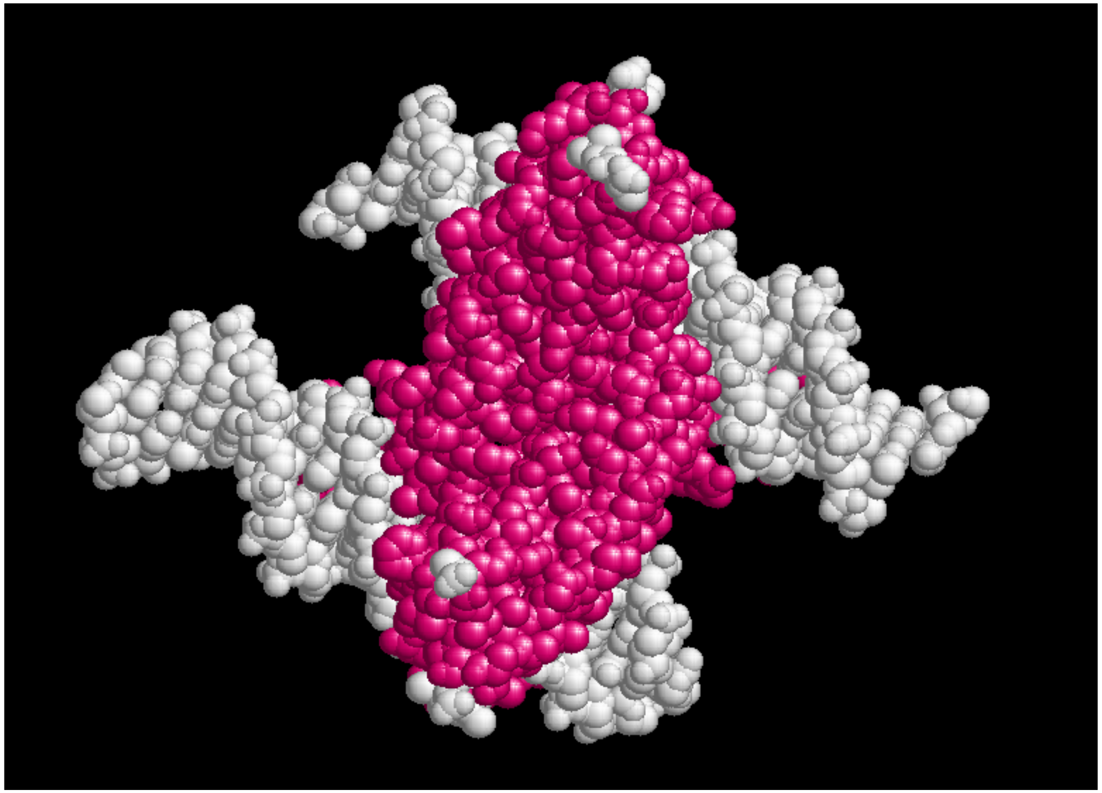
In silico promoter analysis



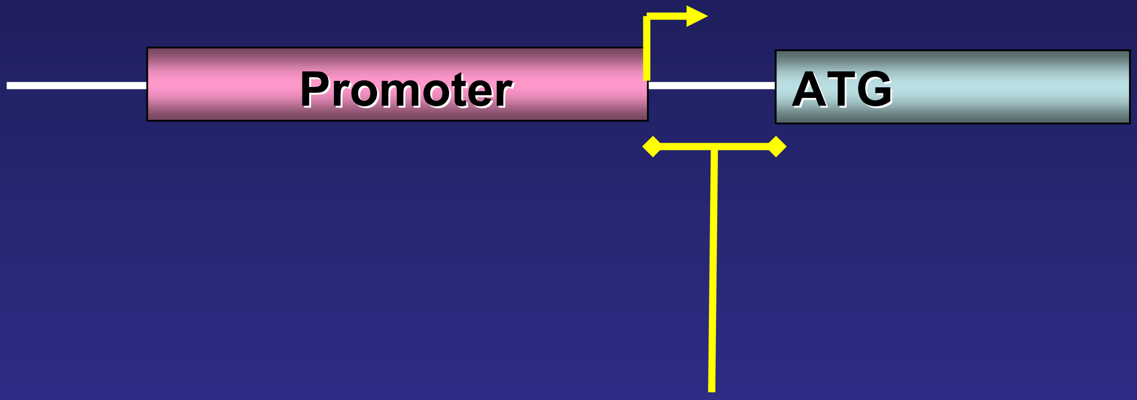
Edward J. Oakeley

Alessandro DiCara, Karsten Schmidt

Summary

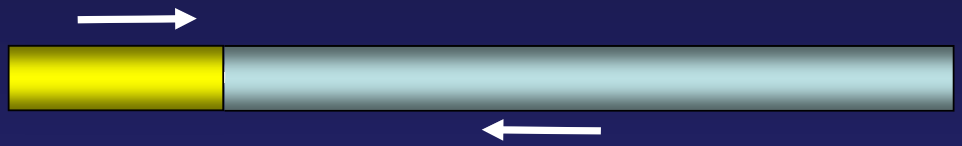
- Where should I look for a promoter?
- Simplifying the analysis problem
- Which transcription factors are expressed?
- Do they dance alone or with friends?
- **Visualisation of the data**

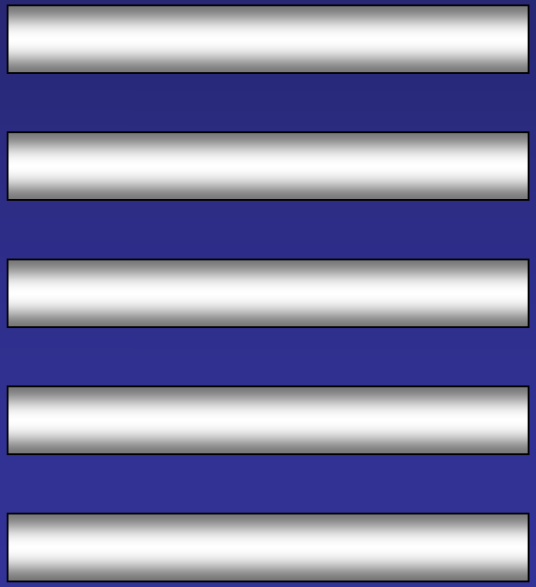
Where should I look?



50 bp?

Oligo capping





26,480 human starts mapped (Genomatix, 2002)

Map start of transcription

Transcription factors

Transfac Pro 7.4.1

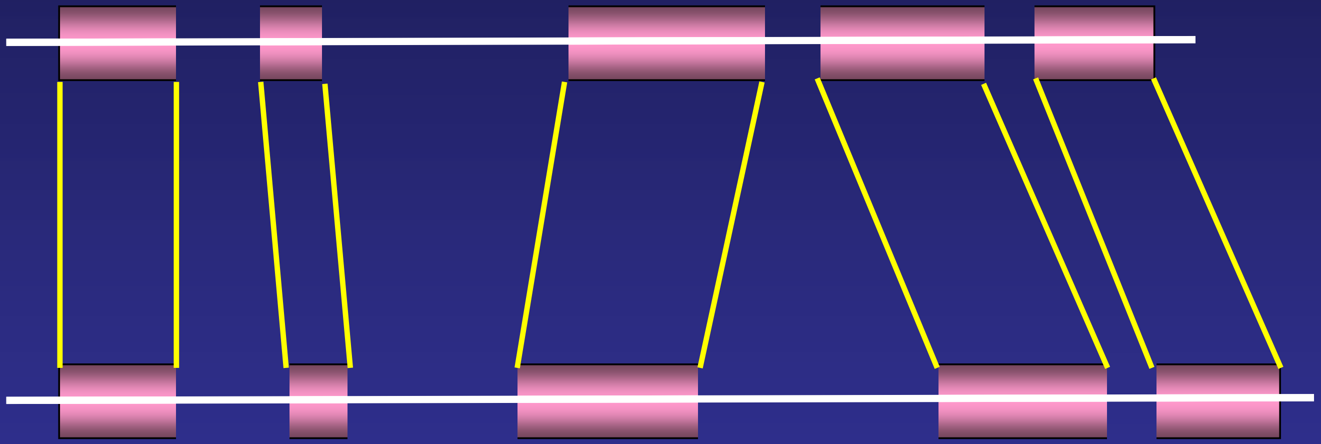
- Minimise false positives 0.2/bp: 400 hits
- Balance false neg/pos 1/bp: 2000 hits
- Minimise false negatives 5/bp: **10000 hits**

FMI

DNA block alignment

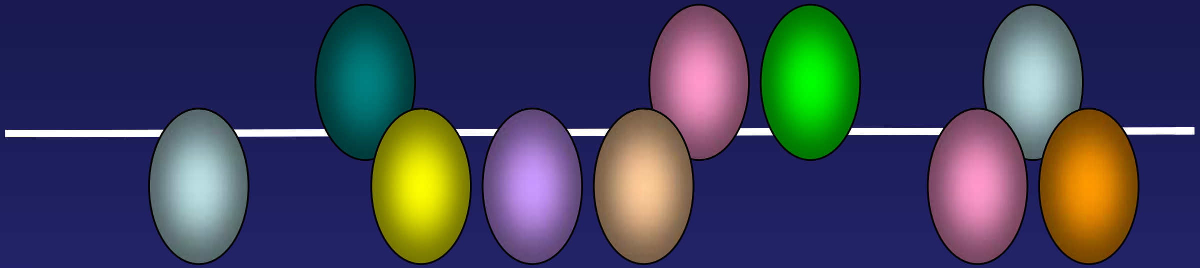
Ewan Birney, European Bioinformatics Institute

Human promoter (from Genomatix)

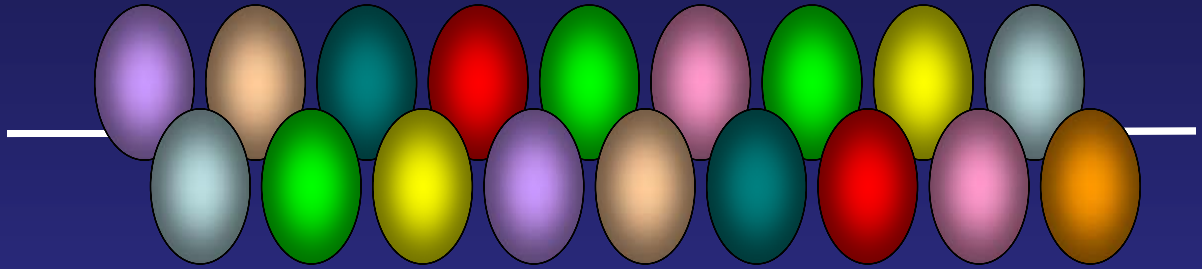


Mouse orthologue promoter
(also from Genomatix or from BAC clone)

Functional promoters



Remove TFs not expressed



Remove TFs by Affymetrix ID (e.g. 123456_at)

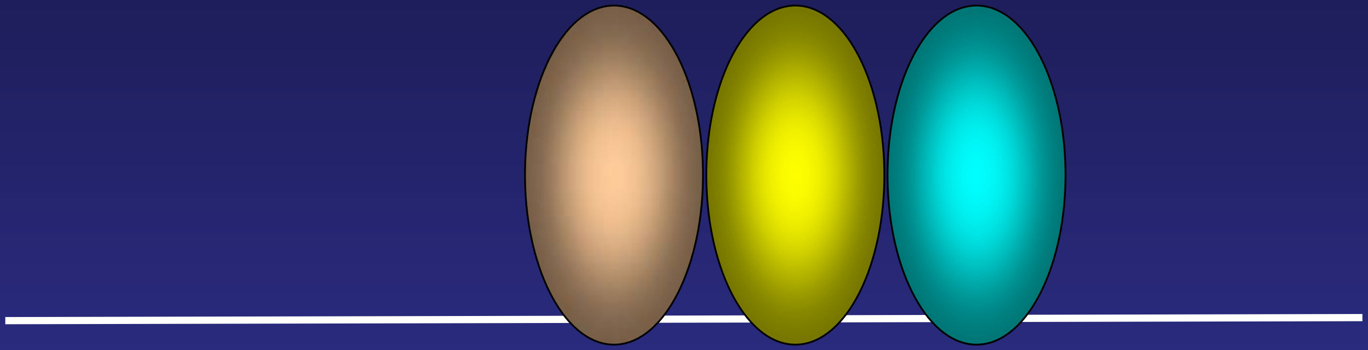
(usually those listed as Absent or Marginal)

Exclude specific Transfac matrices (e.g. V$OCT1_01)

Exclude specific TF names (e.g. OCT1)

Look for patterns

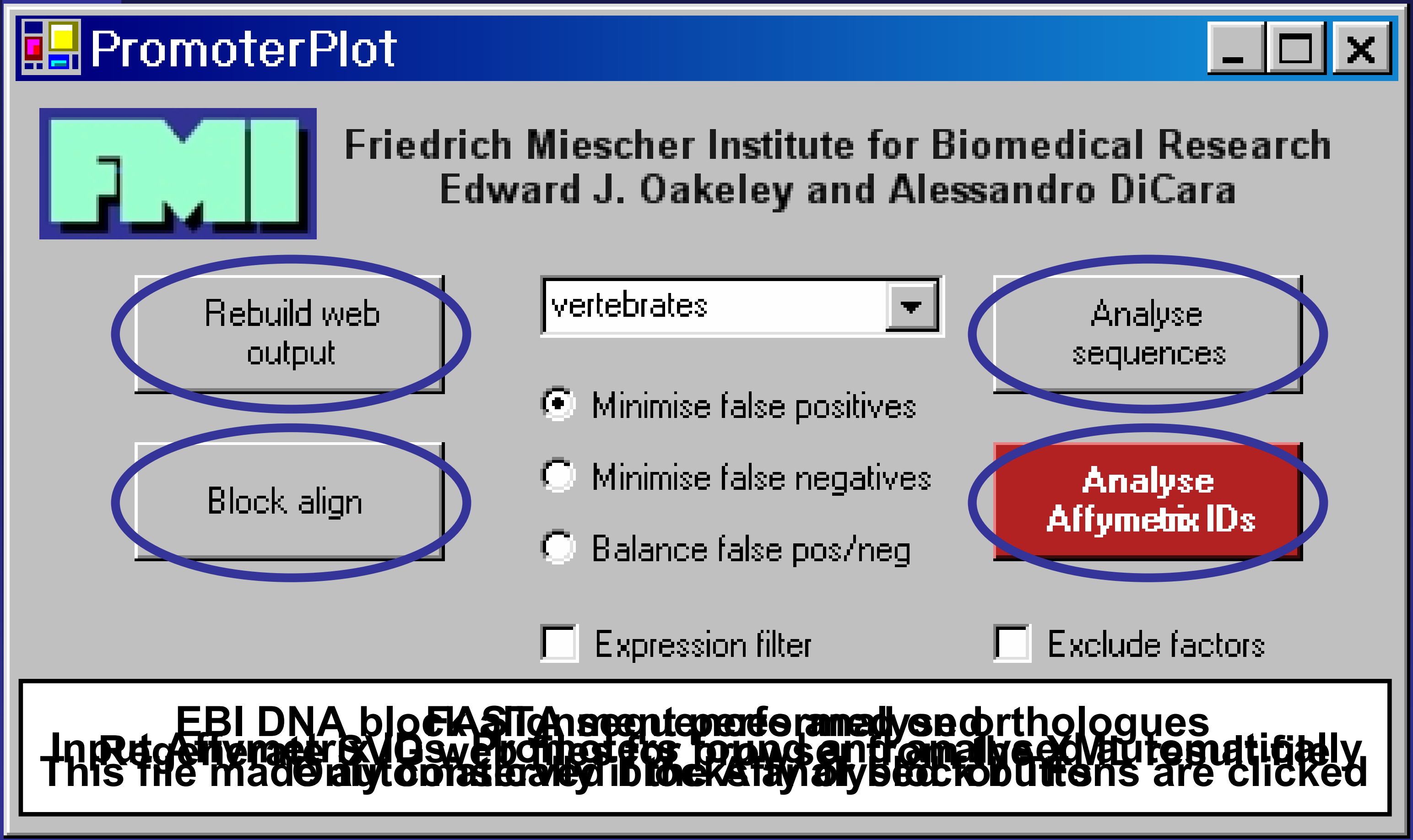
IBM Teiresias pattern discovery algorithm



Patterns are composed of three or more factors each >10bp apart and all expressed.

A pattern must occur in a user defined number of promoters (at least 2)

Automated solutions



Friedrich Miescher Institute for Biomedical Research
Edward J. Oakeley and Alessandro DiCara

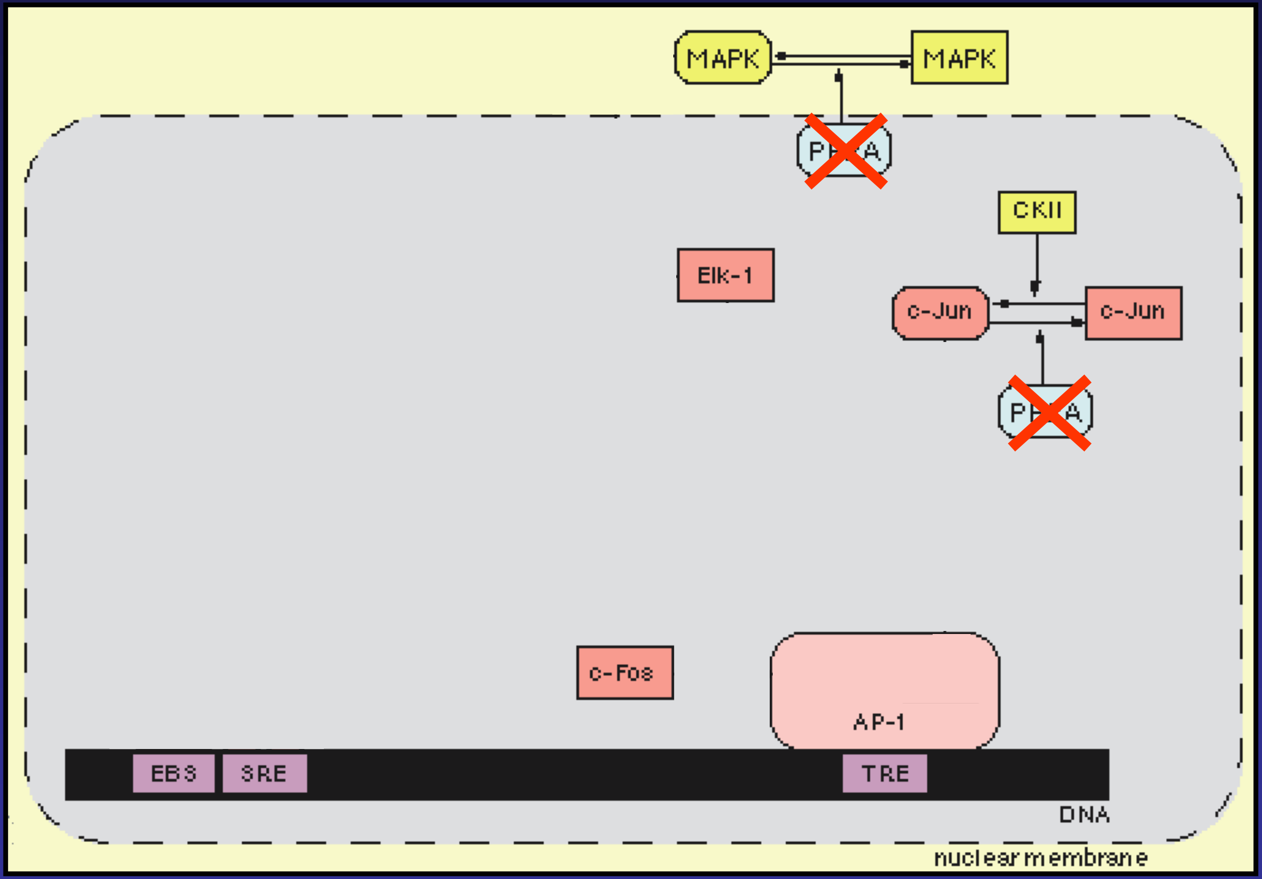
● Minimise false positives
○ Minimise false negatives
○ Balance false pos/neg

☐ Expression filter
☐ Exclude factors

EBI DNA block [illegible] FASTA [illegible] orthologues
Input Affymetrix IDs [illegible] automatically
Regenerate [illegible]
This file made [illegible] buttons are clicked
Only [illegible]

Real science

Inhibit PP2A

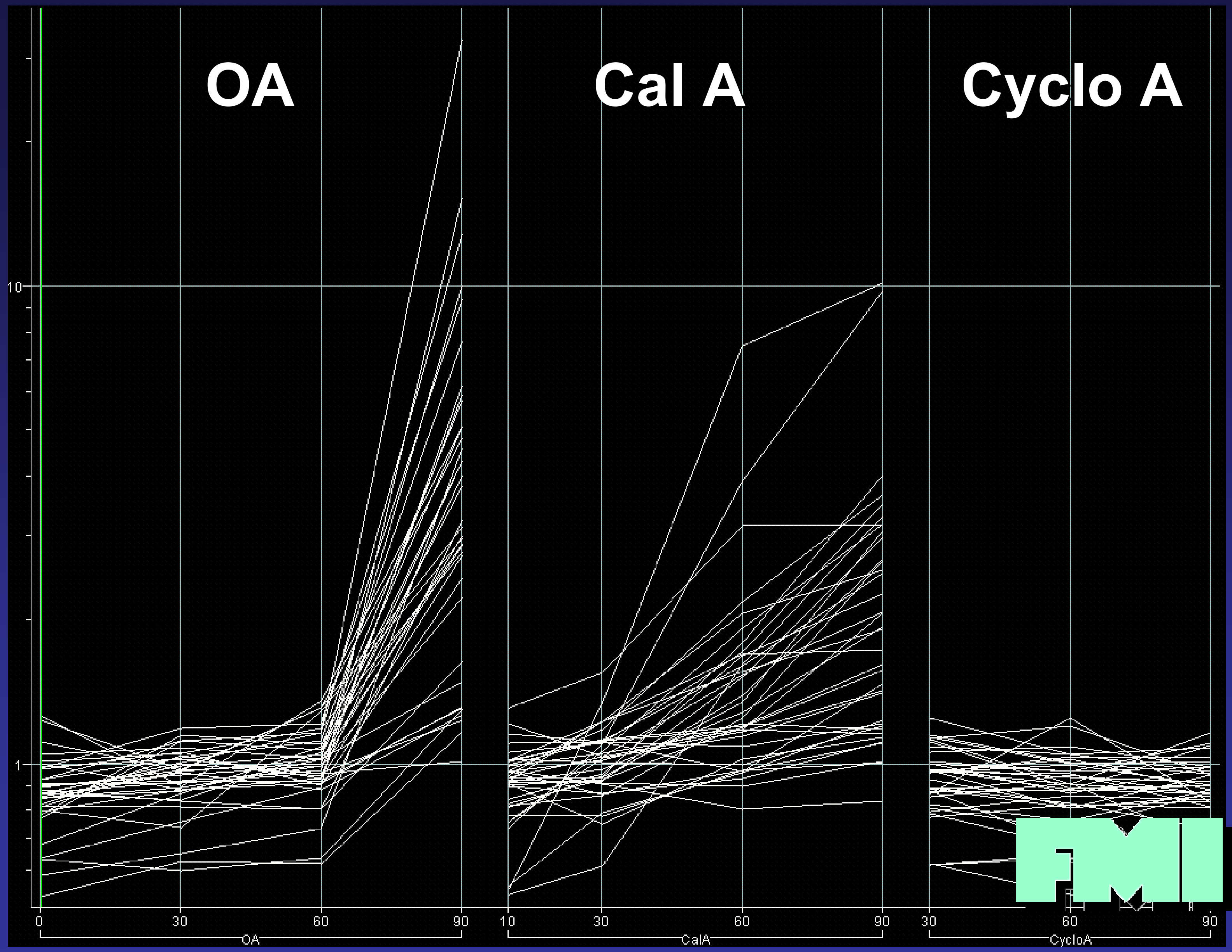


Inhibit PP2A

Treat HEK293 cells (human) with:

- Okadaic acid
- Calyculin A
- Cyclosporin A (control)

Specific inhibitors of PP2A



OA

Cal A

Cyclo A

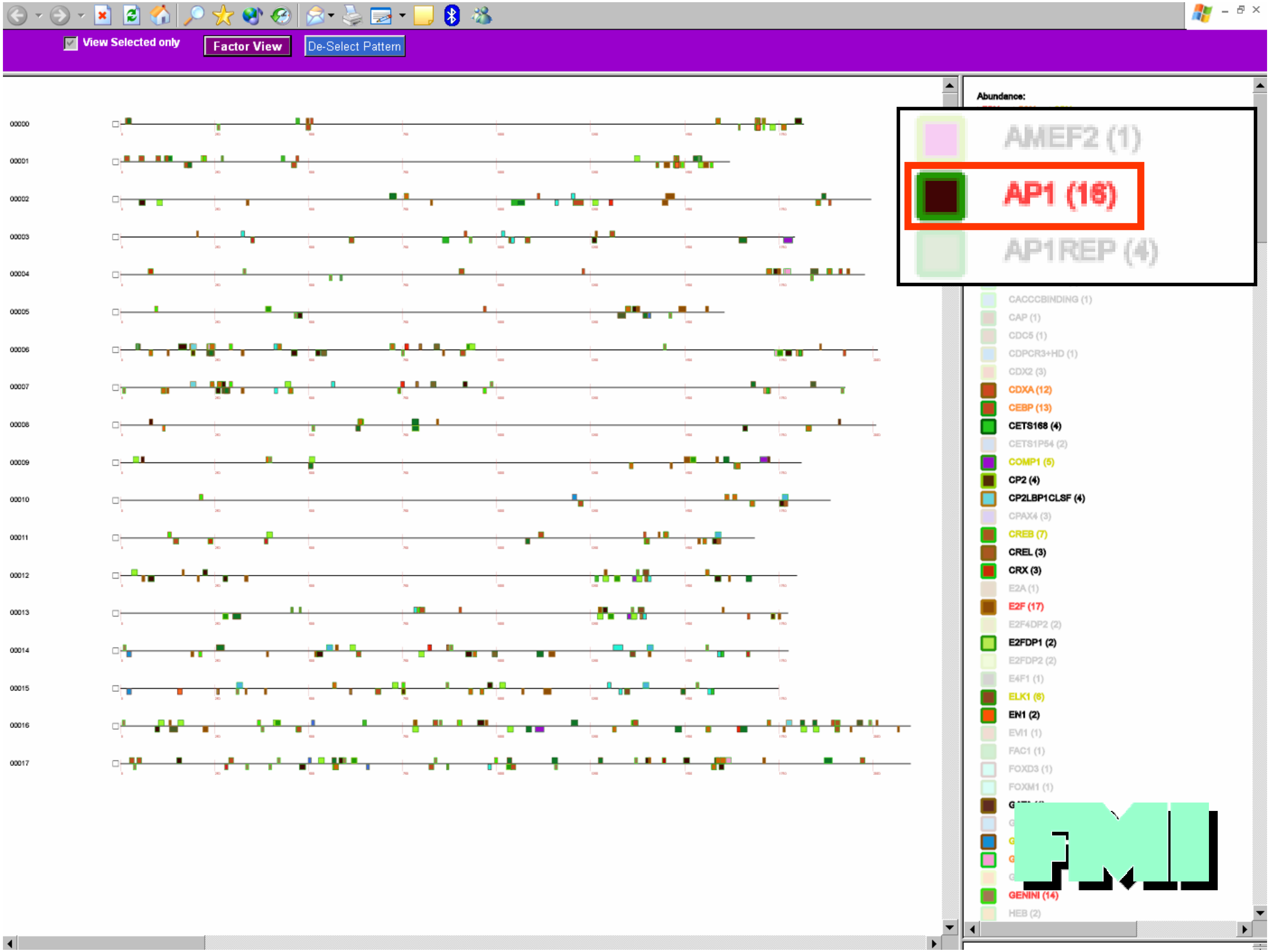
TFs upregulated

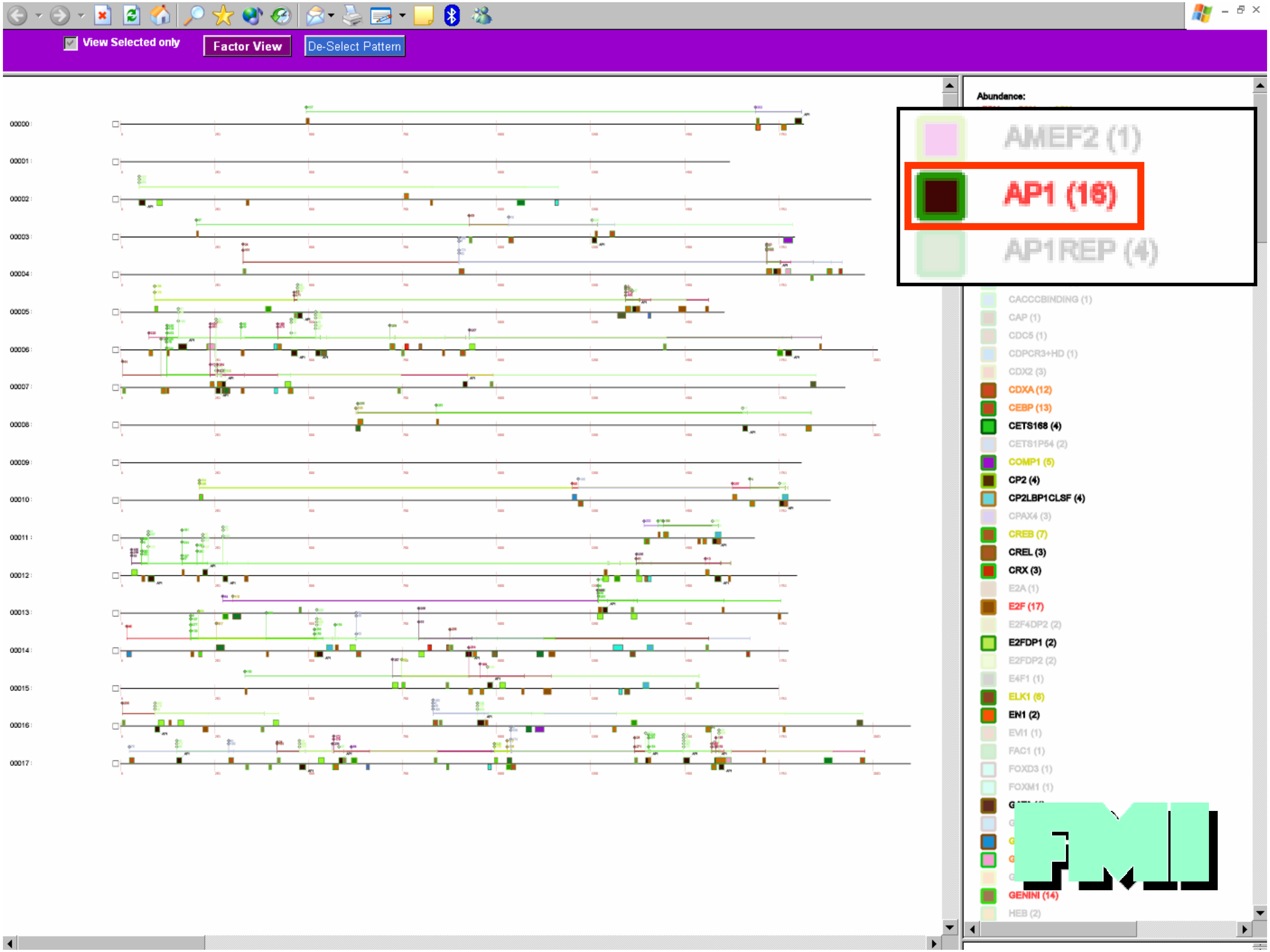
(with OA and Cal.A but not Cyclo.A)

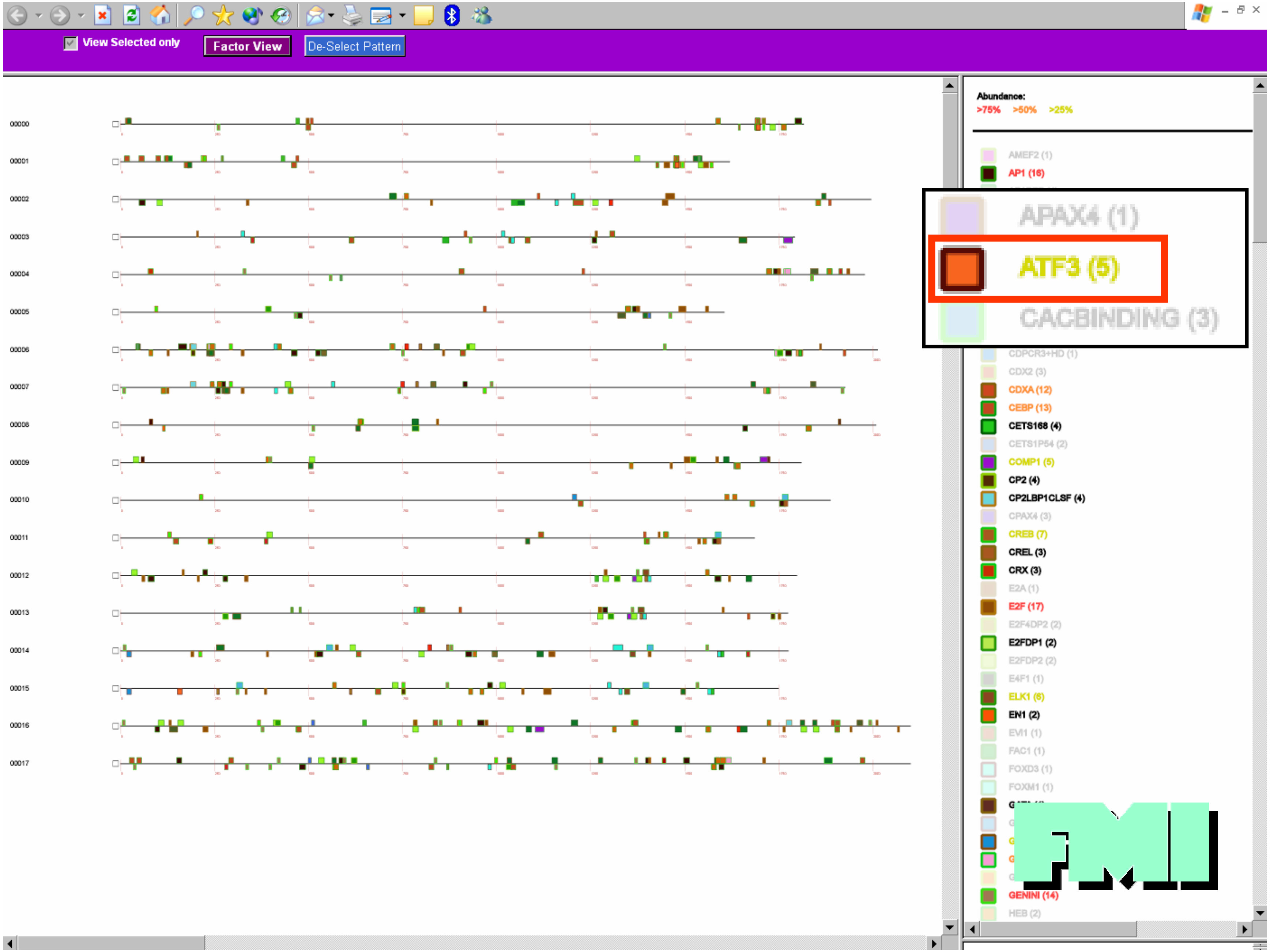
- FOS (75 -> 2851: x38)
- ATF3 (373 -> 2930: x7.9)
- JUN (220 -> 906: x4)

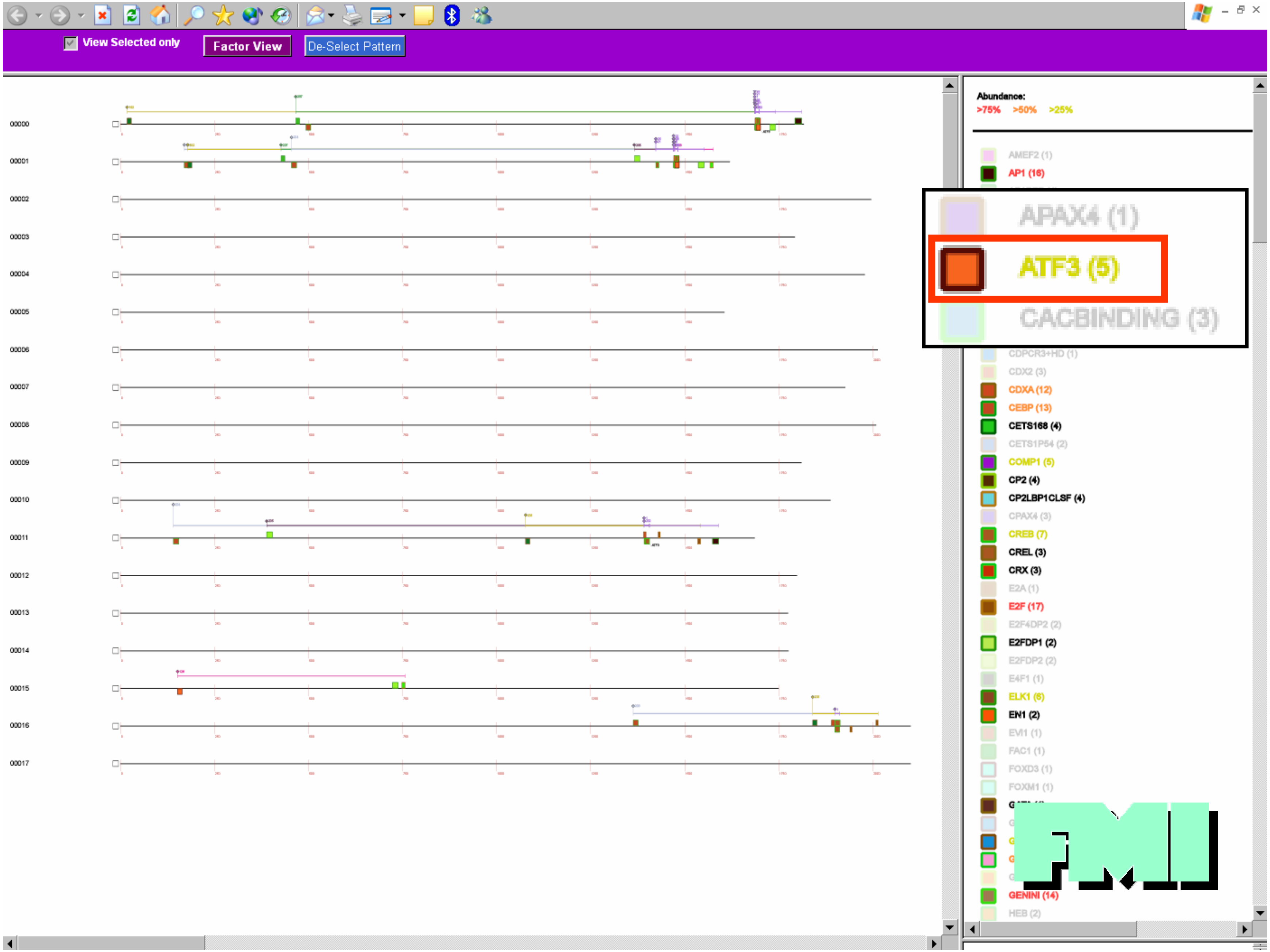
AP1

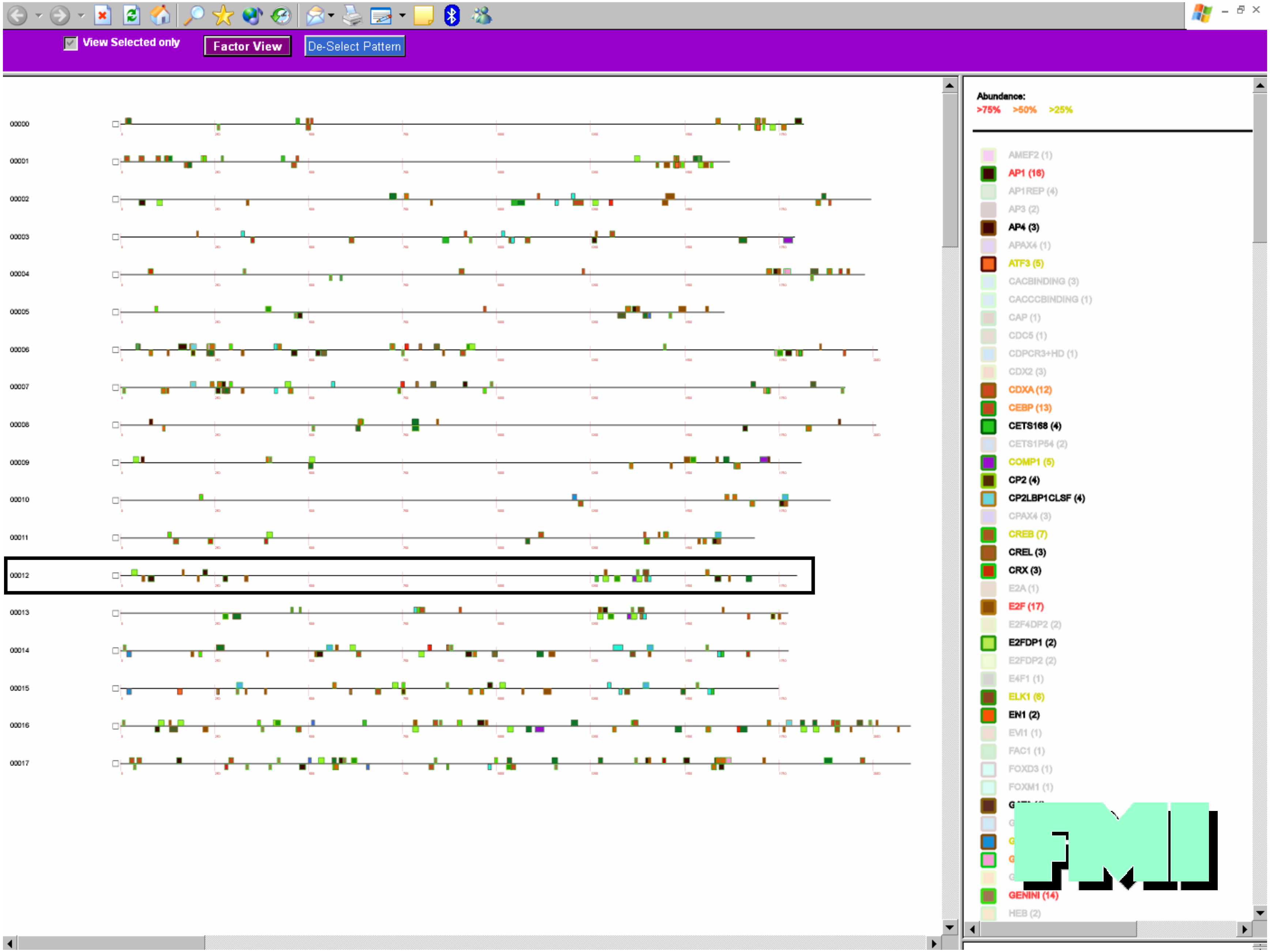
Image visualisation

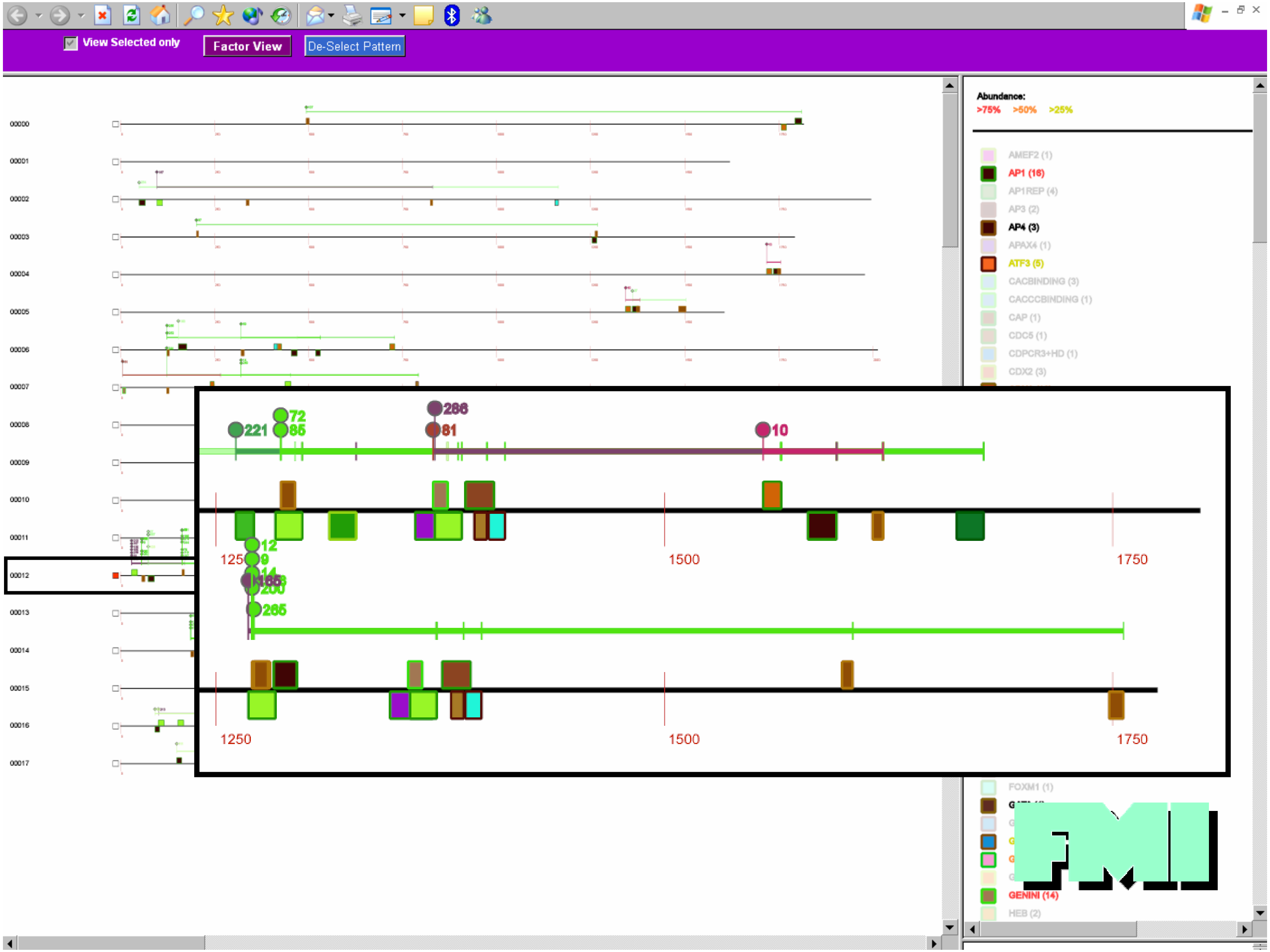

☑ View Selected only
Factor View
De-Select Pattern
00000
00001
00002
00003
00004
00005
00006
00007
00008
00009
00010
00011
00012
00013
00014
00015
00016
00017
Abundance:
AMEF2 (1)
AP1 (18)
AP1REP (4)
CACCCBINDING (1)
CAP (1)
CDC5 (1)
CDPCR3+HD (1)
CDX2 (3)
CDXA (12)
CEBP (13)
CETS168 (4)
CETS1P54 (2)
COMP1 (5)
CP2 (4)
CP2LBP1CLSF (4)
CPAX4 (3)
CREB (7)
CREL (3)
CRX (3)
E2A (1)
E2F (17)
E2F4DP2 (2)
E2FDP1 (2)
E2FDP2 (2)
E4F1 (1)
ELK1 (6)
EN1 (2)
EVI1 (1)
FAC1 (1)
FOXD3 (1)
FOXM1 (1)
GENINI (14)
HEB (2)


☑ View Selected only
Factor View
De-Select Pattern
00000
00001
00002
00003
00004
00005
00006
00007
00008
00009
00010
00011
00012
00013
00014
00015
00016
00017
Abundance:
AMEF2 (1)
AP1 (18)
AP1REP (4)
CACCCBINDING (1)
CAP (1)
CDC5 (1)
CDPCR3+HD (1)
CDX2 (3)
CDXA (12)
CEBP (13)
CETS168 (4)
CETS1P54 (2)
COMP1 (5)
CP2 (4)
CP2LBP1CLSF (4)
CPAX4 (3)
CREB (7)
CREL (3)
CRX (3)
E2A (1)
E2F (17)
E2F4DP2 (2)
E2FDP1 (2)
E2FDP2 (2)
E4F1 (1)
ELK1 (6)
EN1 (2)
EVI1 (1)
FAC1 (1)
FOXD3 (1)
FOXM1 (1)
GENINI (14)
HEB (2)
FMI

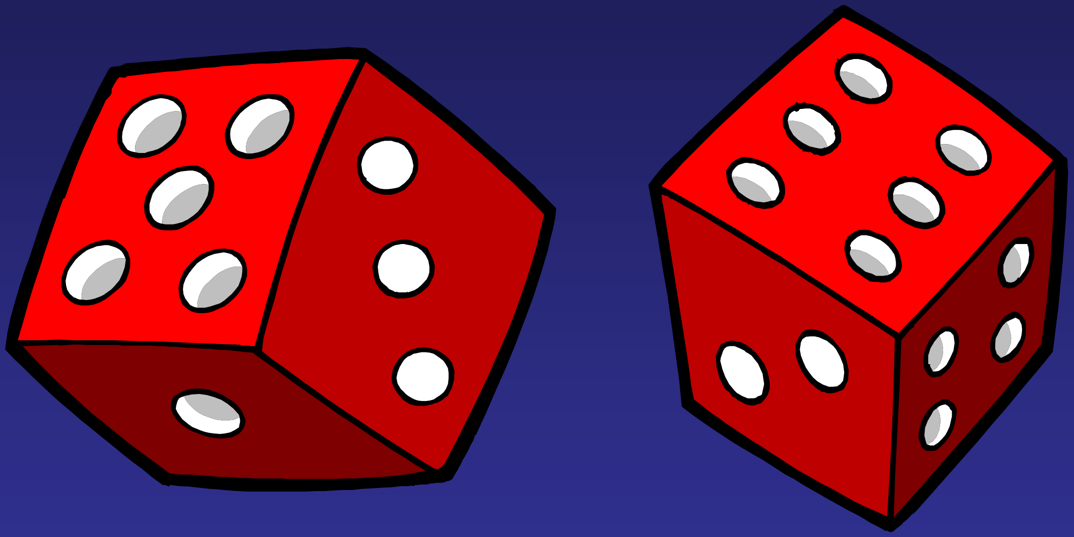

☑ View Selected only
Factor View
De-Select Pattern
00000
00001
00002
00003
00004
00005
00006
00007
00008
00009
00010
00011
00012
00013
00014
00015
00016
00017
Abundance:
>75% >50% >25%
AMEF2 (1)
AP1 (16)
APAX4 (1)
ATF3 (5)
CACBINDING (3)
CDPCR3+HD (1)
CDX2 (3)
CDXA (12)
CEBP (13)
CETS168 (4)
CETS1P54 (2)
COMP1 (5)
CP2 (4)
CP2LBP1CLSF (4)
CPAX4 (3)
CREB (7)
CREL (3)
CRX (3)
E2A (1)
E2F (17)
E2F4DP2 (2)
E2FDP1 (2)
E2FDP2 (2)
E4F1 (1)
ELK1 (6)
EN1 (2)
EVI1 (1)
FAC1 (1)
FOXD3 (1)
FOXM1 (1)
GENINI (14)
HEB (2)

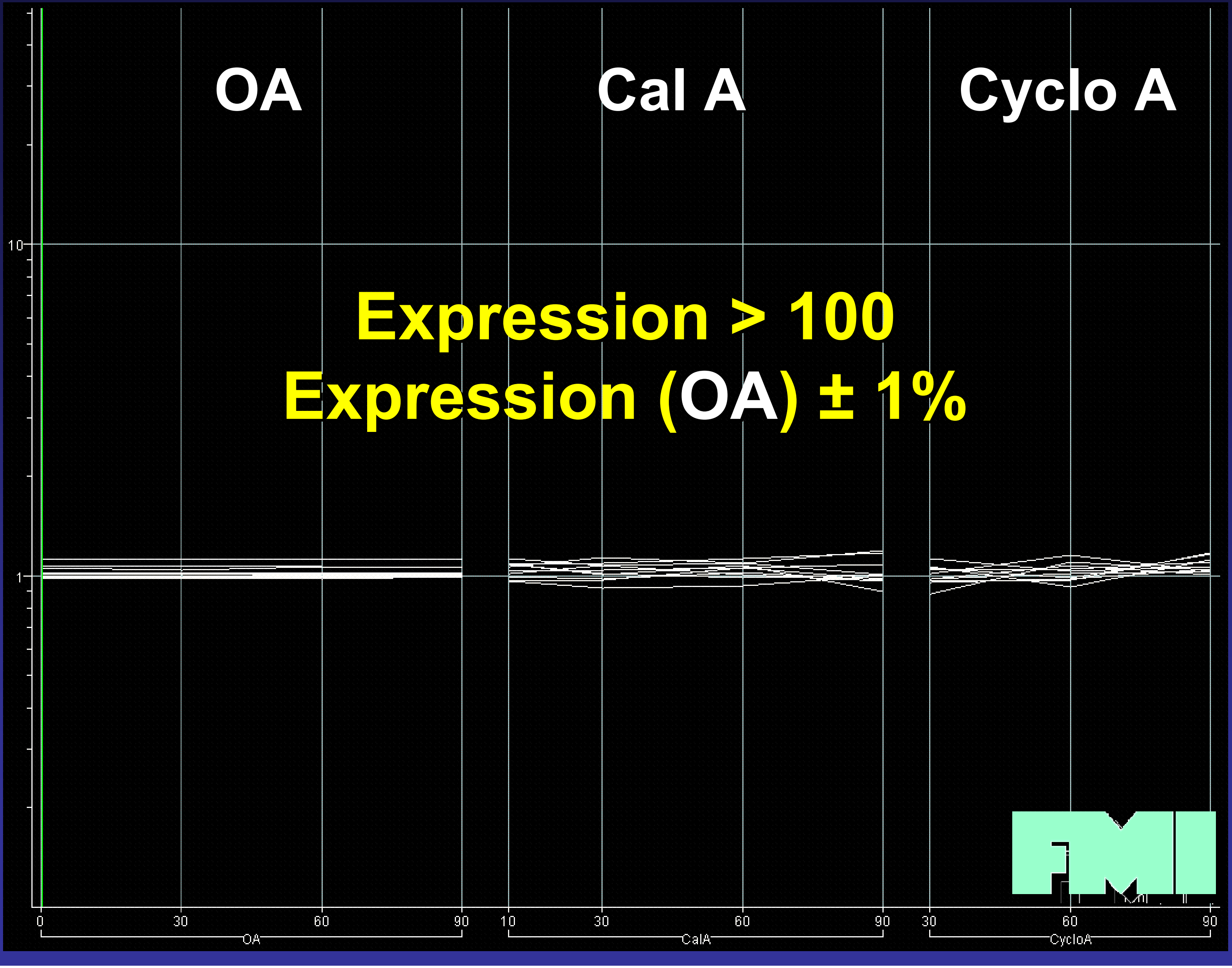

☑ View Selected only
Factor View
De-Select Pattern
00000
00001
00002
00003
00004
00005
00006
00007
00008
00009
00010
00011
00012
00013
00014
00015
00016
00017
Abundance:
>75% >50% >25%
AMEF2 (1)
AP1 (16)
APAX4 (1)
ATF3 (5)
CACBINDING (3)
CDPCR3+HD (1)
CDX2 (3)
CDXA (12)
CEBP (13)
CETS168 (4)
CETS1P54 (2)
COMP1 (5)
CP2 (4)
CP2LBP1CLSF (4)
CPAX4 (3)
CREB (7)
CREL (3)
CRX (3)
E2A (1)
E2F (17)
E2F4DP2 (2)
E2FDP1 (2)
E2FDP2 (2)
E4F1 (1)
ELK1 (6)
EN1 (2)
EVI1 (1)
FAC1 (1)
FOXD3 (1)
FOXM1 (1)
GENINI (14)
HEB (2)


☑ View Selected only
Factor View
De-Select Pattern
00000
00001
00002
00003
00004
00005
00006
00007
00008
00009
00010
00011
00012
00013
00014
00015
00016
00017
Abundance:
>75% >50% >25%
AMEF2 (1)
AP1 (16)
AP1REP (4)
AP3 (2)
AP4 (3)
APAX4 (1)
ATF3 (5)
CACBINDING (3)
CACCCBINDING (1)
CAP (1)
CDC5 (1)
CDPCR3+HD (1)
CDX2 (3)
CDXA (12)
CEBP (13)
CETS168 (4)
CETS1P54 (2)
COMP1 (5)
CP2 (4)
CP2LBP1CLSF (4)
CPAX4 (3)
CREB (7)
CREL (3)
CRX (3)
E2A (1)
E2F (17)
E2F4DP2 (2)
E2FDP1 (2)
E2FDP2 (2)
E4F1 (1)
ELK1 (6)
EN1 (2)
EVI1 (1)
FAC1 (1)
FOXD3 (1)
FOXM1 (1)
GENINI (14)
HEB (2)
FMI


☑ View Selected only
Factor View
De-Select Pattern
00000
00001
00002
00003
00004
00005
00006
00007
00008
00009
00010
00011
00012
00013
00014
00015
00016
00017
Abundance:
>75% >50% >25%
AMEF2 (1)
AP1 (16)
AP1REP (4)
AP3 (2)
AP4 (3)
APAX4 (1)
ATF3 (5)
CACBINDING (3)
CACCCBINDING (1)
CAP (1)
CDC5 (1)
CDPCR3+HD (1)
CDX2 (3)
FOXM1 (1)
GENINI (14)
HEB (2)
221
72
85
286
81
10
12
9
14
165
200
265
1250
1500
1750
FMI

How specific is this?





OA

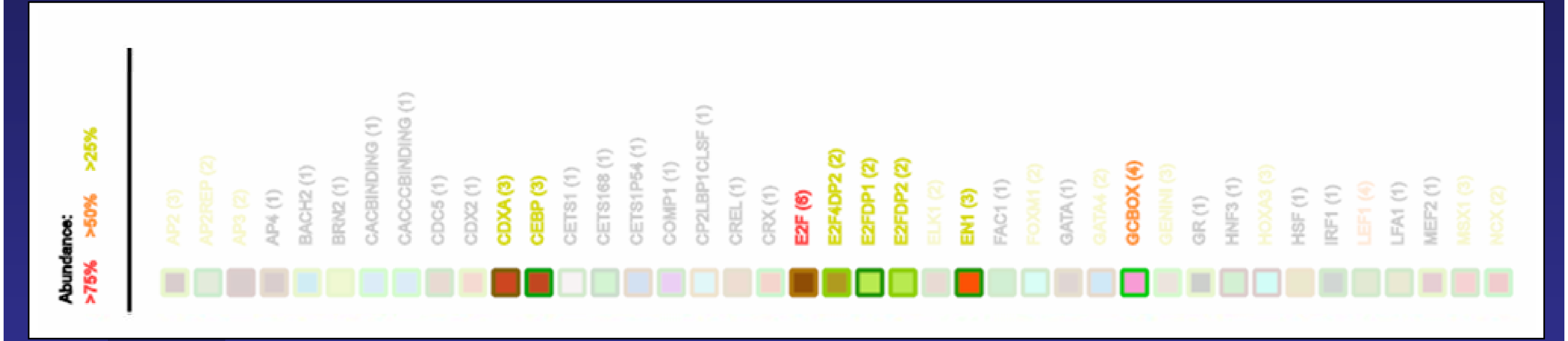
Cal A

Cyclo A

Expression > 100

Expression (OA) ± 1%

How specific is this?



No AP1 or ATF3 patterns

Summary

- Analyse AffyID (via Genomatix) or FASTA
- Exclude factors by AffyID, common names or Transfac matrix IDs
- Optional block alignment
- IBM Teiresias pattern analysis
- Interactive web output
- http://www.fmi.ch/members/edward.oakeley/promoter_plot.zip

Acknowledgements

Alessandro DiCara
Karsten Schmidt
Herbert Angliker