

# Detection of transcriptional regulatory motifs specific to plant gene responses in stress conditions

Margot Correa

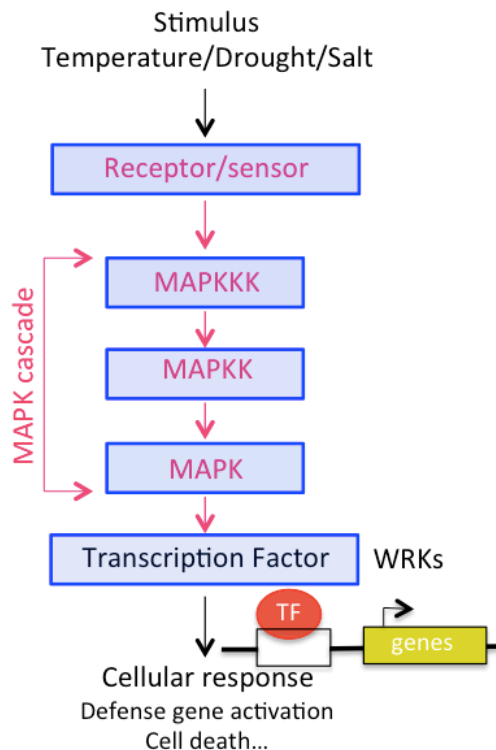
Team : Genomic Networks (IPS2) - Statistics and Genome (LaMME)

PEPI IBIS, Paris

6 June 2019

## Objective

- Plant triggered signaling pathways to adapt and survive to both biotic and abiotic stresses



- Objective of the study
  - Are there specific genes which could be regulated by one stress category ?
  - To compare genes involved in biotic and abiotic stress responses
  - Detect transcriptional regulatory motifs specific to one stress category
  - Characterization of specific motifs
  - Analyze GO term enrichments of stress response genes

# Transcription Factor Binding Sites (TFBS)

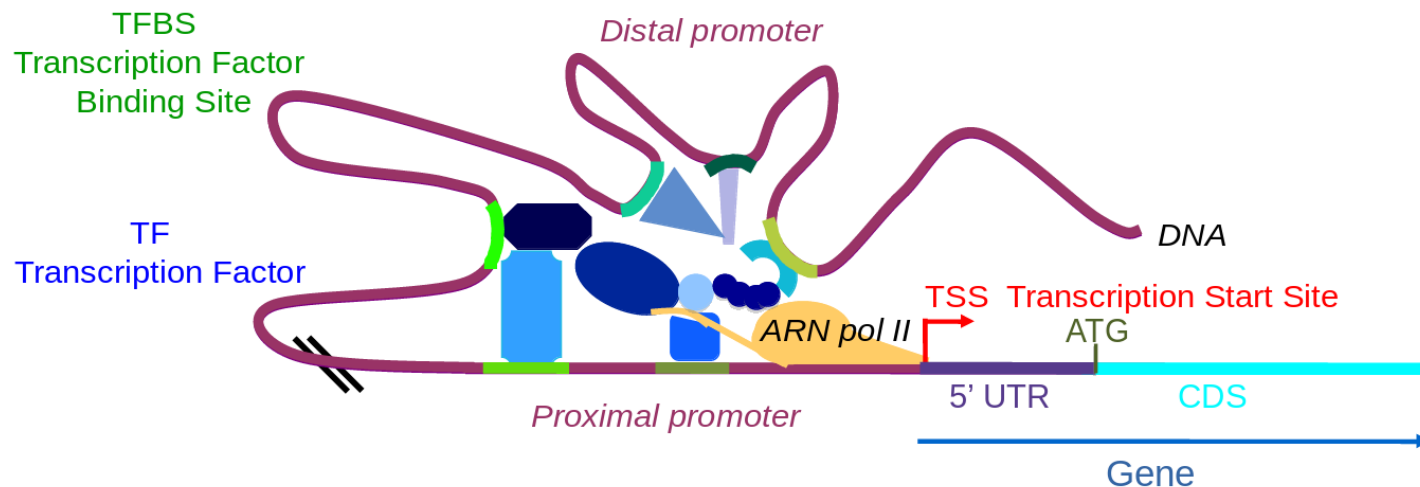
Transcriptional regulation is one of the fundamental processes controlling gene expression.

What Is a Transcription Factor?

- Bind to the promoter region upstream of gene and activate or inhibit gene transcription
- Activated in response to environmental stimuli

What Is a Transcription Factor Binding Sites?

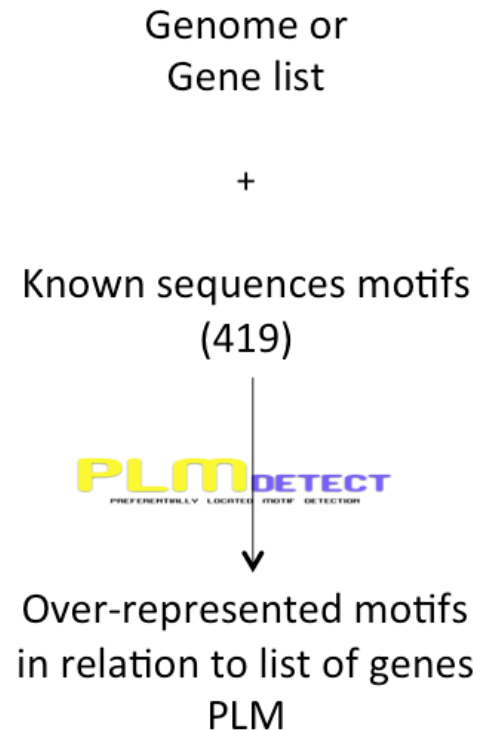
- Regulatory motifs involved in gene transcription
- TFBSs are short DNA sequences (5-15 pb) located in the promoter regions



## PLMdetect (PLM for Preferentially Located Motifs) V.Bernard et al., 2010

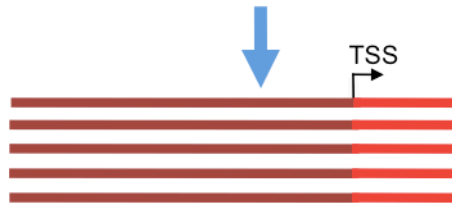
Detect regulatory motifs involved in the transcription with an over-representation in the proximal promoter region. TSS has to be well defined.

What Is PLMdetect approach?



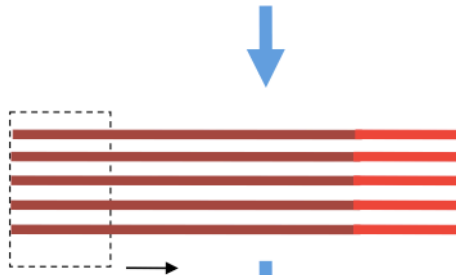
# PLMdetect (PLM for Preferentially Located Motifs) V. Bernard et al., 2010

Extract promoters sequences

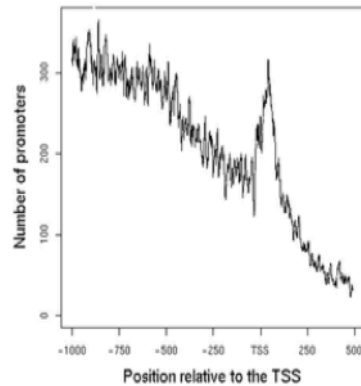


- ① Alignment of promoters  
[-1000,+500] relative to the TSS

For each motif,



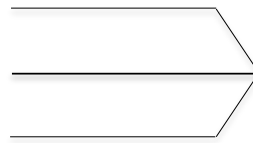
- ② Count the number of occurrence of  
the motif along sequence



- ③ Get motif distribution  
Peak detection ?  
**Preferentially Located Motif (PLM)**

# PLMdetect (PLM for Preferentially Located Motifs)

AGRISdb Arabidopsis (90)  
PLACEdb Plant database (400)  
V. Bernard et al., 2010 (15)

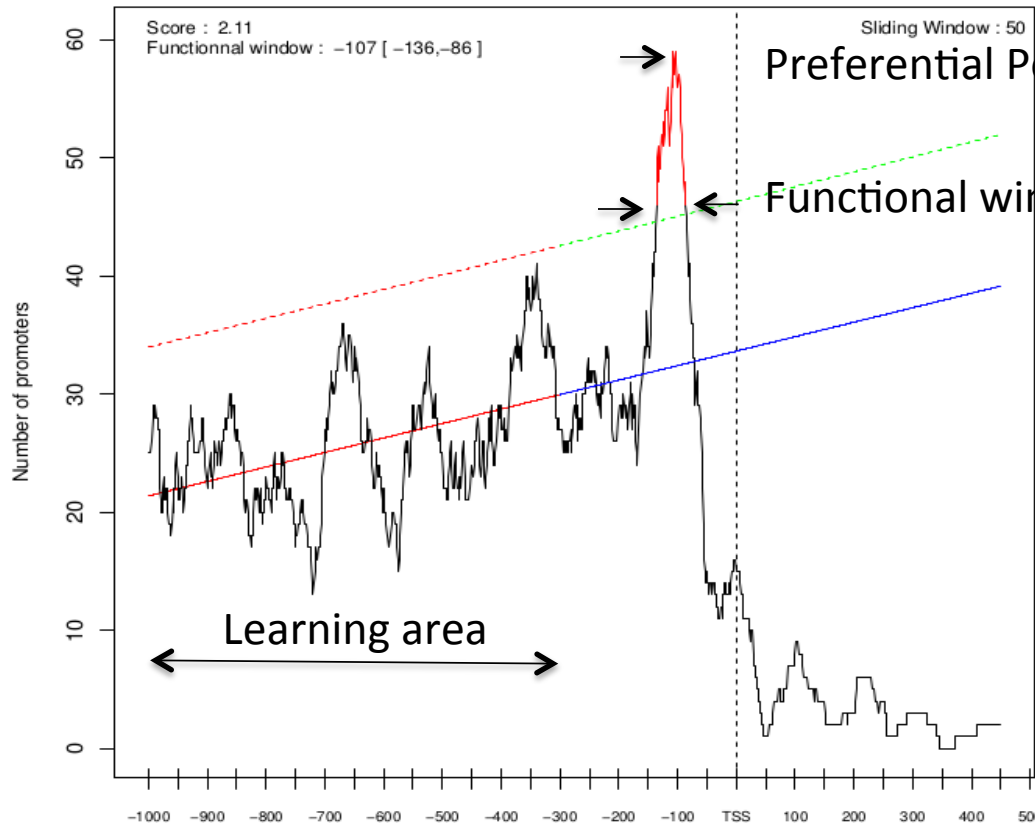


Remove

Size too long (length >16 pb)  
Too degenerate (N, DVHB, YRWSKM)

PLMdb  
419 motifs

### Distribution of CAATTATTA



Gene list containing the motif in their promoter

# Transcriptomic data - Results



387 experimental comparisons in *Arabidopsis Thaliana* stress conditions



18 stress categories

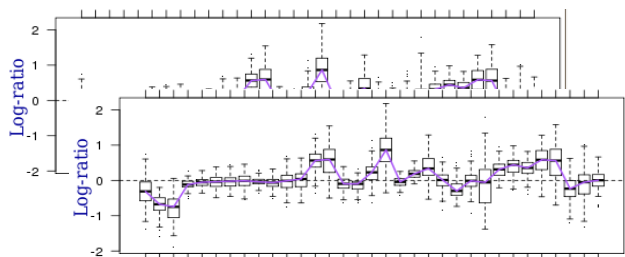


17264 genes differentially expressed

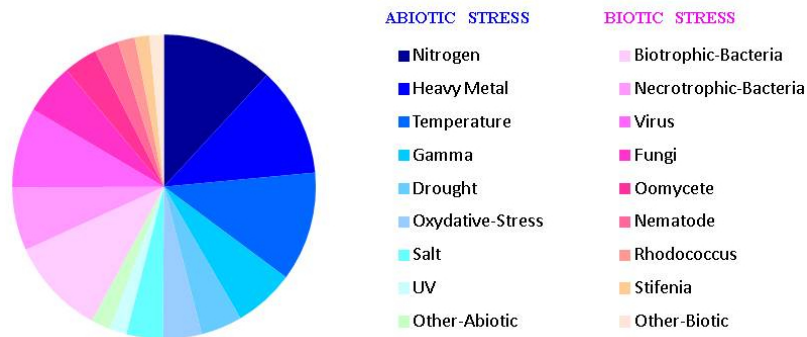


Clustering of DEGs (MFDR control)

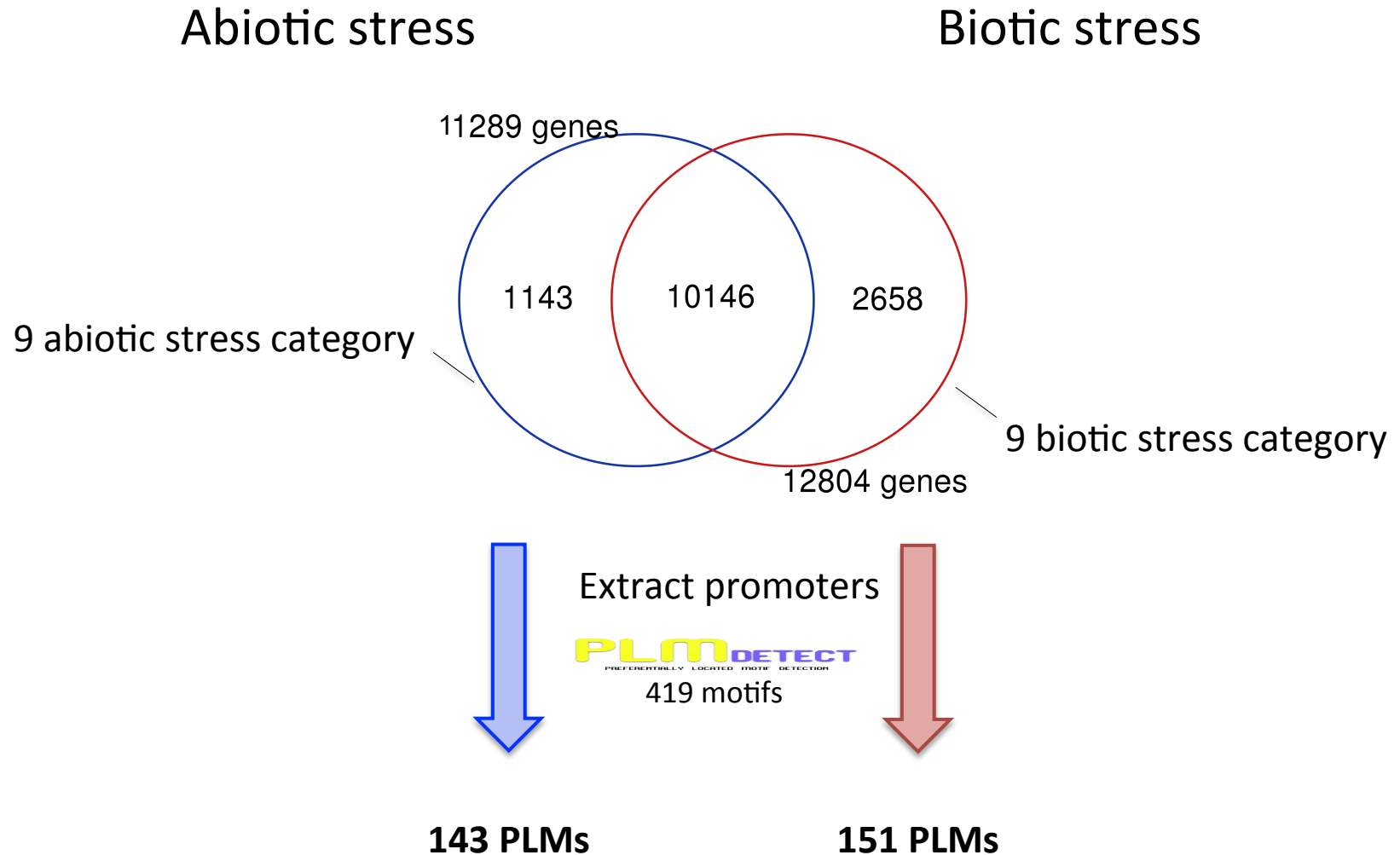
13947 genes in 681 co-expression clusters



<http://tools.ips2.u-psud.fr/CATdb>



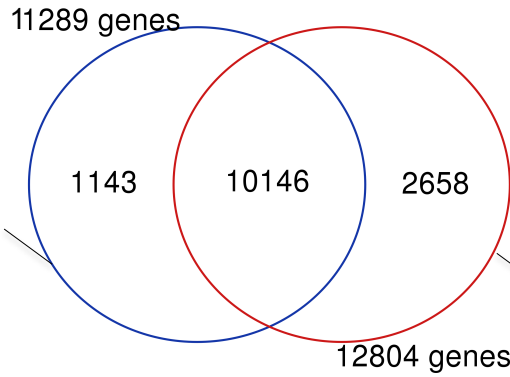
Aim : Detection of transcriptional regulatory motifs specific to *At* gene responses in stress conditions





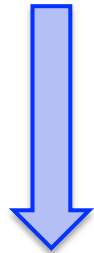
# Abiotic stress

# Biotic stress



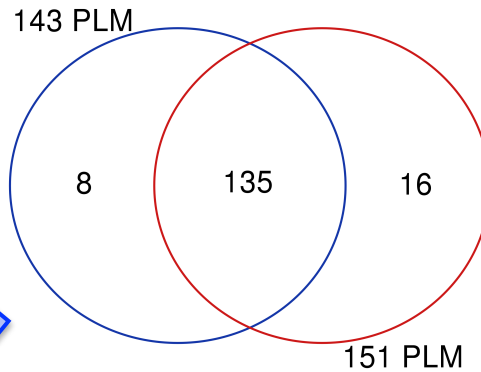
9 abiotic stress category

9 biotic stress category



**PLM**DETECT  
PREFERENTIALLY LOCATED MOTIF DETECTION

419 motifs



**8 specific PLMs**

**16 specific PLMs**



989 genes

5870 genes

PLM characterization

Identify associated TFs

GO terms enrichment

## Results - Motifs characterization sequences

16 specific PLMs to abiotic stress, 8 specific PLMs to biotic stress

➔ Potential TFBS

Motif seq.	Nb of genes	Motif seq.	Nb of genes
CANNTG	3254	AACGTT	254
AMCWAMC	3026	TATTAA	251
AAATTAAA	2072	GAAAAA	131
ACNGCT	1777	CAATNATTG	87
CTGACY	1091	MTCCWACC	39
CCGTCC	283	TGACGTAA	31
AACAAAC	261	CACGCAAT	17
CAAACAC	244	TACCGACAT	11
CAAT	221		
ACCGAGA	102		
GTCGAC	92		
TTTCCCGC	74		
GCGGGAAA	72		
CAATTATTA	52		
ATACGTGT	48		
CATCCAACG	16		

M{A,C} W{A,T} Y{C,T}

ACCGAGA / TACCGACAT

DRE promoter motif

CAATTATTA / CAATNATTG

ATHB6 et ATHB5 exhibit identical DNA binding specificities

## Results - Motifs characterization via PLMdb

### 8 specific PLMs to biotic stress

Motif seq.	Description
AACGTT	T-box; one of the ACGT elements
TATTAA	RNA polymerase binding motif in eukaryotes
GAAAAA	GT-1 motif found in the promoter of soybean
CAATNATTG	Consensus binding sequence for At, class I HDzip protein
MTCCWACC	Myb1 binding site motif
TGACGTAA	TGA-box#1 in putative auxin-responsive element (AUXRE) of Soybean GH3 promoter
CACGCAAT	Sequence found in Soybean GH3 gene promoter
TACCGACAT	Binding site of dreb1 and dreb2; DRE promoter motif

### 16 specific PLMs to abiotic stress

Motif seq.	Description
CANNTG	Myc consensus
AMCWAMC	MYB4 binding site motif
AAATTTAA	-
ACNGCT	VRE found in some stress genes (e.g. TRX8 and MYB44)
CTGACY	W box
CCGTCC	Box A
AACAAAC	Core of AACA motifs found in rice glutelin genes
CAAACAC	Conserved in many storage-protein gene promoters
CAAT	Promoter consensus sequence found in legA gene of pea
ACCGAGA	DRE1 core found in maize rab17 gene promoter
GTCGAC	Core CRT/DRE motif
TTTCCCGC	E2Fa element found in the At CDC6 gene promoter
GCGGGAAA	Binding site of OsE2F1 and OsE2F2
CAATTATTA	Consensus binding sequence for At HDzip protein, ATHB6
ATACGTGT	Sequence found in the At chlorophyll a/b binding protein gene promoter
CATCCAACG	Nonamer motif found in promoter of wheat histone genes H3 and H4

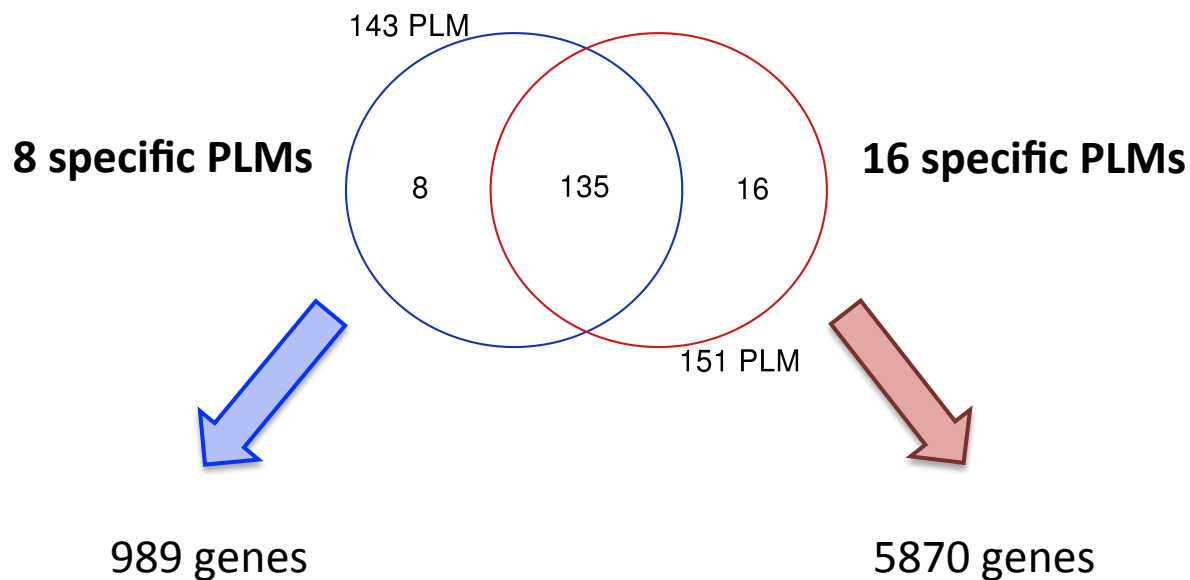
- ⇒ Motifs are binding site of known TFs (wrk, Myb, DRE) and involved in stress response
- ⇒ According to literature, some TFs are involved in abiotic stress response, but there TFs are found on both biotic and abiotic stress response (Myc, Myb, DRE)
- ⇒ Our information : specific motifs located in a preferred position

## Results - GO term enrichments

### Target genes characterization

GO term enrichments analysis of biotic and abiotic stress response genes

Is there an over-representation of certain functions for stress-specific genes compared to *Arabidopsis* genes?



topGO

## Results – GO term enrichments

topGO

GO term enrichment of biotic and abiotic stress response genes compared to all *Arabidopsis* genes - Fisher exact with P-value P-value < 0.01

Reference : 25855 *Arabidopsis* genes

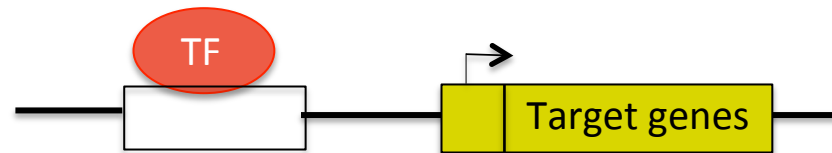
	989 genes stress biotic	5870 genes stress abiotic
CC	Golgi, cytoplasm, phosphatase complex	Cytoplasm, cell
BP	Cellular process, development process and biological regulation	Response to chemical, response to abiotic stimulus, response to oxygen-containing compound, response to stimulus
MF	Anion binding, protein kinase activity, carbohydrate derivative binding,	Catalytic activity, omega peptidase activity, peptide receptor activity and oxidoreductase activity

⇒ The results obtained are consistent between the 3 ontology (CC,BP,MF)

⇒ The biotic stress response genes are enriched in protein kinase activity, that are enzymes early implied in process of transduction signal in biotic stress responses

## Summary and next to come

- We identified 8 specific motifs to biotic stress and 16 specific motifs to abiotic stress as potential TFBS of each category
- Our results revealed these specific motifs are binding site of known TFs



- Which are TFs that bind on TFBS?  
DAPseqdb (DNA affinity purification sequencing),  
PBMdb (Protein Binding Microarray)

Thanks to Genomic Networks team

And thank for your attention!

# Counting motifs

Automatic identification :

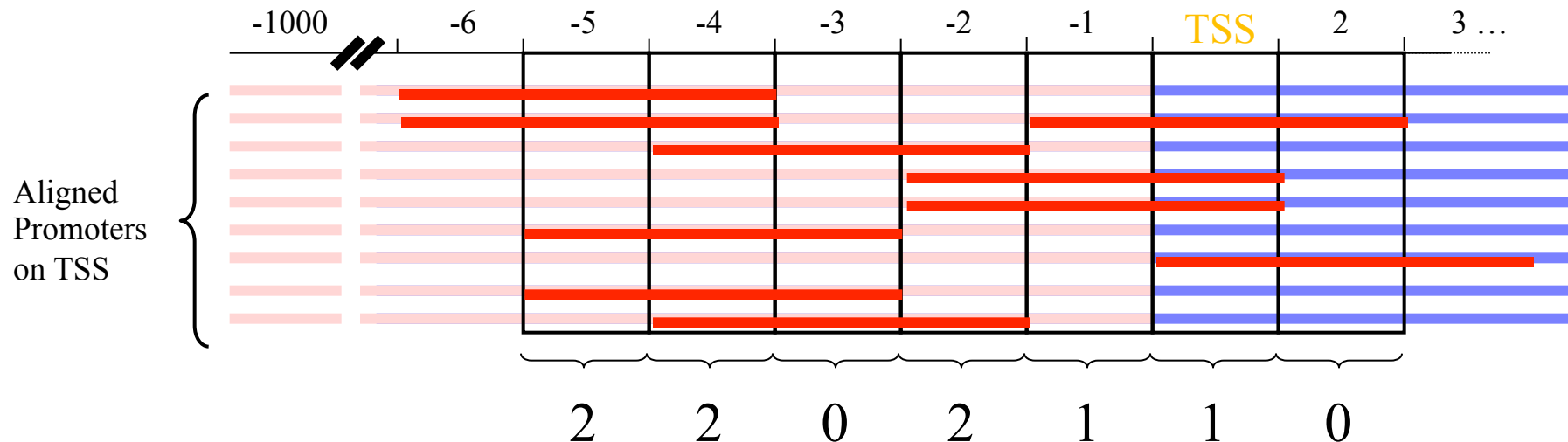
interval [-1000, ATG]

Sliding windows of N bases through the interval

At each positions, count number of promoters by start position

Motif size=3 bases

Sliding window size= 1 base





## Results - TFs bind to TFBSs

Which are the TFs that bind on its TFBSs? Are they known?

- DAP-Seq (DNA affinity purification sequencing) : uses in-vitro expression of TFs to identify their BS. ampDAP-Seq, modified version of DAP-Seq, remove methylation known to affect TFs binding on TFBS (76% of *Arabidopsis* TFs are sensitive to methylation in their motifs)

DAPseqdb - release may 2016

529 TFs (30% of *Arabidopsis* TFs)

- PBM

PBM (Protein Binding Microarray) : DNA microarray technology to identify TFBSs. UniPROBE (Universal PBM Resource for Oligonucleotide Binding Evolution) database contains data generated by PBM

PBMdb -

709 non-redundant proteins and complexes from a diverse organisms motifs are given in the form of logos