 5.5e-26 |
| TGTAA | 472/3115 | 237/3115 | 2.4e-21 | 1.9e-16 |
| TGTAC | 412/3115 | 273/3115 | 1.0e-8 | 8.2e-4 |

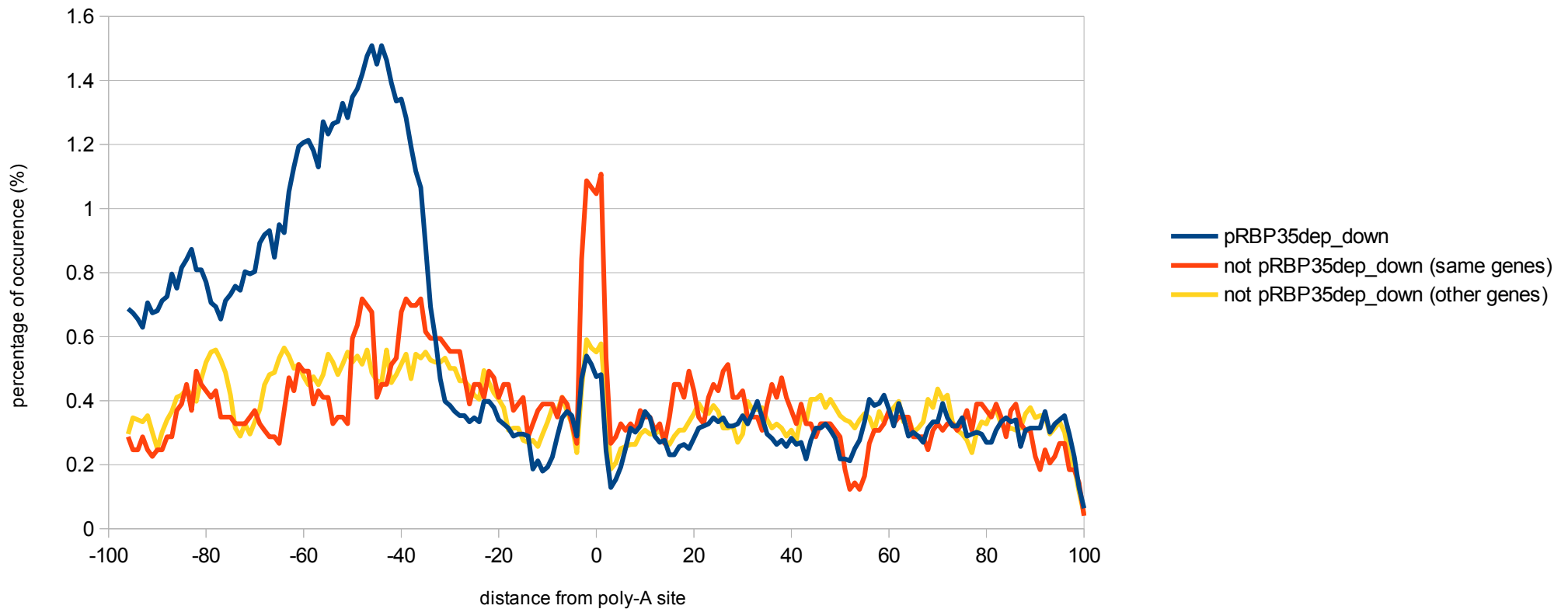
- Output of DREME, pRBP35dep as positive sequences list versus not-pRBP35dep negative list

UGUAH motif – pRBP35dep vs not pRBP35dep

- In the first graph, we want to show how poly(A) sites dependent from *RBP35* display a different profile for the UGUAAH motif in the respect to “regular” poly(A) sites
- We therefore plot down-regulated RBP35 dependent poly(A) sites against two groups of poly(A) not dependent from RBP35 of the same size, one group of poly(A) sites belonging to the same genes and one group of poly(A) sites belonging to other genes

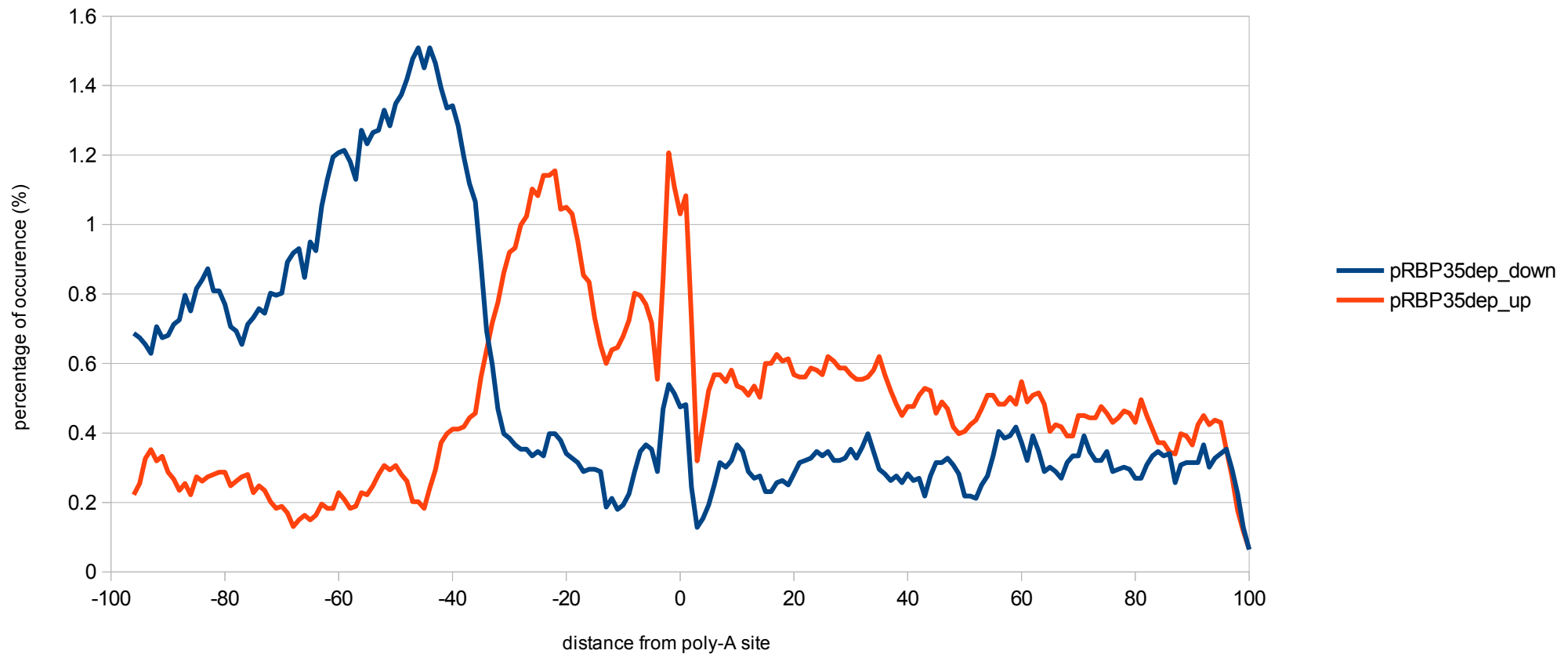
UGUAH is enriched at -45 in poly(A) sites dependent from *RBP35*

UGUAH motif - down-regulated RBP35 dependent poly(A) sites



UGUAH motif – RBP35 dependent poly(A) sites (up vs down regulated)

UGUAH motif - up&down-regulated RBP35 dependent poly(A) sites



UAGA– pRBP35dep vs not pRBP35dep

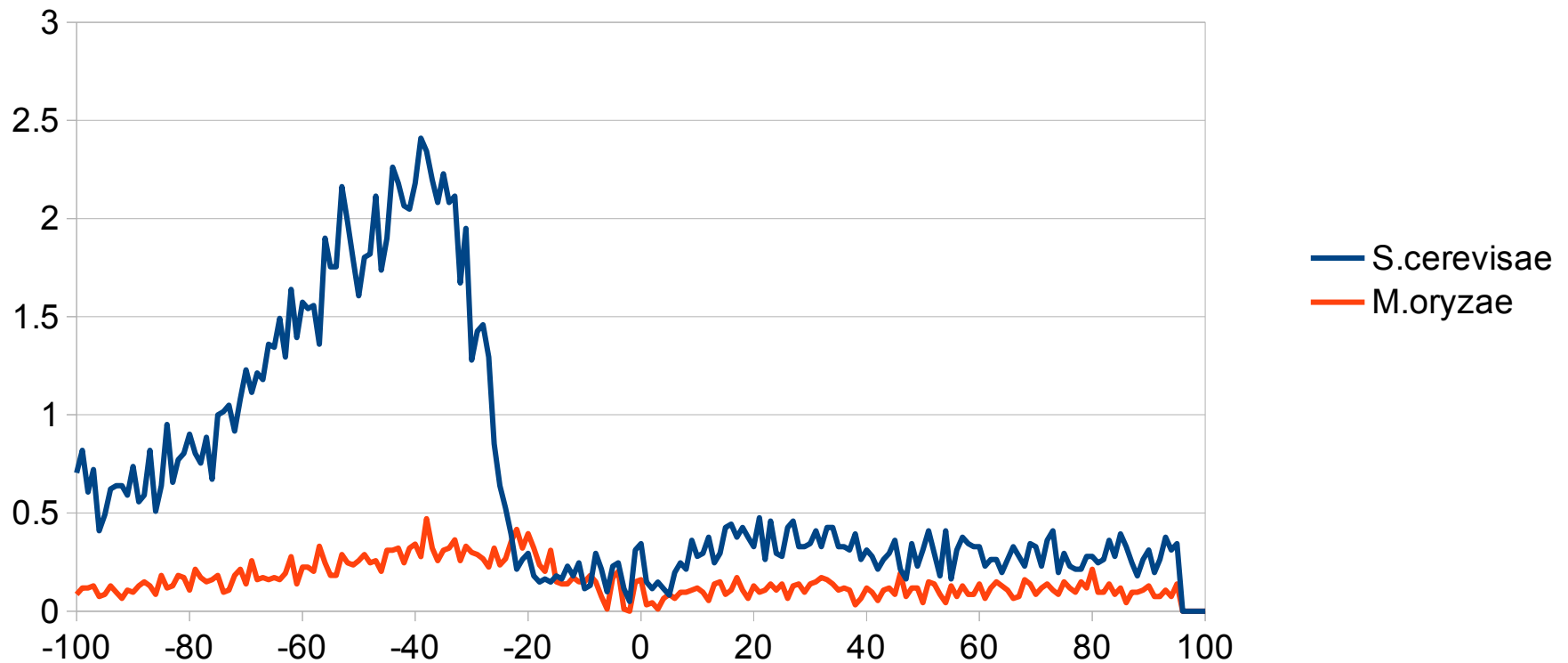
UAGA motif - P1 vs not P1



UAGA is shifted in RBP35dep poly(A) sites, probably due to UGUA enrichment

The HRP1 binding motif TAYRTA from *S.cerevisiae* is not found in *M.oryzae*

M.oryzae vs S.cerevisiae TAYRTA motif



Maybe in M.oryzae CFIm binds to UGUA and Hrp1 binds to UAGA?

