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?

Promoter

ATG

50 bp?
Oligo capping

26,480 human
starts mapped
(Genomatix, 2002)
Map start of transcription

2kb

100bp
Transcription factors

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

Input Affymetrix IDs. Promoters found and analysed automatically.

EBI DNA block alignment performed on orthologues. Only conserved blocks analysed for TFs.

FASTA sequences analysed. This file made automatically if the Affy or block buttons are clicked.

Regenerate SVG web files for browser from the XML result file.
Real science
Inhibit PP2A

Treat HEK293 cells (human) with:

- Okadaic acid
- Calyculin A
- Cyclosporin A (control)
TFs upregulated
(with OA and Cal.A but not Cyclo.A)

- **FOS** (75 -> 2851: x38)
- **ATF3** (373 -> 2930: x7.9)
- **JUN** (220 -> 906: x4)
Image visualisation
AMEF2 (1)
AP1 (18)
AP1REP (4)
How specific is this?
Expression $> 100$

Expression (OA) $\pm 1\%$
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