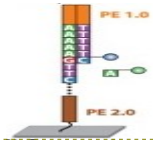
Sequencing resume

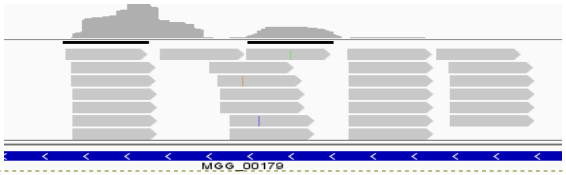
- 2 strains (*WT, 2D4*) x 4 conditions (*CM, MM, -N, -C*) x 3 replicates
- 4751592 – 11517077 total reads database
- ~63% - ~83% successfully mapped reads
- 43 bp mean read length
- ~92% - ~98% replicates correlation
- ~100bp mean pair ended distance
- ~400x coverage for poly(A) site*

** assuming an amount of 22000 mRNA molecules per cell*

Workflow

Program and details	Procedure	% of data left
	Illumina sequencing	100%
fasta-mcf, *adapter trimming*	Data prepping	96%
gsnap, *default parameters*	Alignment	96%
• *Low quality mapping (<30))* • *High A/T content (>90%)* • *Internal priming*	Filtering	75%
Assign reads to known features, Including all intergenic space up to 400 bp from gene end	Assignment	65%
Clustering of most significant poly(A) sites (p-value < 0.05 over global gene expression) at 33 bp	poly(A) sites detection	27%

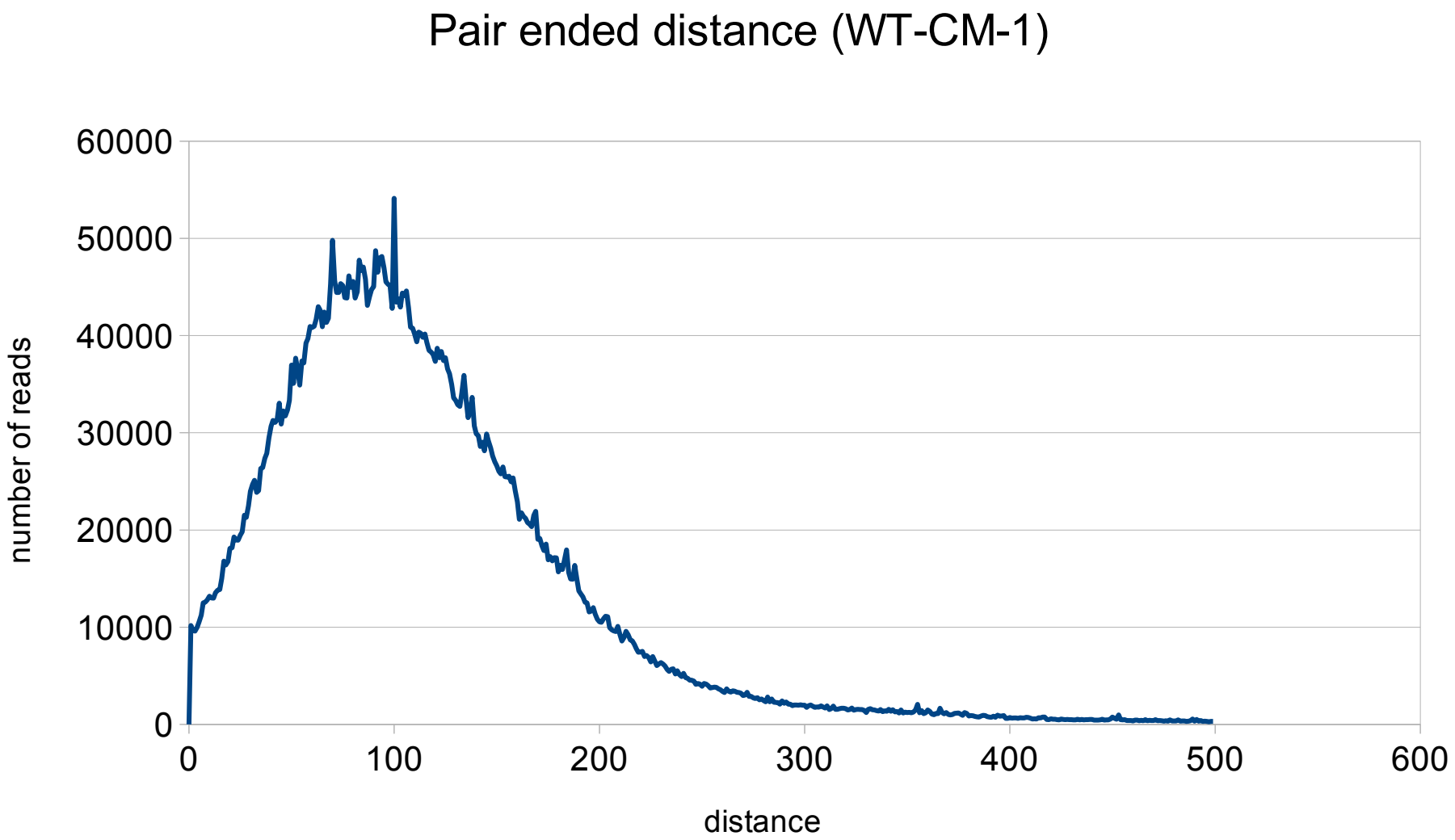




*Example: **Actin,** single cut poly(A):*

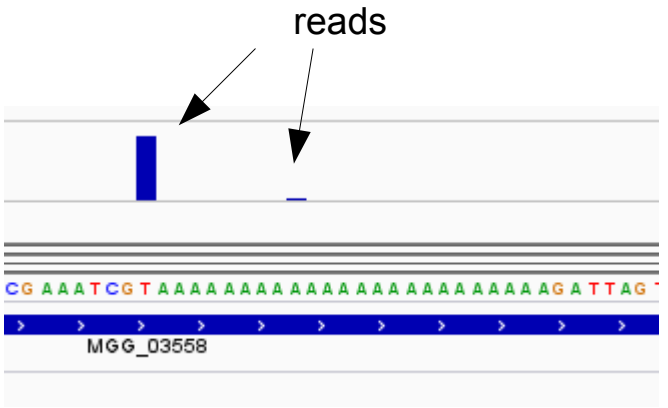
- whole gene expression ~7000 reads
- poly(A) site expression ~3000 reads

Pair ended reads distance

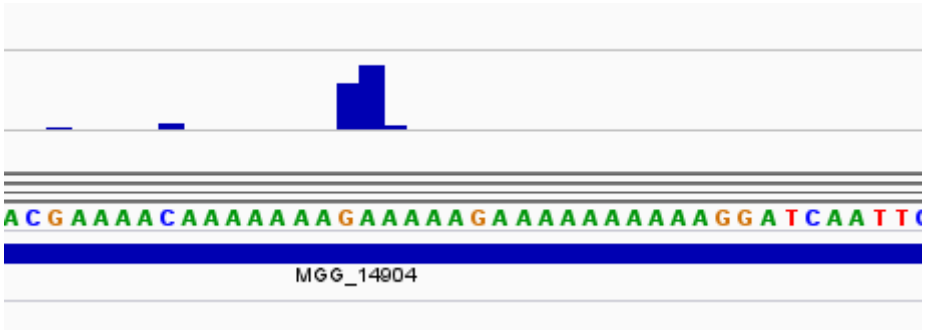


~2.5% of poly(A) sites are internal priming

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



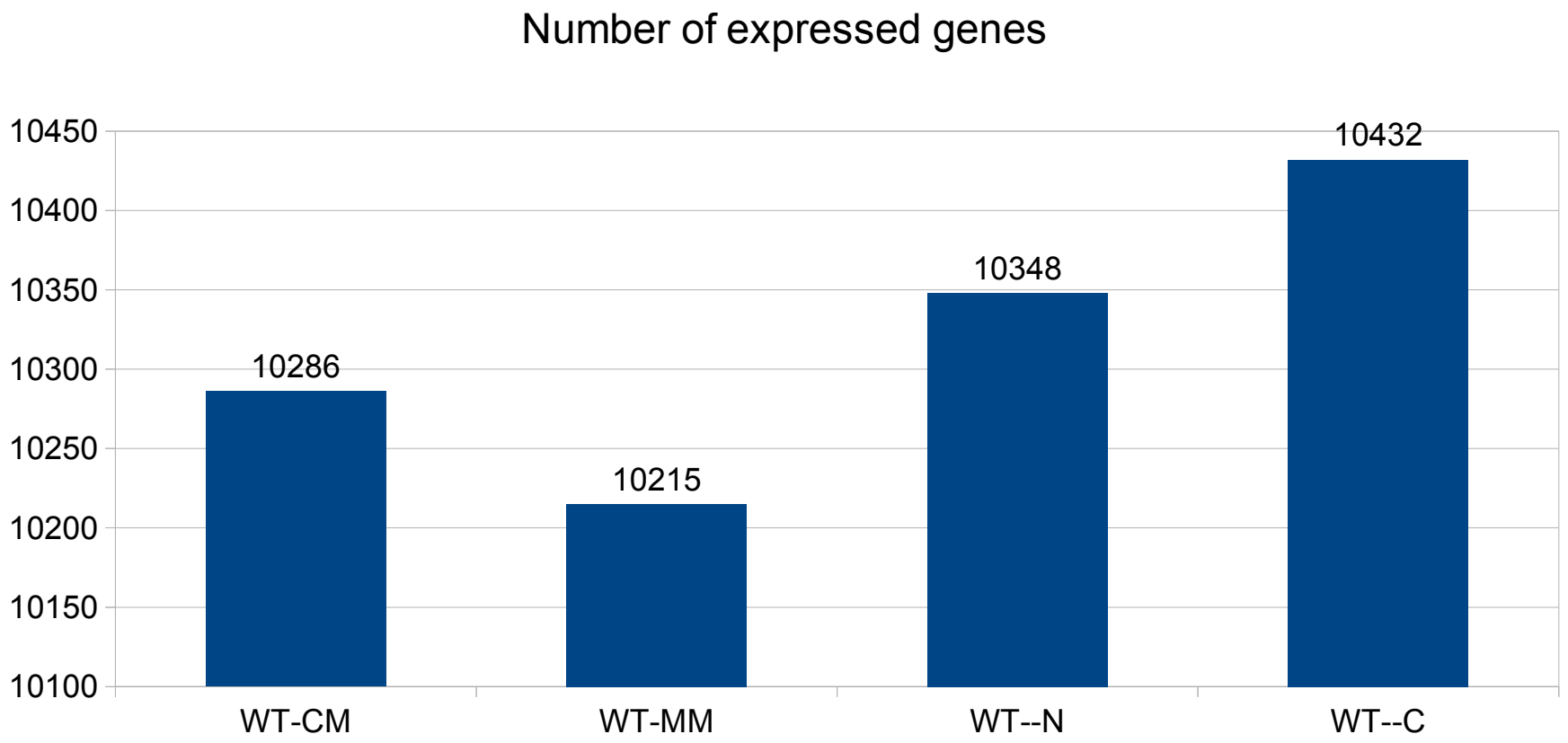
stretches of As



Results of the polyadenylatoin sites by Illumina sequencing

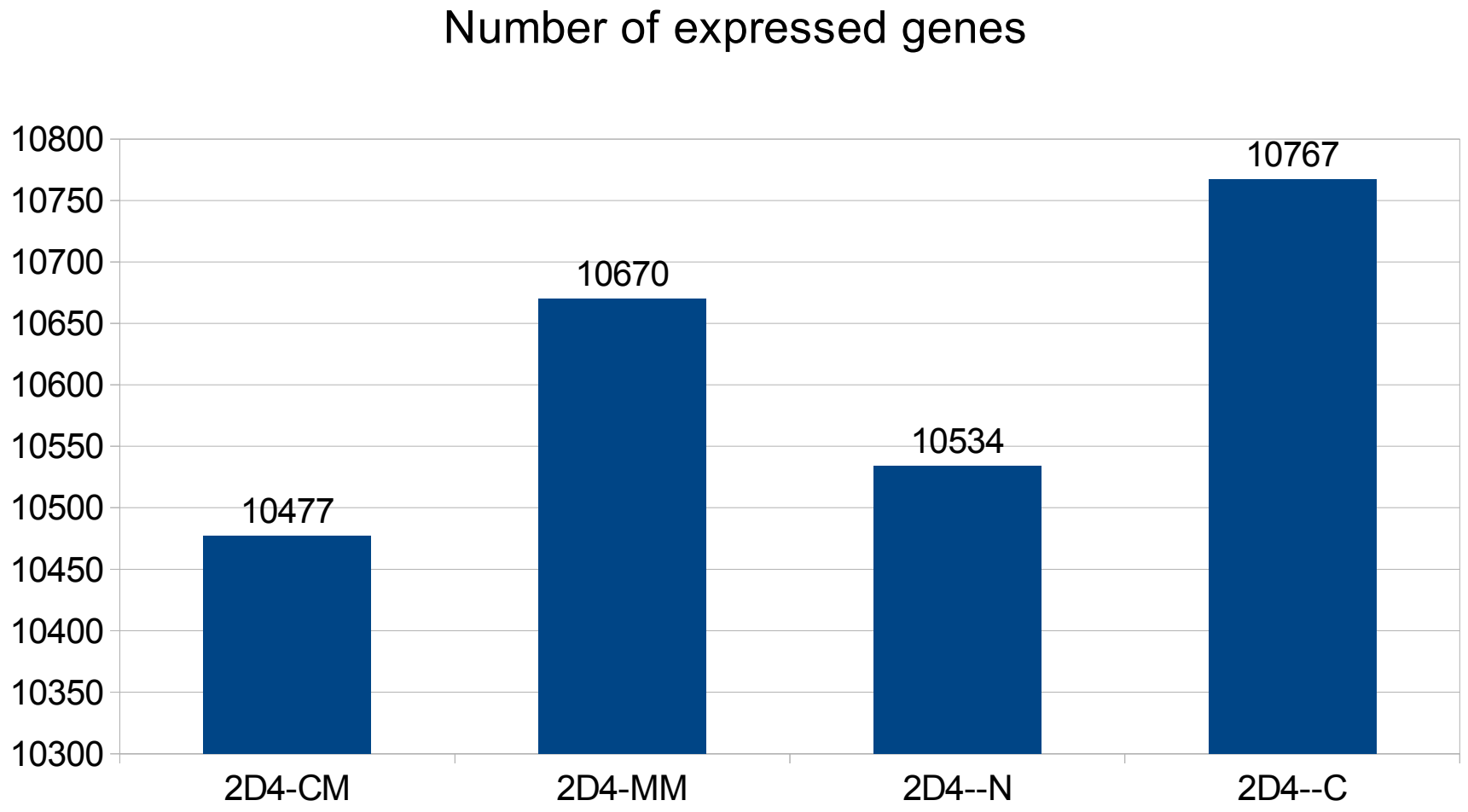
	CM *WT/Δrbp35*	**MM** *WT/Δrbp35*	**MM-N** *WT/Δrbp35*	**MM-C** *WT/Δrbp35*
Number of genes with a recognizable poly(A) site	7511 / 7715	7524 / 8091	7629 / 7888	8007 / 8281
Genes with APA in CDS	56 / 71	53 / 90	75 / 94	78 / 87
Genes with APA in 5'UTR	66 / 82	66 / 81	77 / 89	63 / 97
Genes with APA in 3'UTR	1429 / 1600	1271 / 1753	1342 / 1608	1463 / 1725
Number of genes with APA in WT but not in Δrbp35	633	509	621	682
Number of genes with APA in Δrbp35 but not in WT	833	1038	906	994

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



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

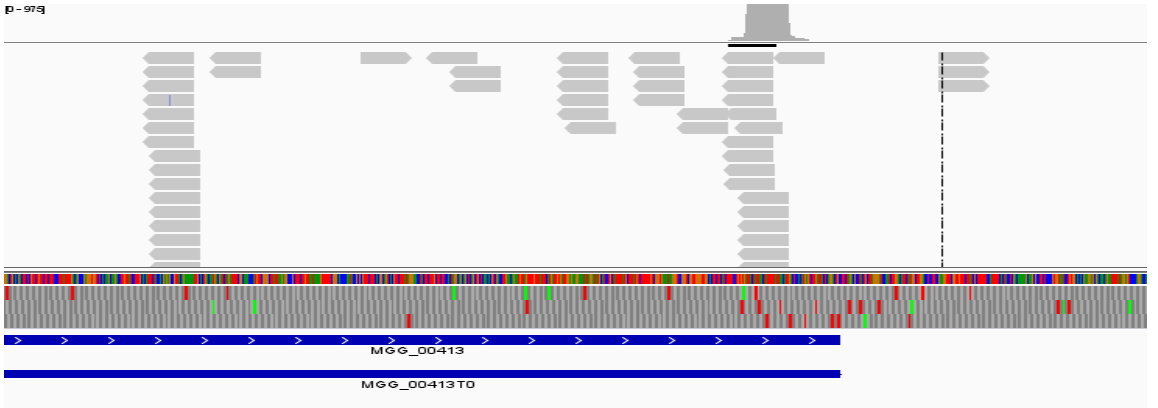
~10000 genes are expressed, out of a total of 13218 annotated genes (2D4)



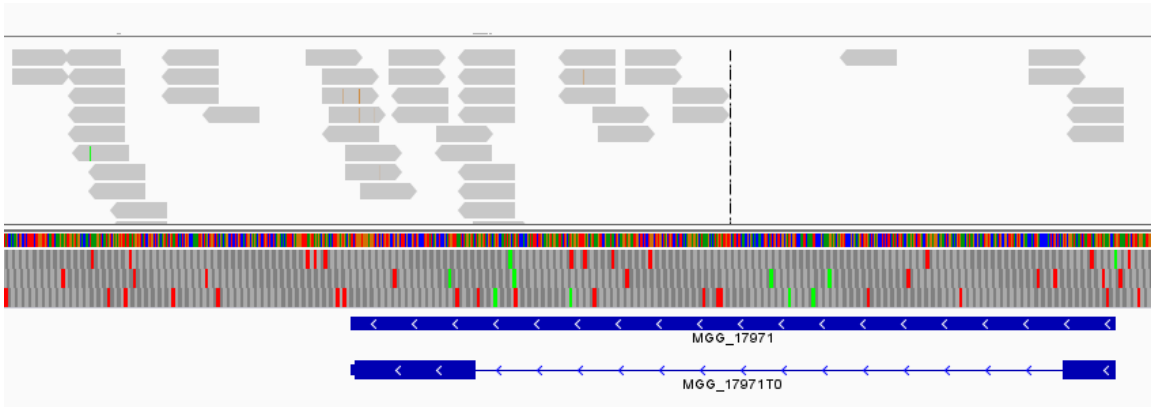
- 9918 genes are expressed in every condition (2D4 only)
- 1964 genes are never expressed (2D4 only)

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

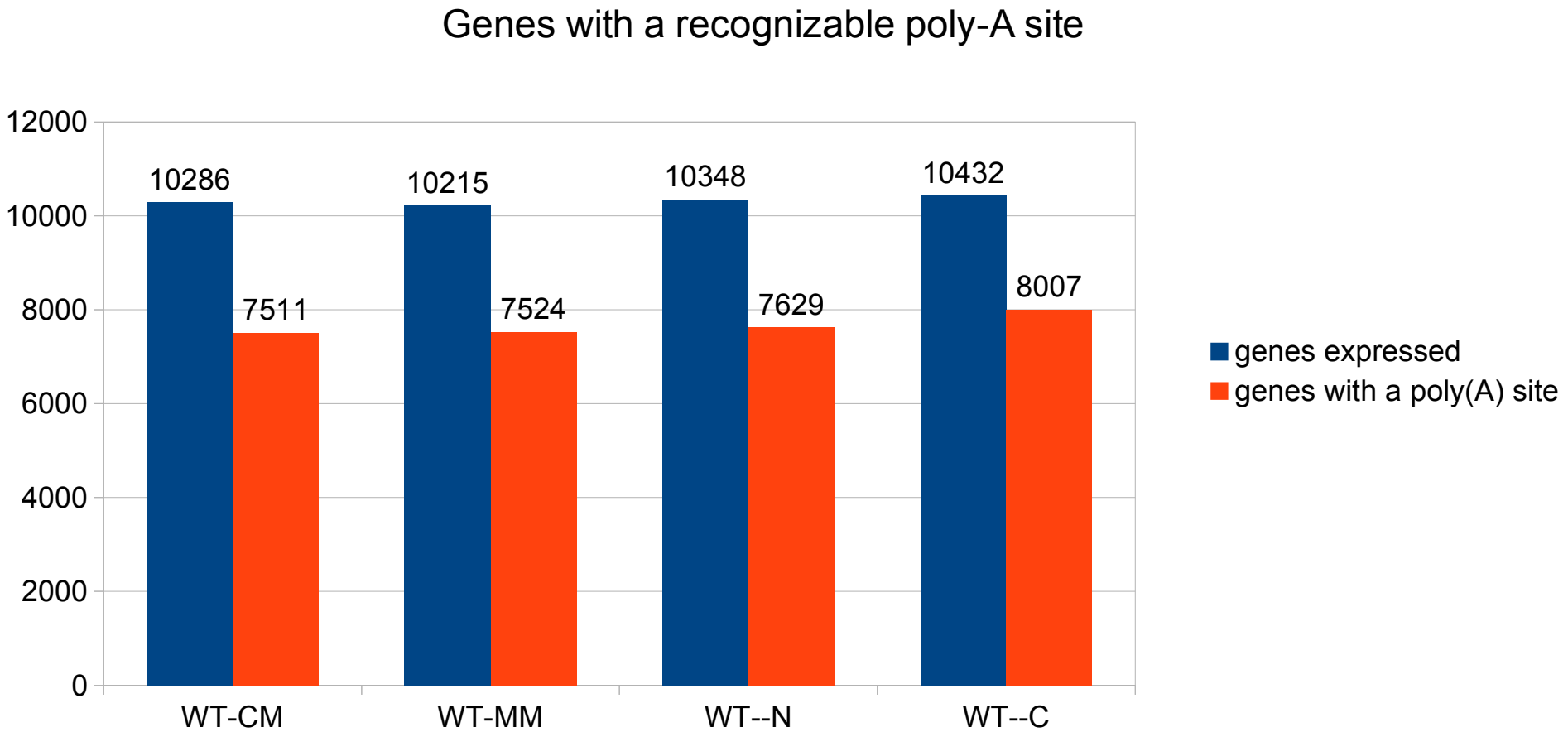
Expressed gene <u>with</u> a recognizable poly(A) site:



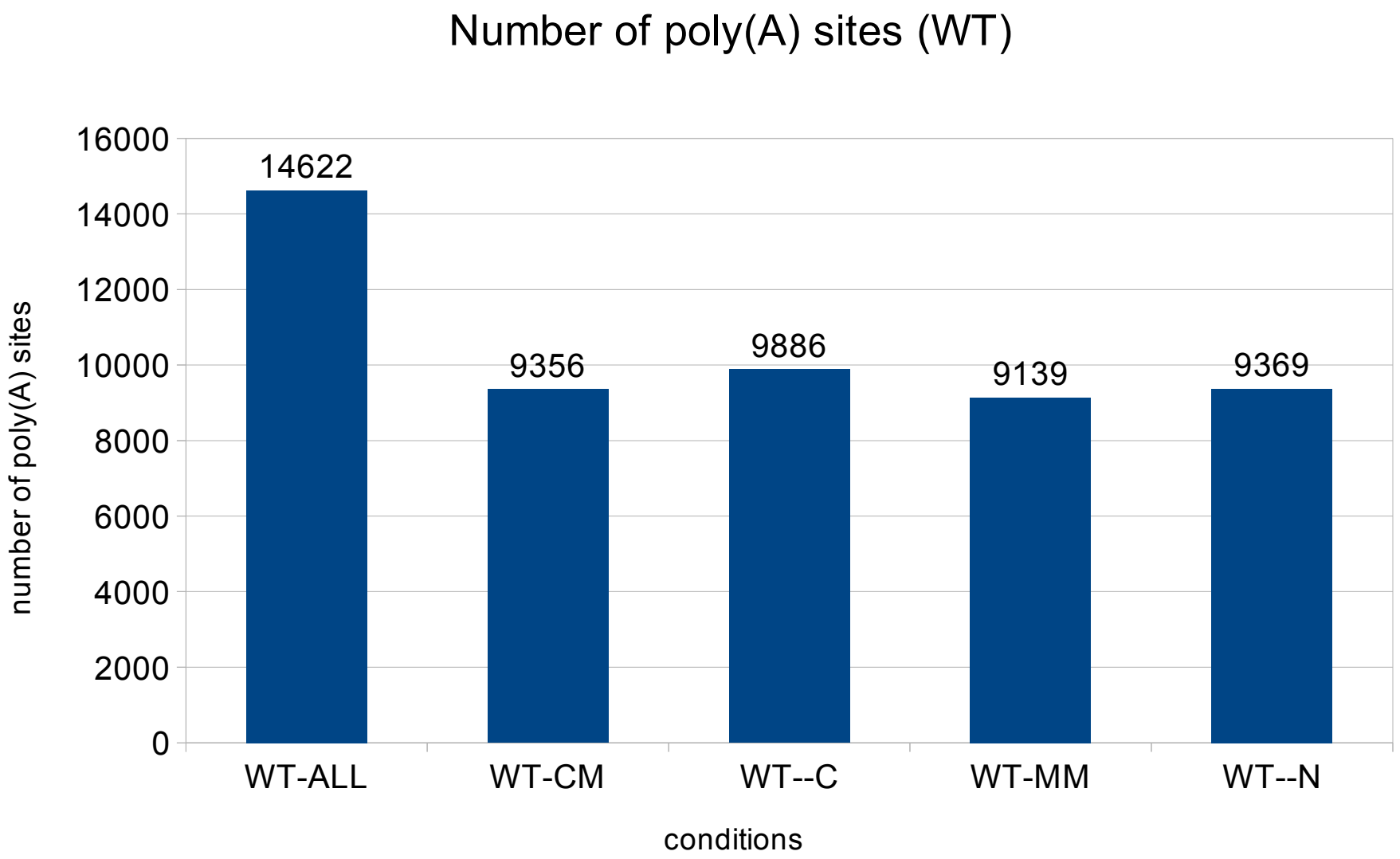
Expressed gene <u>without</u> a recognizable poly(A) site:



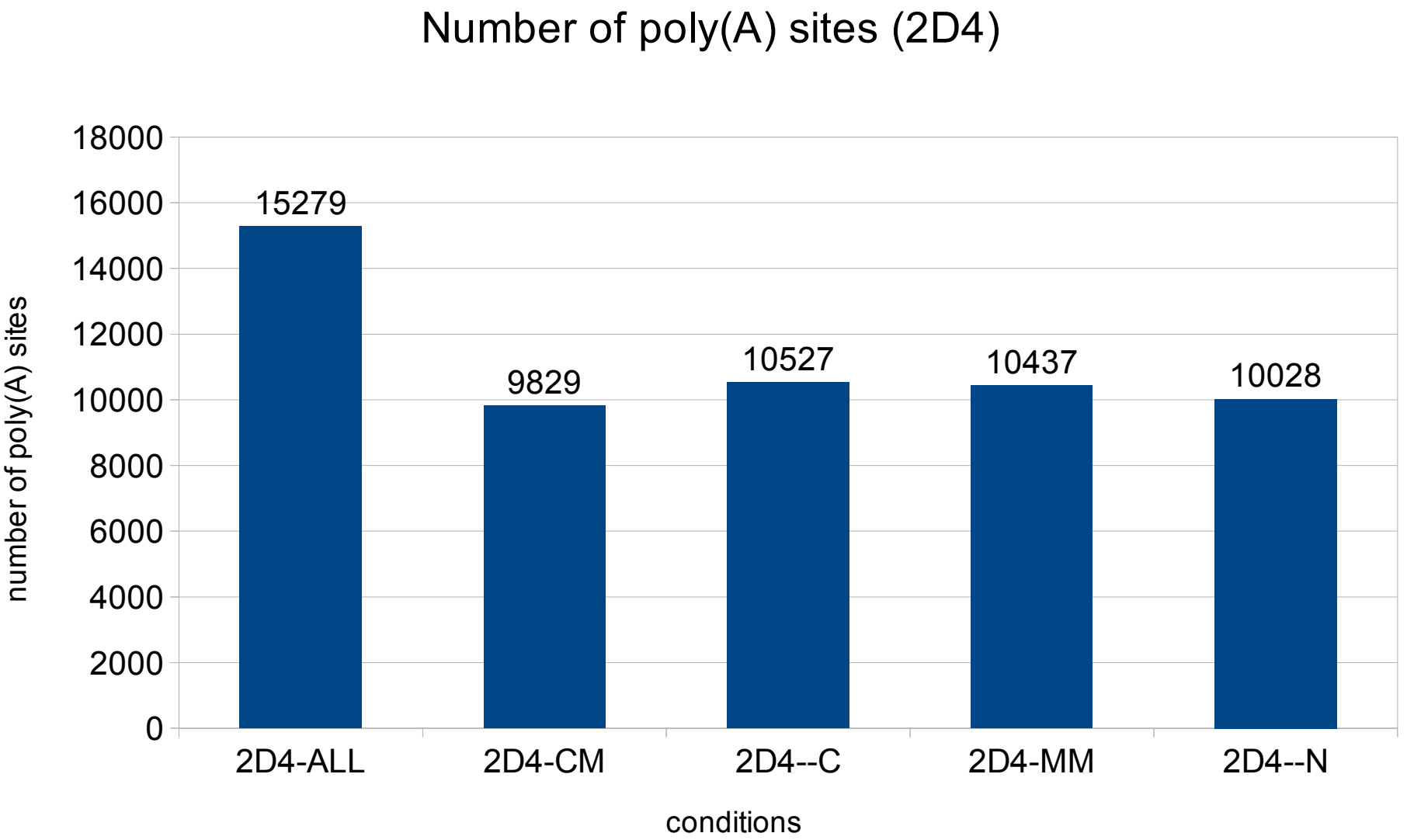
Only ~75% of genes expressed have a recognizable poly(A) site



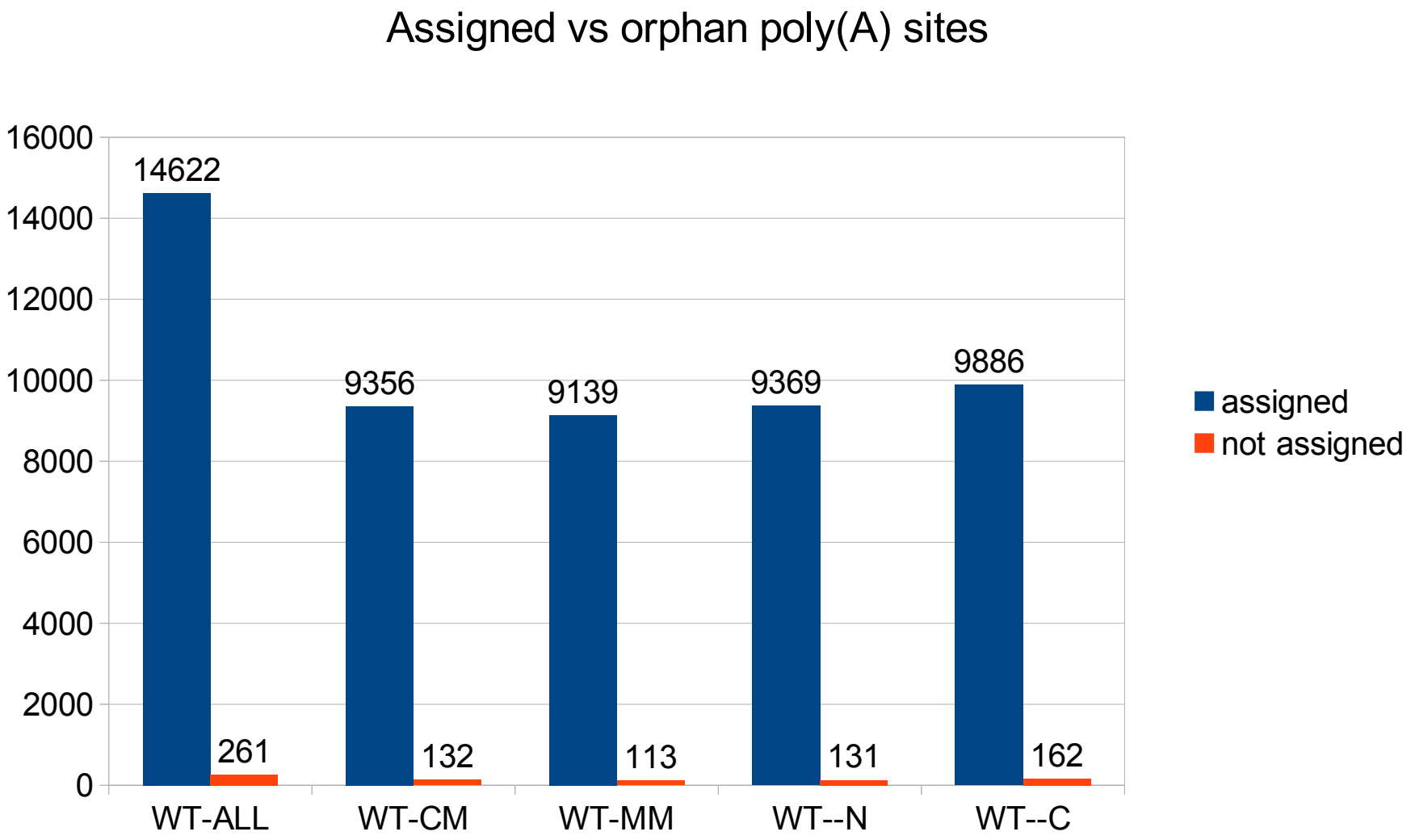
~15000 poly(A) site could be assigned to annotated genes



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



17%-19% of poly(A) sites could not be assigned to any annotated gene

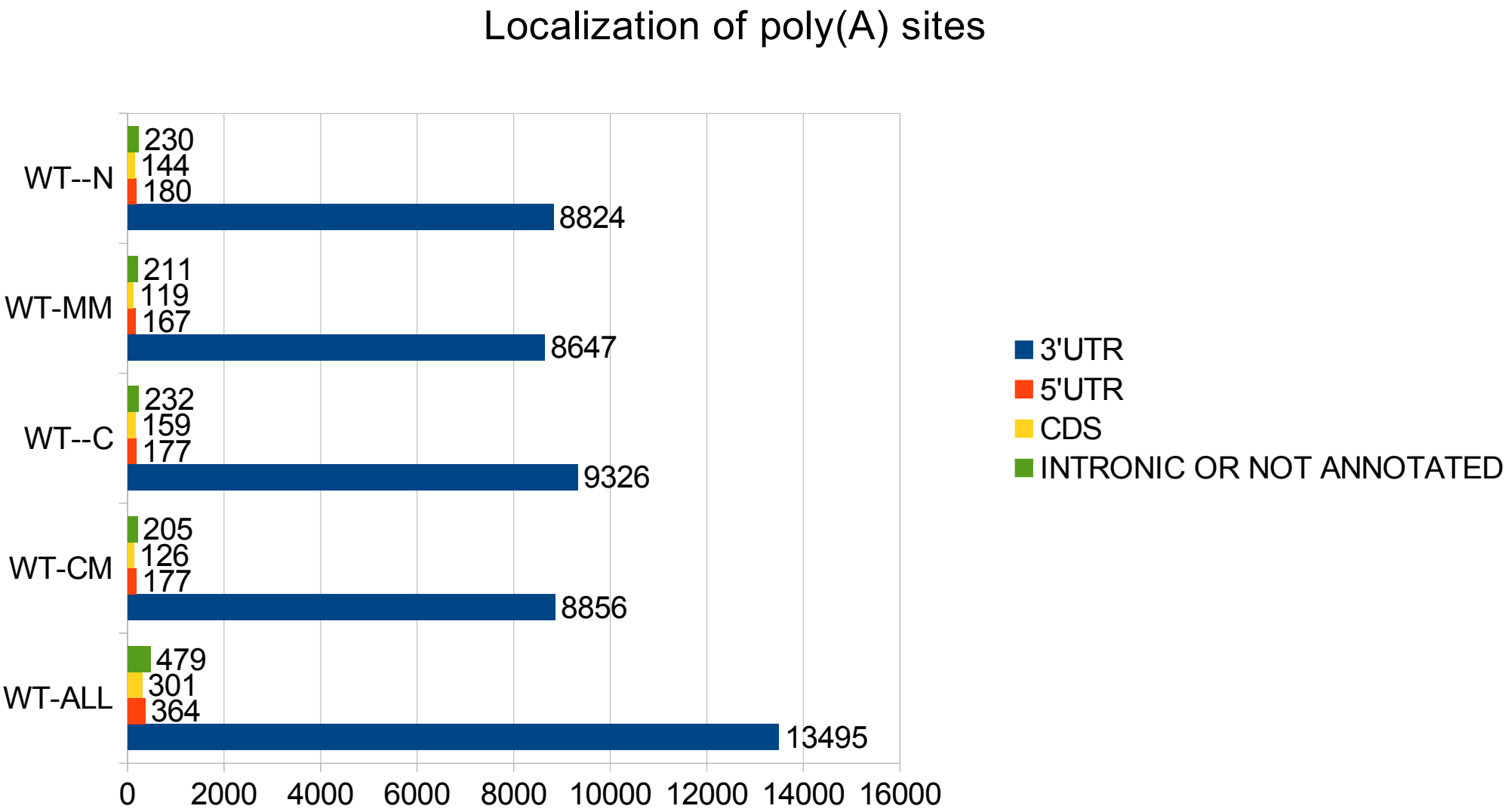


261 orphans poly(A) sites highly expressed in WT (>100 reads)

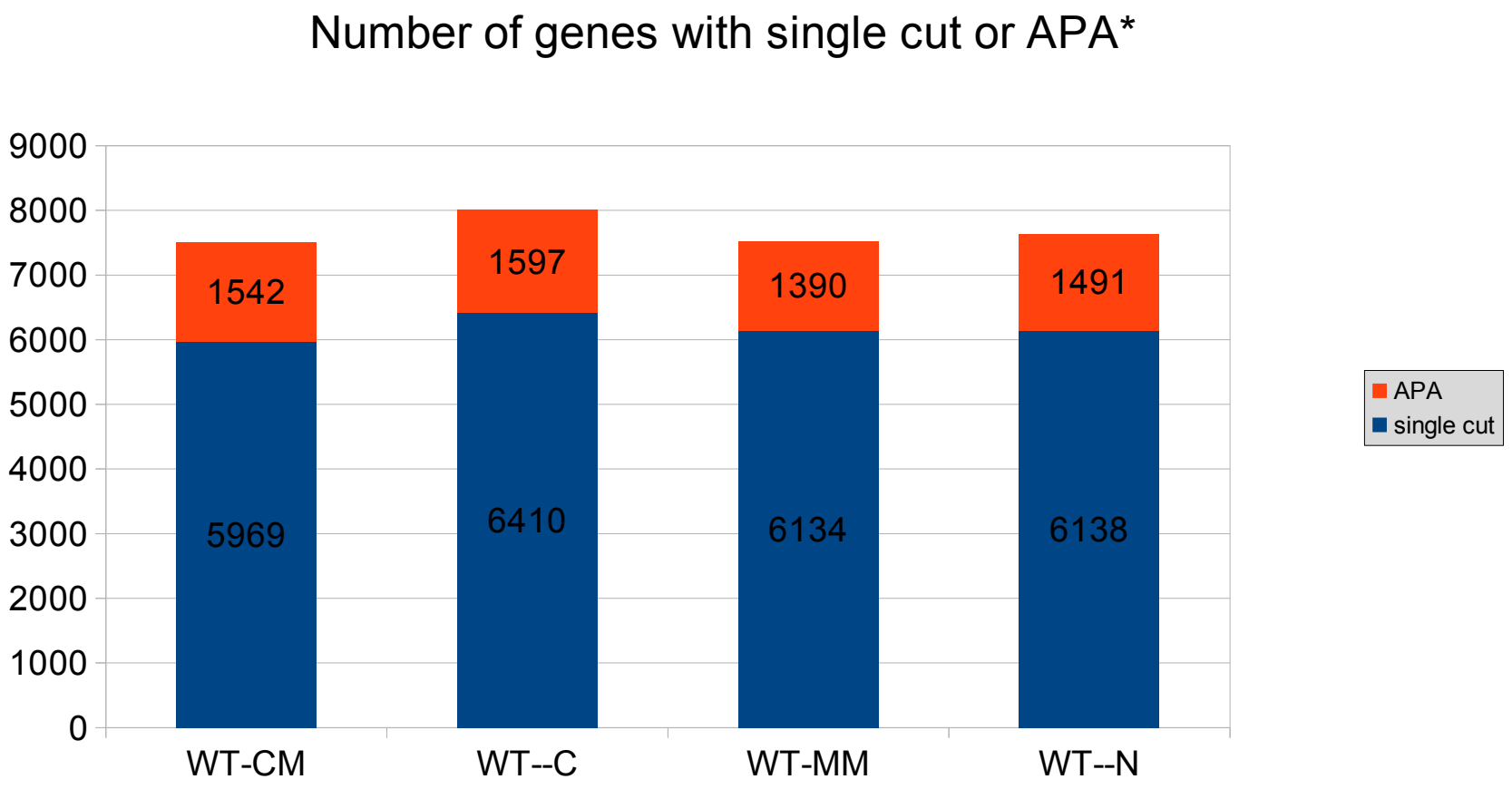
- 46 hits against Uniprot nt/nr database
- 4 hits against Rfam(ncRNA) database
- 80 overlapping annotated genes antisense

DIFFERENTIALLY EXPRESSED ORPHANS IN THE WT	
CM → MM-C	137
CM → MM	6
CM → MM-N	24
MM → MM-C	91
MM → MM-N	0
MM-N → MM-C	55

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

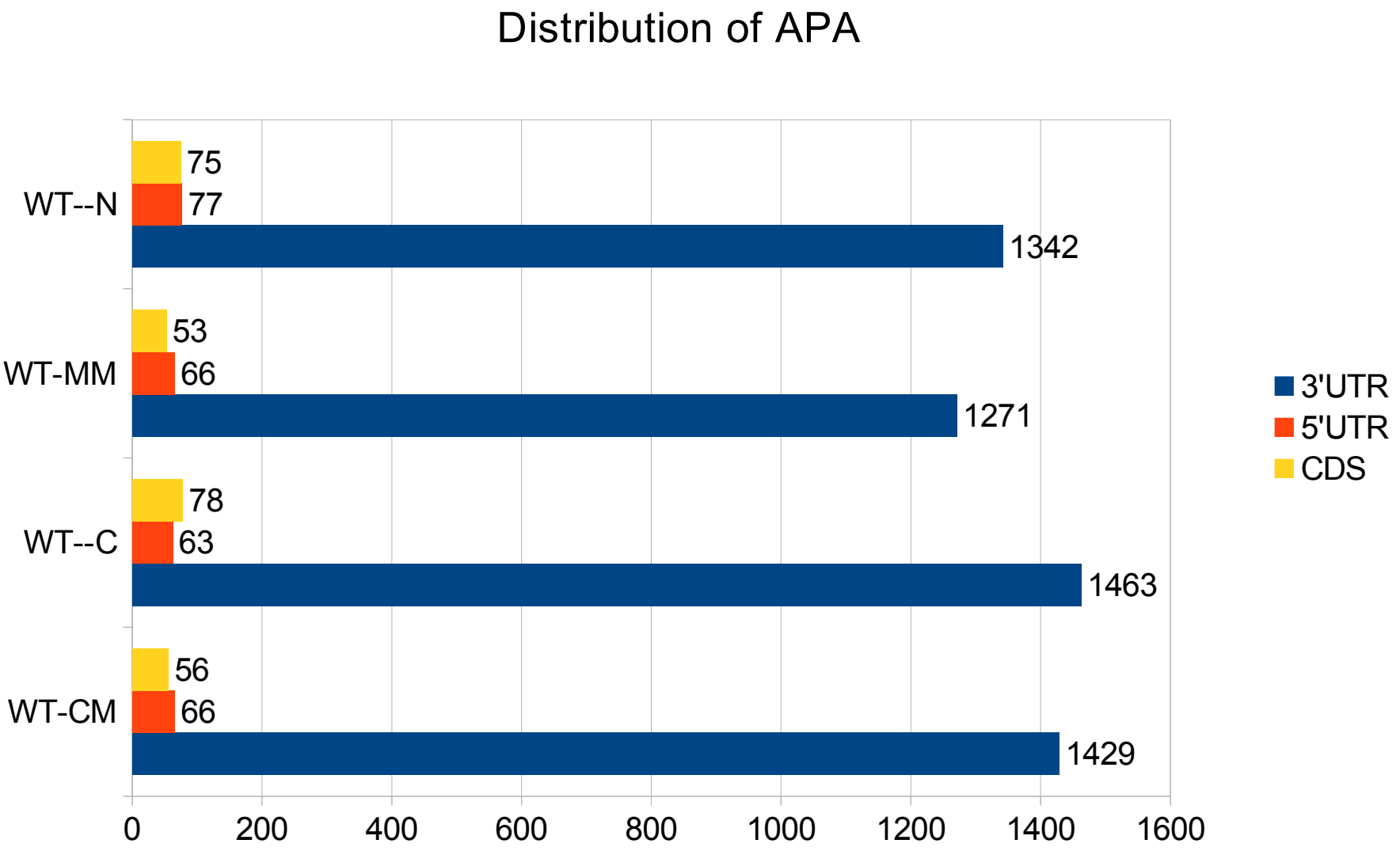


~20% of genes are alternatively polyadenilated

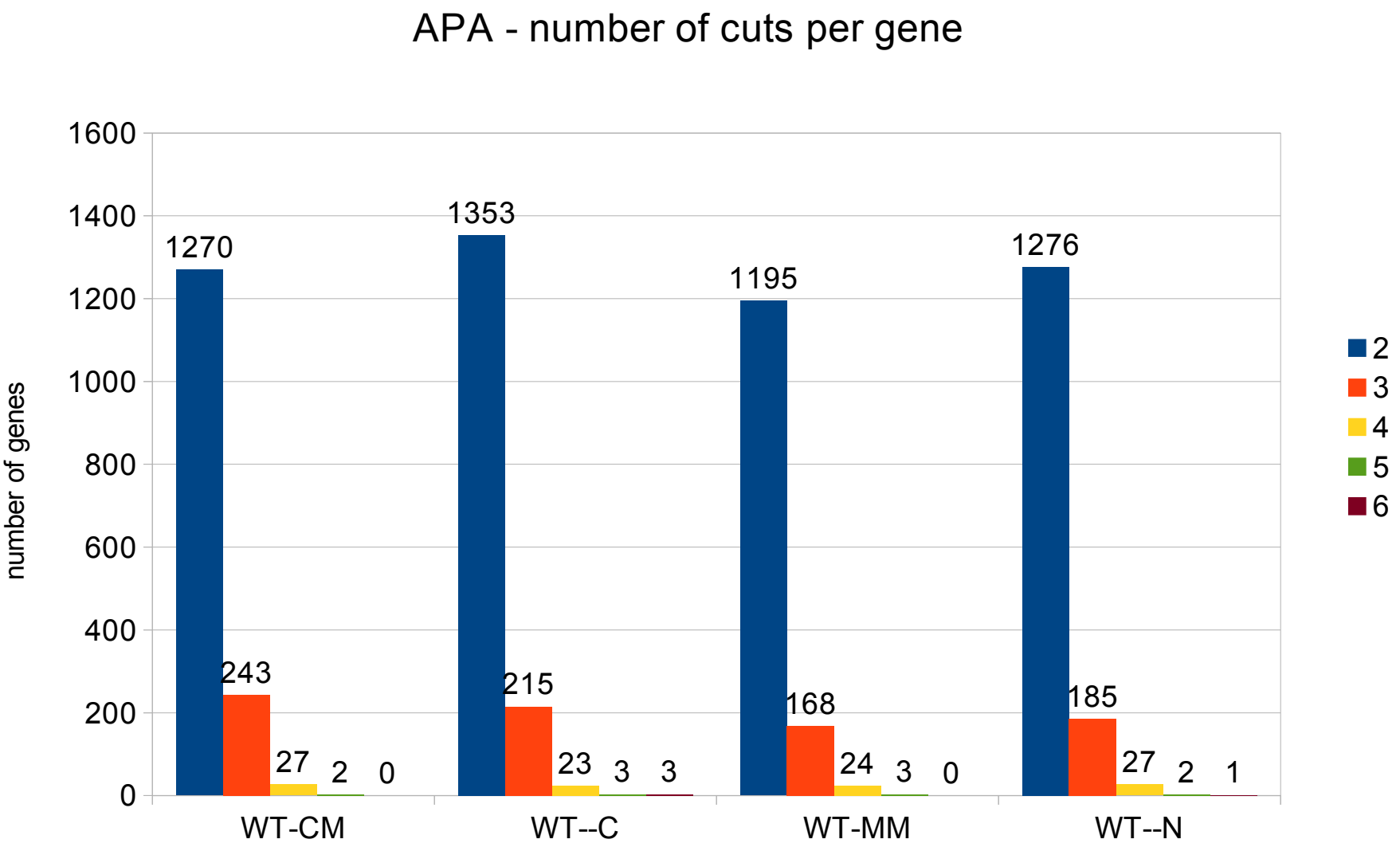


** calculated over the global number of expressed genes with a recognizable poly(A) site*

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

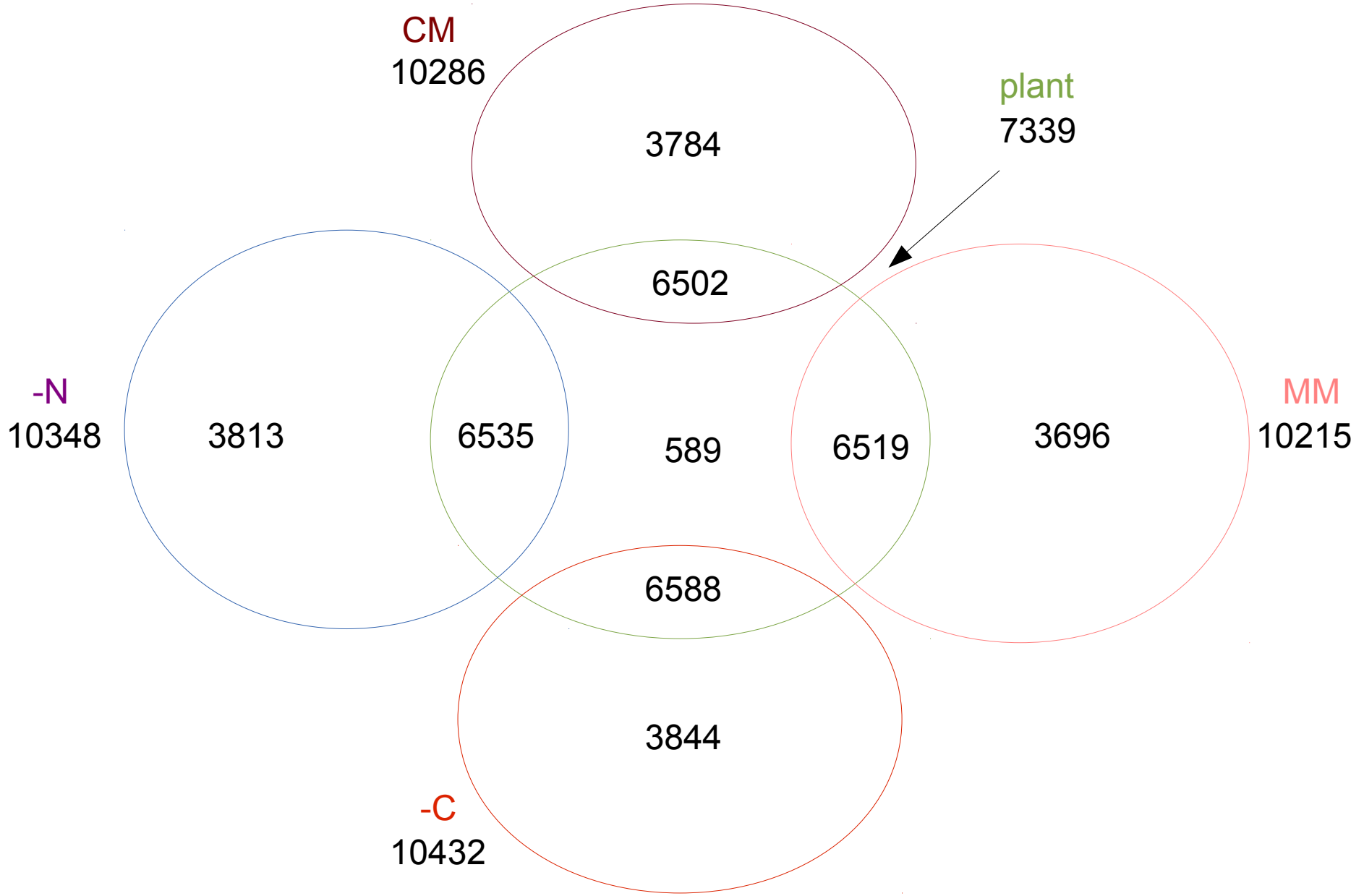


>80% of APA is composed of two cleveage sites

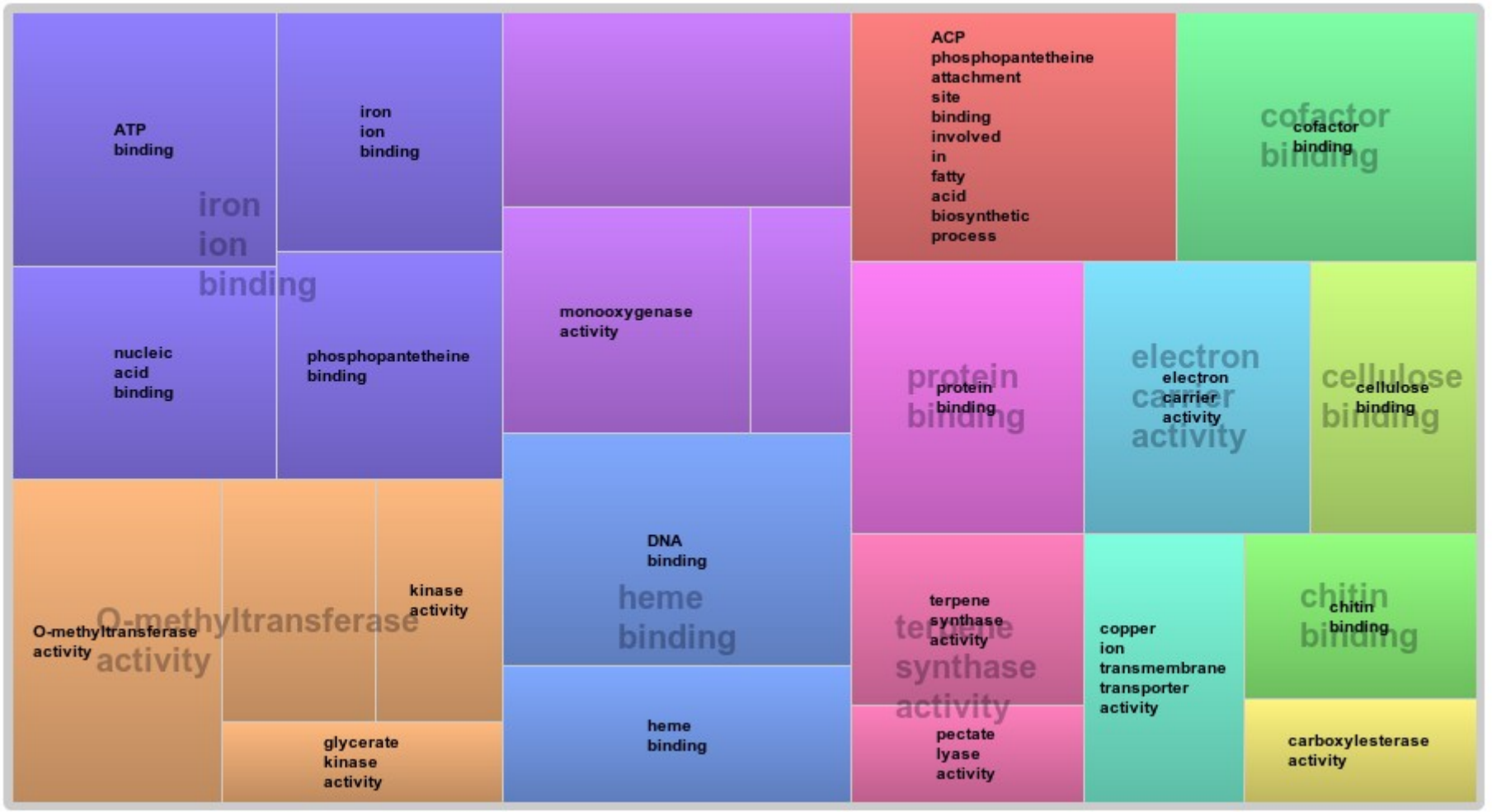


Gene expression between WT conditions

589 genes expressed in plant are never expressed in vitro



Plant specific genes functional analysis

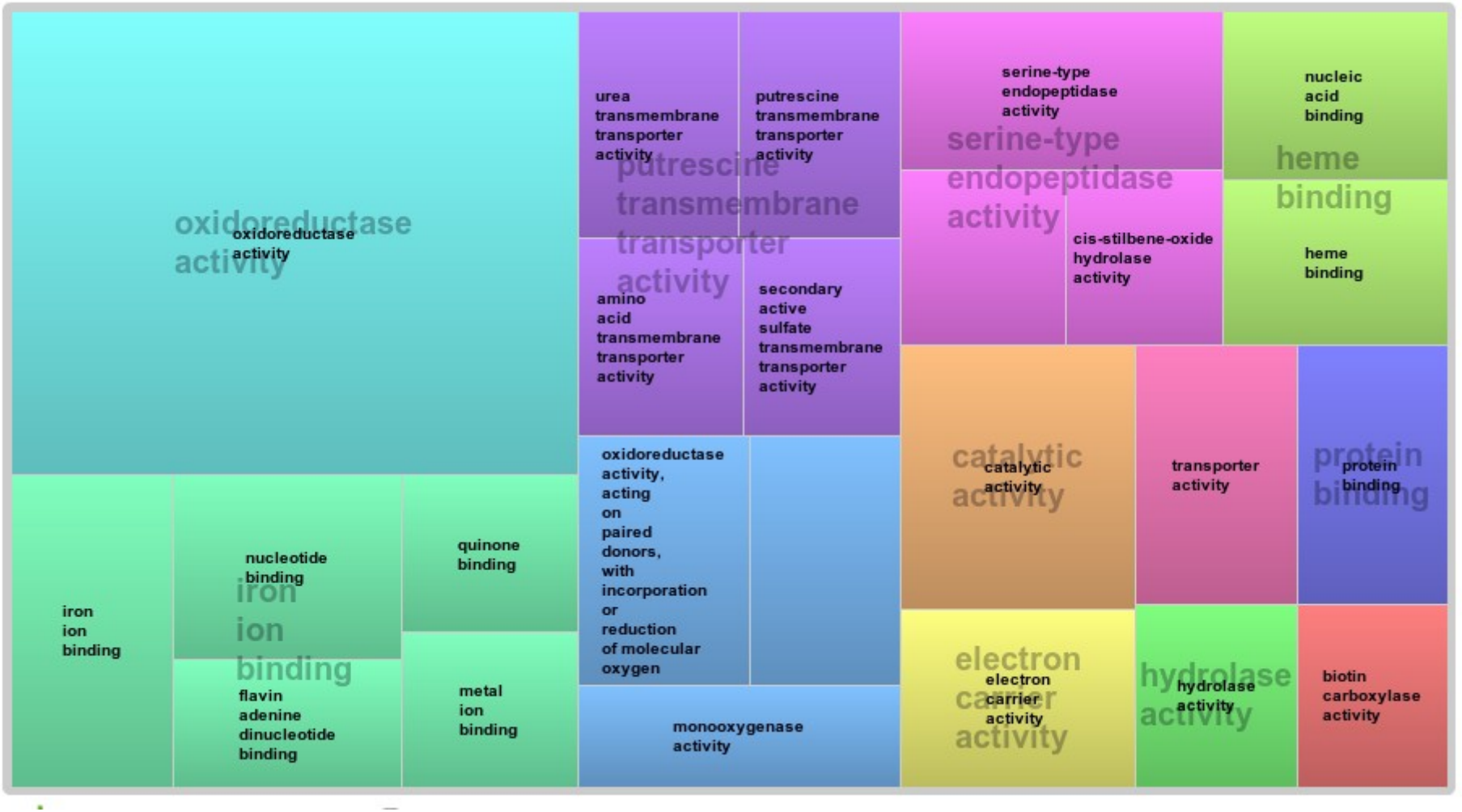


p-value = 0.05

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

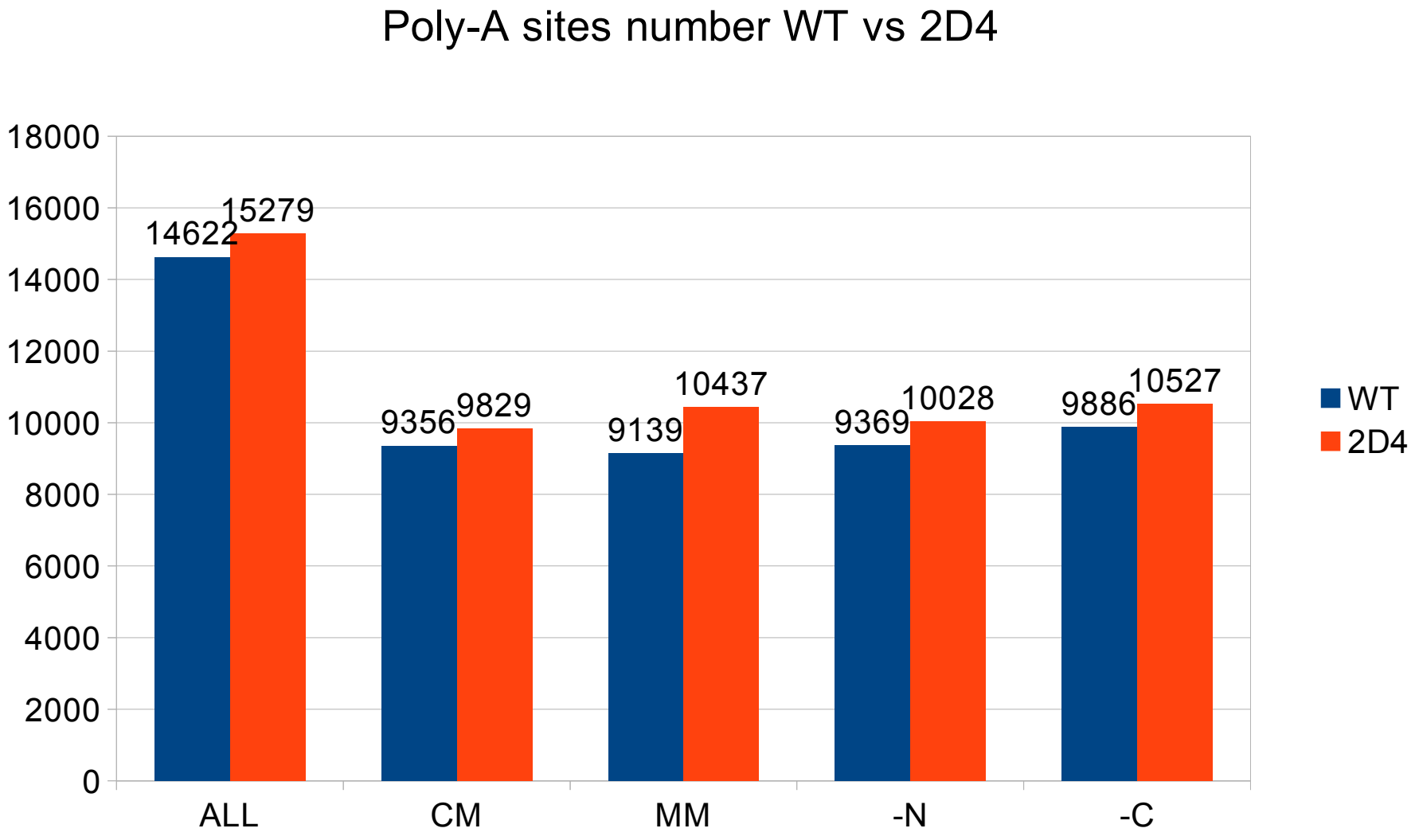
DIFFERENTIALLY EXPRESSED GENES IN THE WT			
	DOWN	UP	TOTAL
CM → MM-C	1586	1617	3203
CM → MM	151	285	436
CM → MM-N	177	347	524
MM → MM-C	601	608	1209
MM → MM-N	1	5	6
MM-N → MM-C	216	313	529

CM → MM functional analysis



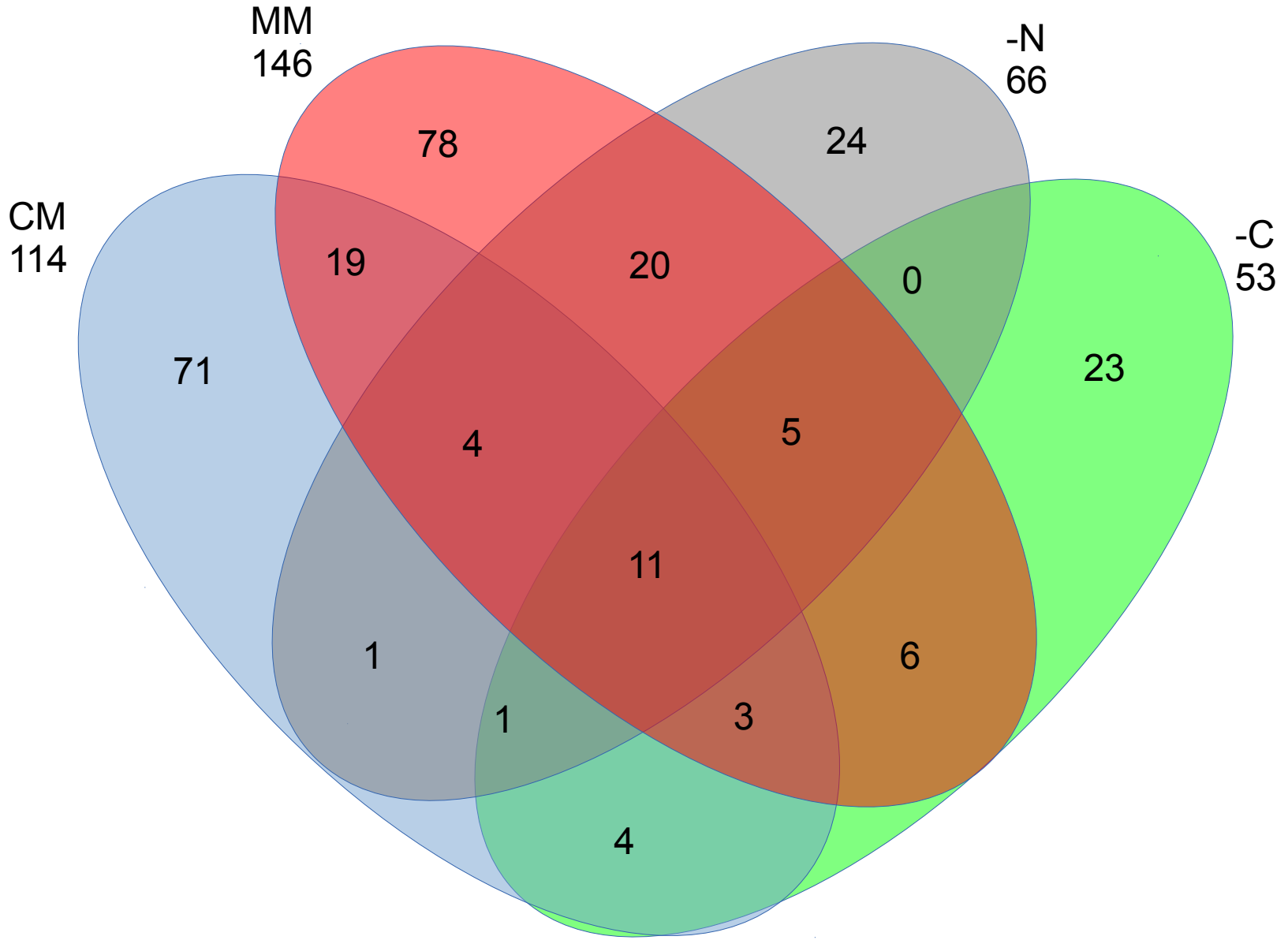
p-value = 0.1

Δrbp35 affects poly(A) sites number

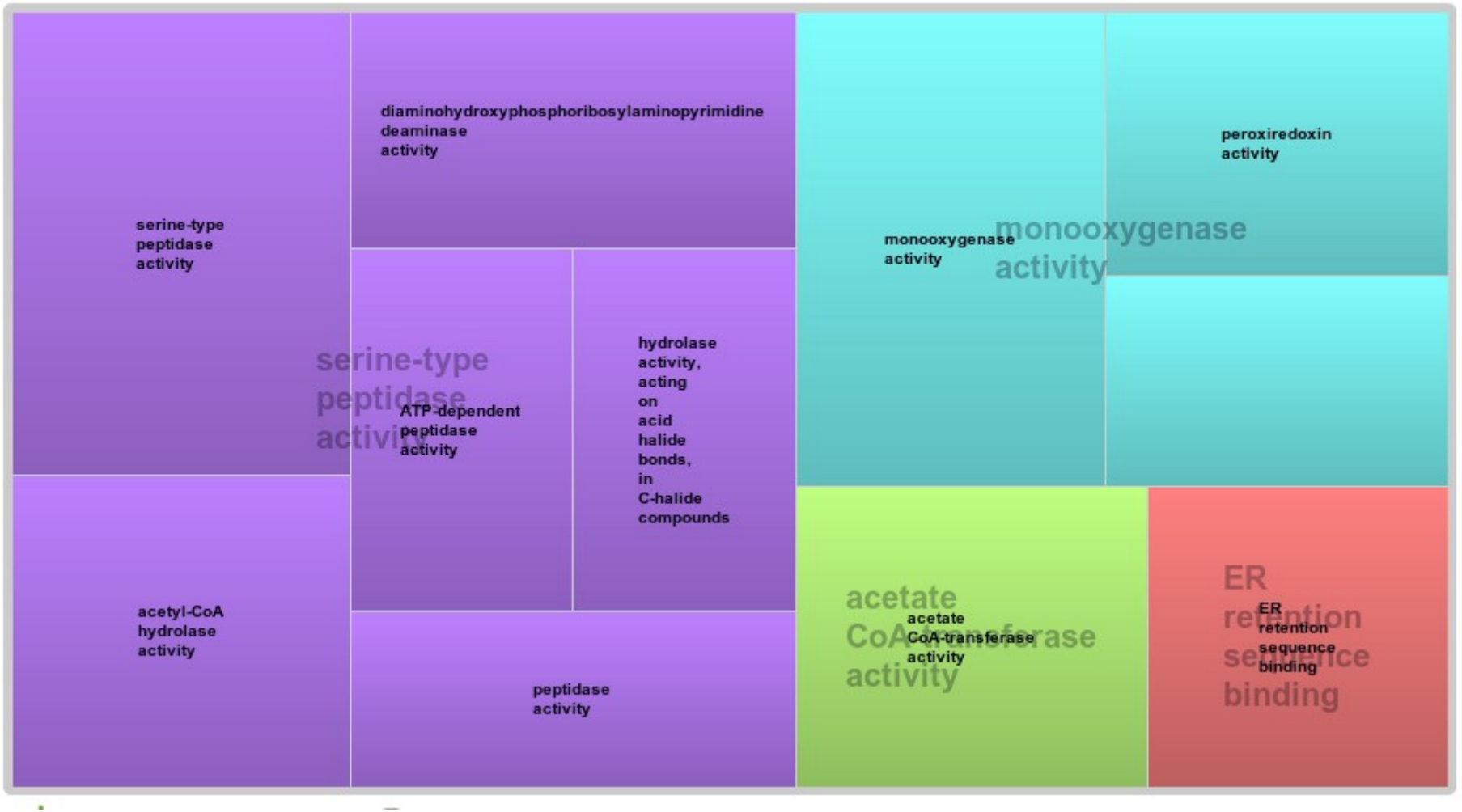


Δrbp35 affects number of cut sites per gene

Differentially expressed genes WT → Δ*rbp35*



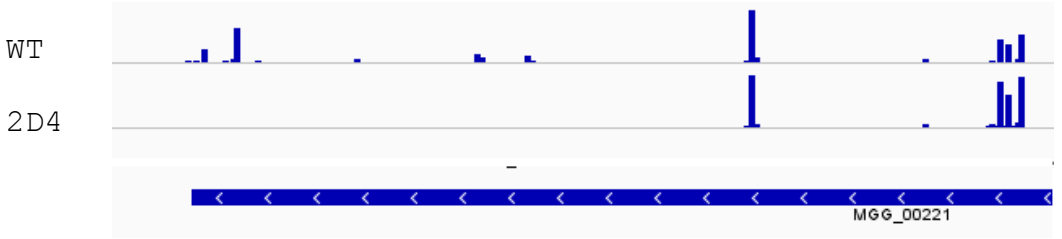
WT → 2D4 functional analysis



p-value = 0.05

Terminology - 1

- **group P1**: Down-regulated poly(A) sites in Δrbp35 (polyA site RBP35 dependent):



- **Group G1**: genes containing group P1 poly(A) sites

- **Group P1-2D4**: every poly(A) sites in G1 genes from 2D4

Terminology - 2

- **group P2**: Up-regulated poly(A) sites in Δrbp35 (unspecific poly(A) site-not RBP35 dependent):

- **Group G2**: genes containing group P2 poly(A) sites

- **Group P2-WT**: every poly(A) sites in G2 genes from WT

Terminology - 3

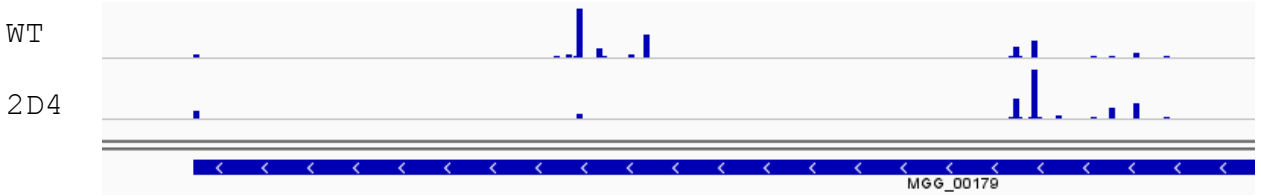
Group P3: poly(A) sites union from P1 and P2 groups

Group G3: Union of G1 and G2 genes groups, genes whose poly(A) sites are up or down-regulated

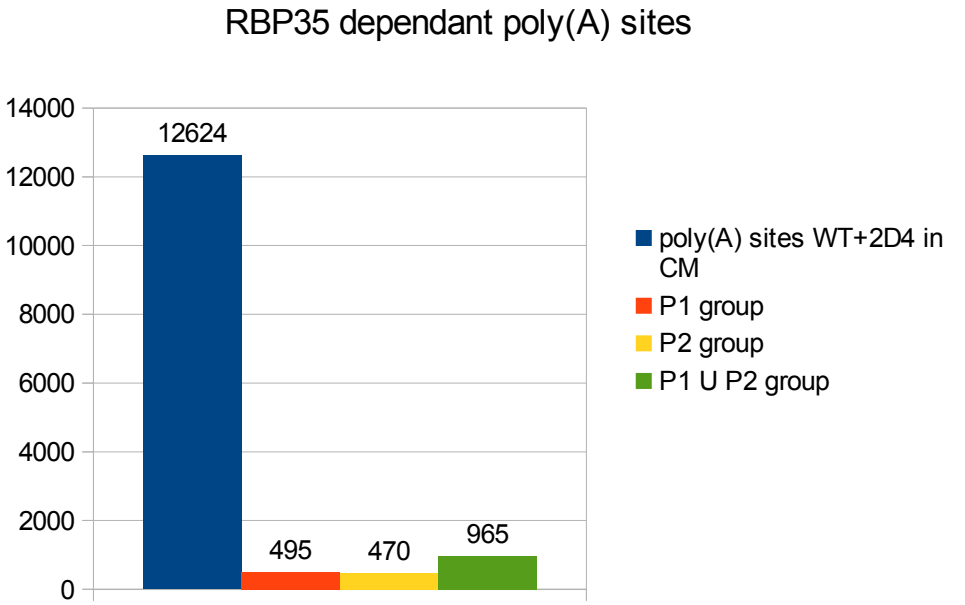
Terminology - 4

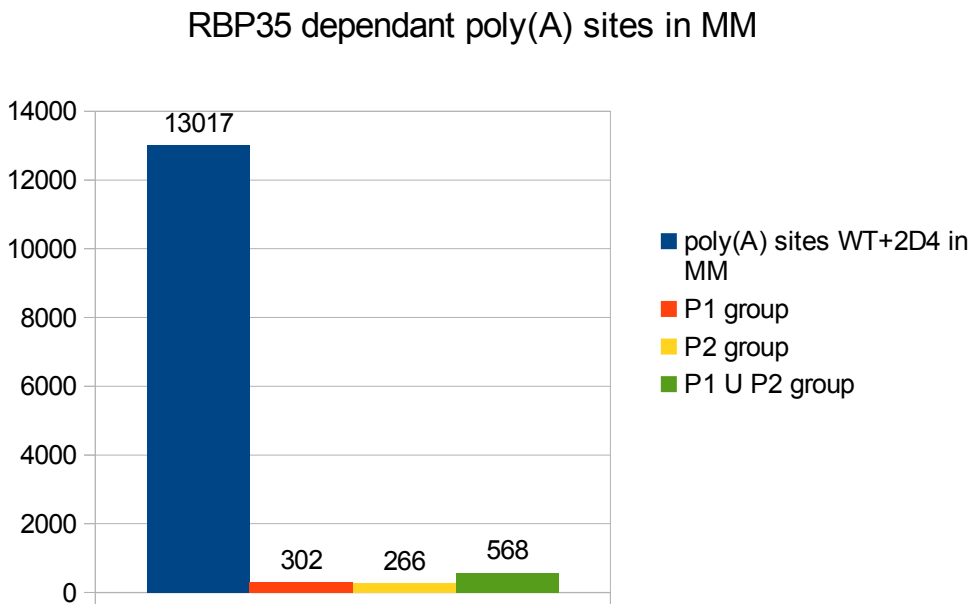
Group P4: poly(A) sites from P1 and P2 groups appearing on the same genes

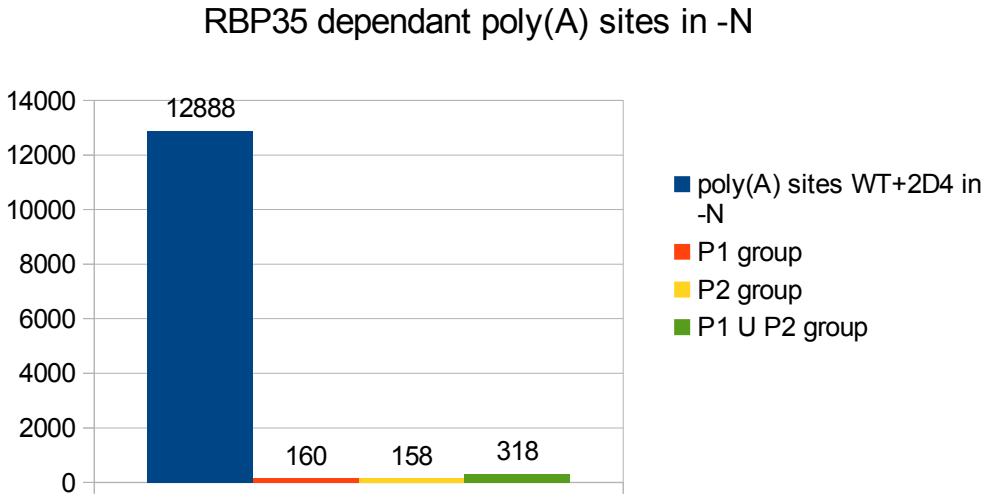
Group G4: Genes with change position (both down and up-regulated) poly(A) sites in Δrbp35 (positional RBP35 dependent), a specific case from G1 and G2 (intersection of G1 and G2):

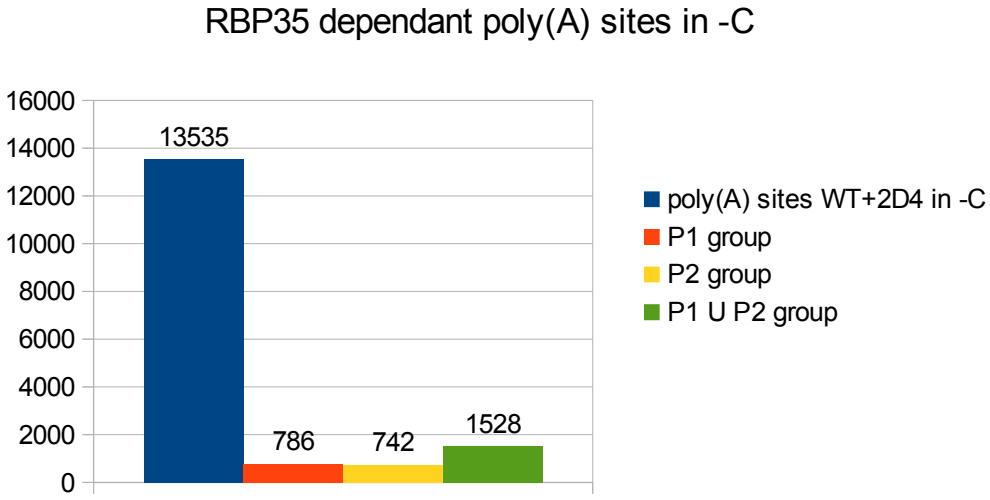


2%-11% of poly(A) sites are targeted by RBP35 in all media



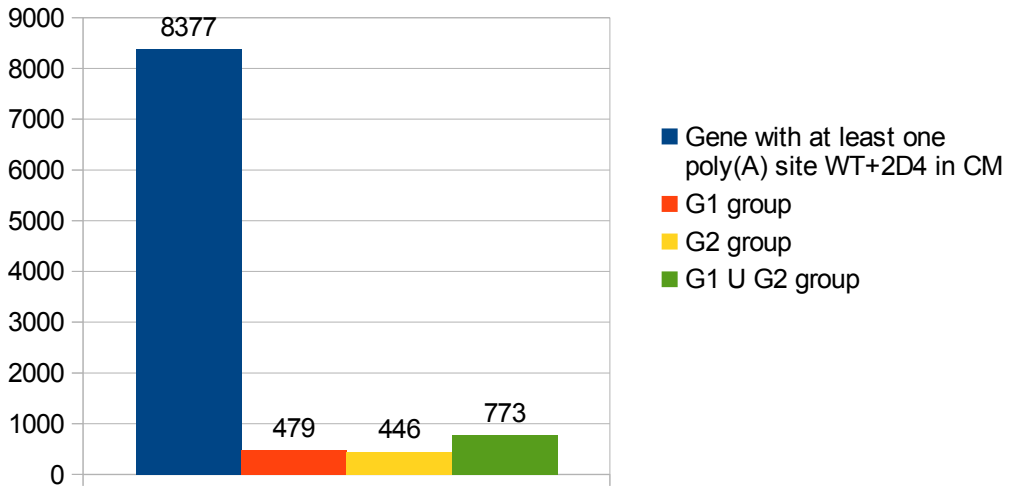




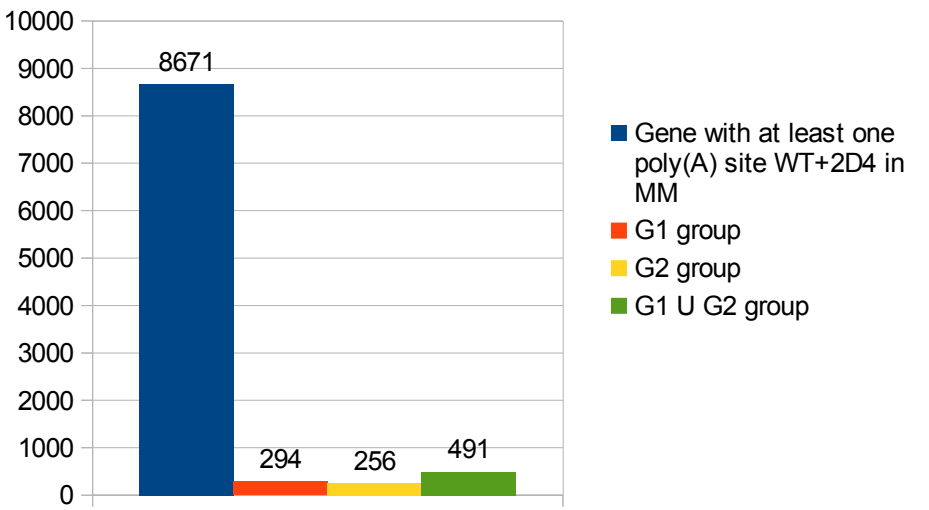


3%-13% of genes are targeted by RBP35 in all media

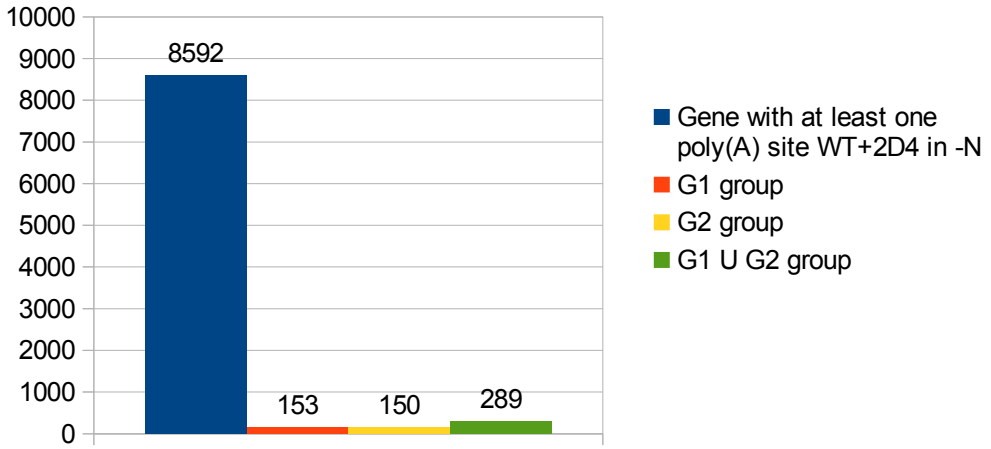
RBP35 dependant genes in CM



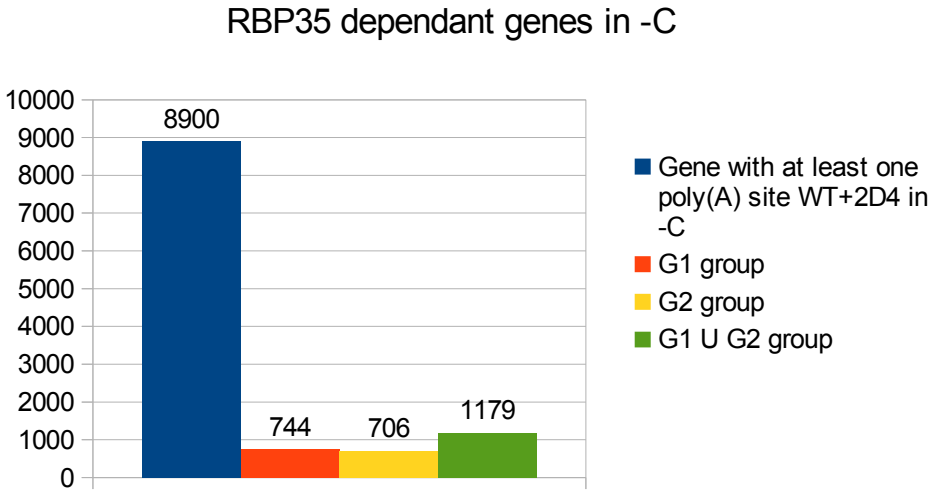
RBP35 dependant genes in MM



RBP35 dependant genes in -N

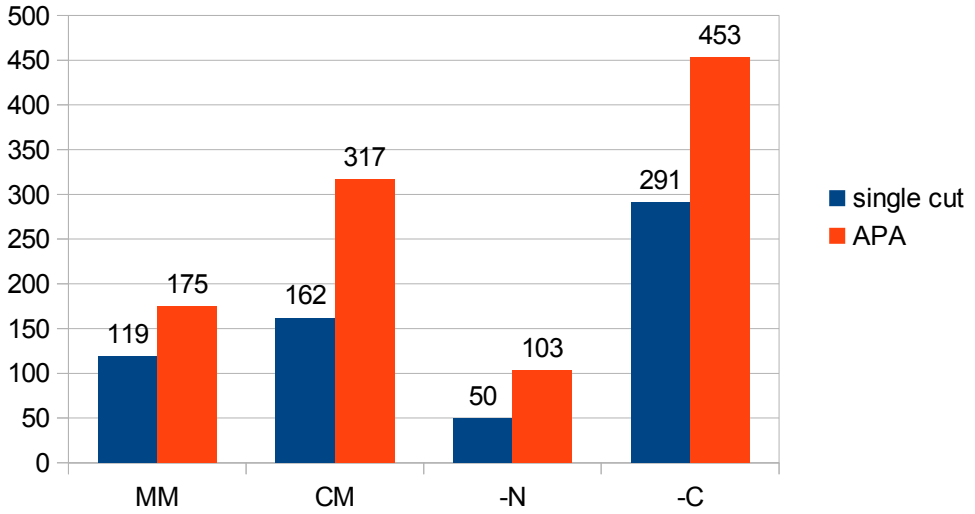


RBP35 dependant genes in -C

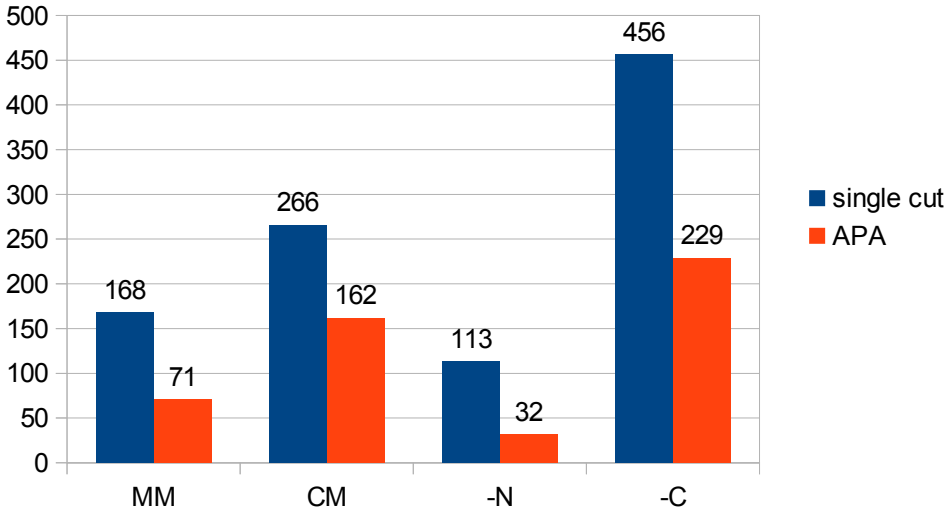


APA in RBP35 dependant genes in WT

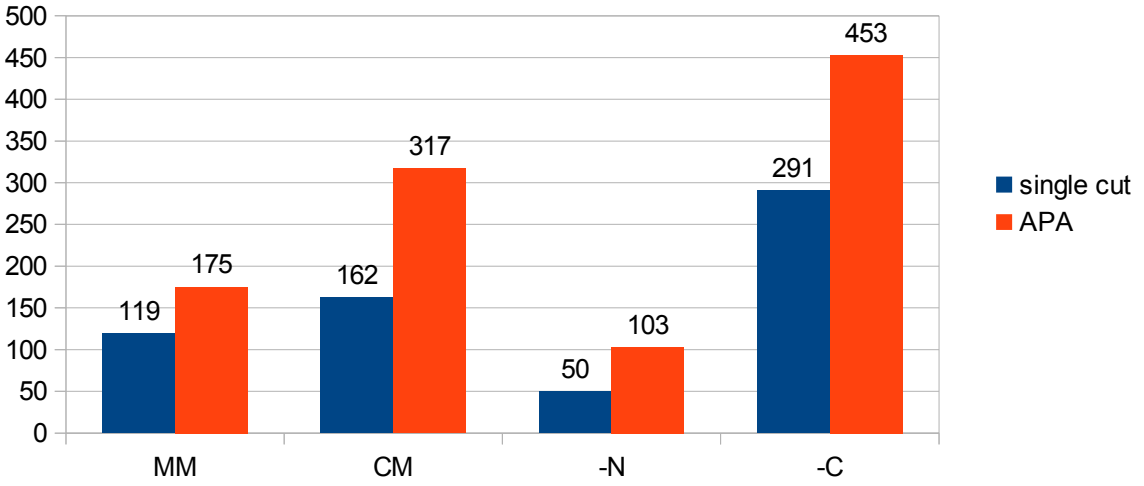
Single cut & APA in G1 group genes



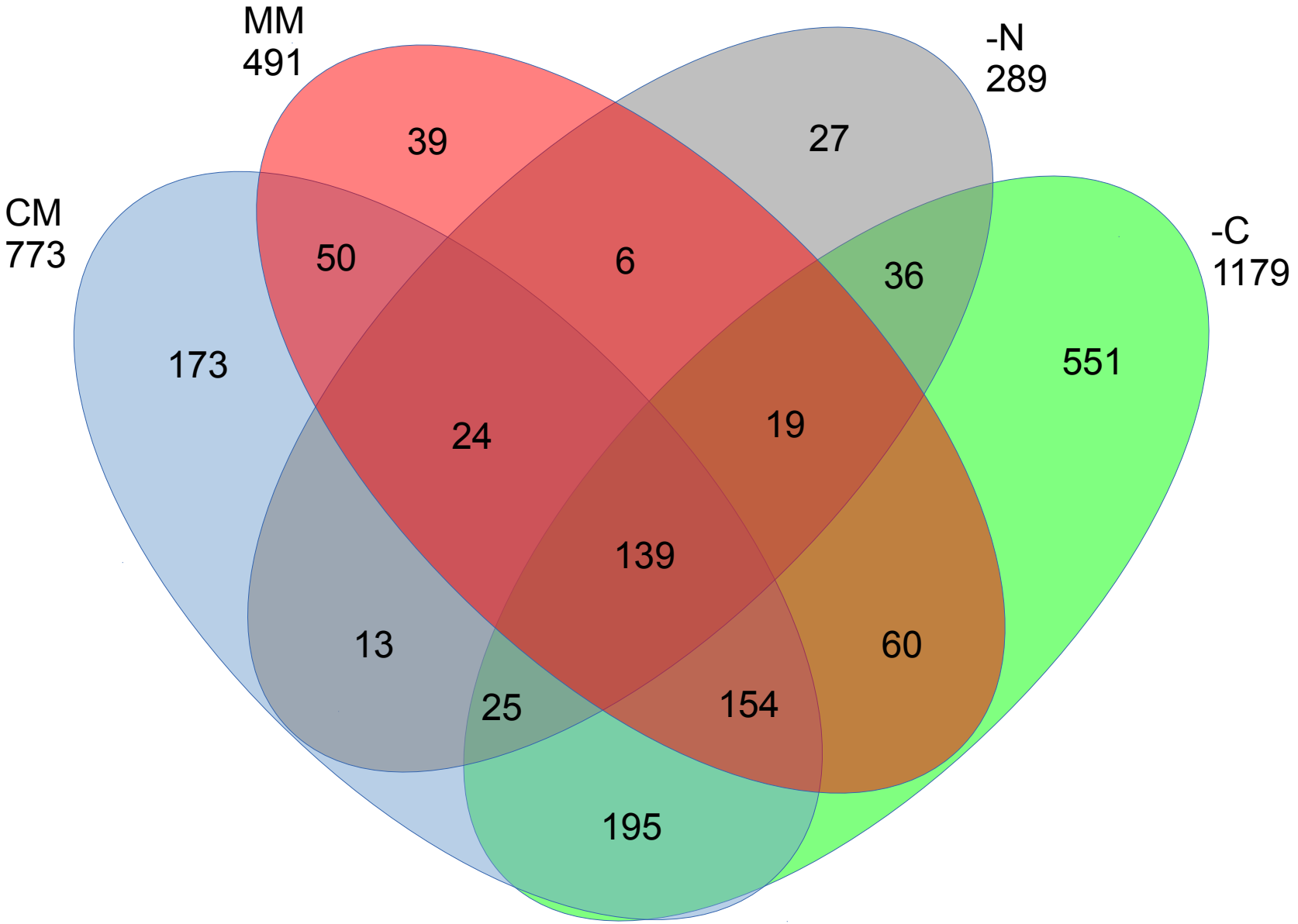
Single cut & APA in G2 group genes



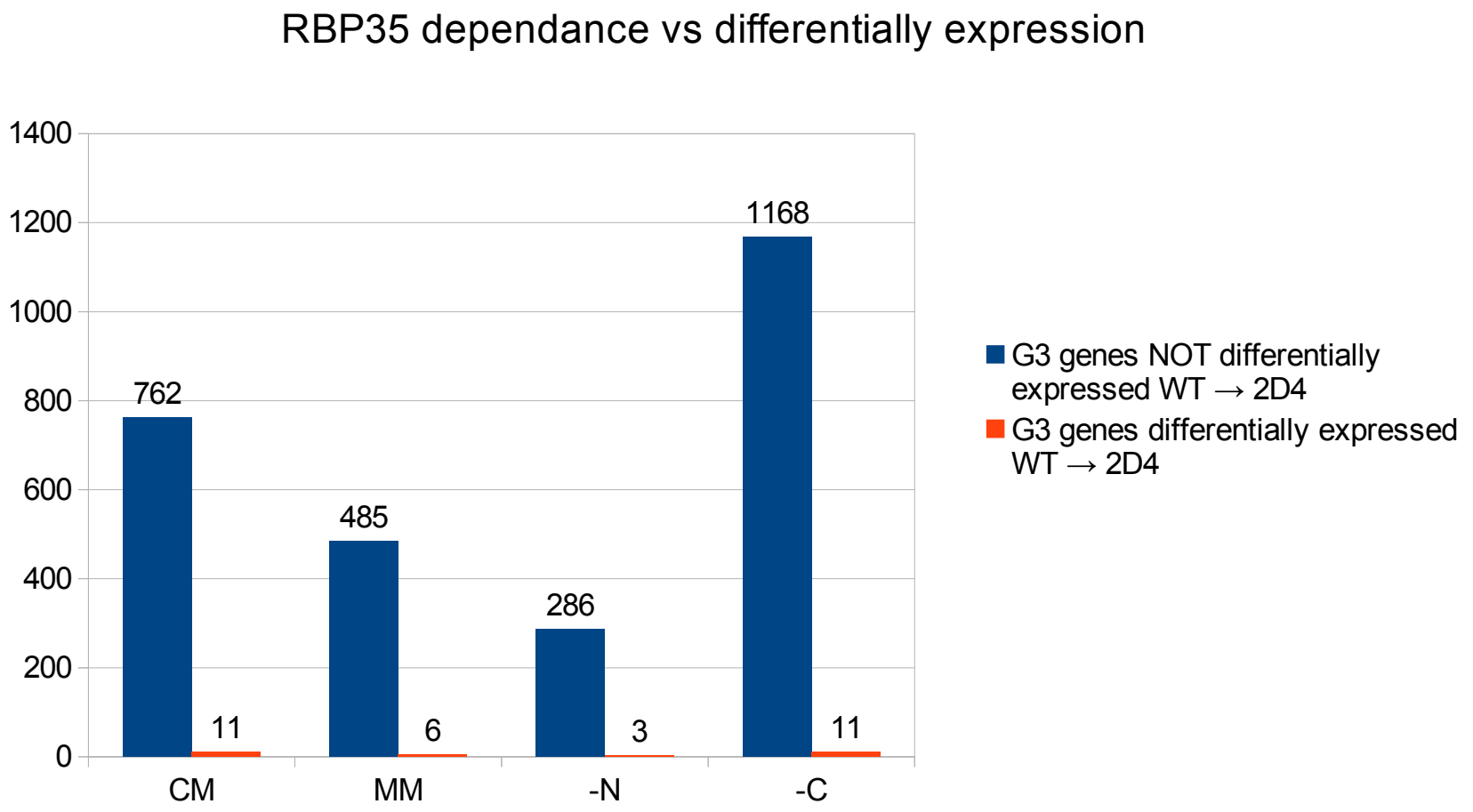
Single cut & APA in G3 group genes



RBP35 dependant genes

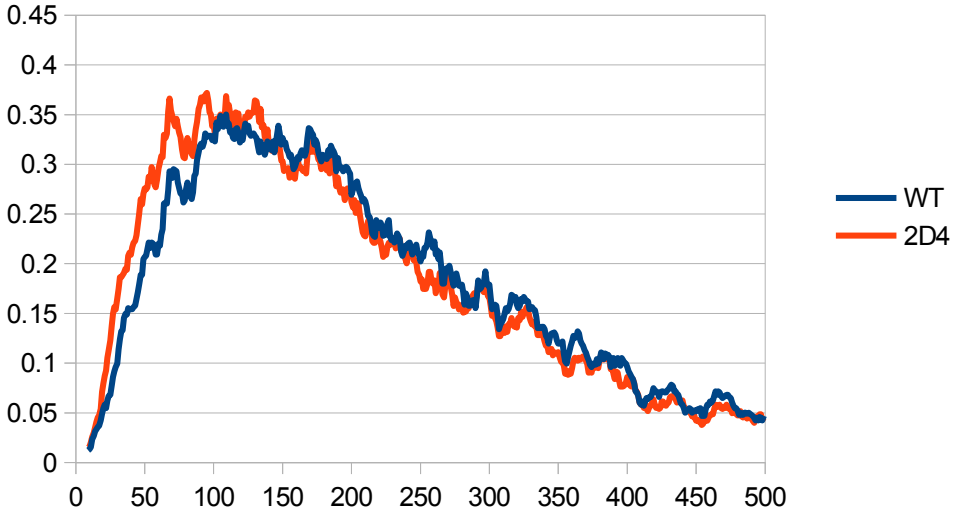


There is no correlation between RBP35 dependance and differential expression

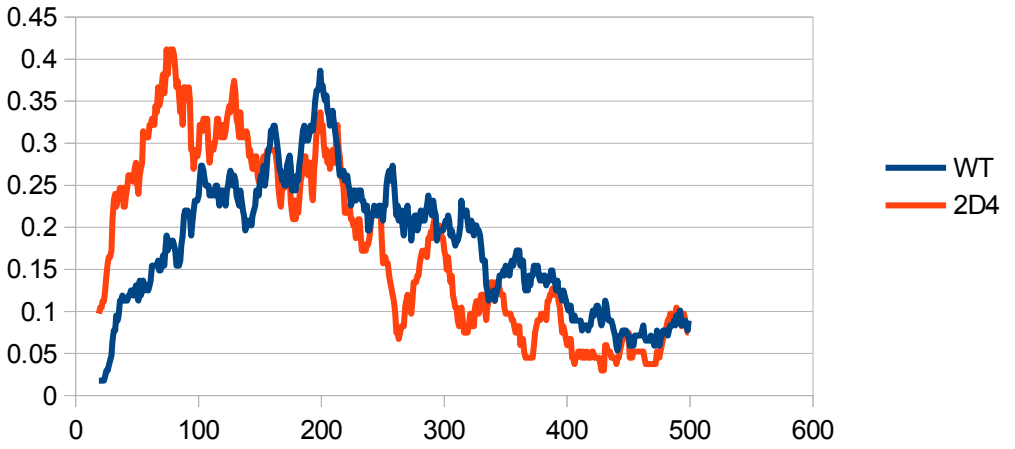


Δrbp35 affects 3'UTR length

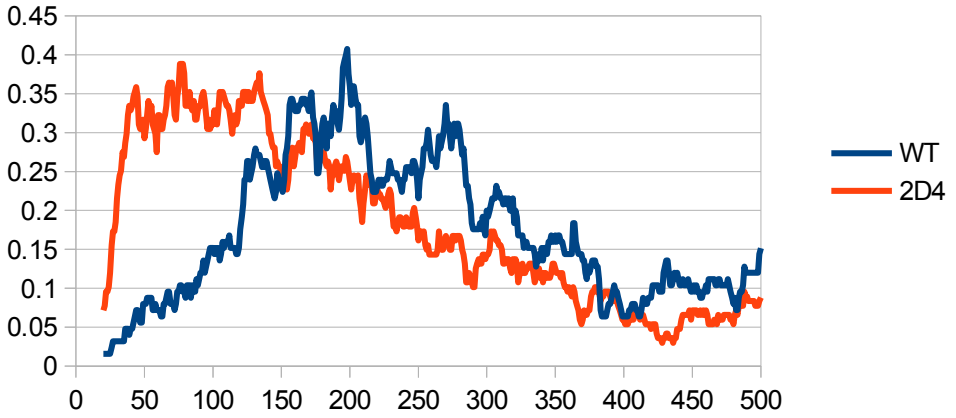
3'UTR length (bp) all genes CM



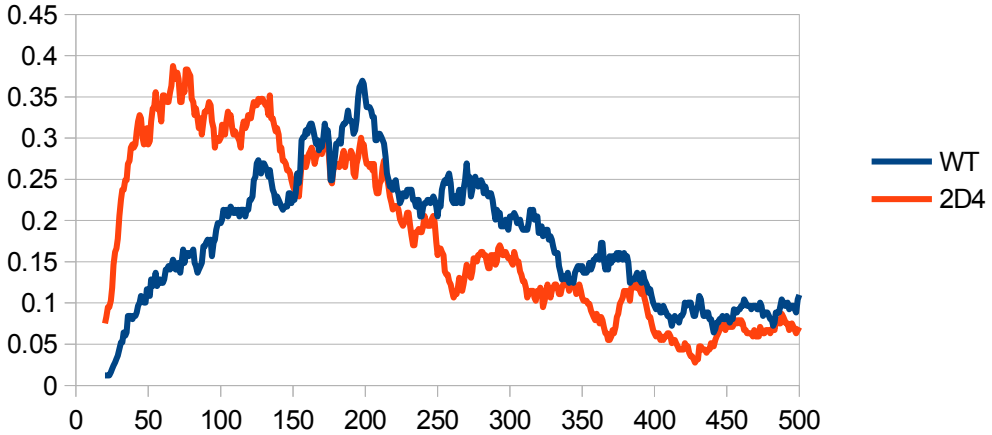
3'UTR length (bp) G1 group only CM



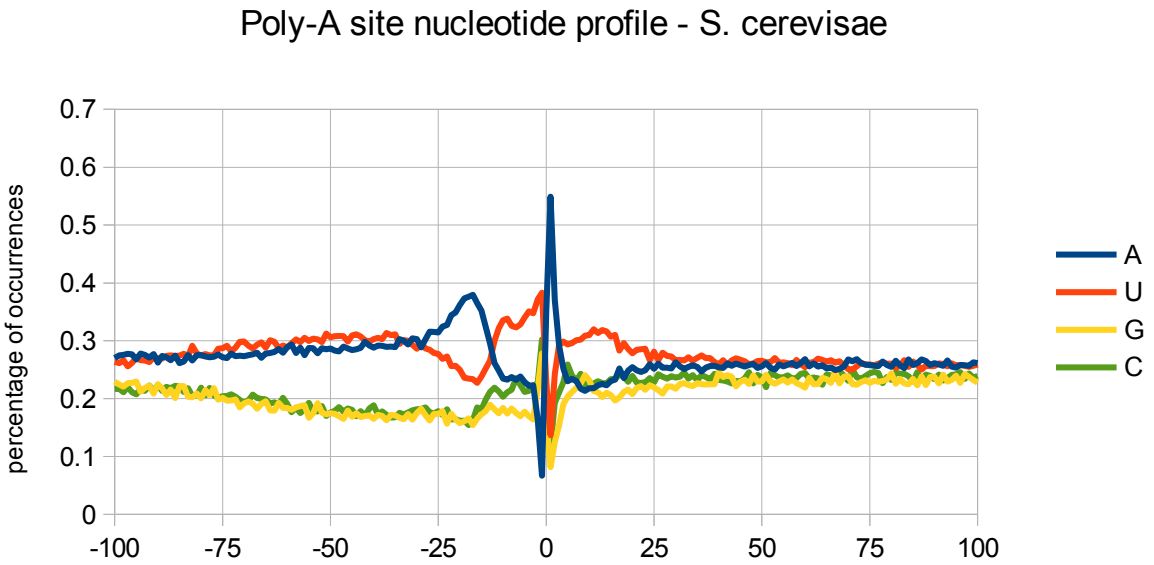
3'UTR length (bp) G2 group only CM

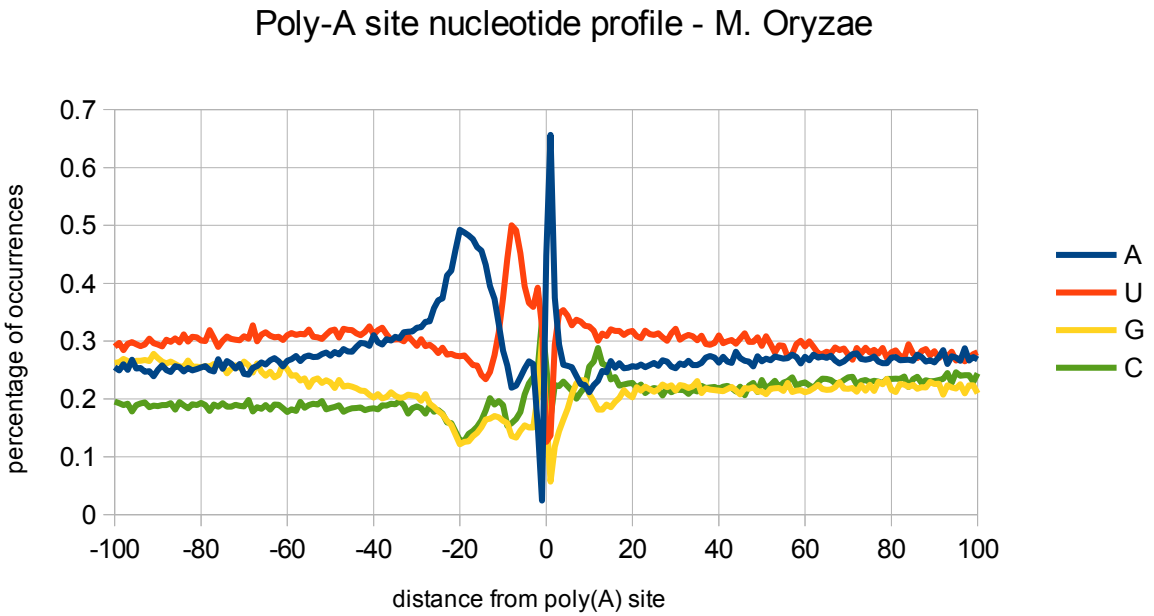


3'UTR length (bp) G3 group only CM

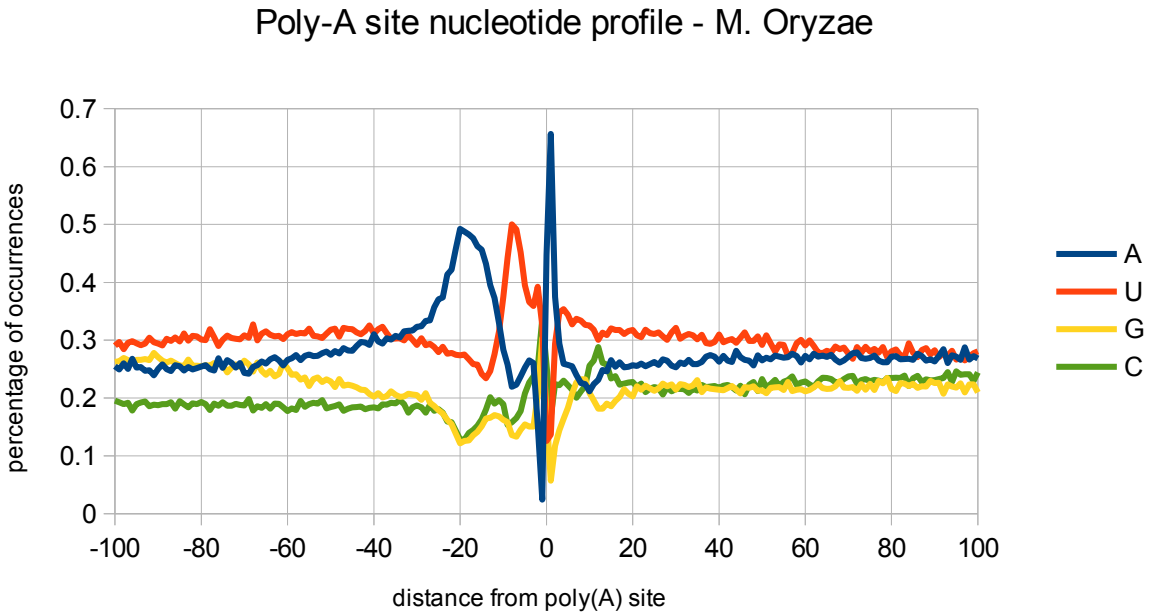


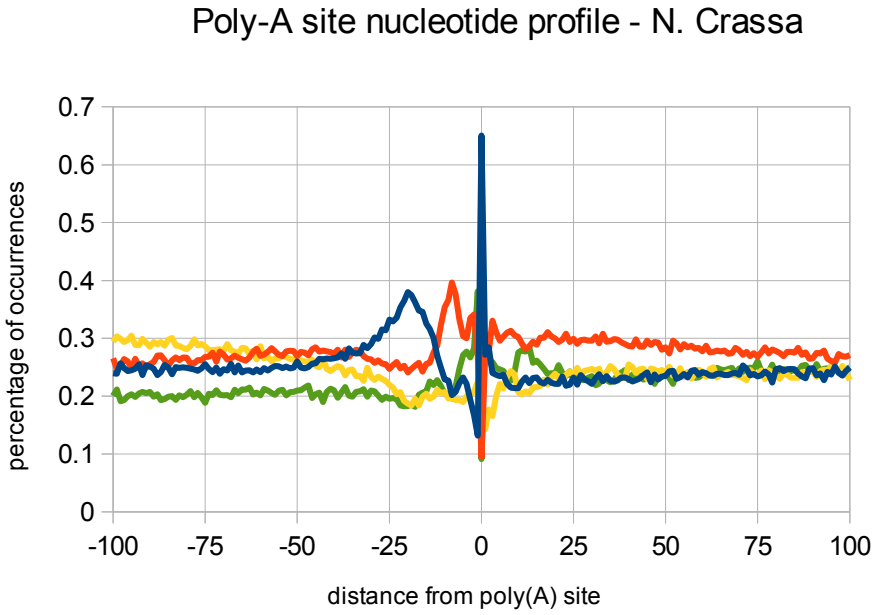
Nucleotides profile of poly(A) sites slighty differs from *S.cerevisae*

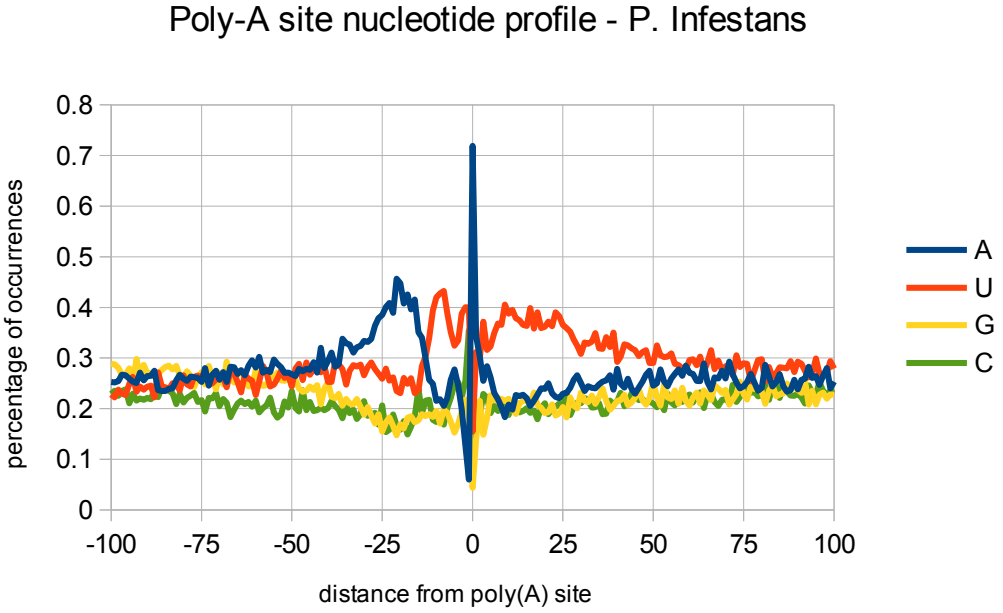




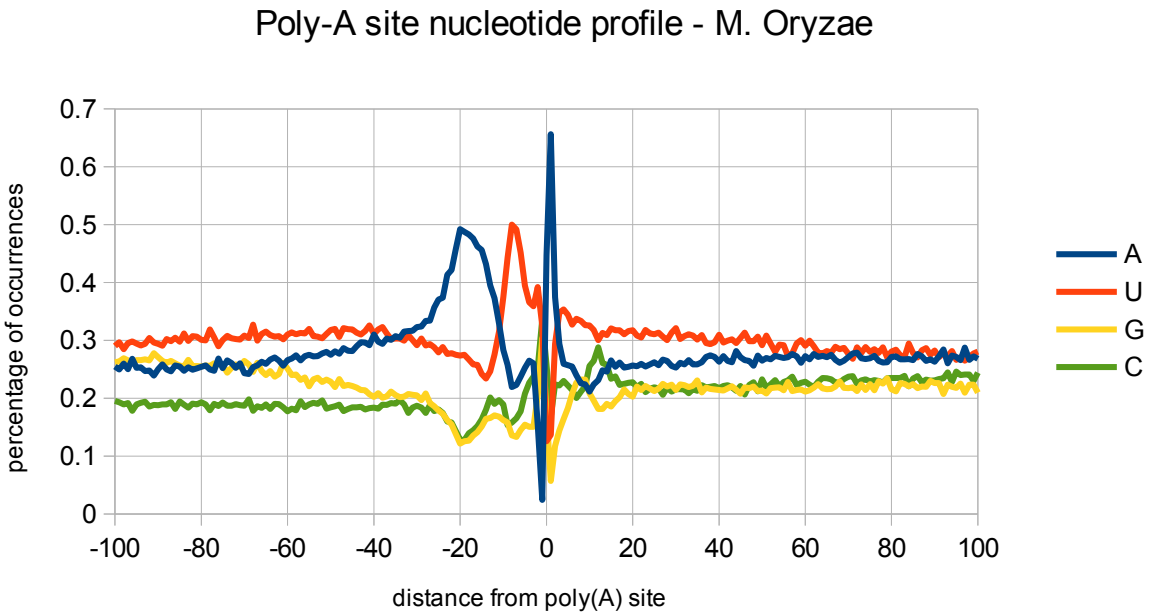
Nucleotides profile of poly(A) sites resembles *N. crassa*

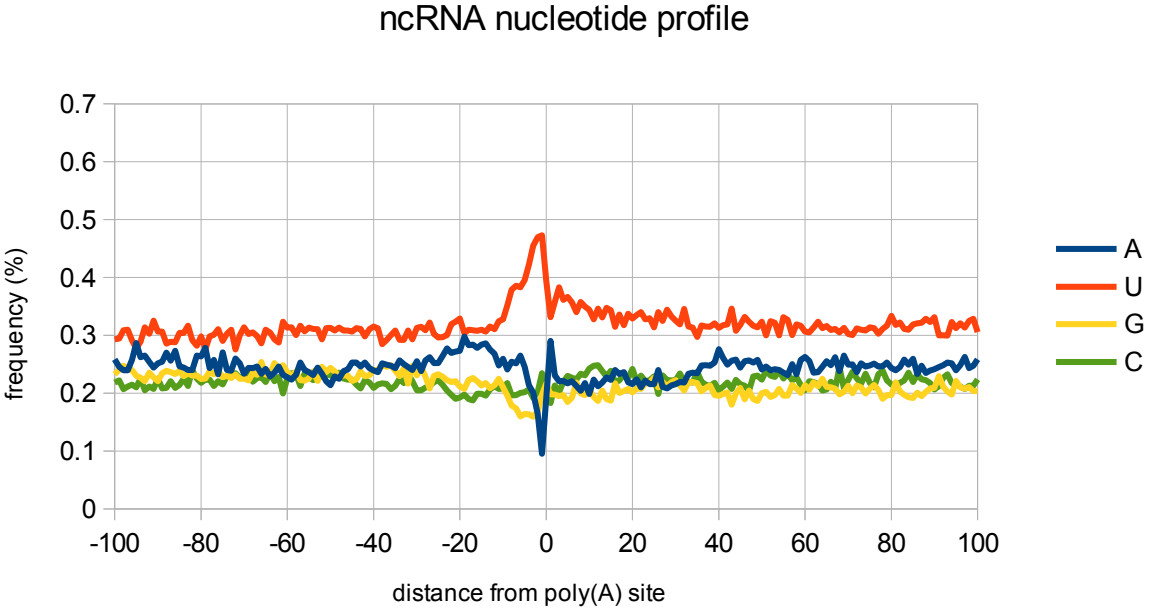






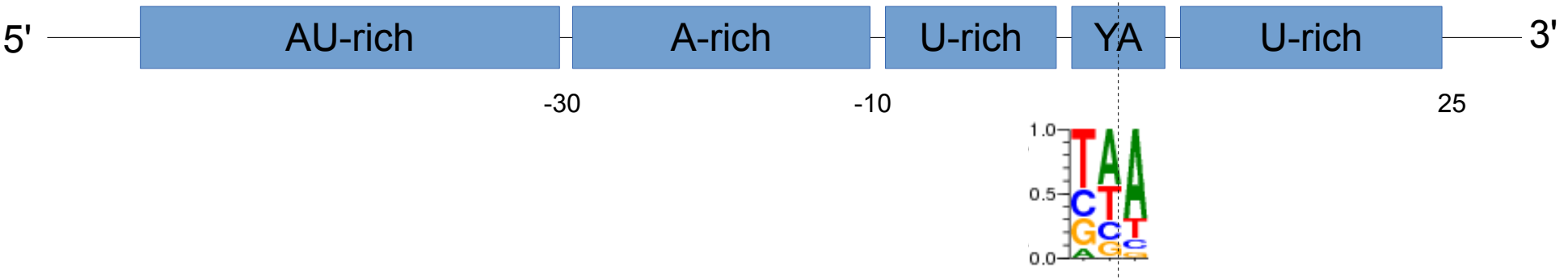
Nucleotides profile of poly(A) of ncRNA is different



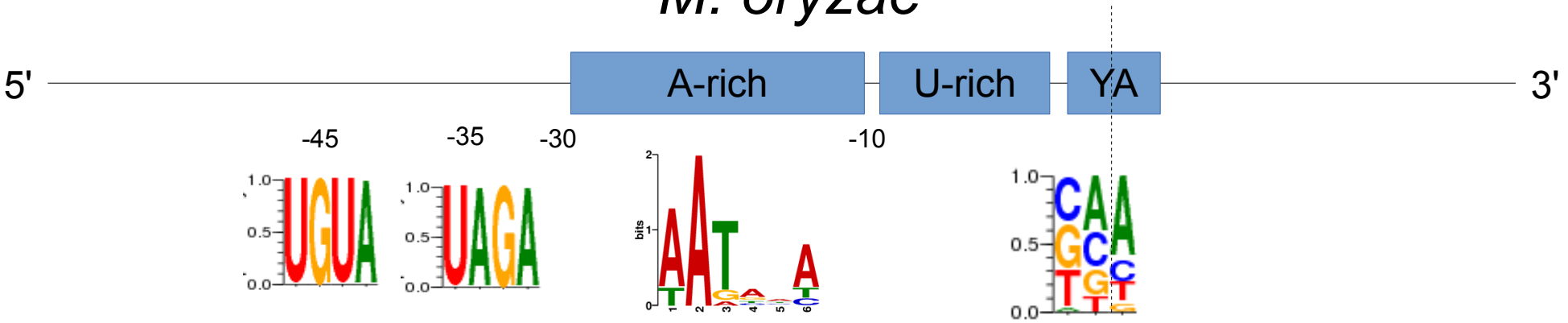


Nucleotides profile of poly(A) sites differs from *S.cerevisae*

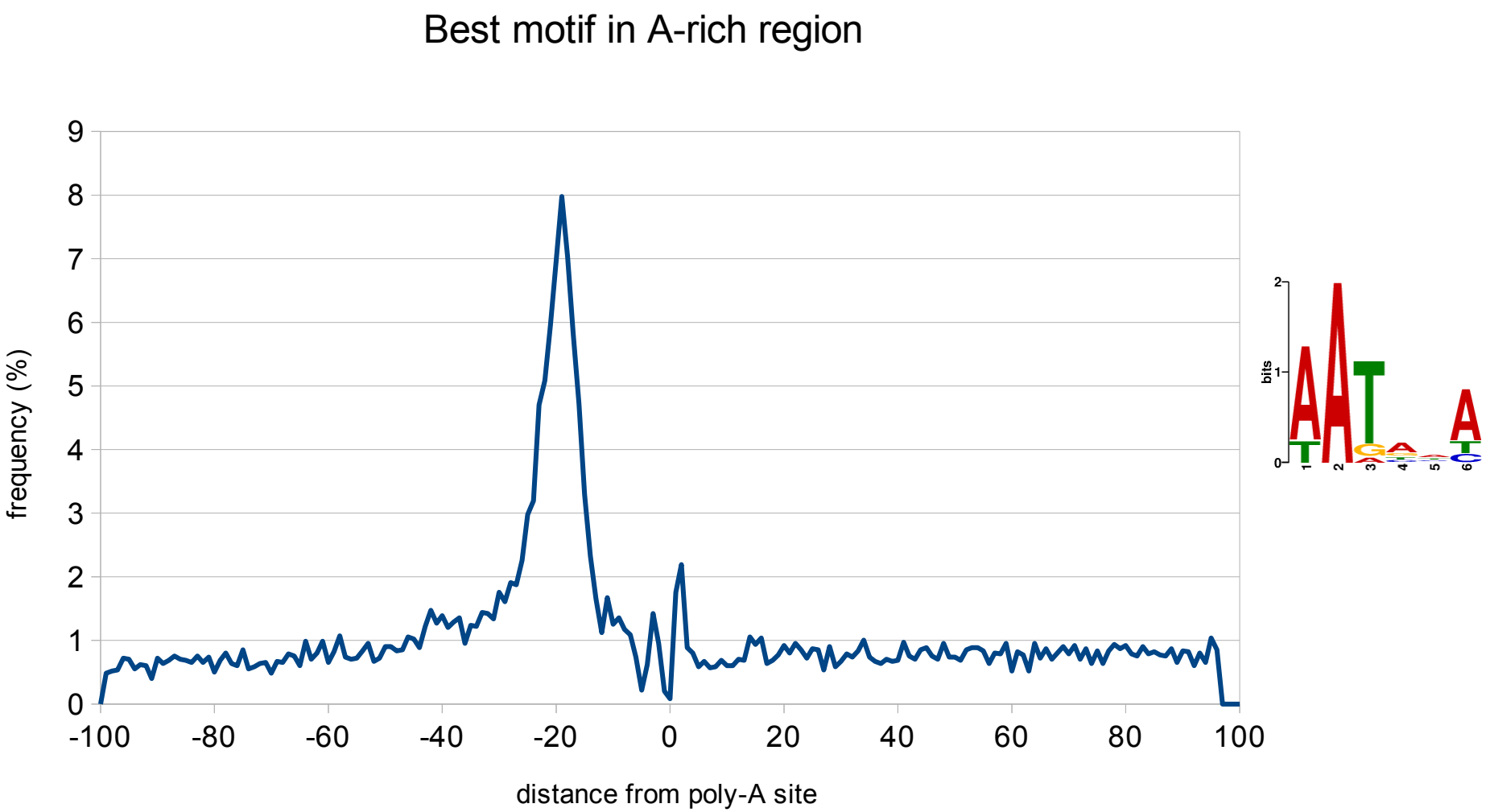
S.cerevisae



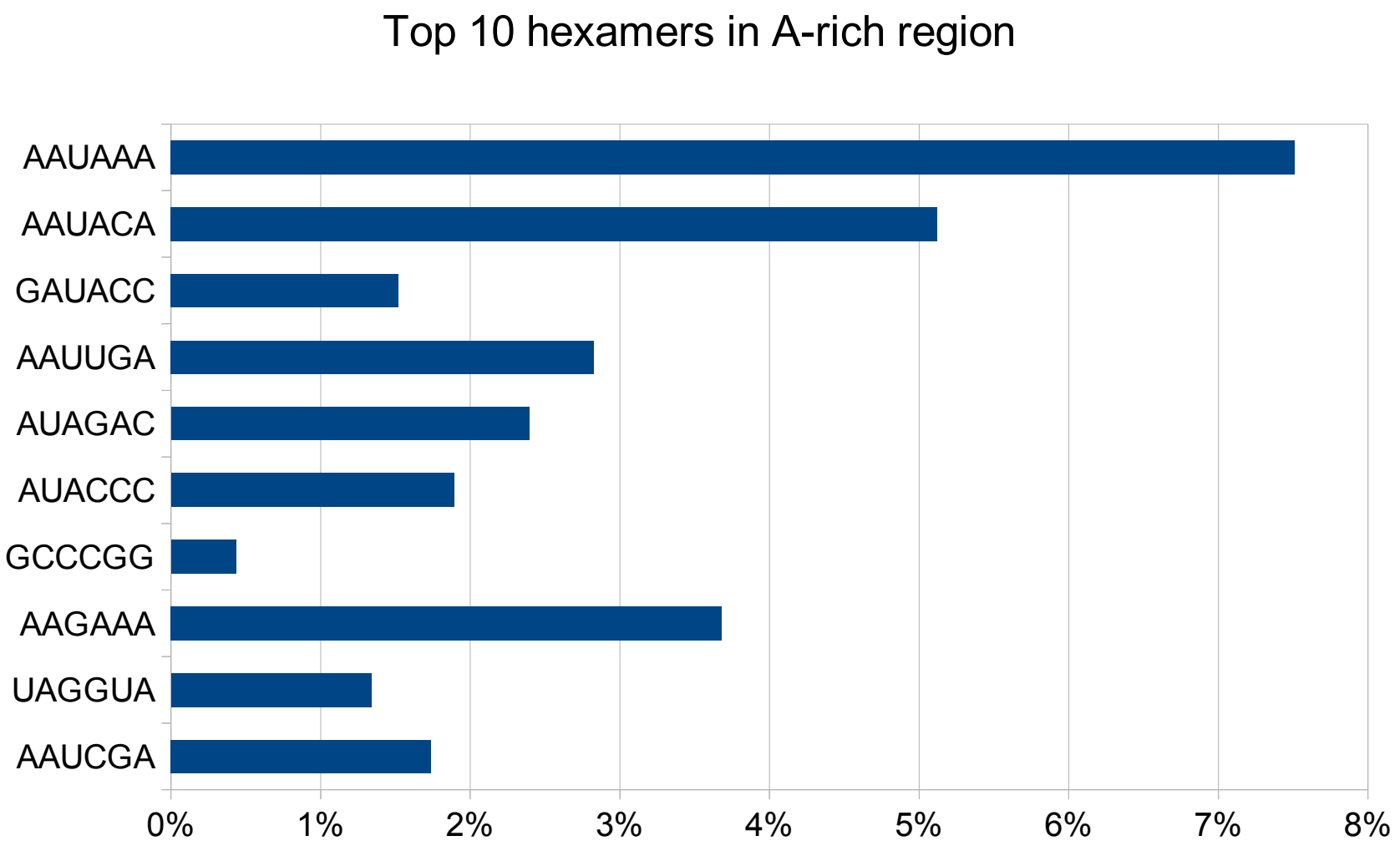
M. oryzae



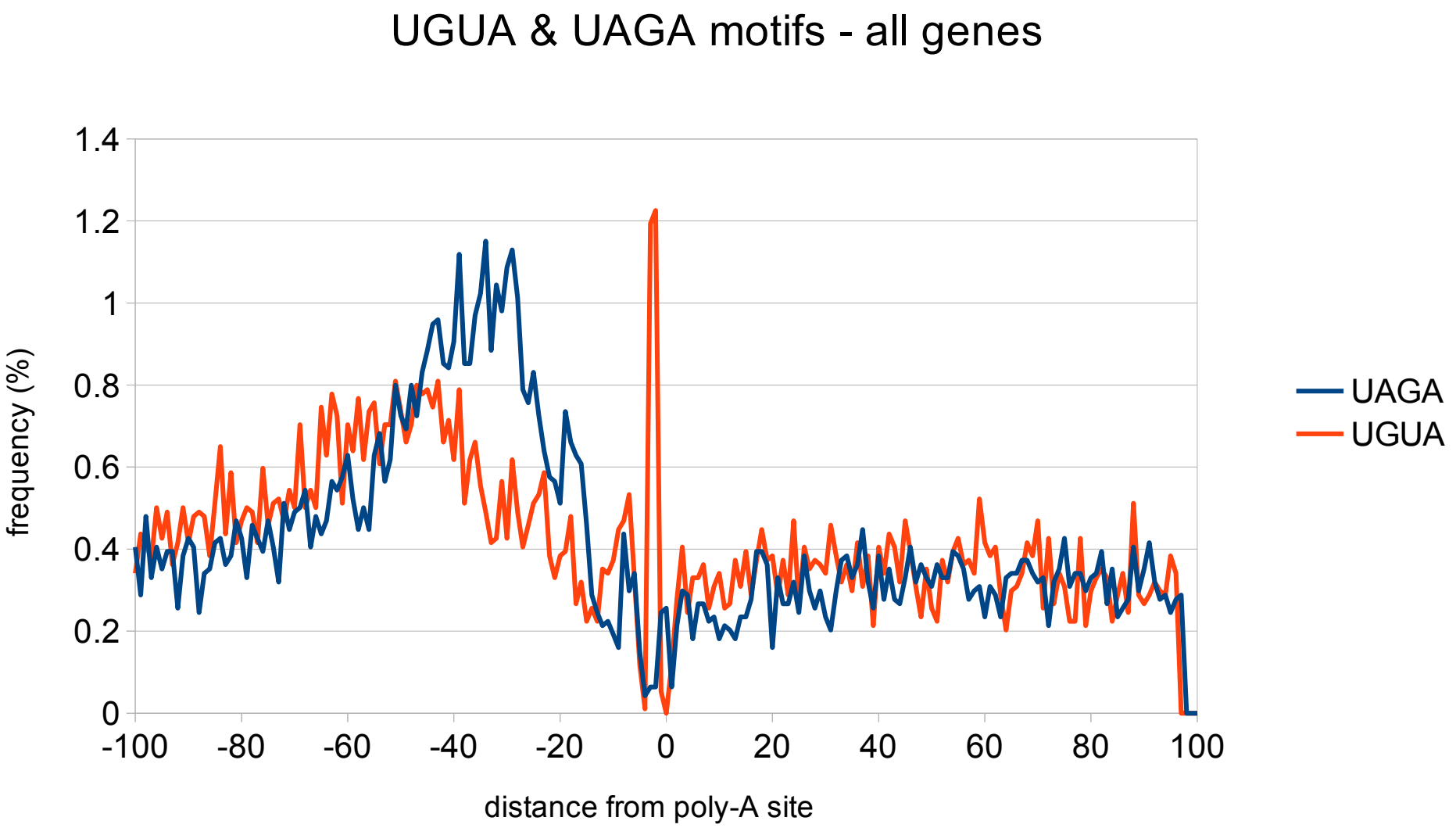
The A-RICH region is located -30 -10 bp upstream



Top 10 hexamers in A-rich region



UAGA & UGUA motifs



Polyadenilation signals in common genes

MPG1

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

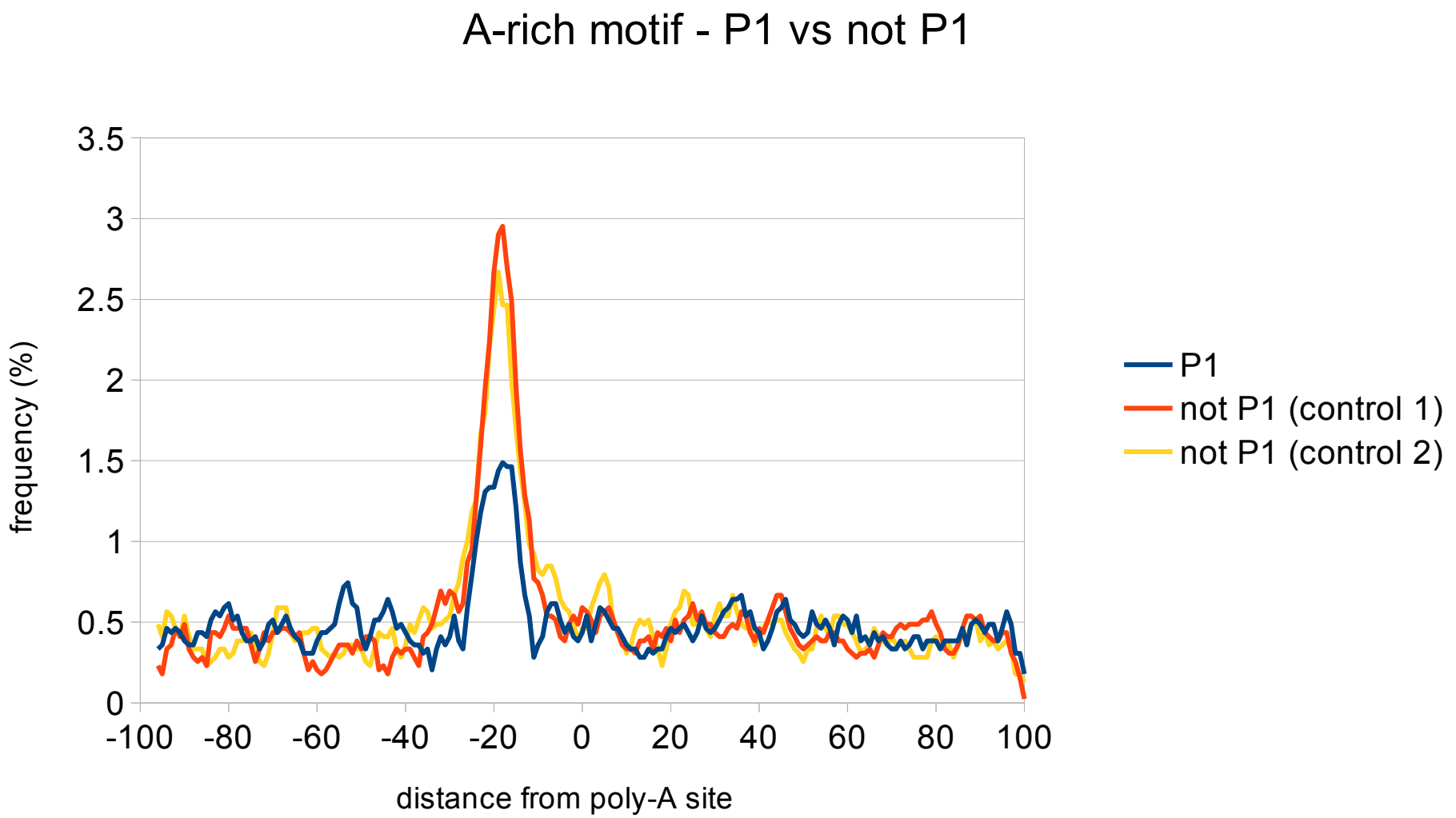
PMK1

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

MST12

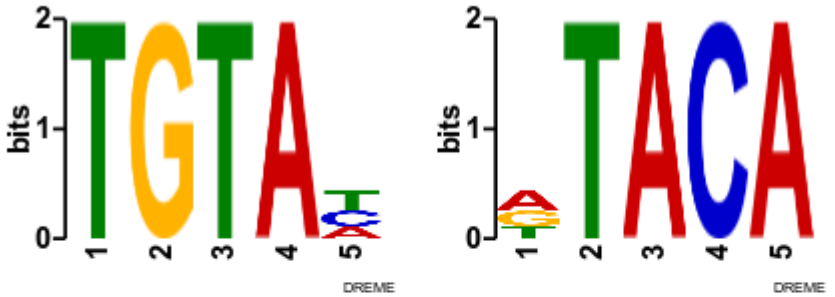
...CAGUGGCAUAAAAUCACAAAAUCUU**UAGA**AAGAUCAC**AGAAAA**CCUUUUGUC**CA**...

poly(A) sites targeted by *RBP35* show a less pronounced A-rich region



UGUAH motif is enriched in poly(A) sites targeted by *RBP35*

1. TGTAH



Details

Positives	Negatives	P-value	E-value	Unerased E-value
506/779	279/779	4.5e-31	2.7e-26	2.7e-26

Enriched Matching Words

Word	Positives	Negatives	P-value	▼E-value
TGTAT	262/779	136/779	1.4e-13	8.0e-9
TGTAC	197/779	101/779	3.6e-10	2.1e-5
TGTAA	156/779	85/779	4.0e-7	2.4e-2
TGTAC	134/779	85/779	2.2e-4	1.3e+1

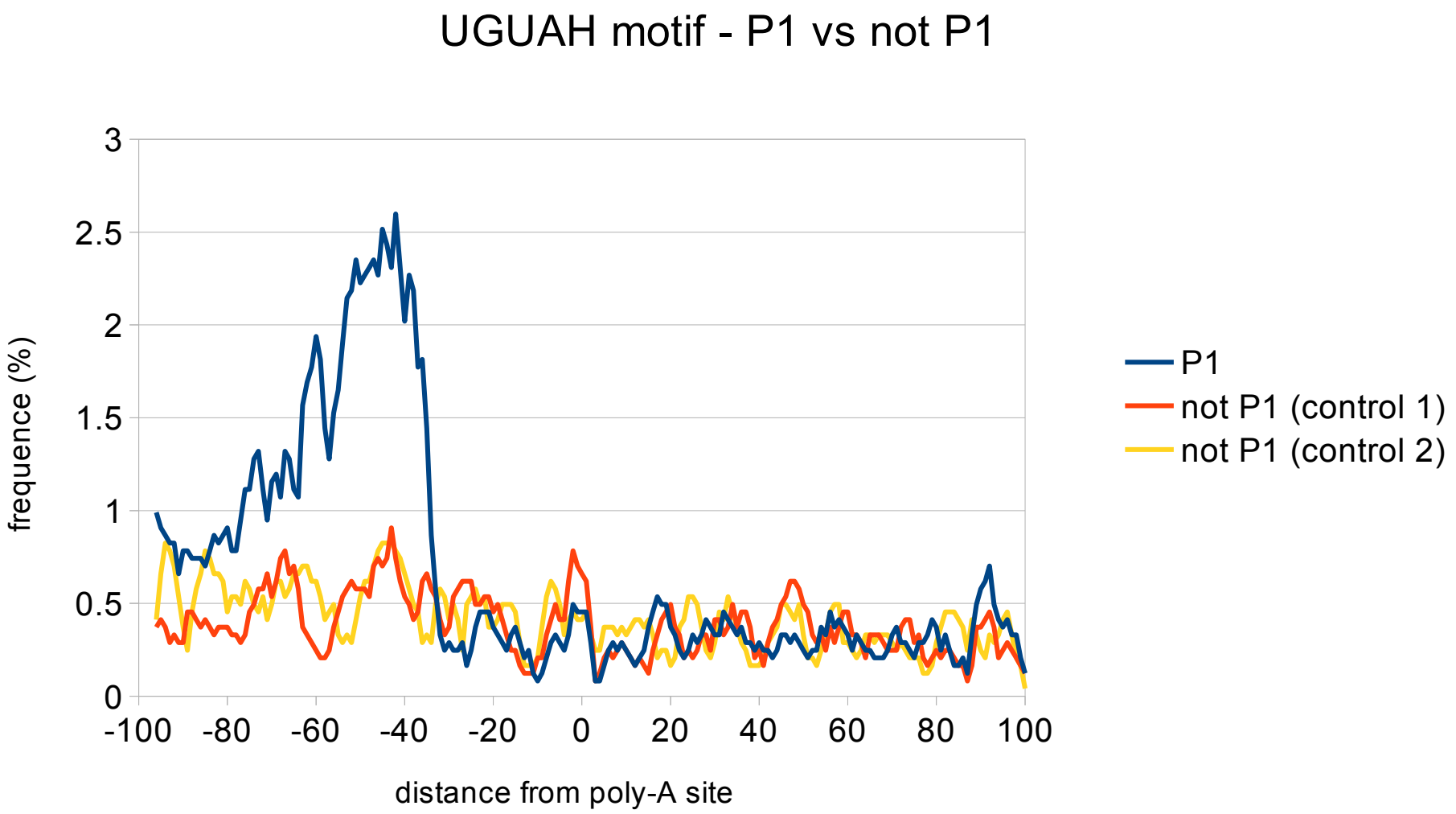
- Output of DREME, P1 as positive sequences list versus not-P1 negative list

UGUAH motif – P1 vs not P1

- In the first graph, we want to show how poly(A) sites targeted by *RBP35* display a different profile for the UGUA motif in the -100 -36 region respect to “regular” poly(A) sites

- We therefore plot P1 group poly(A) sites against two control groups of poly(A) sites of identical size not belonging to P1 group in WT

UGUAH is enriched at -45 in poly(A) sites targeted by *RBP35*

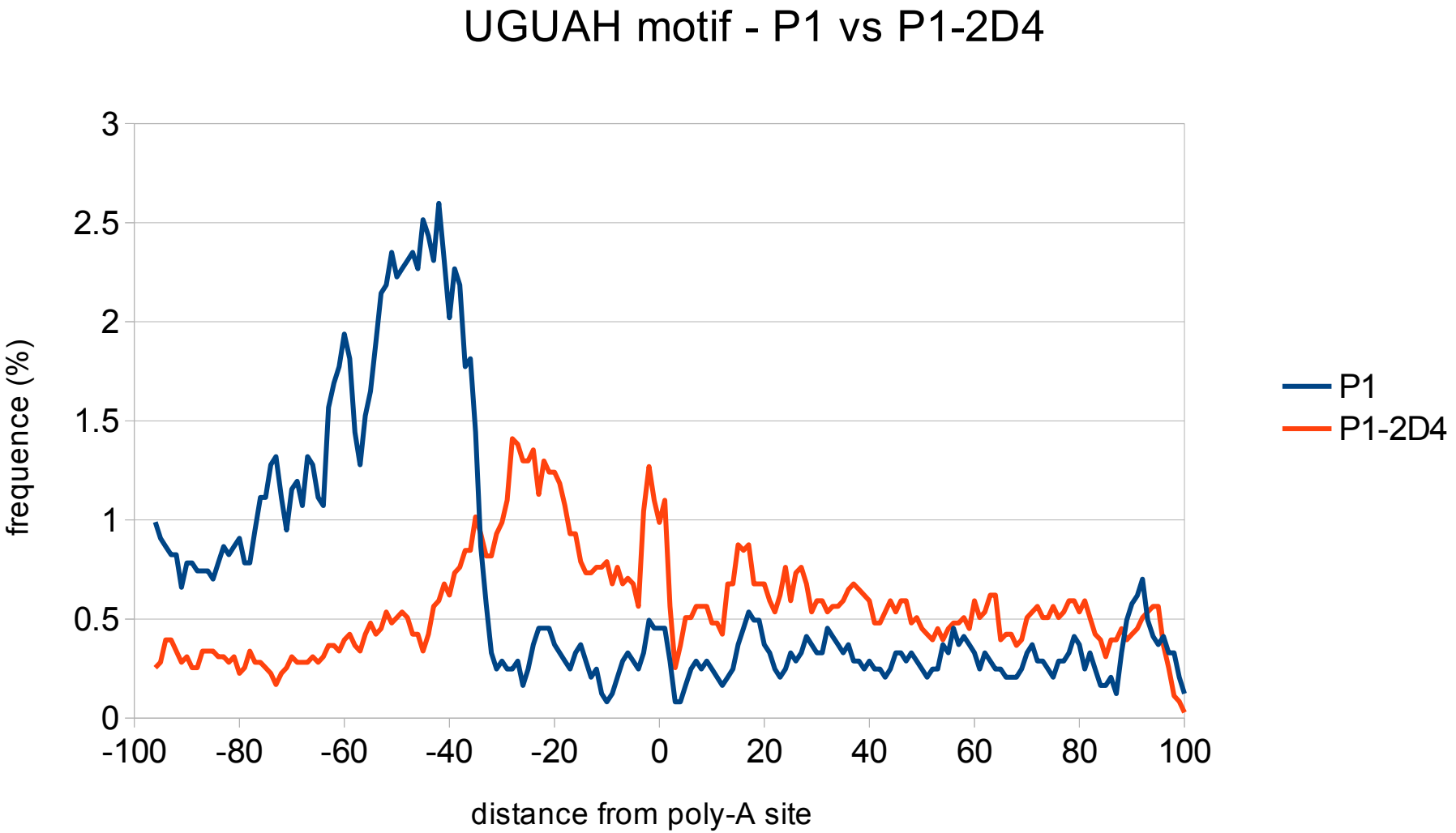


UGUAH motif – P1 vs P1-2D4

- In the second graph, we want to show how genes targeted by *RBP35* are affected in the mutant, using P1 group

- We therefore plot P1 group poly(A) sites against P1-2D4 group poly(A) sites, which are poly(A) sites from 2D4 in the same genes affected of P1 group

UGUAH – P1 vs P1-2D4

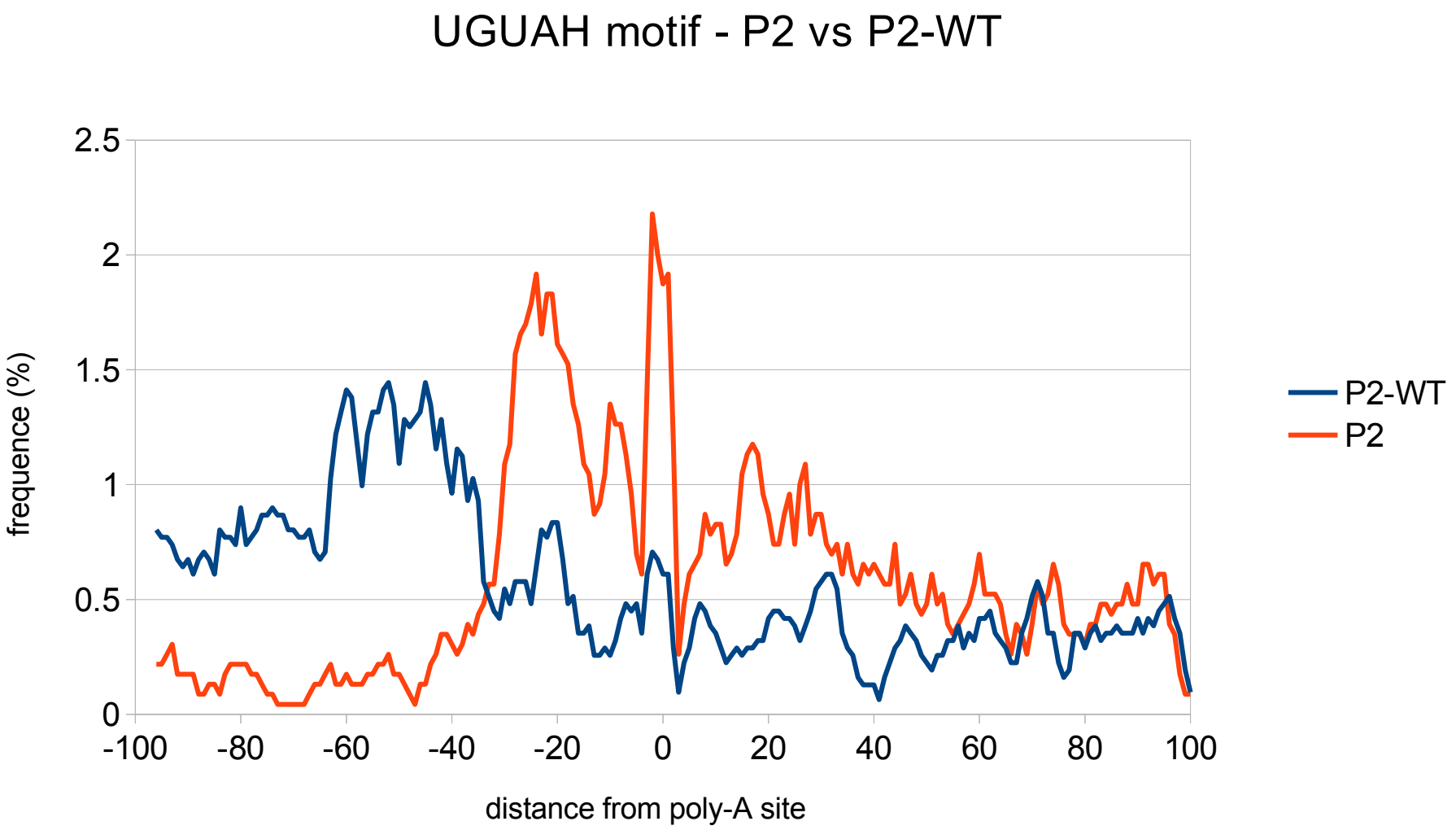


UGUAH motif – P2 vs P2-WT

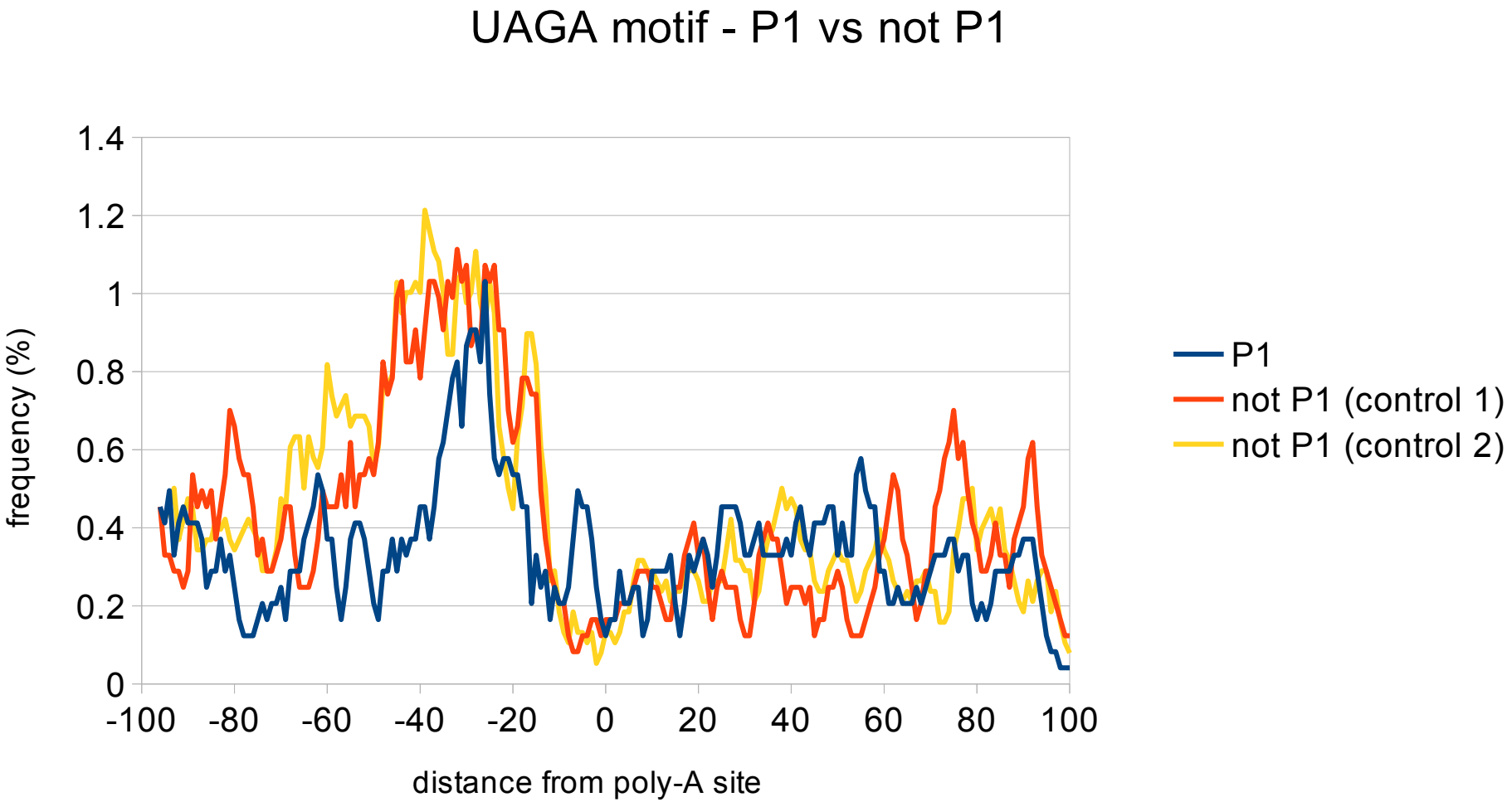
- In the second graph, we want to show how genes targeted by *RBP35* are affected in the mutant, using P2 group

- We therefore plot P2 group poly(A) sites against P2-WT group poly(A) sites, which are poly(A) sites from WT in the same genes affected of P2 group

UGUAH – P2 vs P2-WT

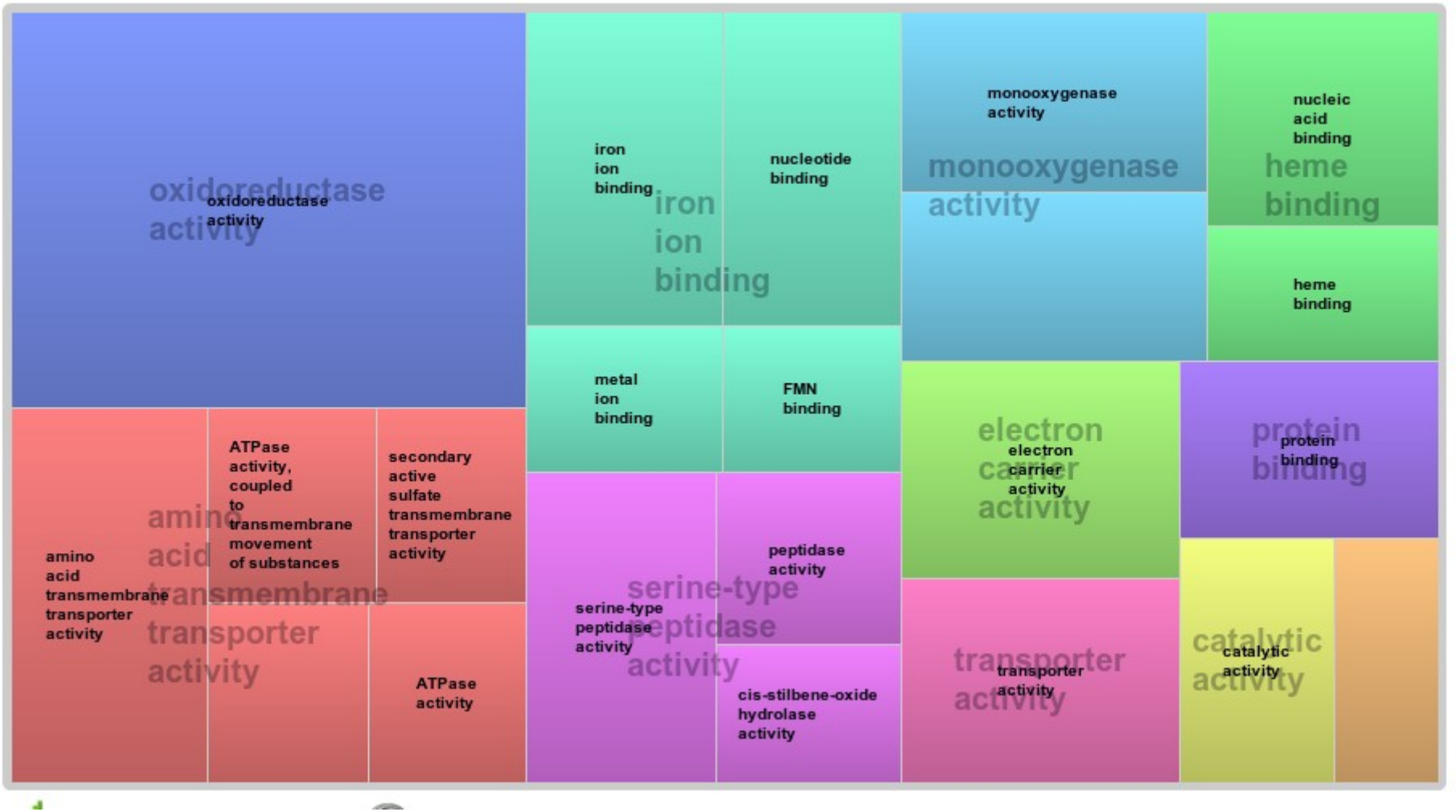


UAGA– P1 vs not P1



The gap in P1 set at -45 is probably due to UGUAH enrichment, as seen before

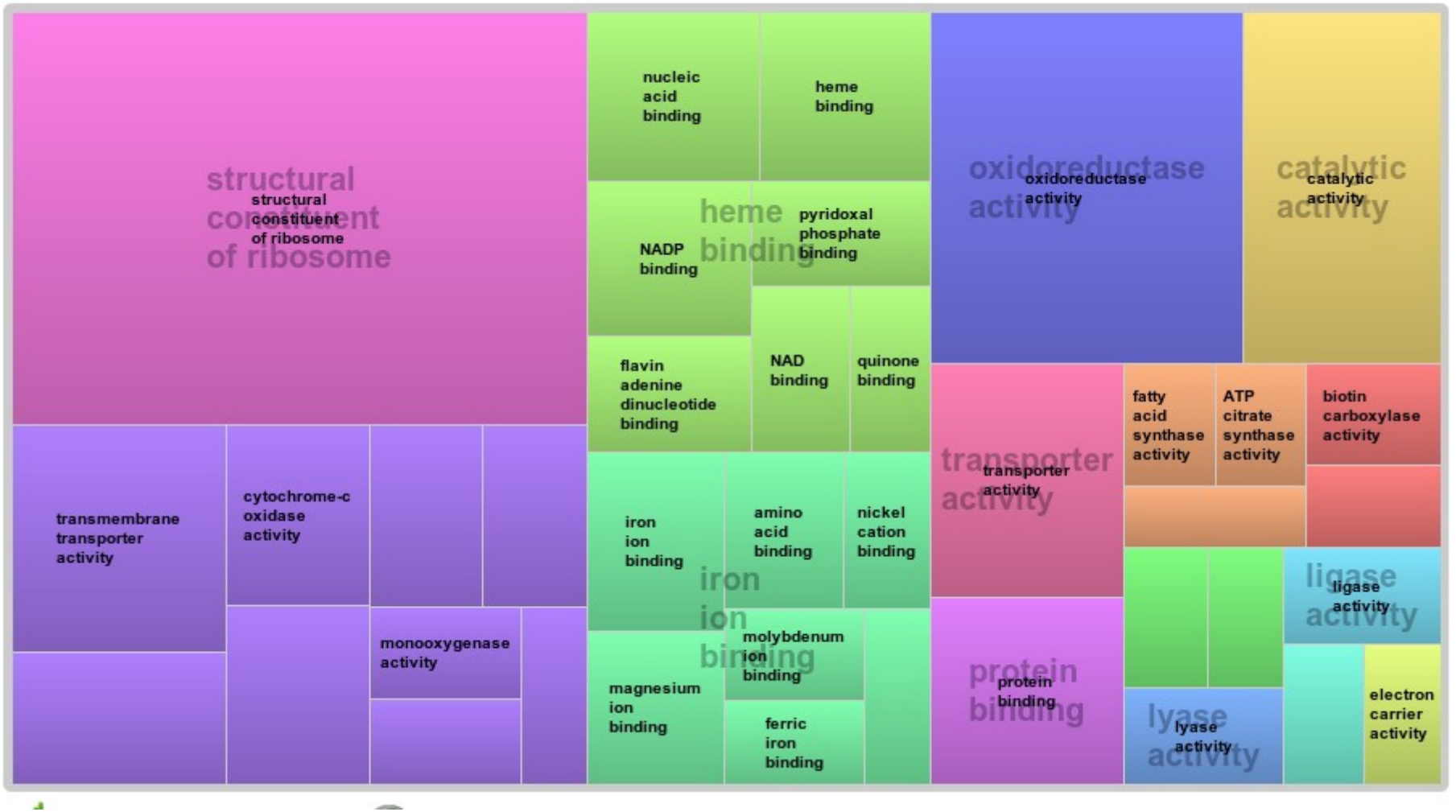
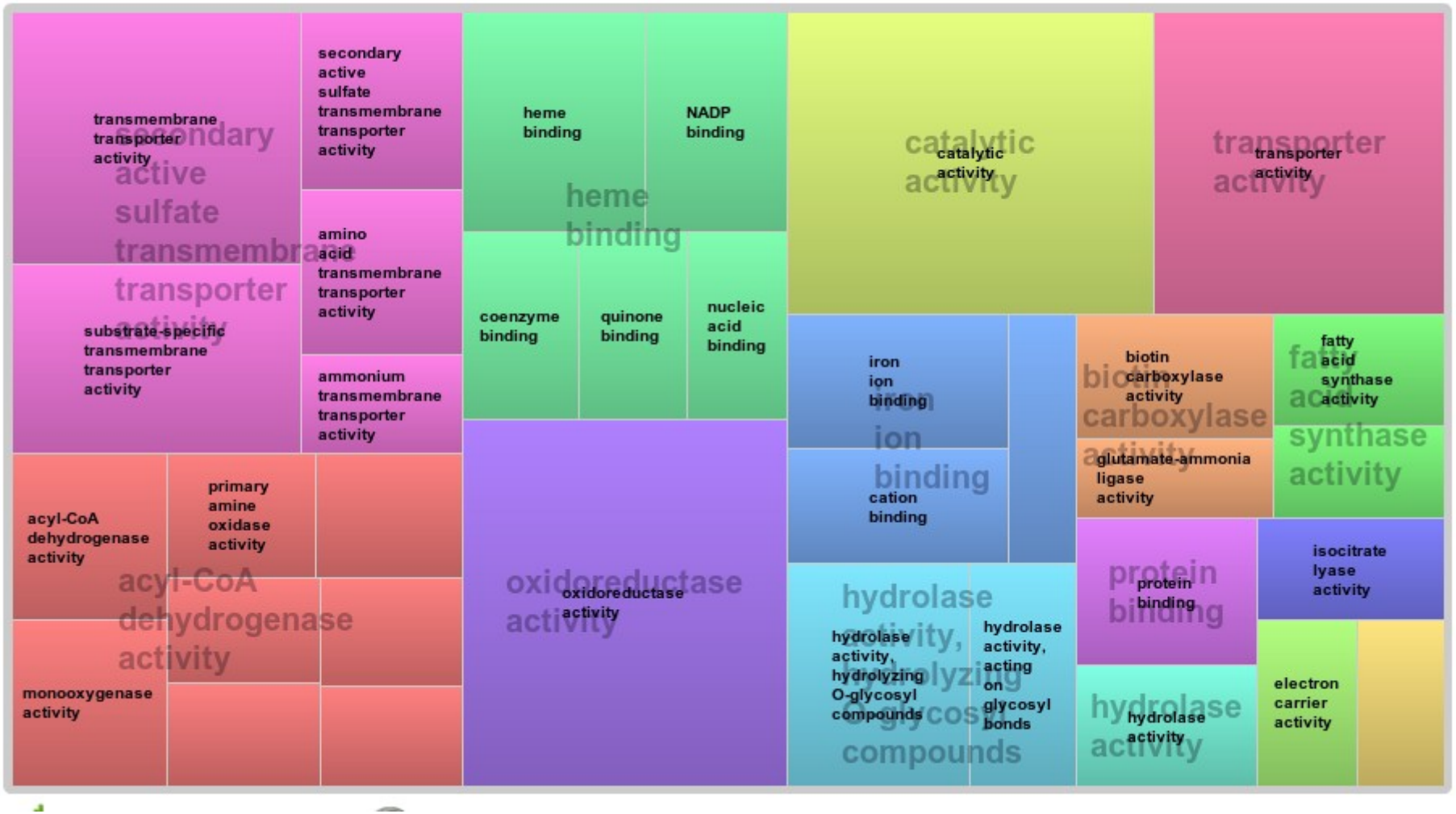
CM → -N functional analysis



CM → -C functional analysis

MM → -N functional analysis

MM → -C functional analysis



-N → -C functional analysis