Pathway plasticity

SW2: Pathway Analysis in Transcriptomics, Proteomics and Metabolomics

Saturday, March 17, 2012
8:00 am – 4:30 pm

Alexander Kel

geneXplain
Wolfenbüttel

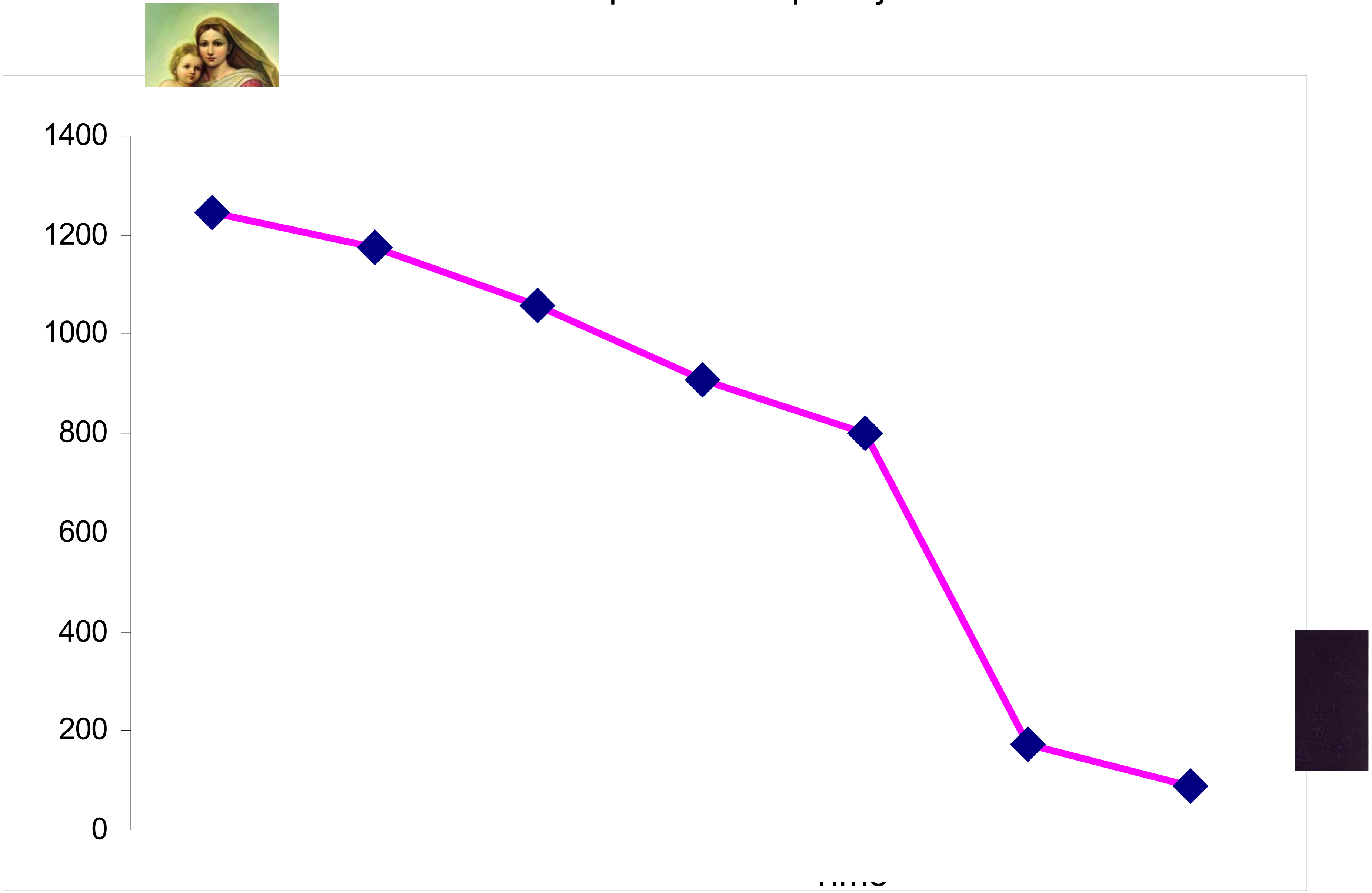
Biosoft.ru, Skolkovo
Moscow

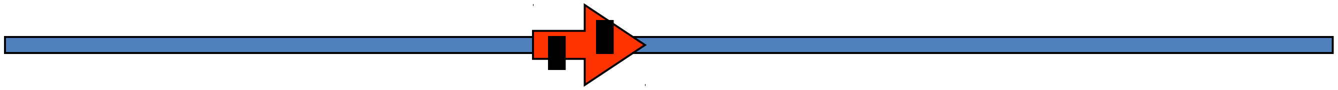
0100100010011101
Institute of Systems Biology
Novosibirsk

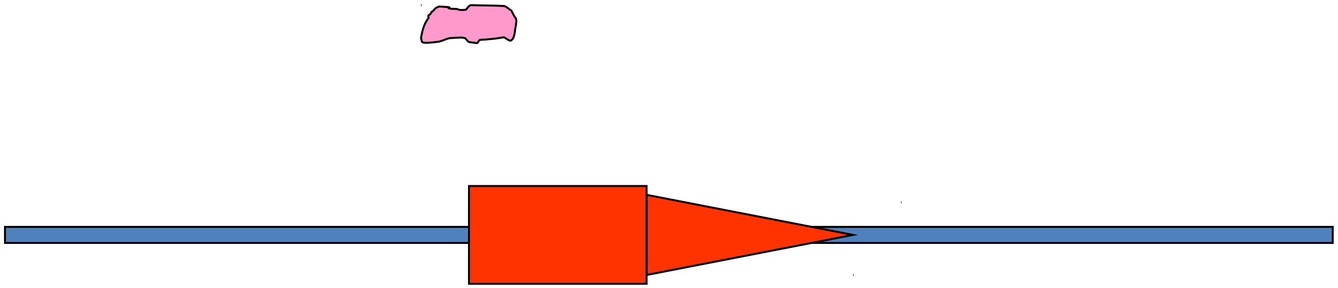
alexander.kel@genexplain.com


JOAN MIRÓ

Size of zip file = complexity







AP-1

Consensus: `TGAgTCA`

Human collagenase (-2013)

```
*******
TGAGTCA
```

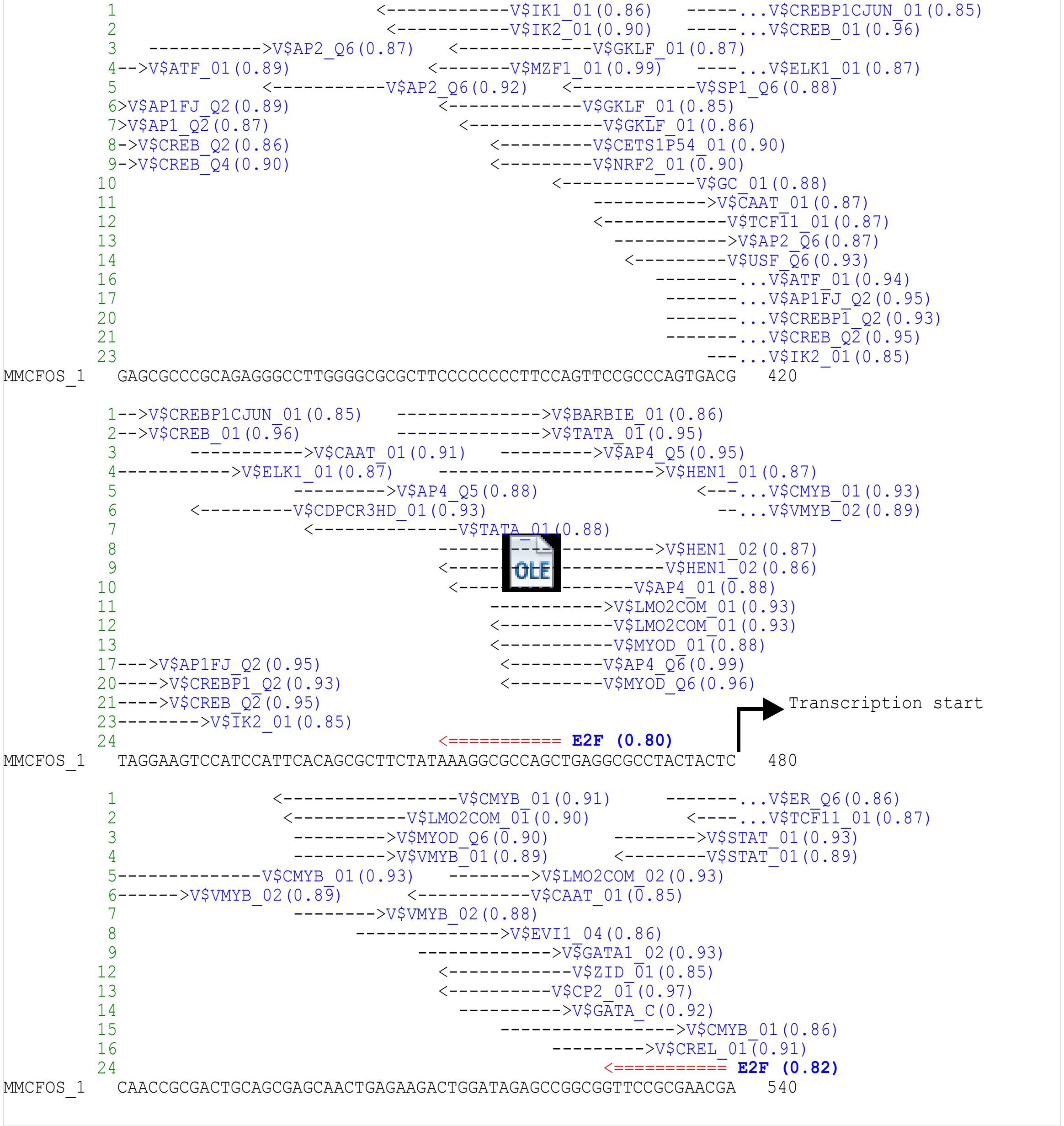
Mouse IL-2 (-143)

```
** ** *
TGTGTAA
```

Mouse TNF-alpha (-82)

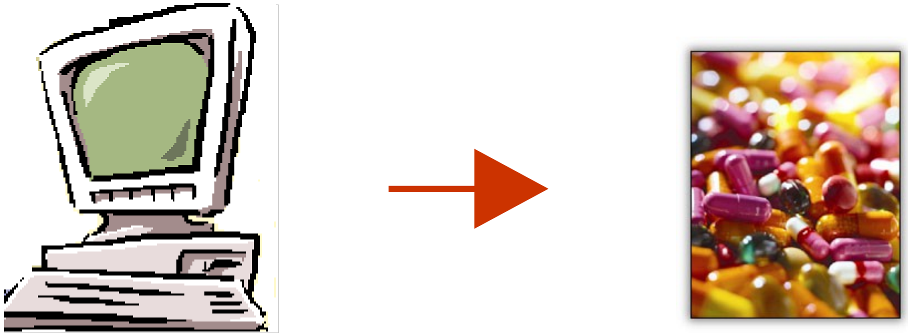
```
*   **
TTTCTCC
```

Mouse c-fos promoter
(Matrix search for TF binding sites)

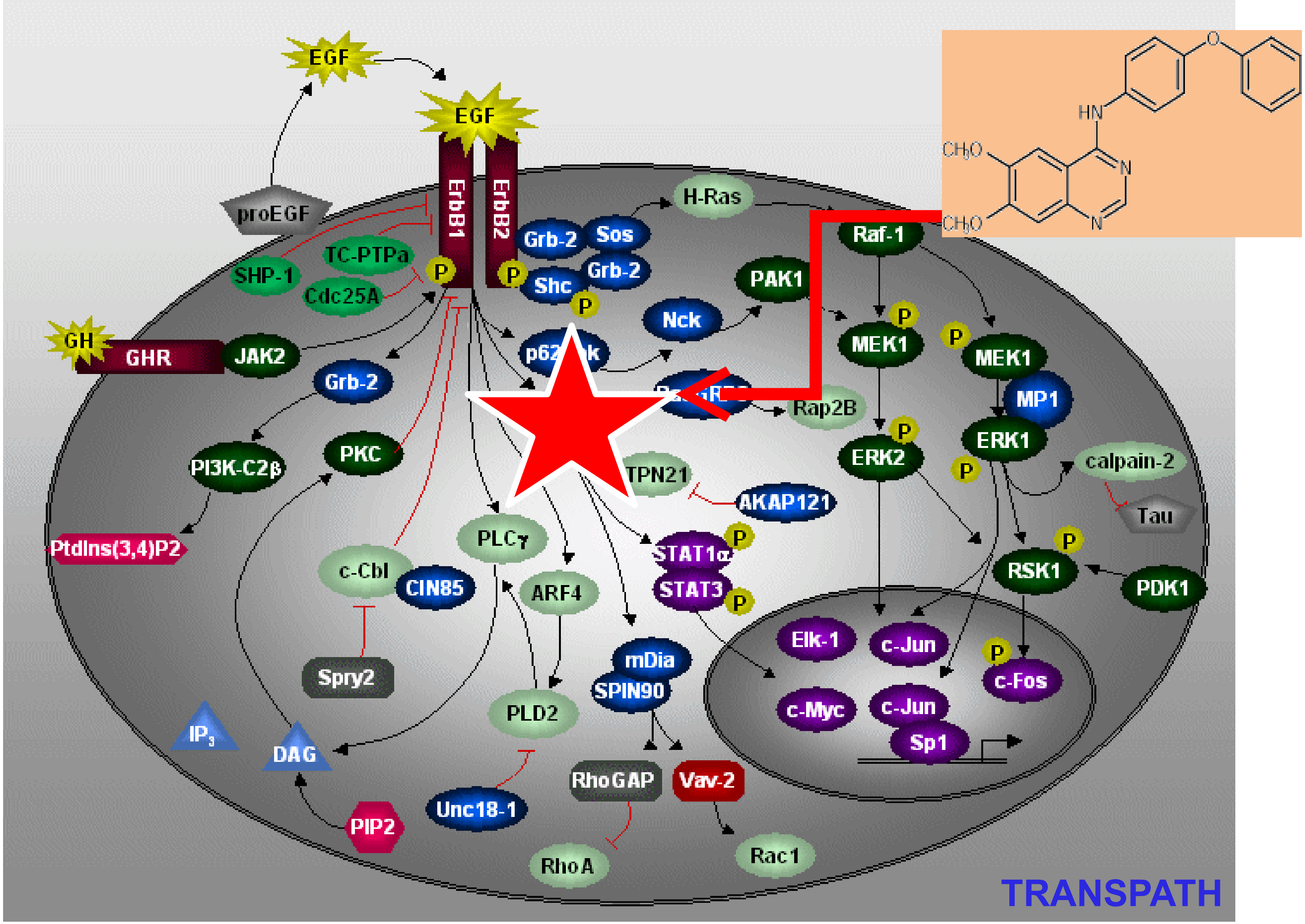


```
          1                         <------------V$IK1_01(0.86)   -----...V$CREBP1CJUN_01(0.85)
          2                          <-----------V$IK2_01(0.90)   -----...V$CREB_01(0.96)
          3   ----------->V$AP2_Q6(0.87)   <-------------V$GKLF_01(0.87)
          4-->V$ATF_01(0.89)             <-------V$MZF1_01(0.99)   ----...V$ELK1_01(0.87)
          5              <-----------V$AP2_Q6(0.92)   <------------V$SP1_Q6(0.88)
          6>V$AP1FJ_Q2(0.89)              <-------------V$GKLF_01(0.85)
          7>V$AP1_Q2(0.87)                  <-------------V$GKLF_01(0.86)
          8->V$CREB_Q2(0.86)                   <---------V$CETS1P54_01(0.90)
          9->V$CREB_Q4(0.90)                   <---------V$NRF2_01(0.90)
         10                                          <-------------V$GC_01(0.88)
         11                                              ----------->V$CAAT_01(0.87)
         12                                              <------------V$TCF11_01(0.87)
         13                                                ----------->V$AP2_Q6(0.87)
         14                                                 <---------V$USF_Q6(0.93)
         16                                                    --------...V$ATF_01(0.94)
         17                                                     -------...V$AP1FJ_Q2(0.95)
         20                                                     -------...V$CREBP1_Q2(0.93)
         21                                                     -------...V$CREB_Q2(0.95)
         23                                                         ---...V$IK2_01(0.85)
MMCFOS_1   GAGCGCCCGCAGAGGGCCTTGGGGCGCGCTTCCCCCCCCTTCCAGTTCCGCCCAGTGACG   420

          1-->V$CREBP1CJUN_01(0.85)   -------------->V$BARBIE_01(0.86)
          2-->V$CREB_01(0.96)         -------------->V$TATA_01(0.95)
          3       ----------->V$CAAT_01(0.91)   --------->V$AP4_Q5(0.95)
          4----------->V$ELK1_01(0.87)    --------------------->V$HEN1_01(0.87)
          5                --------->V$AP4_Q5(0.88)               <---...V$CMYB_01(0.93)
          6       <---------V$CDPCR3HD_01(0.93)                      --...V$VMYB_02(0.89)
          7                  <--------------V$TATA_01(0.88)
          8                               ------      --------->V$HEN1_02(0.87)
          9                               <-----      ----------V$HEN1_02(0.86)
         10                                <----      -------V$AP4_01(0.88)
         11                                    ----------->V$LMO2COM_01(0.93)
         12                                    <-----------V$LMO2COM_01(0.93)
         13                                    <-----------V$MYOD_01(0.88)
         17--->V$AP1FJ_Q2(0.95)                 <---------V$AP4_Q6(0.99)
         20---->V$CREBP1_Q2(0.93)               <---------V$MYOD_Q6(0.96)
         21---->V$CREB_Q2(0.95)                                              Transcription start
         23-------->V$IK2_01(0.85)
         24                               <=========== E2F (0.80)
MMCFOS_1   TAGGAAGTCCATCCATTCACAGCGCTTCTATAAAGGCGCCAGCTGAGGCGCCTACTACTC   480

          1               <-----------------V$CMYB_01(0.91)     -------...V$ER_Q6(0.86)
          2                <-----------V$LMO2COM_01(0.90)         <----...V$TCF11_01(0.87)
          3                 --------->V$MYOD_Q6(0.90)      -------->V$STAT_01(0.93)
          4                 --------->V$VMYB_01(0.89)      <--------V$STAT_01(0.89)
          5--------------V$CMYB_01(0.93)   -------->V$LMO2COM_02(0.93)
          6------>V$VMYB_02(0.89)      <-----------V$CAAT_01(0.85)
          7                 -------->V$VMYB_02(0.88)
          8                       -------------->V$EVI1_04(0.86)
          9                             ------------->V$GATA1_02(0.93)
         12                               <------------V$ZID_01(0.85)
         13                               <----------V$CP2_01(0.97)
         14                                 ---------->V$GATA_C(0.92)
         15                                     ----------------->V$CMYB_01(0.86)
         16                                          --------->V$CREL_01(0.91)
         24                                               <=========== E2F (0.82)
MMCFOS_1   CAACCGCGACTGCAGCGAGCAACTGAGAAGACTGGATAGAGCCGGCGGTTCCGCGAACGA   540
```

Use computers to discover drugs

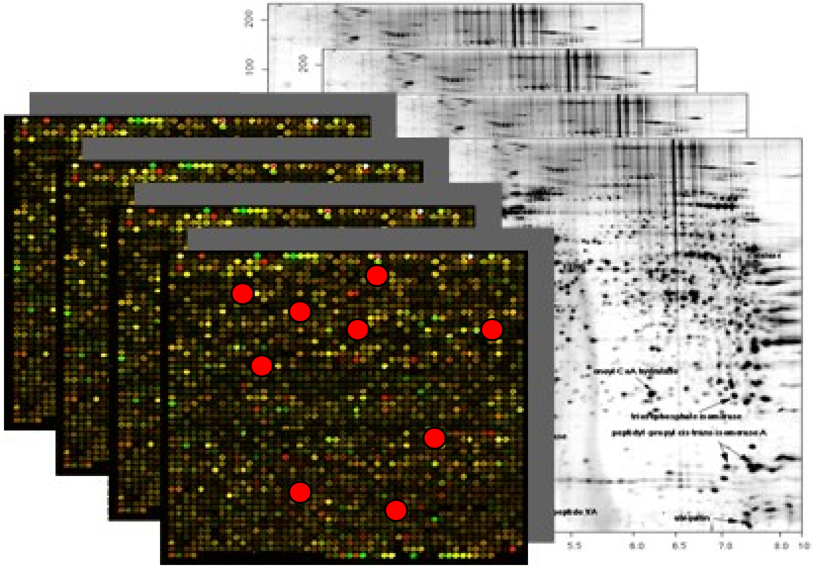


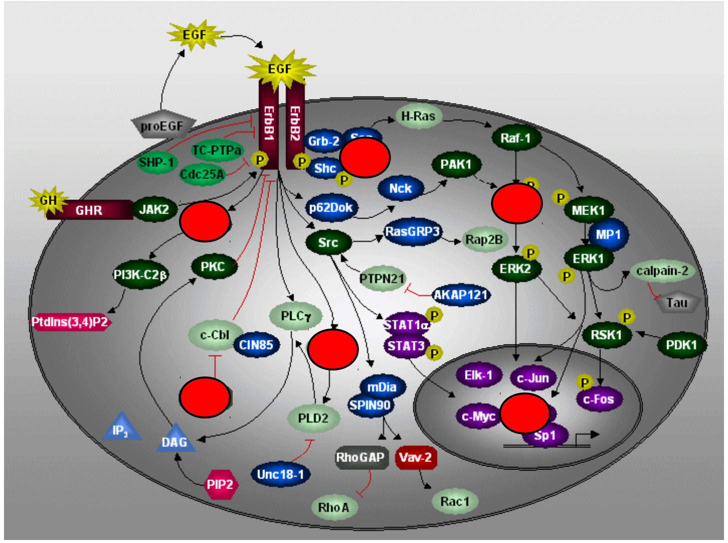
We should find a key pathway of a disease, select a good <u>target</u> and inhibit it.



TRANSPATH

Pathway mapping

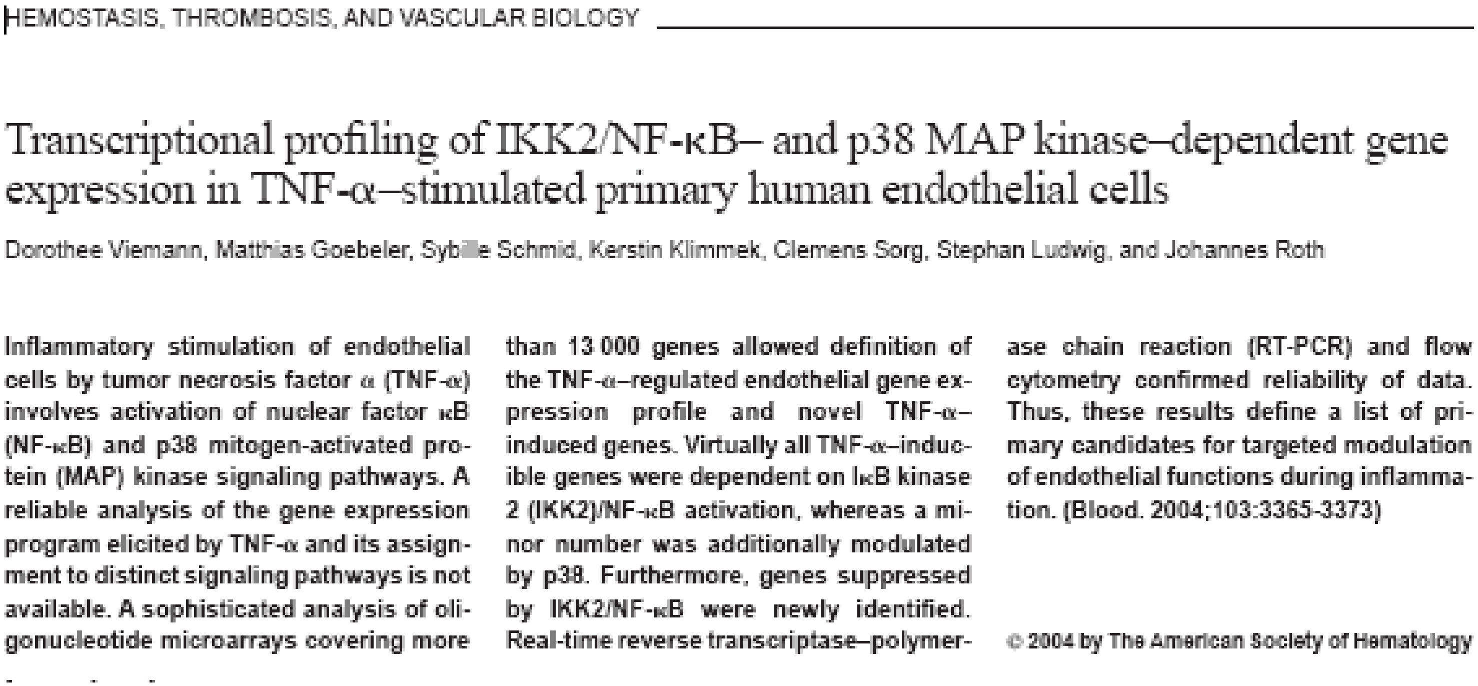




Differentially expressed genes/proteins

Mapping on pathways

Cause of disease ??



HEMOSTASIS, THROMBOSIS, AND VASCULAR BIOLOGY

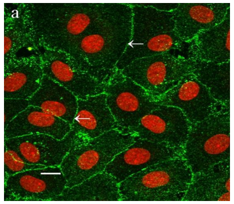
Transcriptional profiling of IKK2/NF-κB– and p38 MAP kinase–dependent gene expression in TNF-α–stimulated primary human endothelial cells

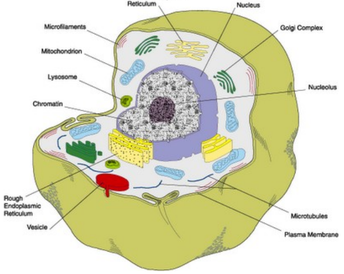
Dorothee Viemann, Matthias Goebeler, Sybille Schmid, Kerstin Klimmek, Clemens Sorg, Stephan Ludwig, and Johannes Roth

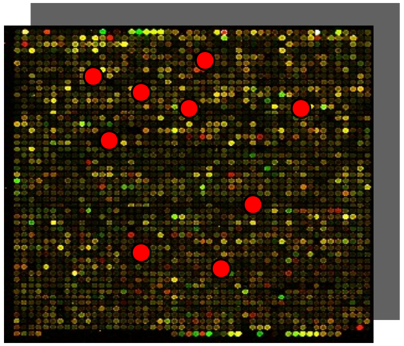
Inflammatory stimulation of endothelial cells by tumor necrosis factor α (TNF-α) involves activation of nuclear factor κB (NF-κB) and p38 mitogen-activated protein (MAP) kinase signaling pathways. A reliable analysis of the gene expression program elicited by TNF-α and its assignment to distinct signaling pathways is not available. A sophisticated analysis of oligonucleotide microarrays covering more than 13 000 genes allowed definition of the TNF-α–regulated endothelial gene expression profile and novel TNF-α–induced genes. Virtually all TNF-α–inducible genes were dependent on IκB kinase 2 (IKK2)/NF-κB activation, whereas a minor number was additionally modulated by p38. Furthermore, genes suppressed by IKK2/NF-κB were newly identified. Real-time reverse transcriptase–polymerase chain reaction (RT-PCR) and flow cytometry confirmed reliability of data. Thus, these results define a list of primary candidates for targeted modulation of endothelial functions during inflammation. (Blood. 2004;103:3365-3373)

TNF-α

117 differentially expressed genes

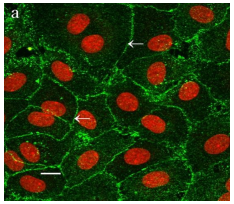


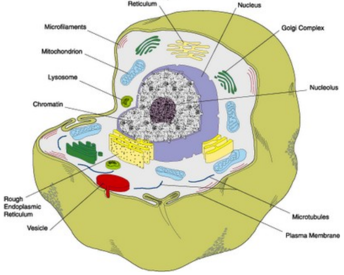




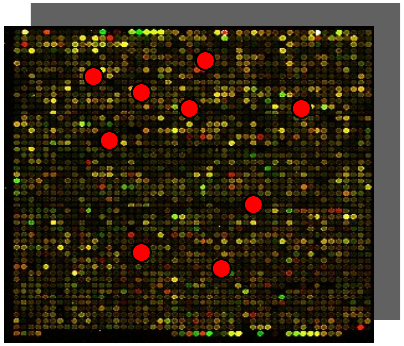
Can we predict TNF pathway?

?

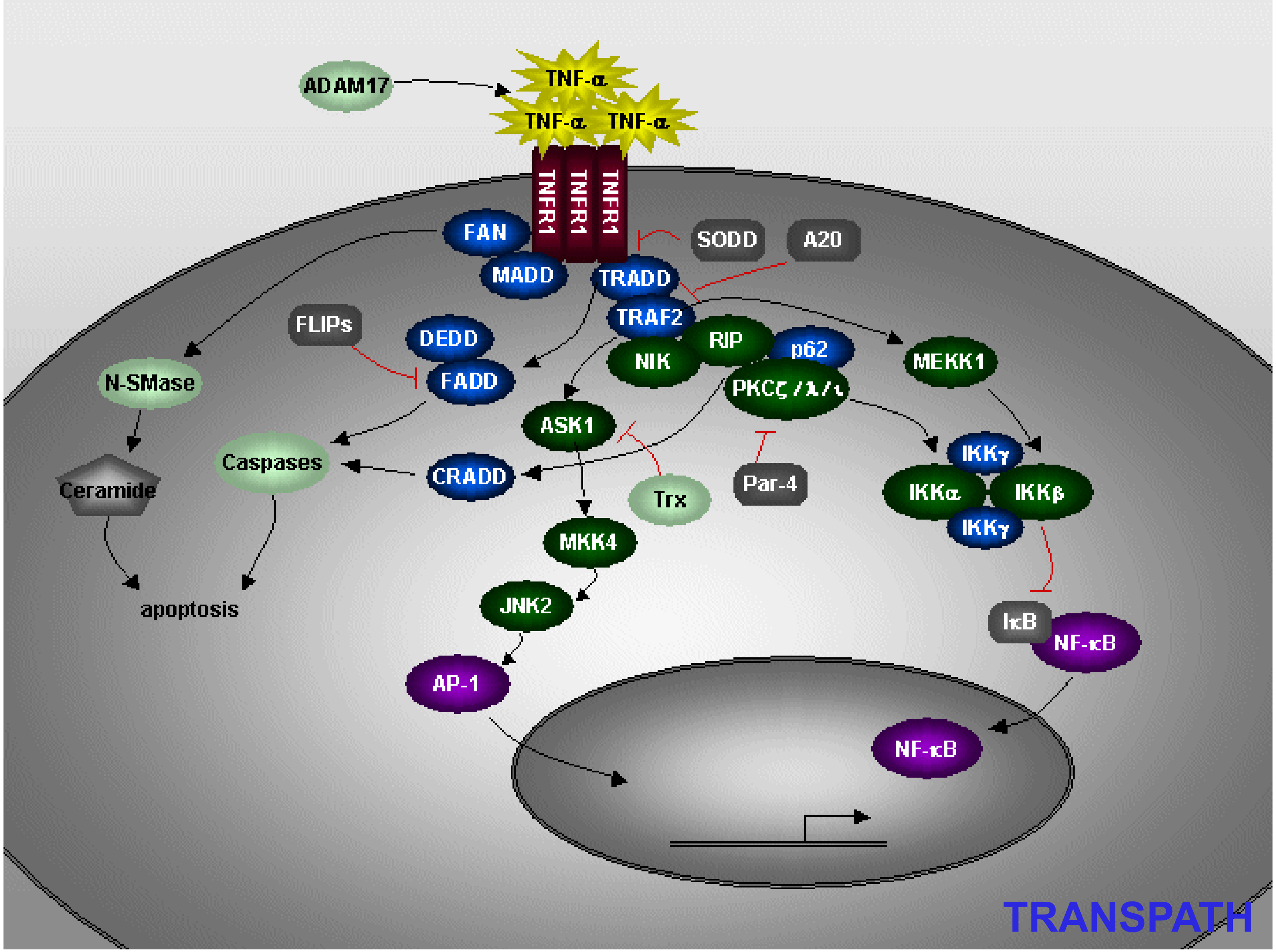




117 differentially expressed genes



Canonical TNF pathway



Lets do mapping the differentially expressed genes on canonical pathways.

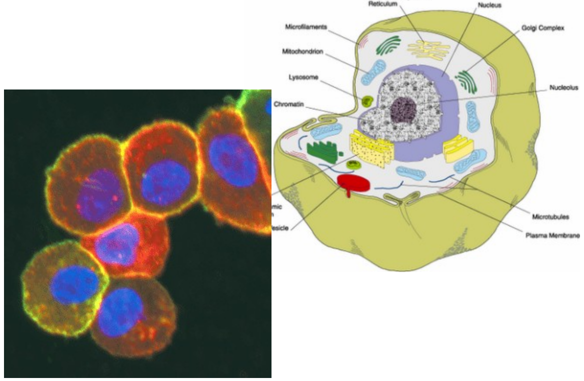
Pathway name	Hits	Pathway_id	Hit names	Pathway siz	p-value
M-CSF ---> c-Ets-2	2	CH000000060	ETS2; CSF1	5	3.07E-03
IFNalpha, IFNbeta, IFNgamma ---> Rap1	3	CH000000595	IFNGR1; TYK2; IFNGR2	19	4.34E-03
Epo ---Lyn---> STAT5A	2	CH000000524	STAT5A; LYN	6	4.56E-03
activin A ---> Smad3	2	CH000000680	INHBA; SMAD3	10	1.31E-02
IFN pathway	3	CH000000740	IFNGR1; TYK2; IFNGR2	29	1.44E-02
Sonic Hedgehog pathway	2	CH000001022	MTSS1; PTCH	19	4.48E-02
hypoxia pathways	2	CH000000987	CDKN1B; NRIP1	21	5.38E-02
EDAR pathway	2	CH0000007[illegible]9	NFKBIA; CYLD	27	8.40E-02
Epo pathway	2	CH000000741	STAT5A; LYN	32	1.12E-01
TGFbeta pathway	3	CH000000711	BMP2; INHBA; SMAD3	72	1.39E-01
IL-22 pathway	1	CH000000762	TYK2	9	1.51E-01
IL-10 pathway	1	CH000000761	TYK2	9	1.51E-01
VEGF-A pathway	2	CH000000723	NOS3; VEGFA	42	1.75E-01
TLR3 pathway	2	CH000000820	TANK; IKBKE	44	1.88E-01
IL-8 pathway	2	CH000000786	CXCL1; IL8	46	2.01E-01
TNF-alpha pathway	2	CH000000772	NFKBIA; OSIL	53	2.48E-01
p38 pathway	2	CH000000849	MAP2K3; DUSP8	55	2.61E-01

Not significant

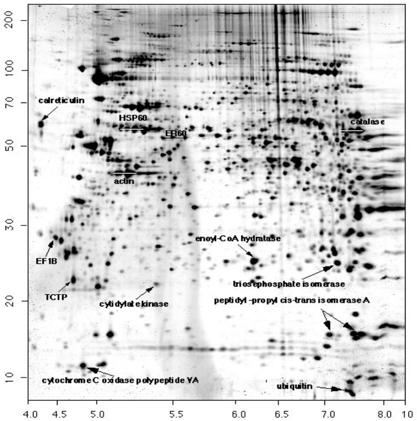
TNF pathway can not be found by direct maping on canonical pathways....

Human epidermoid carcinoma A431 cells treated by epidermal growth factor (EGF)

<u>EGF</u>



320 differntially expressed proteins



Mapping differentially expressed proteins to canonical signal transduction pathways

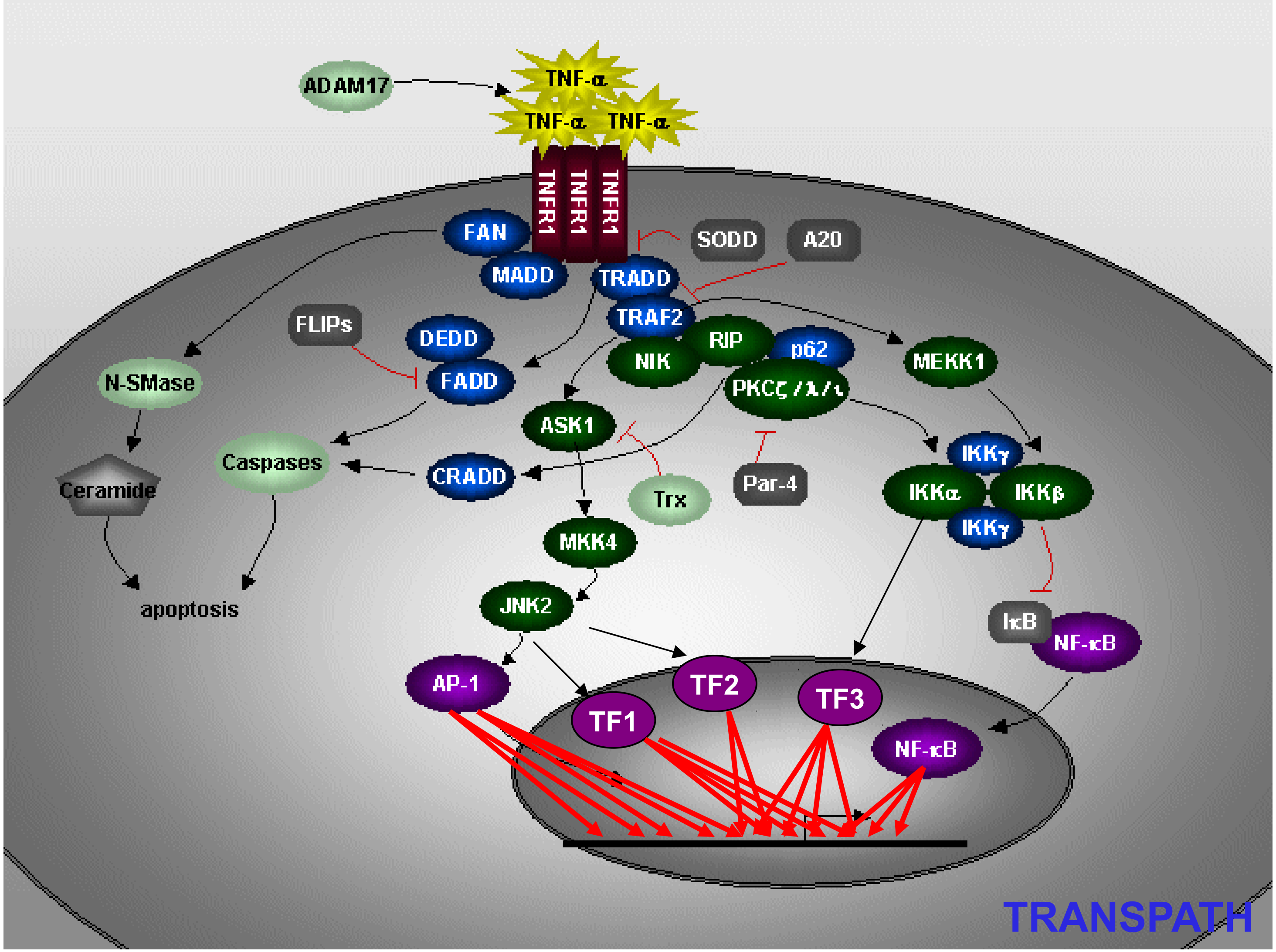
Pathway name	**#Hits in group**	**Hit names**	**Group size**	**p-value**
Caspase network	6	K18; E1; Cytochrome C; Hsp10; Ku70; Cdc42	104	0.00201348
CHIP ---/ Pael-R	2	E1; Hsc70	12	0.01177937
p53 pathway	4	E1; L23; Cytochrome C; Ku70	79	0.02072214
beta-catenin ---/ KAI1	1	Reptin52	5	0.06701759
Aurora-A cell cycle regulation	2	Ubc5B; E1	34	0.07924485
JNK pathway	3	E1; 14-3-3zeta; Trx1	75	0.0813304
parkin associated pathways	2	E1; Hsc70	40	0.10447487
beta-catenin:E-cadherin complex phosphorylation and dissociation	1	alpha-catenin	9	0.11739049
stress-associated pathways	3	E1; 14-3-3zeta; Trx1	100	0.15476
hypoxia pathways	1	Trx1	24	0.2849595
TNF-alpha pathway	1	Trx1	36	0.39594524
EGF pathway	1	E1	103	0.57615756

Mapping on pathways does not work (even in such a simple cases)

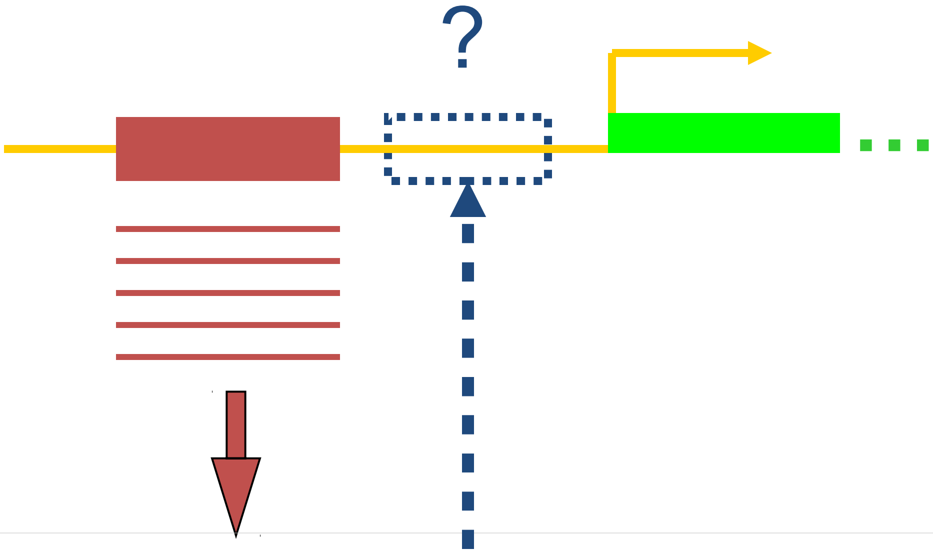
Why ?

Pathways are far from being fully undersood.

BIG gap of knowledge on interactions between TF and their target sites in DNA



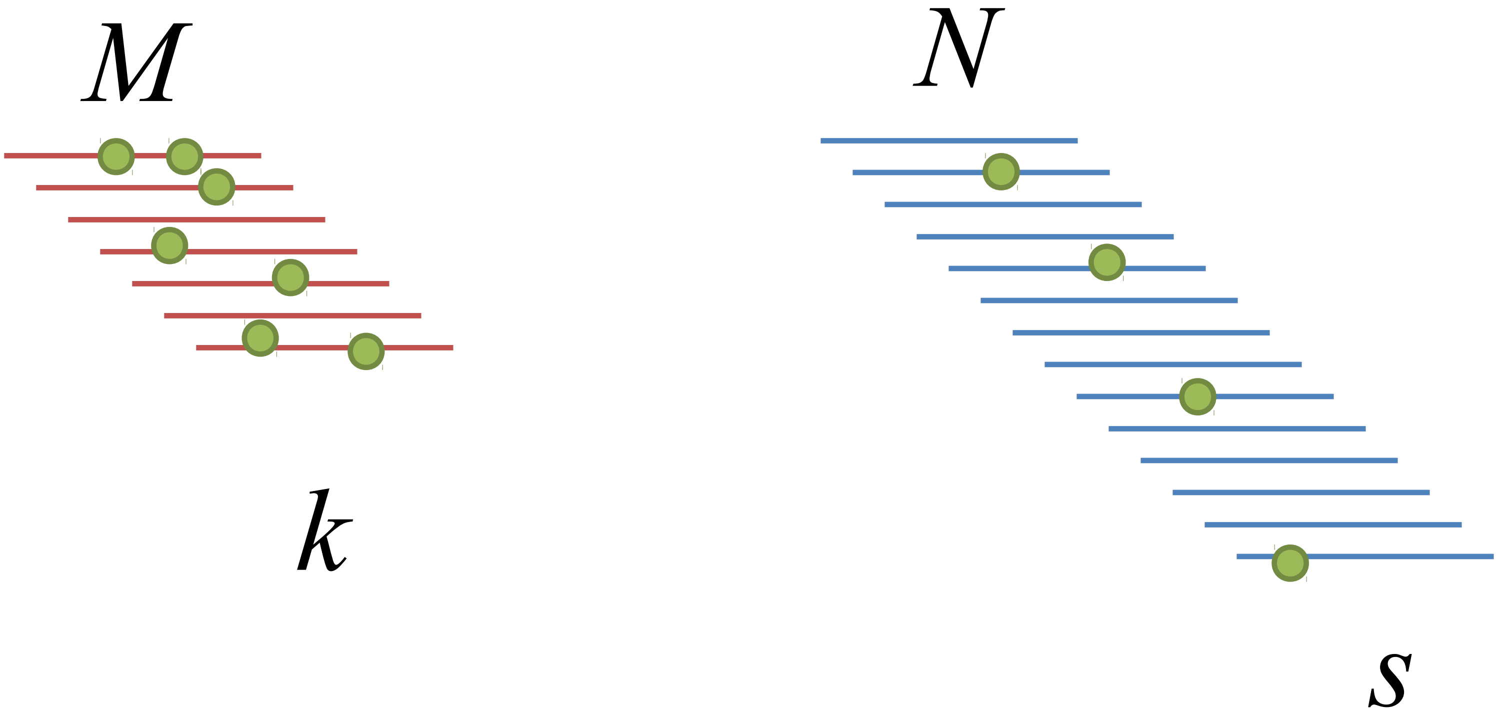
Search for new TF binding sites with PWMs



A	9	2	1	0	1	0	0	0	0	1	**15**	**13**	**13**	7
C	8	3	1	1	**13**	3	**29**	0	**22**	**8**	**9**	1	4	8
G	4	2	2	2	**15**	**26**	0	**29**	7	**17**	3	**7**	**9**	8
T	8	**22**	**25**	**26**	0	0	0	0	0	3	2	**8**	3	6
	N	**T**	**T**	**T**	**S**	**G**	**C**	**G**	**C**	**S**	**M**	**D**	**R**	**N**

$$q = \frac{\sum_{i=1}^{l} I(i)\, f(b_i, i) - \sum_{i=1}^{l} I(i)\, f^{\min}(i)}{\sum_{i=1}^{l} I(i)\, f^{\max}(i)} \qquad (1)$$

$$I(i) = \sum_{b \in \{A,T,G,C\}} f(b,i) \ln(4 f(b,i)) \qquad (2)$$

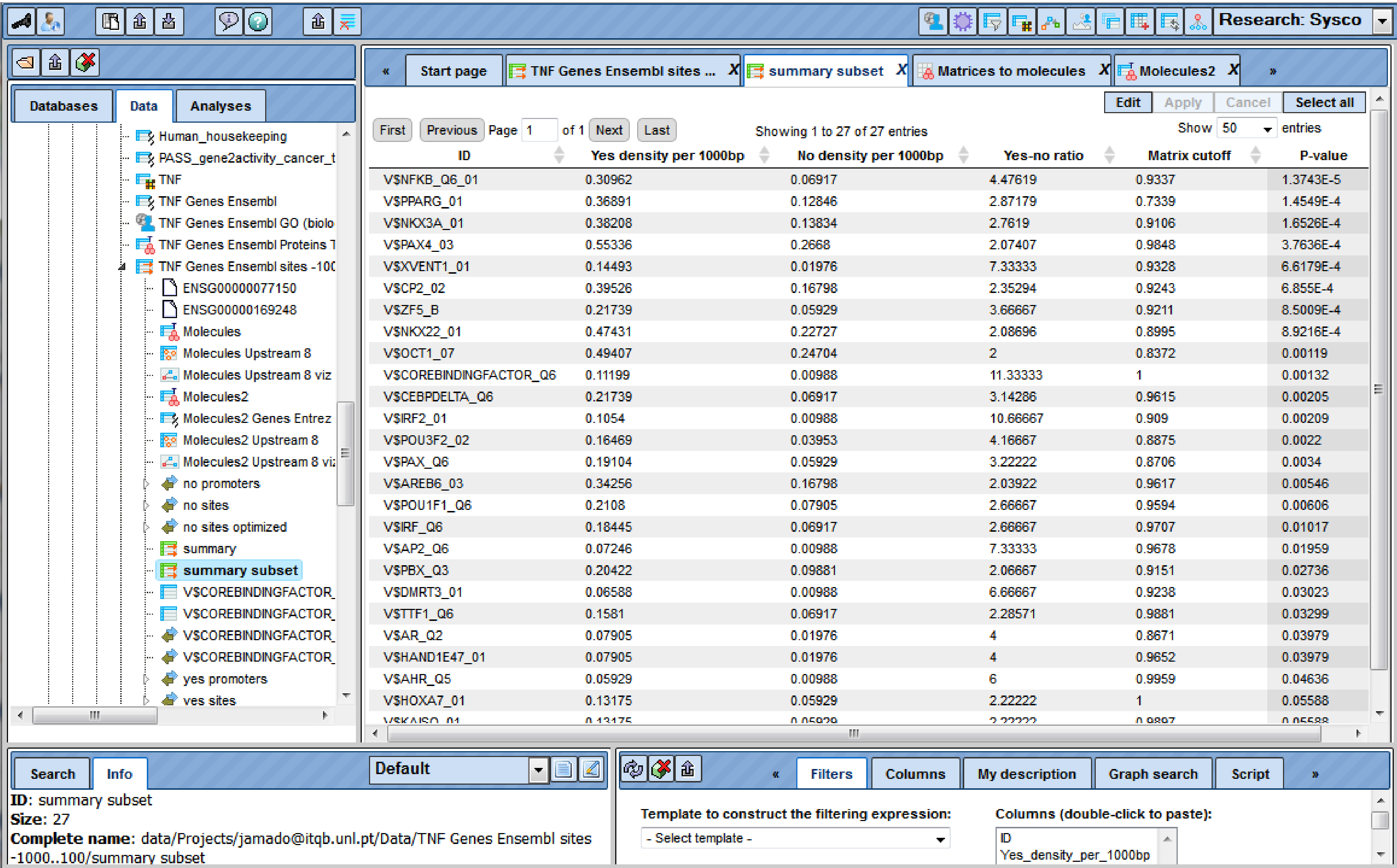


$p = M/N$

$n = k + s$

$$\text{p} - \text{value} = \sum_{i=k}^{n} \binom{n}{i} p^{i} (1 - \text{p})^{n-i}$$

Overrepresented TFs in TNF-alpha regulated promoters



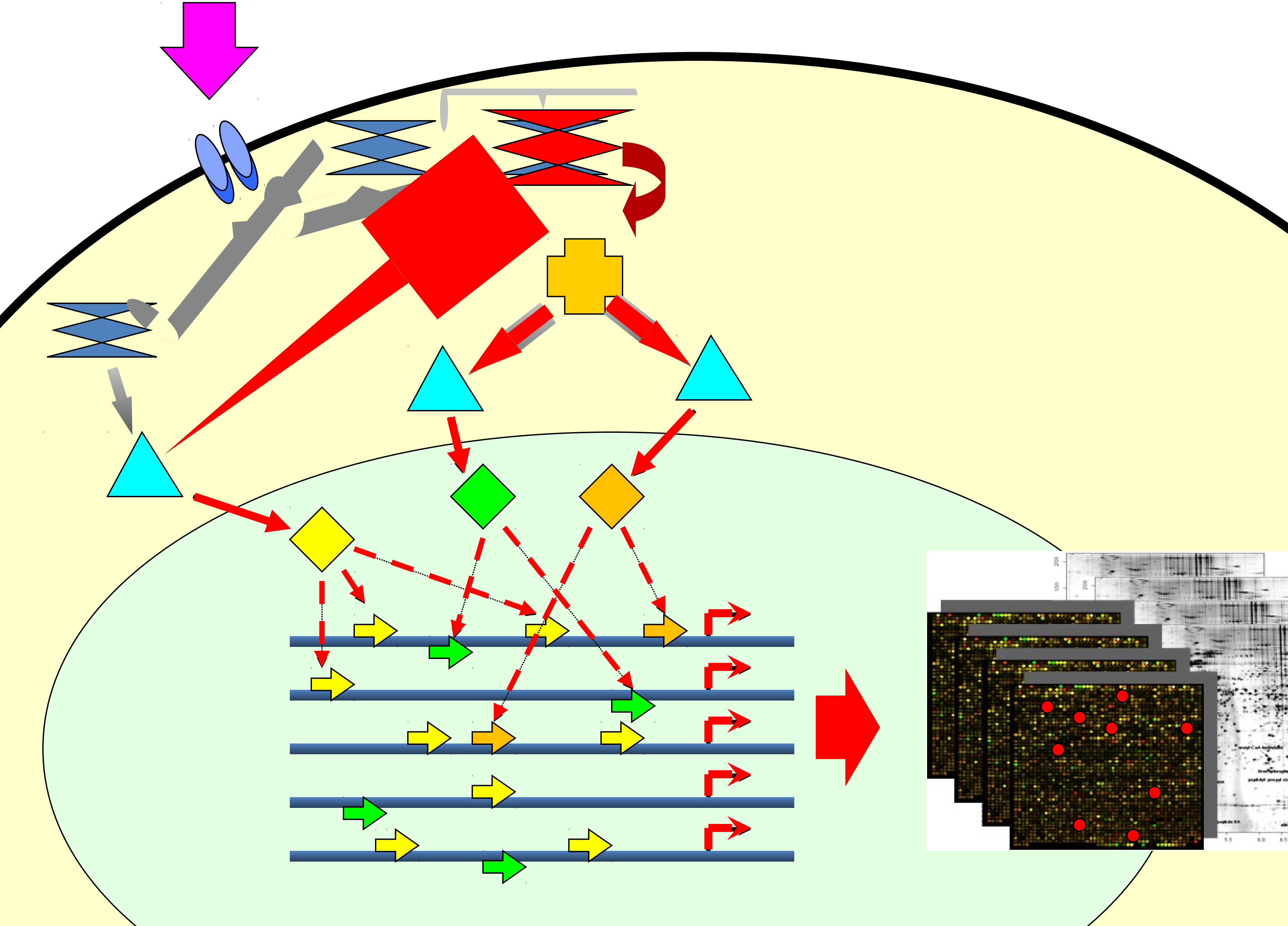
Showing 1 to 27 of 27 entries

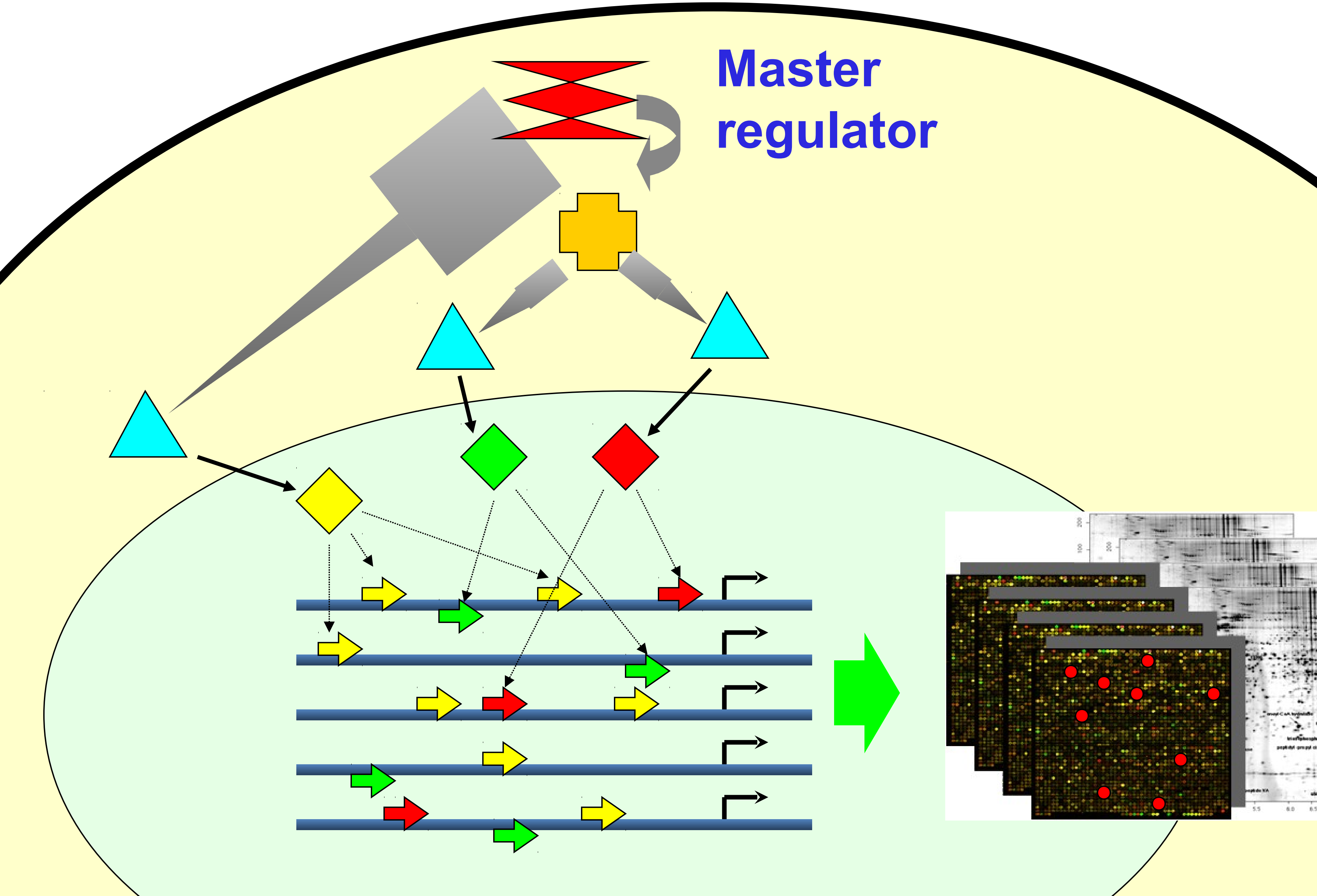
ID	Yes density per 1000bp	No density per 1000bp	Yes-no ratio	Matrix cutoff	P-value
V$NFKB_Q6_01	0.30962	0.06917	4.47619	0.9337	1.3743E-5
V$PPARG_01	0.36891	0.12846	2.87179	0.7339	1.4549E-4
V$NKX3A_01	0.38208	0.13834	2.7619	0.9106	1.6526E-4
V$PAX4_03	0.55336	0.2668	2.07407	0.9848	3.7636E-4
V$XVENT1_01	0.14493	0.01976	7.33333	0.9328	6.6179E-4
V$CP2_02	0.39526	0.16798	2.35294	0.9243	6.855E-4
V$ZF5_B	0.21739	0.05929	3.66667	0.9211	8.5009E-4
V$NKX22_01	0.47431	0.22727	2.08696	0.8995	8.9216E-4
V$OCT1_07	0.49407	0.24704	2	0.8372	0.00119
V$COREBINDINGFACTOR_Q6	0.11199	0.00988	11.33333	1	0.00132
V$CEBPDELTA_Q6	0.21739	0.06917	3.14286	0.9615	0.00205
V$IRF2_01	0.1054	0.00988	10.66667	0.909	0.00209
V$POU3F2_02	0.16469	0.03953	4.16667	0.8875	0.0022
V$PAX_Q6	0.19104	0.05929	3.22222	0.8706	0.0034
V$AREB6_03	0.34256	0.16798	2.03922	0.9617	0.00546
V$POU1F1_Q6	0.2108	0.07905	2.66667	0.9594	0.00606
V$IRF_Q6	0.18445	0.06917	2.66667	0.9707	0.01017
V$AP2_Q6	0.07246	0.00988	7.33333	0.9678	0.01959
V$PBX_Q3	0.20422	0.09881	2.06667	0.9151	0.02736
V$DMRT3_01	0.06588	0.00988	6.66667	0.9238	0.03023
V$TTF1_Q6	0.1581	0.06917	2.28571	0.9881	0.03299
V$AR_Q2	0.07905	0.01976	4	0.8671	0.03979
V$HAND1E47_01	0.07905	0.01976	4	0.9652	0.03979
V$AHR_Q5	0.05929	0.00988	6	0.9959	0.04636
V$HOXA7_01	0.13175	0.05929	2.22222	1	0.05588
V$KAISO_01	0.13175	0.05929	2.22222	0.9897	0.05588

ID: summary subset
Size: 27
Complete name: data/Projects/jamado@itqb.unl.pt/Data/TNF Genes Ensembl sites -1000..100/summary subset

Template to construct the filtering expression:
- Select template -

Columns (double-click to paste):
ID
Yes_density_per_1000bp

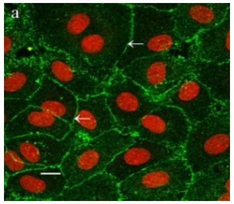

200
100
200
5.5
6.0
6.5

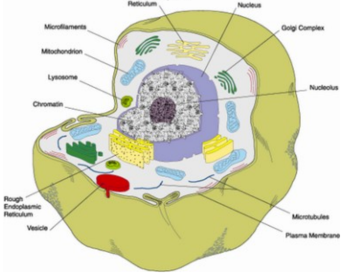


Master regulator

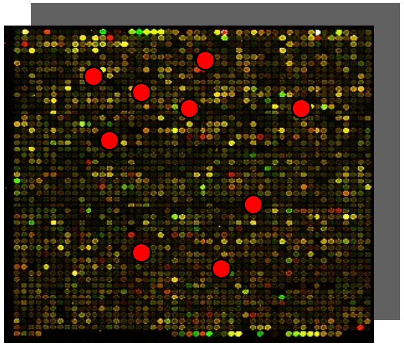
Can we predict TNF pathway?

?



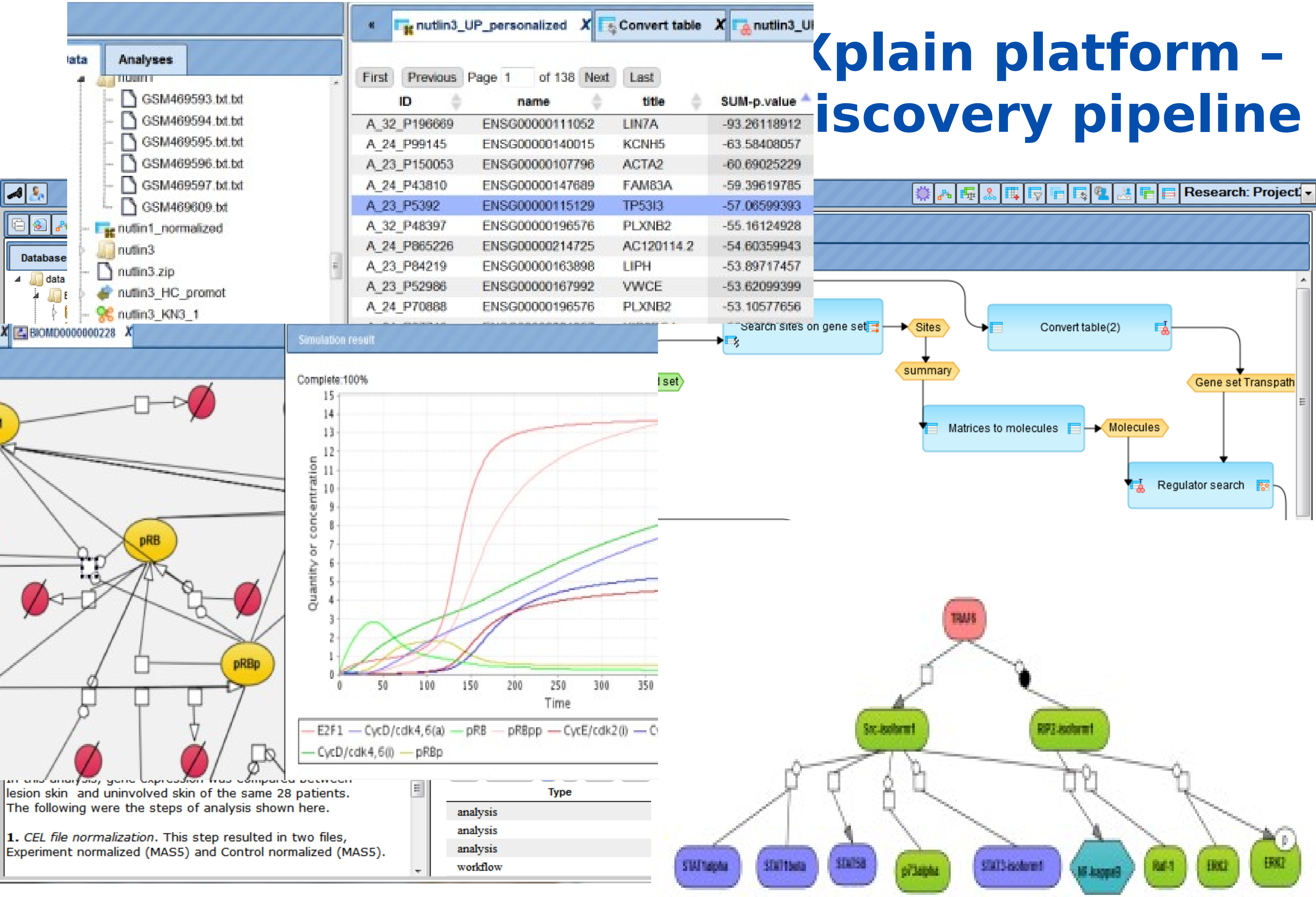


117 differntially expressed genes



3/17/12

xplain platform –
iscovery pipeline



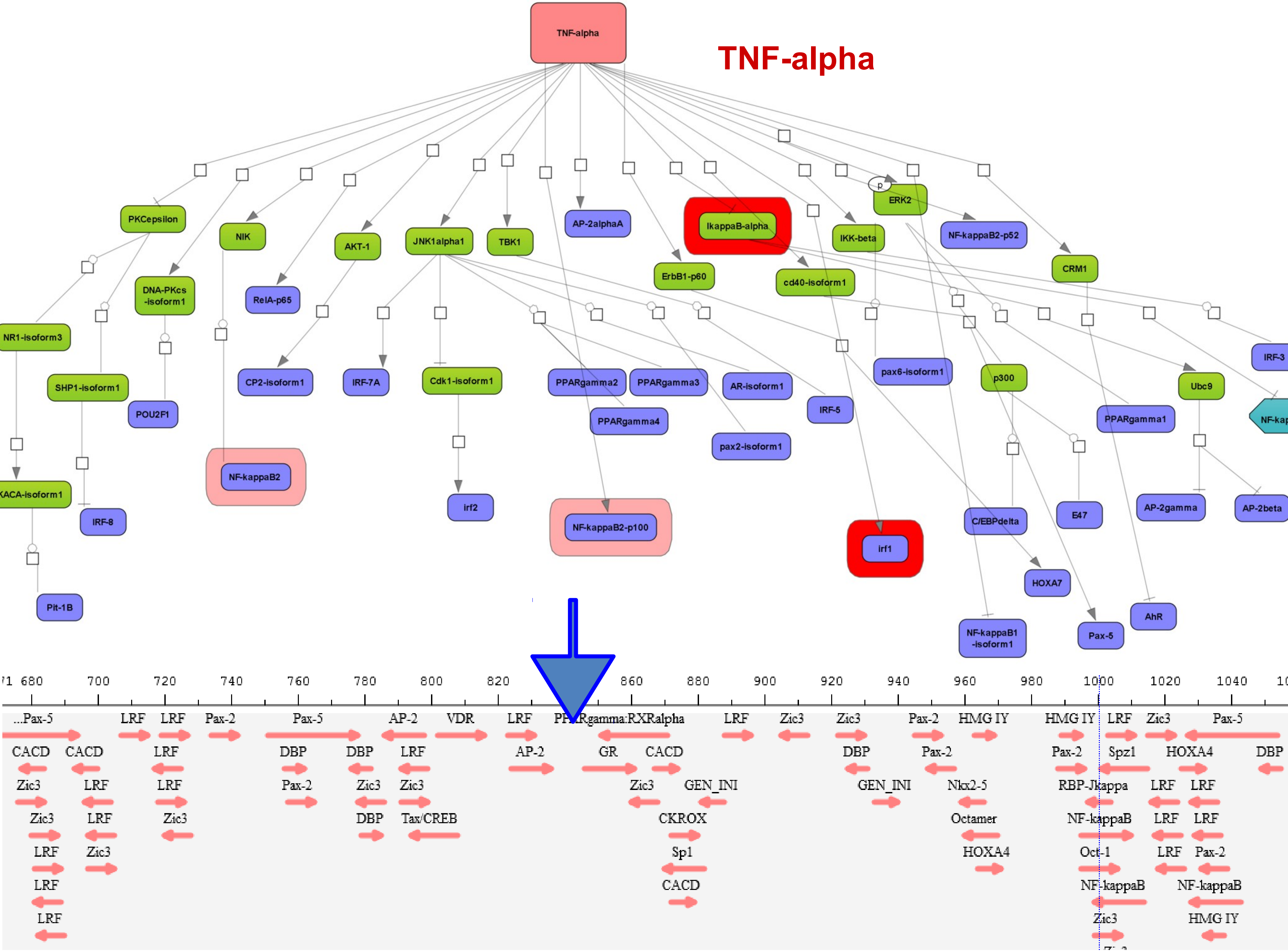
ID	name	title	SUM-p.value
A_32_P196669	ENSG00000111052	LIN7A	-93.26118912
A_24_P99145	ENSG00000140015	KCNH5	-63.58408057
A_23_P150053	ENSG00000107796	ACTA2	-60.69025229
A_24_P43810	ENSG00000147689	FAM83A	-59.39619785
A_23_P5392	ENSG00000115129	TP53I3	-57.06599393
A_32_P48397	ENSG00000196576	PLXNB2	-55.16124928
A_24_P865226	ENSG00000214725	AC120114.2	-54.60359943
A_23_P84219	ENSG00000163898	LIPH	-53.89717457
A_23_P52986	ENSG00000167992	VWCE	-53.62099399
A_24_P70888	ENSG00000196576	PLXNB2	-53.10577656

In this analysis, gene expression was compared between lesion skin and uninvolved skin of the same 28 patients. The following were the steps of analysis shown here.

1. *CEL file normalization*. This step resulted in two files, Experiment normalized (MAS5) and Control normalized (MAS5).

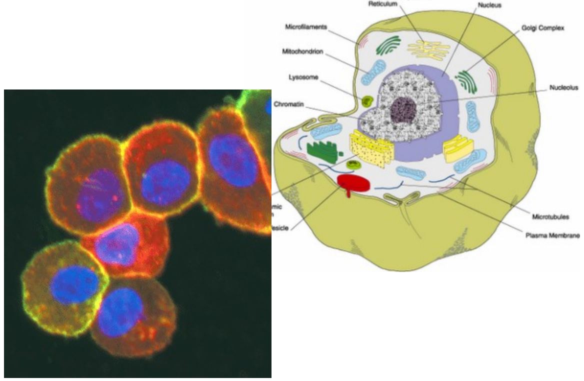
Type
analysis
analysis
analysis
workflow

TNF-alpha

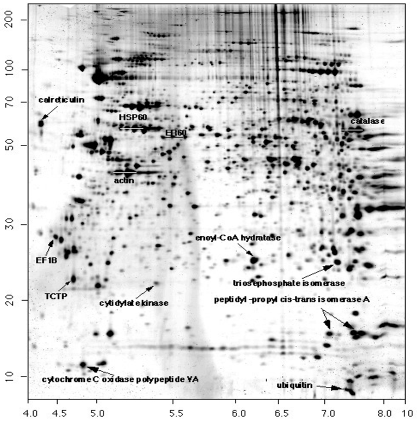

TNF-alpha
PKCepsilon
NIK
AKT-1
JNK1alpha1
TBK1
AP-2alphaA
IkappaB-alpha
ERK2
p
IKK-beta
NF-kappaB2-p52
CRM1
ErbB1-p60
cd40-isoform1
DNA-PKcs
-isoform1
RelA-p65
NR1-isoform3
SHP1-isoform1
POU2F1
CP2-isoform1
IRF-7A
Cdk1-isoform1
PPARgamma2
PPARgamma3
AR-isoform1
pax6-isoform1
p300
Ubc9
IRF-3
NF-kap
PPARgamma4
IRF-5
PPARgamma1
pax2-isoform1
KACA-isoform1
IRF-8
NF-kappaB2
irf2
NF-kappaB2-p100
irf1
C/EBPdelta
E47
AP-2gamma
AP-2beta
HOXA7
Pit-1B
NF-kappaB1
-isoform1
Pax-5
AhR
71
680
700
720
740
760
780
800
820
860
880
900
920
940
960
980
1000
1020
1040
10
...Pax-5
LRF
LRF
Pax-2
Pax-5
AP-2
VDR
LRF
PPARgamma:RXRalpha
LRF
Zic3
Zic3
Pax-2
HMG IY
HMG IY
LRF
Zic3
Pax-5
CACD
CACD
LRF
DBP
DBP
LRF
AP-2
GR
CACD
DBP
Pax-2
Pax-2
Spz1
HOXA4
DBP
Zic3
LRF
LRF
Pax-2
Zic3
Zic3
Zic3
GEN_INI
GEN_INI
Nkx2-5
RBP-Jkappa
LRF
LRF
Zic3
LRF
Zic3
DBP
Tax/CREB
CKROX
Octamer
NF-kappaB
LRF
LRF
LRF
Zic3
Sp1
HOXA4
Oct-1
LRF
Pax-2
LRF
CACD
NF-kappaB
NF-kappaB
LRF
Zic3
HMG IY

Human epidermoid carcinoma A431 cells treated by epidermal growth factor (EGF)

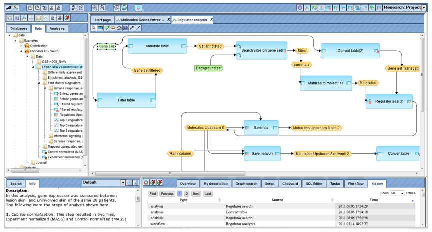
<u>EGF</u> ?



320 differntially expressed proteins



Master regulator analysis



EGF was still not in the list !

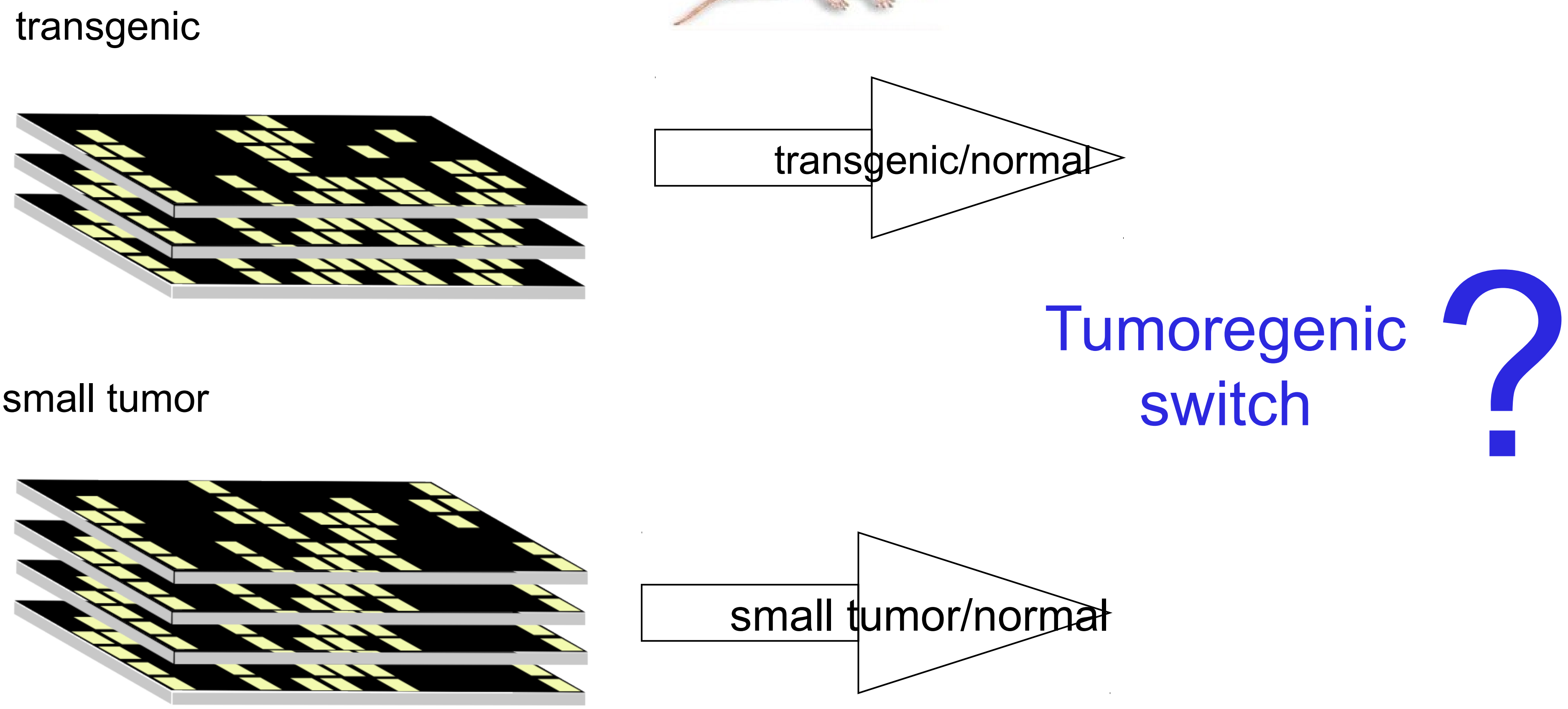
Pathways are far .....far....far from being fully undersood!

Network plasticity

Epidermal Growth Factor induced Carcinogenicity

Philip Stegmaier1, Alexander Kel1, Edgar Wingender1,2, and Jürgen Borlak3

Hepatocellular transcriptome data of IgEGF-overexpressing mice



transgenic

transgenic/normal

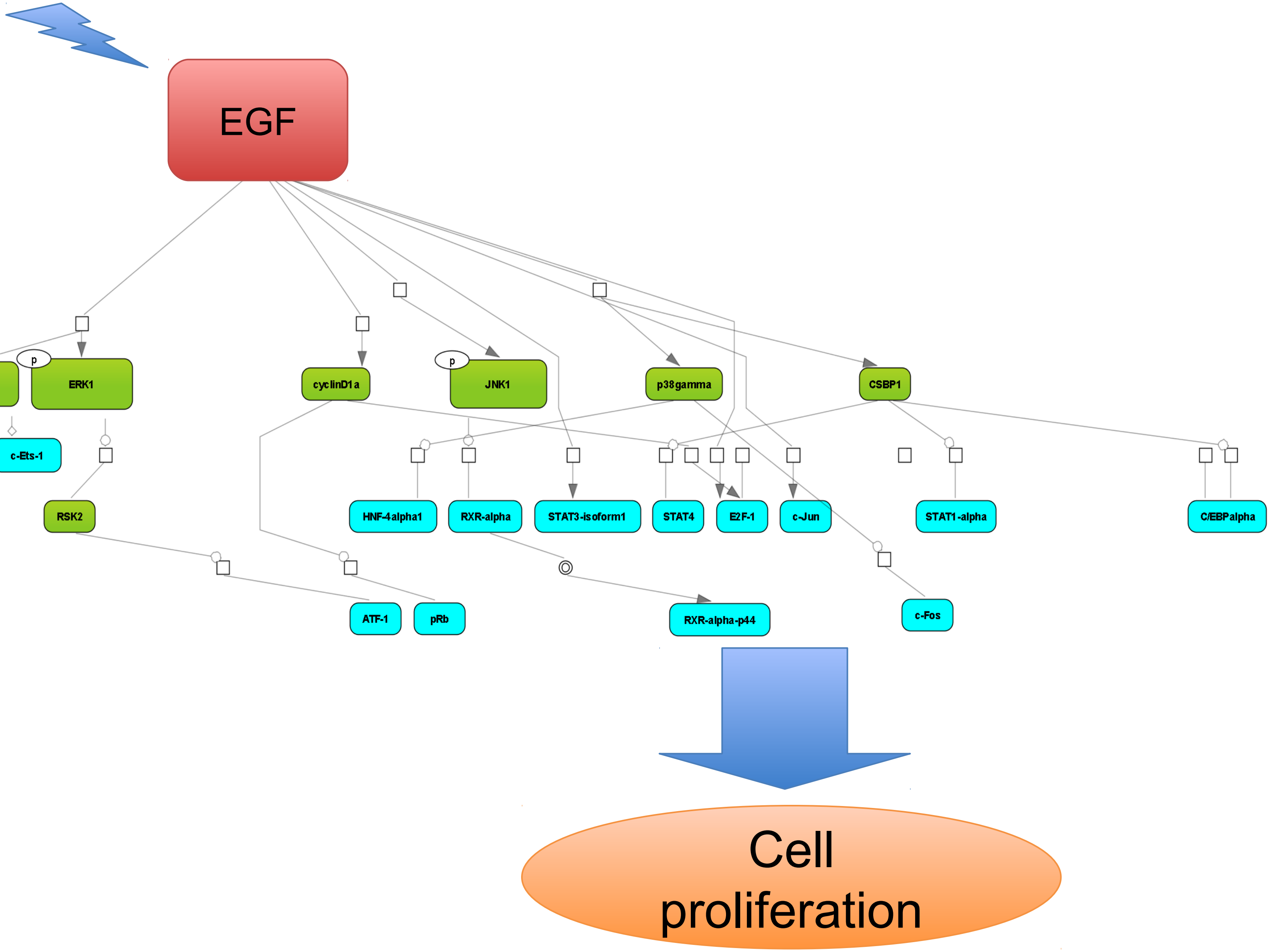
Tumoregenic switch ?

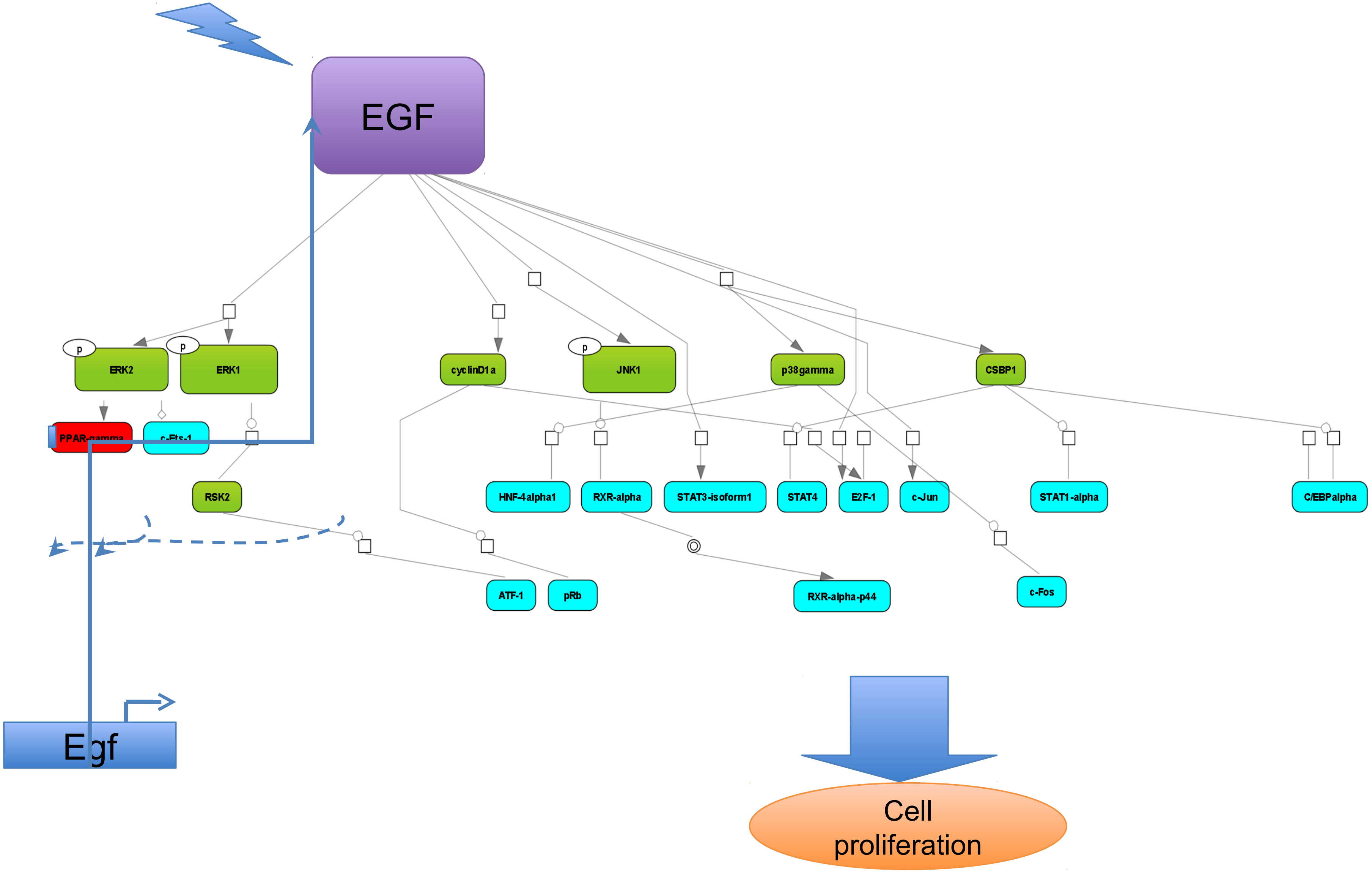
small tumor

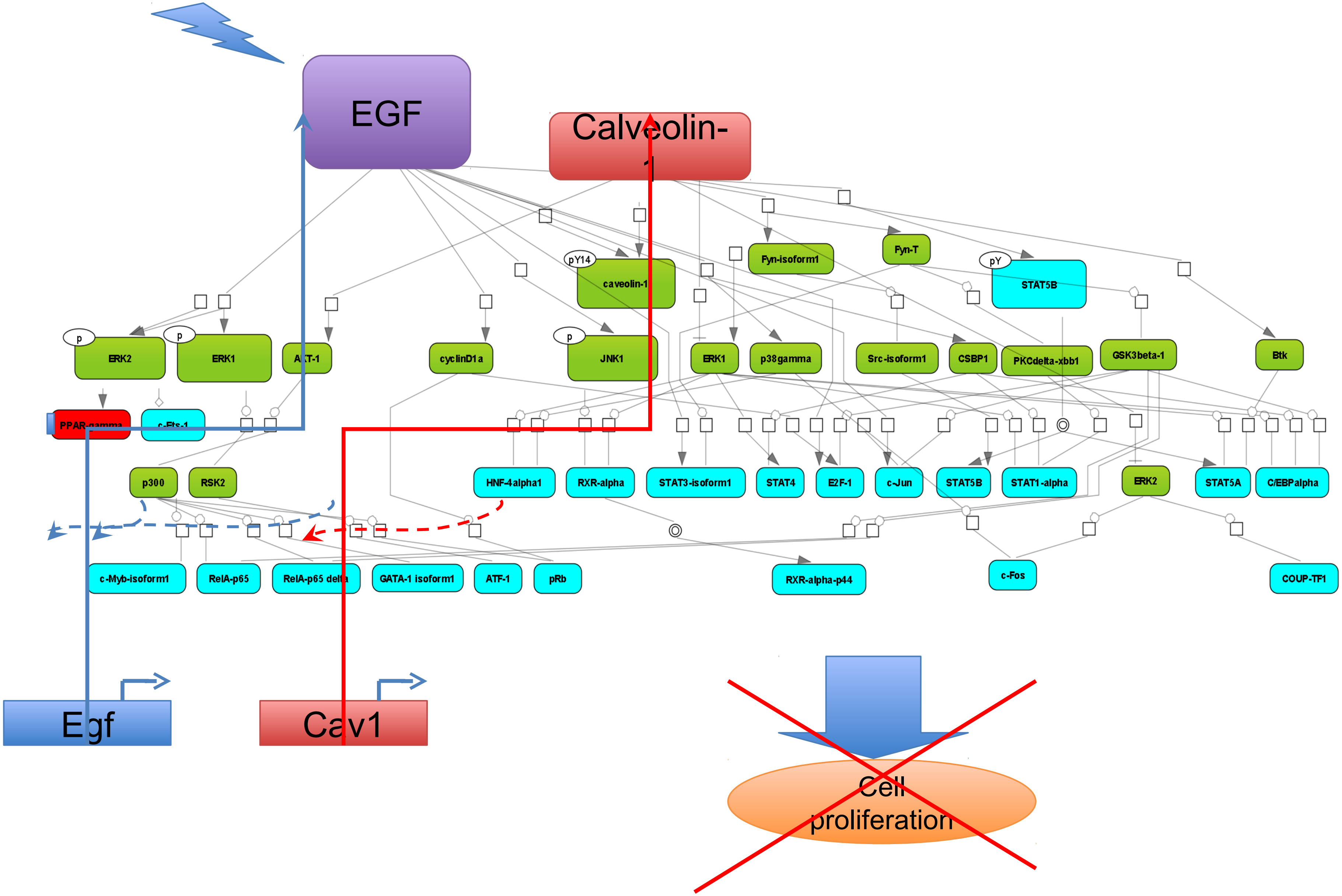
small tumor/normal

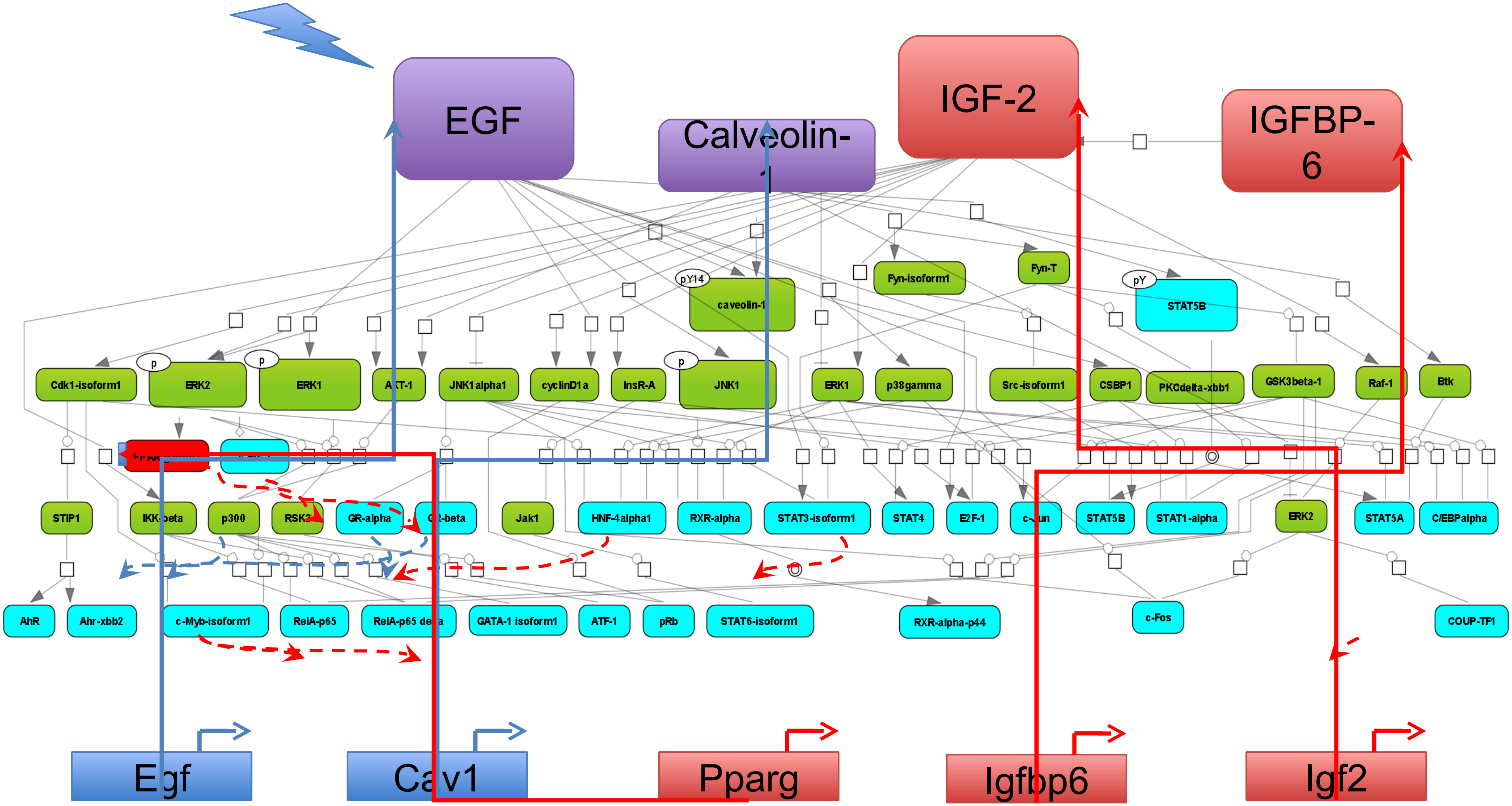
Experimental validation by EMSA

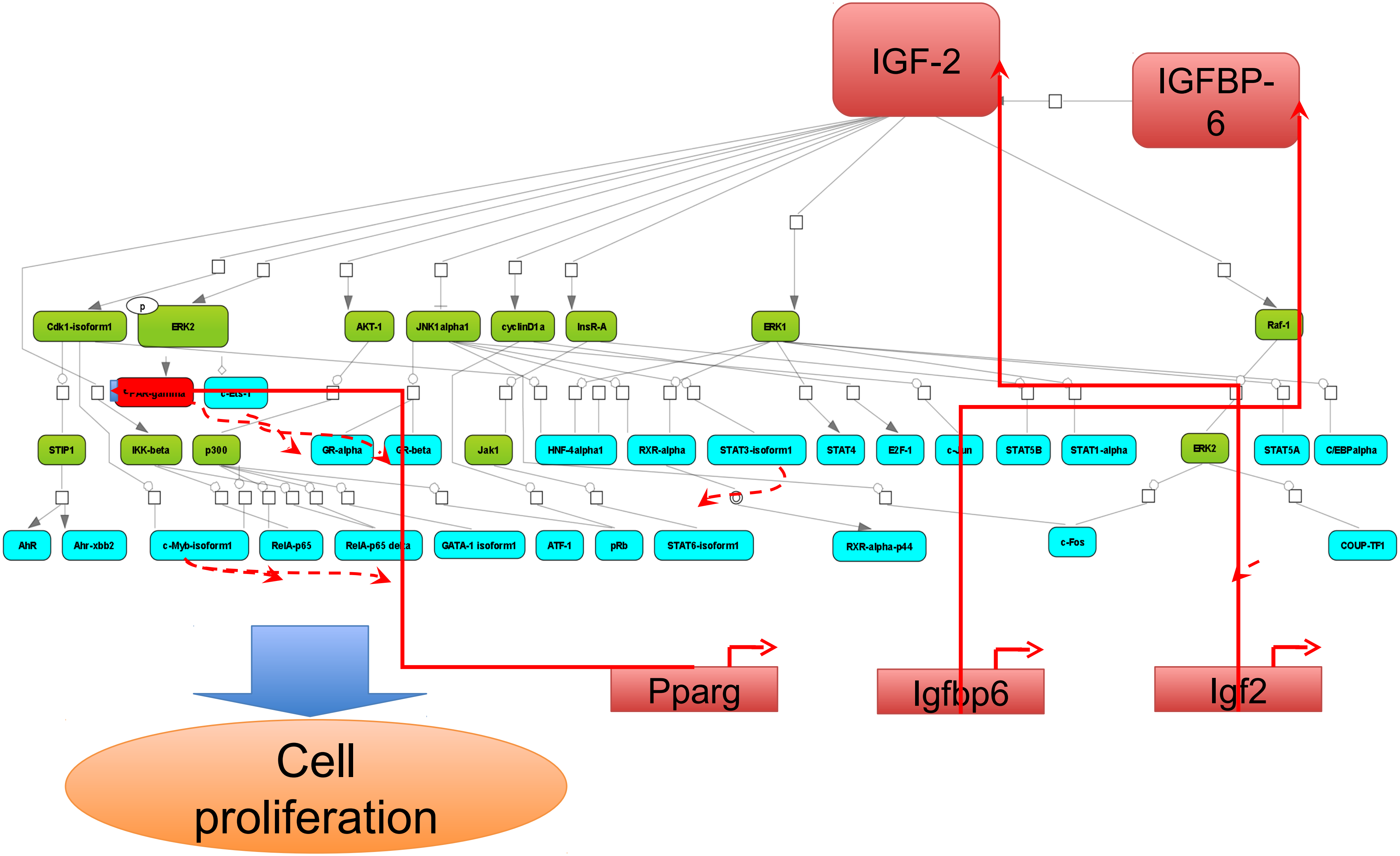
Gene	Fold change Transgenic	Fold change Tumor	PPAR-gamma	c-Myb	c-Ets-2	STAT5B	GATA-1	Mef2a	p53	HNF-4 gamma (alpha -antibody)	GR	C/EBP alpha
Igf2	2.90	25.98	-	+	-			-				
Il1rn	2.75	8.49		-	-	+						
Igfbp6	0.83	7.84	+	-	+	+						
Pparg	0.40	5.76		+	-	+	+	-				
Bmp7	0.92	4.64		+	-	-						
Zbtb7b	3.69	1.11							+			+
Foxc1	4.67	1.43								+	-	+
Xlr	2.90	0.90						-	+			
Erbb3	6.99	2.35	-						+			
Itga4	2.76	1.03							-			+
Th	4.08	1.59						+	-			+
Nr2f1	6.26	2.62							+	+		
Defcr6	13.86	6.05								+	-	+
Nr3c1	3.27	1.44							-		-	+
Cav1	5.35	2.49							-		+	+


EGF
p
ERK2
p
ERK1
cyclinD1a
p
JNK1
p38gamma
CSBP1
PPAR-gamma
c-Ets-1
RSK2
HNF-4alpha1
RXR-alpha
STAT3-isoform1
STAT4
E2F-1
c-Jun
STAT1-alpha
C/EBPalpha
ATF-1
pRb
RXR-alpha-p44
c-Fos
Cell proliferation

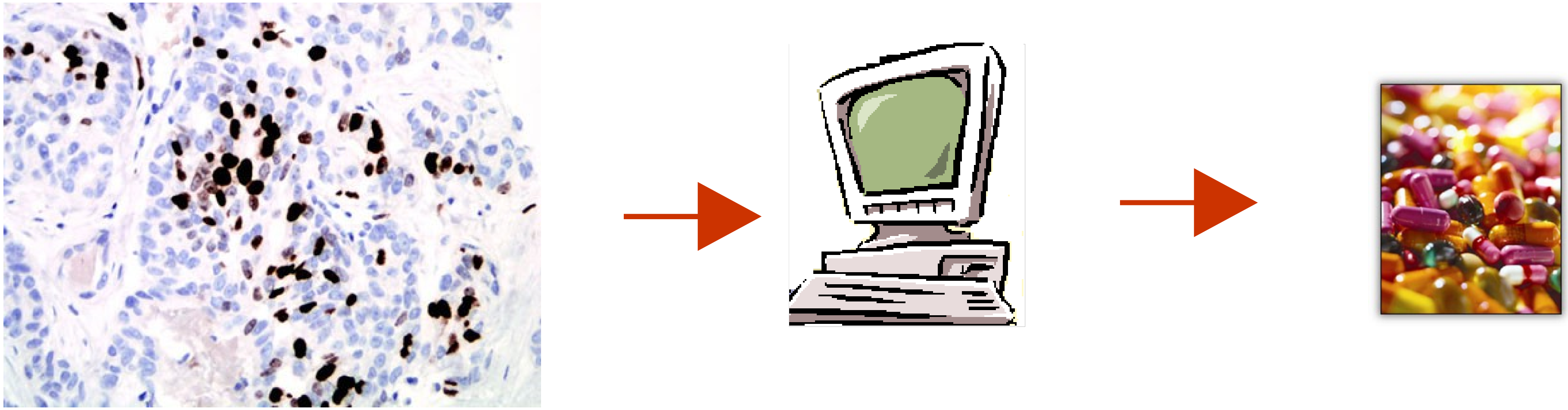

EGF
p
ERK2
p
ERK1
cyclinD1a
p
JNK1
p38gamma
CSBP1
PPAR-gamma
c-Ets-1
RSK2
HNF-4alpha1
RXR-alpha
STAT3-isoform1
STAT4
E2F-1
c-Jun
STAT1-alpha
C/EBPalpha
ATF-1
pRb
RXR-alpha-p44
c-Fos
Egf
Cell proliferation


EGF
Calveolin-1
pY14
caveolin-1
Fyn-isoform1
Fyn-T
pY
STAT5B
p
ERK2
p
ERK1
AKT-1
cyclinD1a
p
JNK1
ERK1
p38gamma
Src-isoform1
CSBP1
PKCdelta-xbb1
GSK3beta-1
Btk
PPAR-gamma
c-Ets-1
p300
RSK2
HNF-4alpha1
RXR-alpha
STAT3-isoform1
STAT4
E2F-1
c-Jun
STAT5B
STAT1-alpha
ERK2
STAT5A
C/EBPalpha
c-Myb-isoform1
RelA-p65
RelA-p65 delta
GATA-1 isoform1
ATF-1
pRb
RXR-alpha-p44
c-Fos
COUP-TF1
Egf
Cav1
Cell proliferation

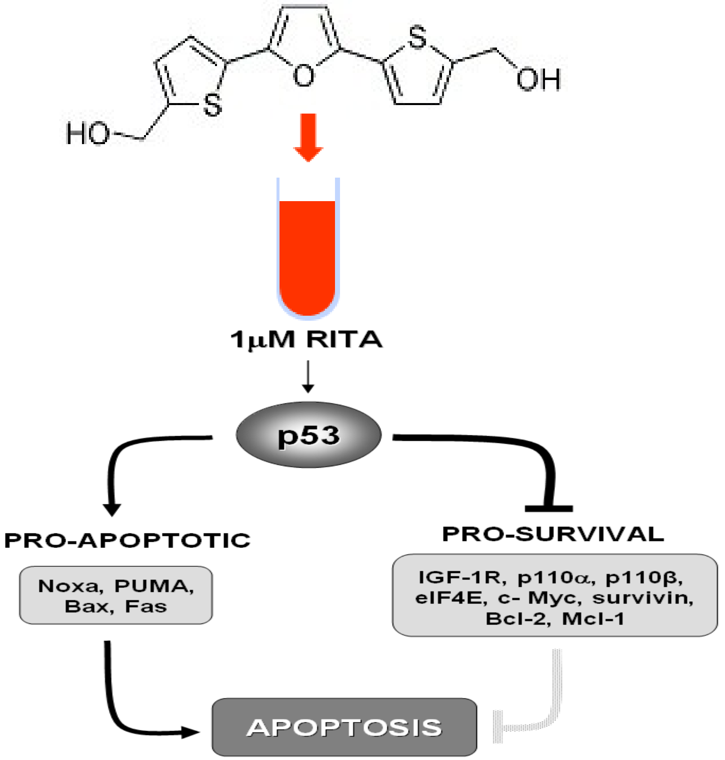

EGF
Calveolin-1
IGF-2
IGFBP-6
pY14
caveolin-1
Fyn-isoform1
Fyn-T
pY
STAT5B
p
p
p
Cdk1-isoform1
ERK2
ERK1
AKT-1
JNK1alpha1
cyclinD1a
InsR-A
JNK1
ERK1
p38gamma
Src-isoform1
CSBP1
PKCdelta-xbb1
GSK3beta-1
Raf-1
Btk
STIP1
IKK-beta
p300
RSK2
GR-alpha
GR-beta
Jak1
HNF-4alpha1
RXR-alpha
STAT3-isoform1
STAT4
E2F-1
c-Jun
STAT5B
STAT1-alpha
ERK2
STAT5A
C/EBPalpha
AhR
Ahr-xbb2
c-Myb-isoform1
RelA-p65
RelA-p65 delta
GATA-1 isoform1
ATF-1
pRb
STAT6-isoform1
RXR-alpha-p44
c-Fos
COUP-TF1
Egf
Cav1
Pparg
Igfbp6
Igf2


IGF-2
IGFBP-6
Cdk1-isoform1
p
ERK2
AKT-1
JNK1alpha1
cyclinD1a
InsR-A
ERK1
Raf-1
PPAR-gamma
c-Ets-1
STIP1
IKK-beta
p300
GR-alpha
GR-beta
Jak1
HNF-4alpha1
RXR-alpha
STAT3-isoform1
STAT4
E2F-1
c-Jun
STAT5B
STAT1-alpha
ERK2
STAT5A
C/EBPalpha
AhR
Ahr-xbb2
c-Myb-isoform1
RelA-p65
RelA-p65 delta
GATA-1 isoform1
ATF-1
pRb
STAT6-isoform1
RXR-alpha-p44
c-Fos
COUP-TF1
Pparg
Igfbp6
Igf2
Cell proliferation

Cancer



Apoptosis versus survival of cancer



1μM RITA

p53

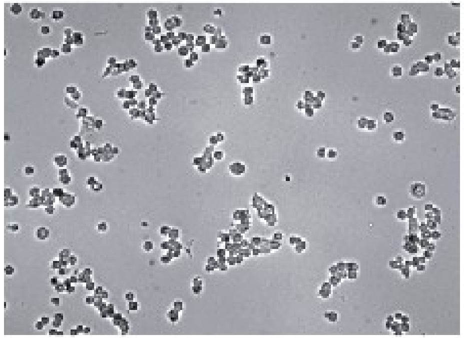
PRO-APOPTOTIC

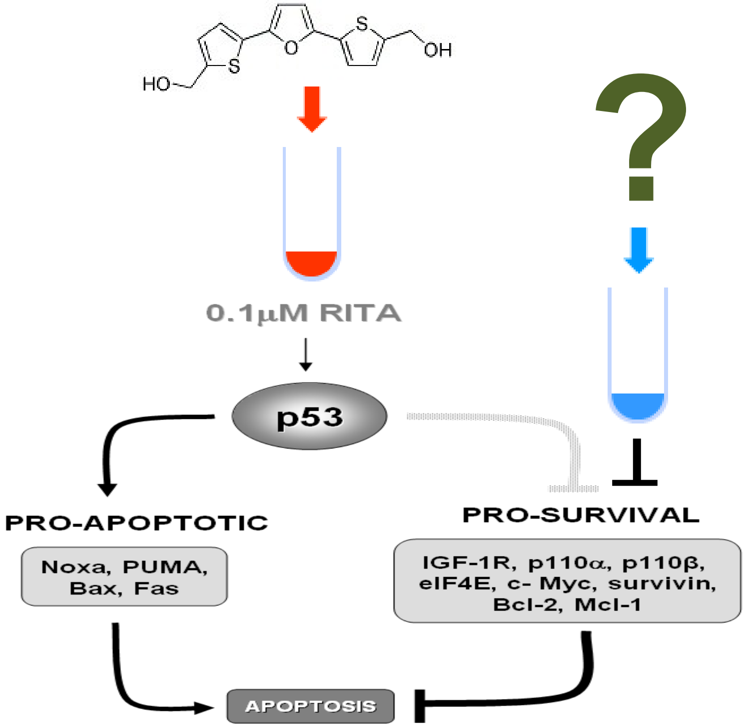
Noxa, PUMA, Bax, Fas

PRO-SURVIVAL

IGF-1R, p110α, p110β, eIF4E, c- Myc, survivin, Bcl-2, Mcl-1

APOPTOSIS





0.1μM RITA

p53

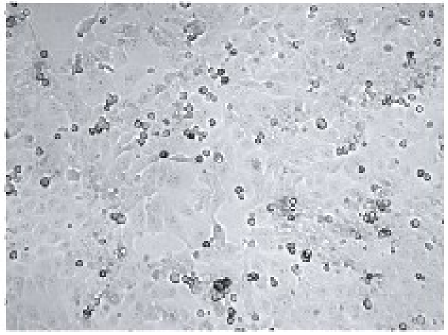
PRO-APOPTOTIC

Noxa, PUMA, Bax, Fas

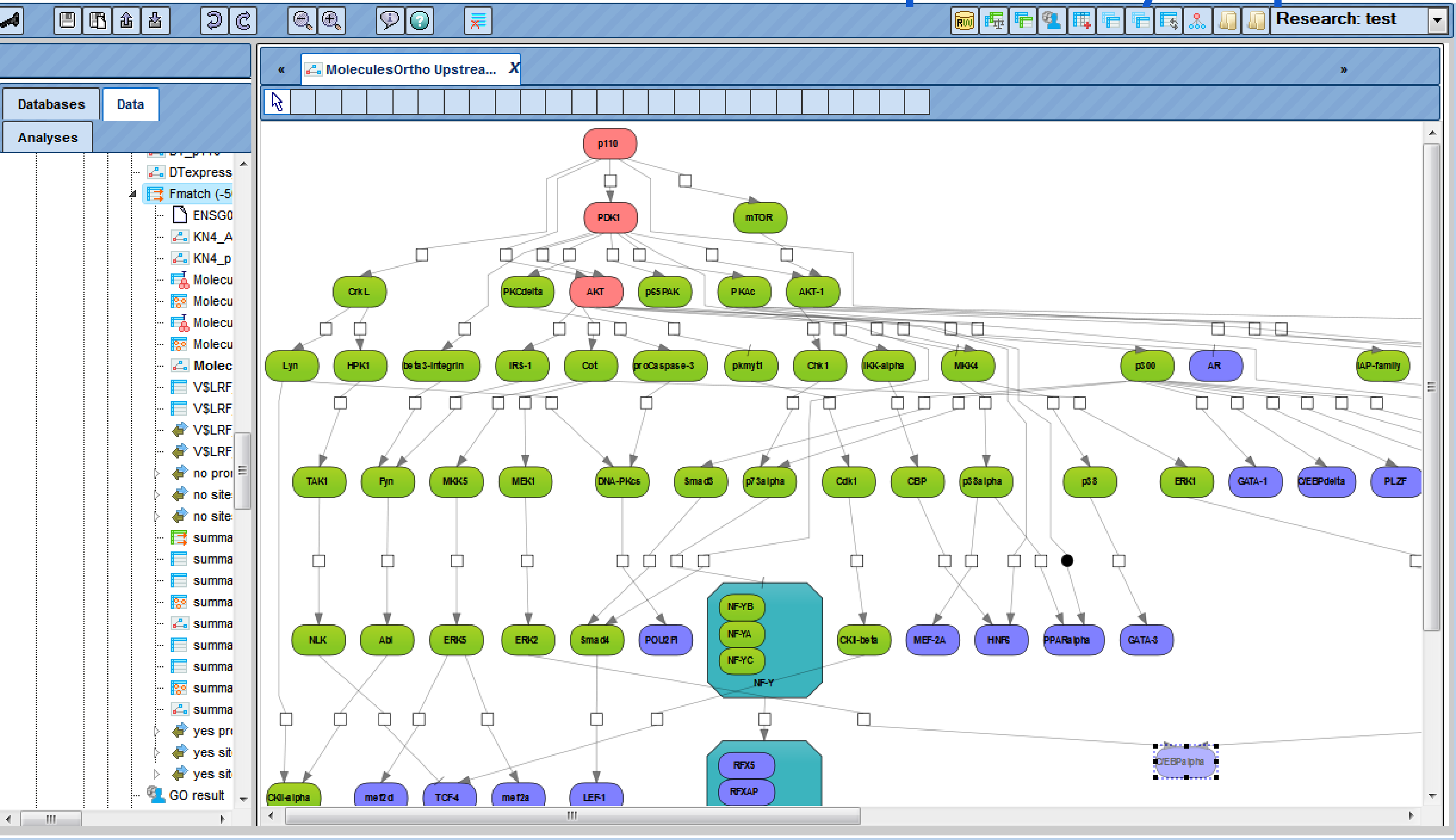
PRO-SURVIVAL

IGF-1R, p110α, p110β, eIF4E, c- Myc, survivin, Bcl-2, Mcl-1

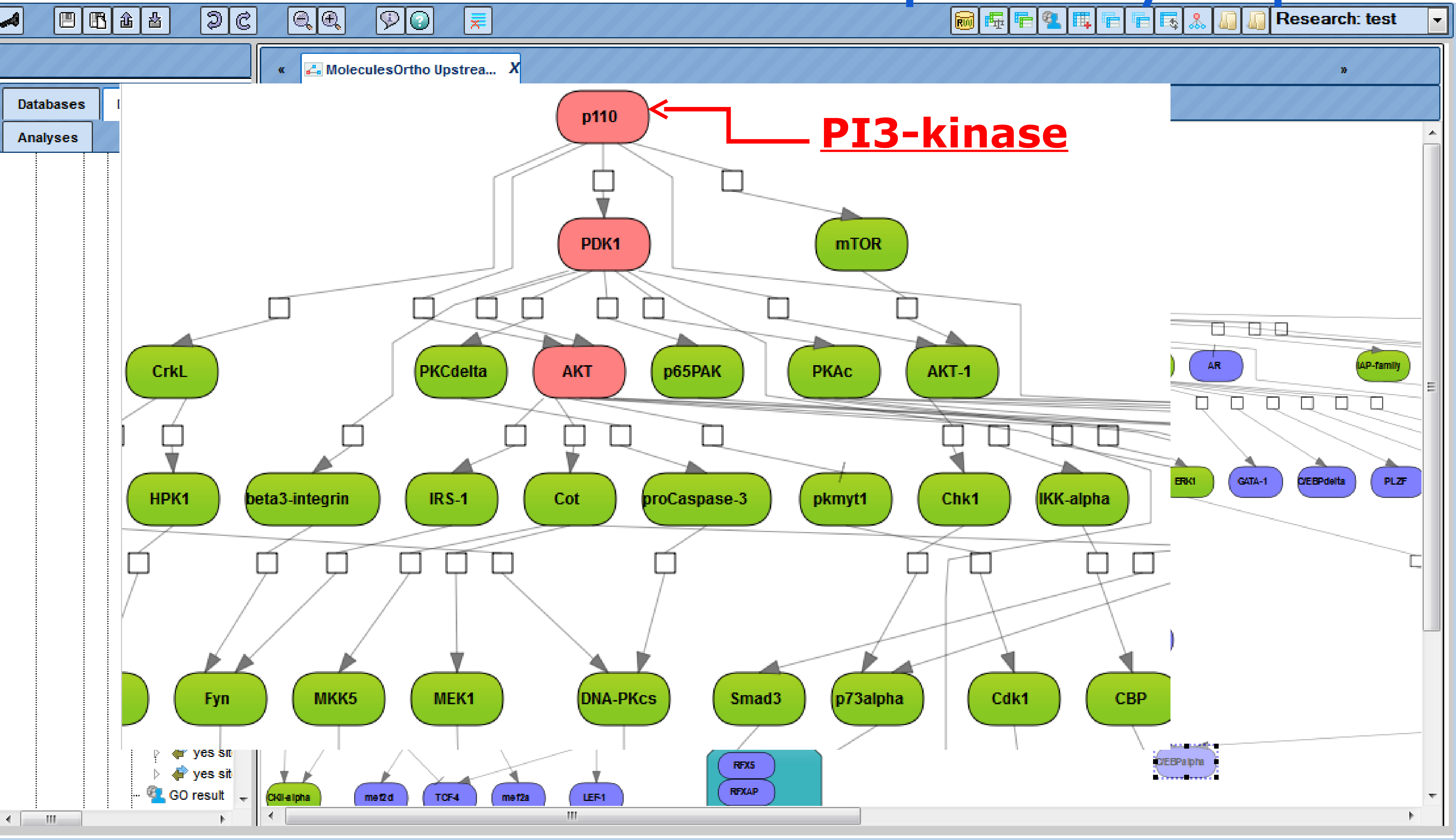
APOPTOSIS



Survival mechanisms of cancer cells upon RITA treatment and potential target proteins for a complementary compound

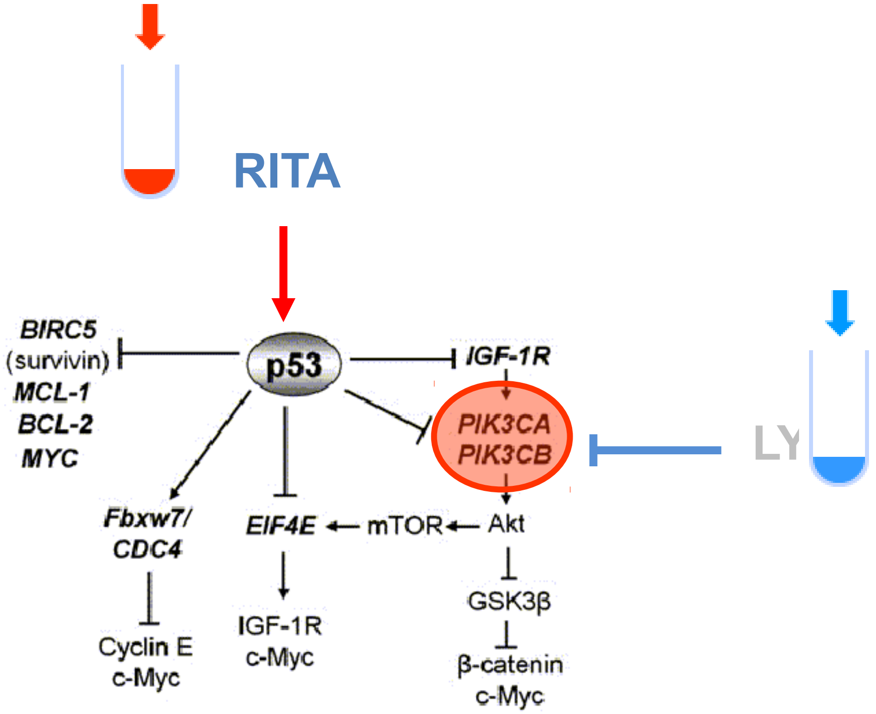


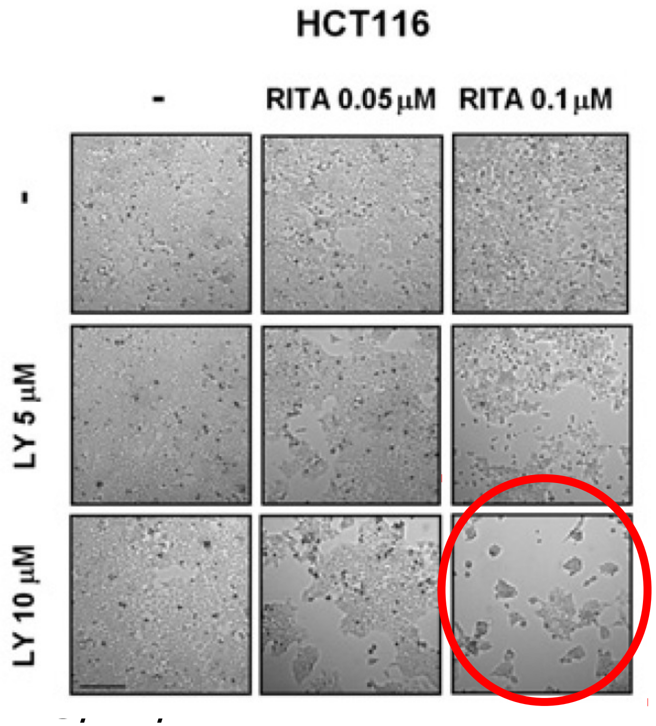
Survival mechanisms of cancer cells upon RITA treatment and potential target proteins for a complementary compound

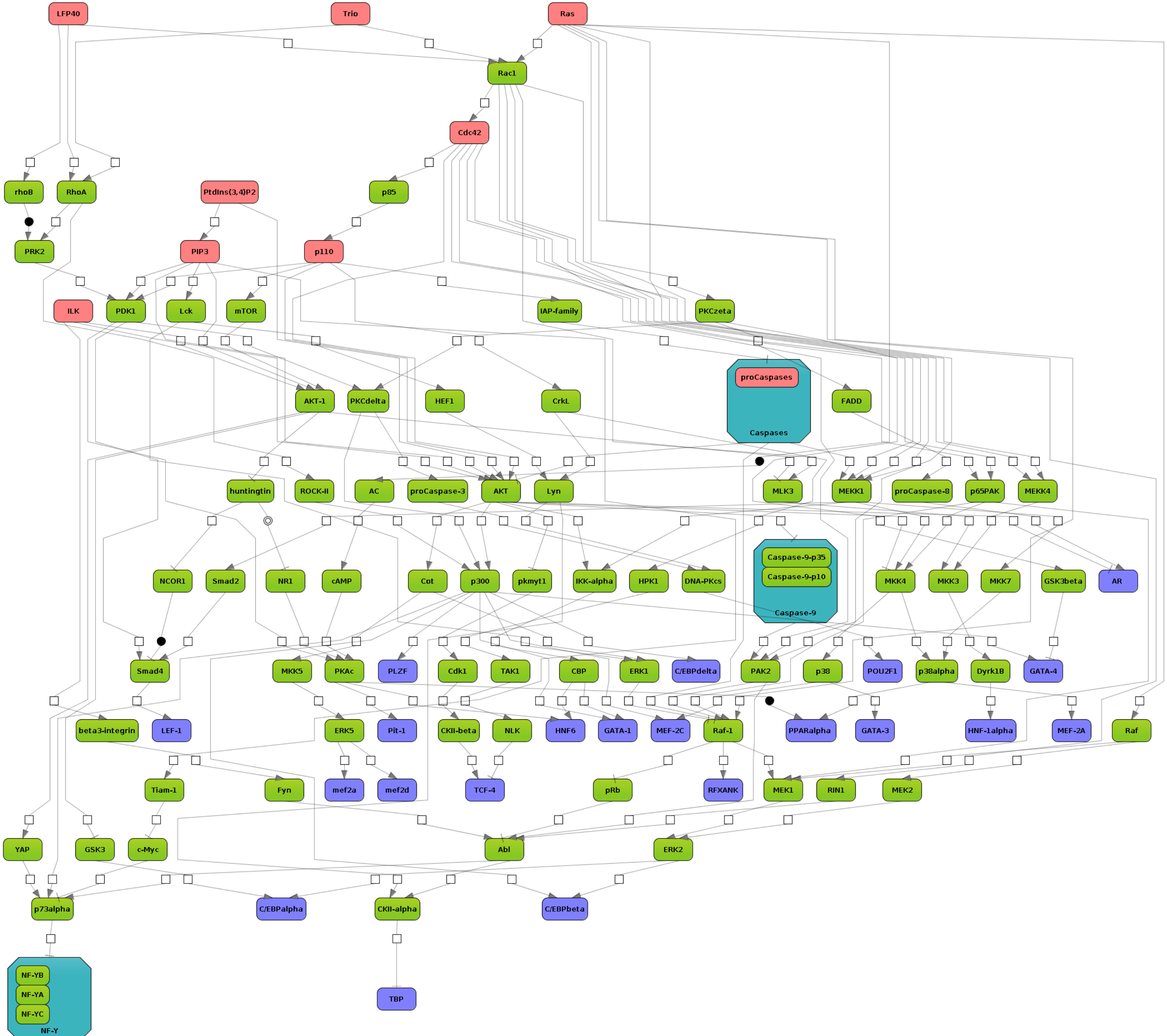


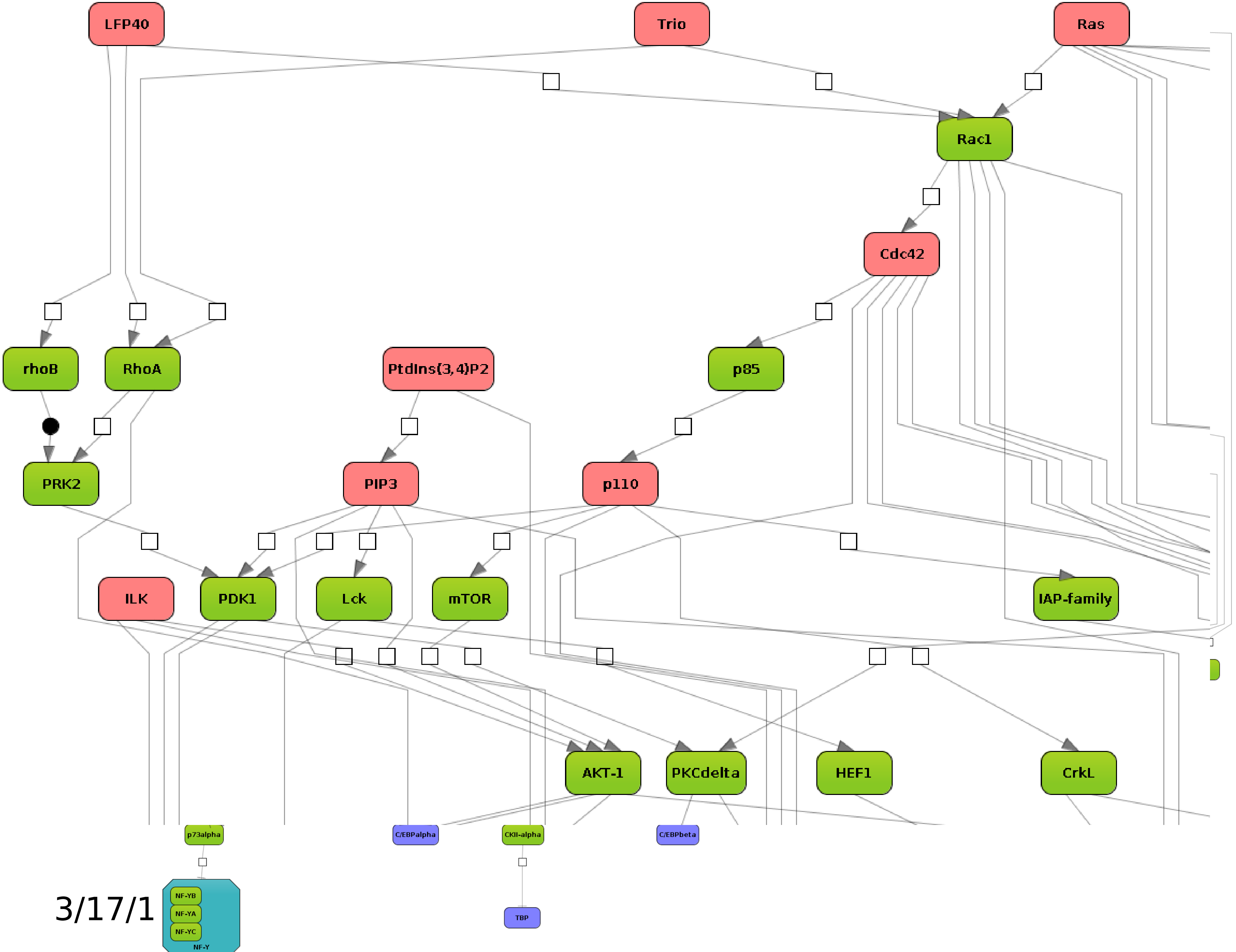
PI3-kinase

Death of Cancer cells treated with 0.1 μM RITA and <u>PI3-kinase</u> inhibitor LY294002






LFP40
Trio
Ras
Rac1
Cdc42
rhoB
RhoA
PtdIns(3,4)P2
p85
PRK2
PIP3
p110
ILK
PDK1
Lck
mTOR
IAP-family
PKCzeta
AKT-1
PKCdelta
HEF1
CrkL
proCaspases
Caspases
FADD
huntingtin
ROCK-II
AC
proCaspase-3
AKT
Lyn
MLK3
MEKK1
proCaspase-8
p65PAK
MEKK4
NCOR1
Smad2
NR1
cAMP
Cot
p300
pkmyt1
IKK-alpha
HPK1
DNA-PKcs
Caspase-9-p35
Caspase-9-p10
Caspase-9
MKK4
MKK3
MKK7
GSK3beta
AR
Smad4
MKK5
PKAc
PLZF
Cdk1
TAK1
CBP
ERK1
C/EBPdelta
PAK2
p38
POU2F1
p38alpha
Dyrk1B
GATA-4
beta3-integrin
LEF-1
ERK5
Pit-1
CKII-beta
NLK
HNF6
GATA-1
MEF-2C
Raf-1
PPARalpha
GATA-3
HNF-1alpha
MEF-2A
Raf
Tiam-1
Fyn
mef2a
mef2d
TCF-4
pRb
RFXANK
MEK1
RIN1
MEK2
YAP
GSK3
c-Myc
Abl
ERK2
p73alpha
C/EBPalpha
CKII-alpha
C/EBPbeta
NF-YB
NF-YA
NF-YC
NF-Y
TBP

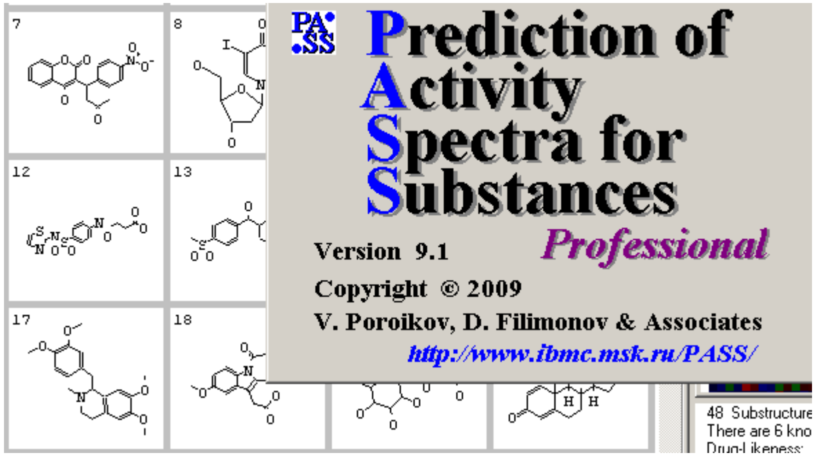

LFP40
Trio
Ras
Rac1
Cdc42
rhoB
RhoA
PtdIns(3,4)P2
p85
PRK2
PIP3
p110
ILK
PDK1
Lck
mTOR
IAP-family
AKT-1
PKCdelta
HEF1
CrkL
p73alpha
C/EBPalpha
CKII-alpha
C/EBPbeta
NF-YB
NF-YA
NF-YC
NF-Y
TBP

3/17/1

ChemNavigator Library

24 million compounds

…



Prediction of Activity Spectra for Substances

Version 9.1 *Professional*

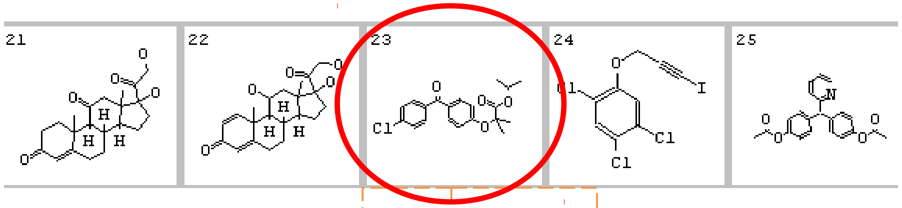
Copyright © 2009

V. Poroikov, D. Filimonov & Associates

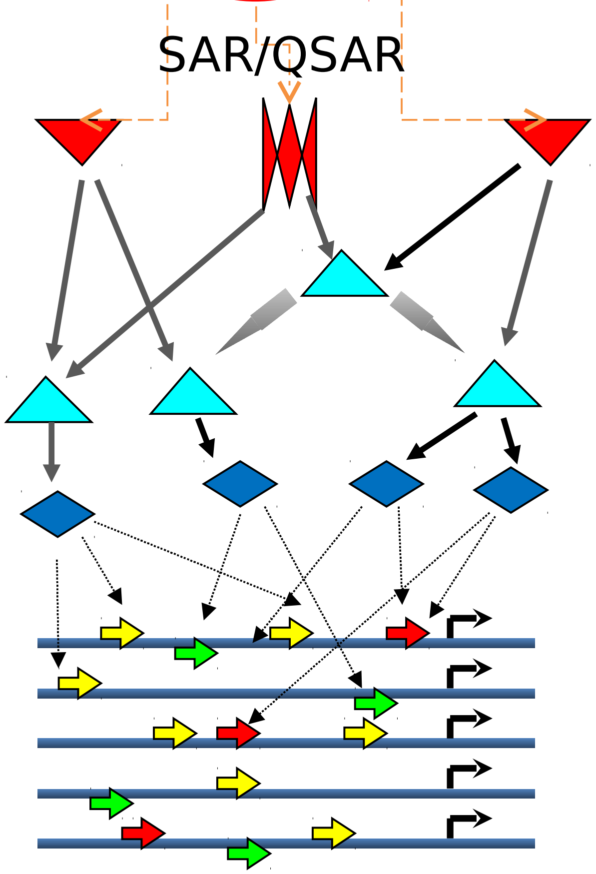
http://www.ibmc.msk.ru/PASS/

Identified

64 novel componds



…



3/17/12

Tested 16 compounds in a panel of several cancer cell lines.

Found active: Compound N15

Hypoxia inducible factor 1 alpha inhibitor

Phosphatidylinositol 3-kinase beta inhibitor

Targets

Showed growth suppression in 3 different breast cancer cell lines. The effect appears to be p53-independent (kills p53-null colon cancer cells) and it does not affect the growth of non-transformed mammary epithelial cells

Found active: Compound N6

Cyclin-dependent kinase 2 inhibitor

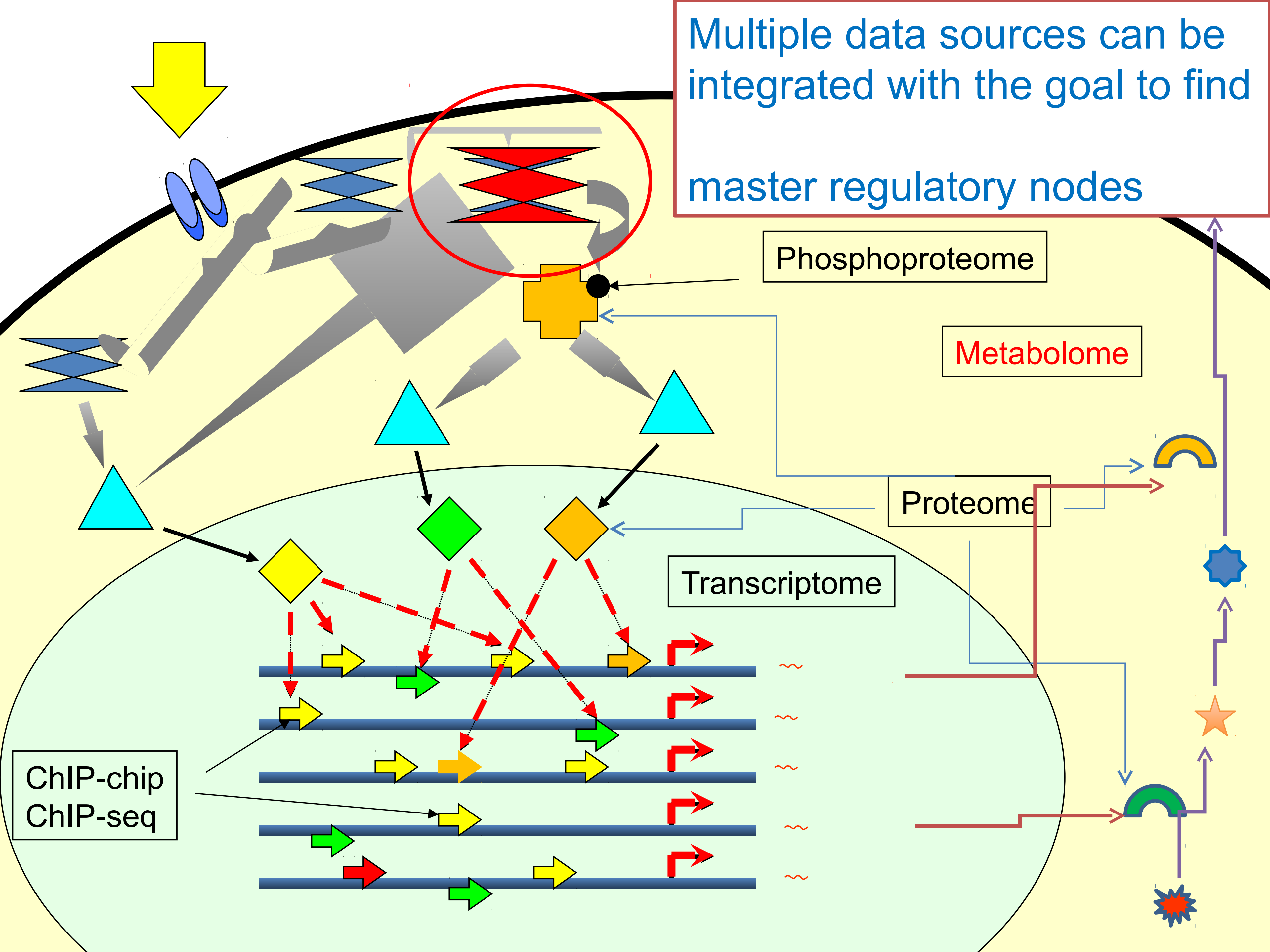
Myc inhibitor

Targets

Out of panel of 7 different cancer lines it killed only melanoma cells without any effects in other cell lines and on control non-transformed mammary epithelial cells.

Multiple data sources can be integrated with the goal to find

master regulatory nodes



Phosphoproteome

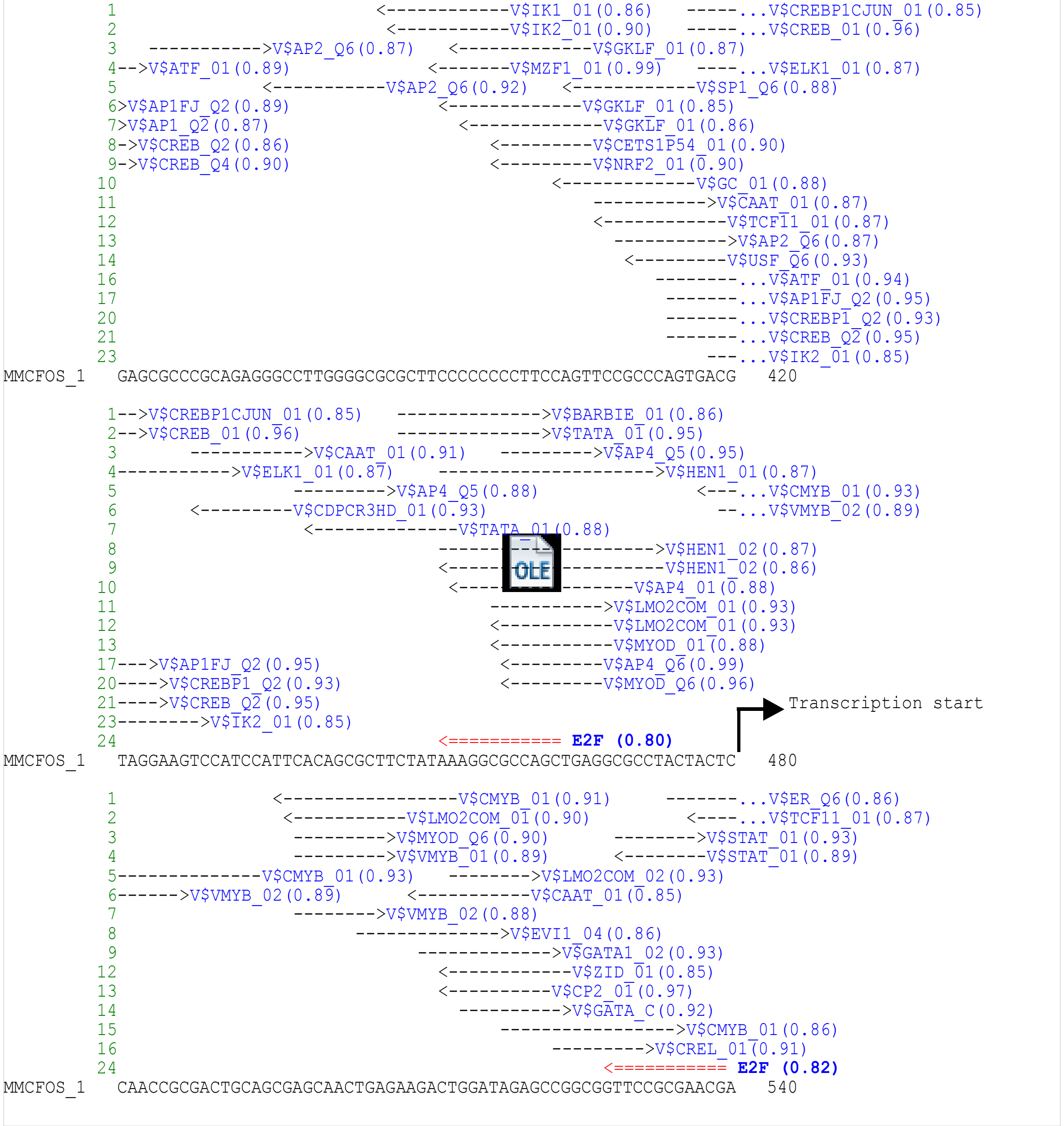
Metabolome

Proteome

Transcriptome

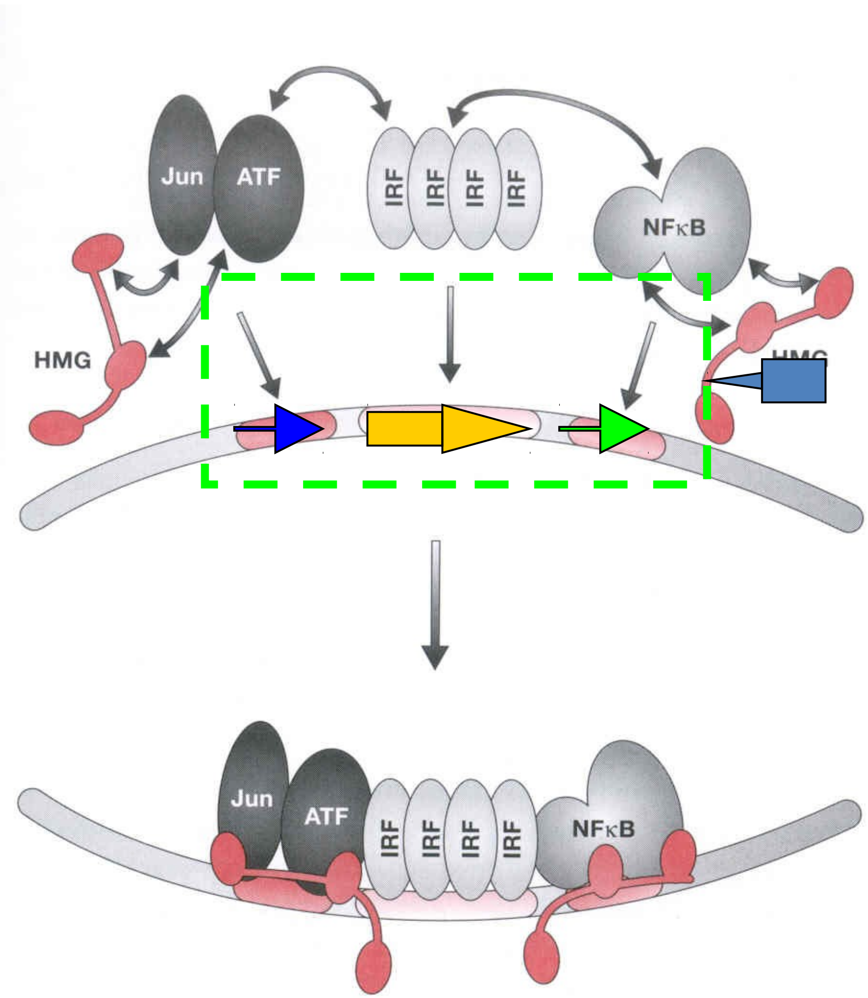
ChIP-chip
ChIP-seq

Mouse c-fos promoter
(Matrix search for TF binding sites)



```
          1                         <------------V$IK1_01(0.86)   -----...V$CREBP1CJUN_01(0.85)
          2                          <-----------V$IK2_01(0.90)   -----...V$CREB_01(0.96)
          3   ----------->V$AP2_Q6(0.87)   <-------------V$GKLF_01(0.87)
          4-->V$ATF_01(0.89)             <-------V$MZF1_01(0.99)   ----...V$ELK1_01(0.87)
          5              <-----------V$AP2_Q6(0.92)   <------------V$SP1_Q6(0.88)
          6>V$AP1FJ_Q2(0.89)              <-------------V$GKLF_01(0.85)
          7>V$AP1_Q2(0.87)                  <-------------V$GKLF_01(0.86)
          8->V$CREB_Q2(0.86)                   <---------V$CETS1P54_01(0.90)
          9->V$CREB_Q4(0.90)                   <---------V$NRF2_01(0.90)
         10                                          <-------------V$GC_01(0.88)
         11                                              ----------->V$CAAT_01(0.87)
         12                                              <------------V$TCF11_01(0.87)
         13                                                ----------->V$AP2_Q6(0.87)
         14                                                 <---------V$USF_Q6(0.93)
         16                                                    --------...V$ATF_01(0.94)
         17                                                     -------...V$AP1FJ_Q2(0.95)
         20                                                     -------...V$CREBP1_Q2(0.93)
         21                                                     -------...V$CREB_Q2(0.95)
         23                                                         ---...V$IK2_01(0.85)
MMCFOS_1   GAGCGCCCGCAGAGGGCCTTGGGGCGCGCTTCCCCCCCCTTCCAGTTCCGCCCAGTGACG   420

          1-->V$CREBP1CJUN_01(0.85)   -------------->V$BARBIE_01(0.86)
          2-->V$CREB_01(0.96)         -------------->V$TATA_01(0.95)
          3       ----------->V$CAAT_01(0.91)   --------->V$AP4_Q5(0.95)
          4----------->V$ELK1_01(0.87)    --------------------->V$HEN1_01(0.87)
          5                 --------->V$AP4_Q5(0.88)               <---...V$CMYB_01(0.93)
          6       <---------V$CDPCR3HD_01(0.93)                      --...V$VMYB_02(0.89)
          7                  <--------------V$TATA_01(0.88)
          8                               ------       -------->V$HEN1_02(0.87)
          9                               <-----       ---------V$HEN1_02(0.86)
         10                                <----       ------V$AP4_01(0.88)
         11                                    ----------->V$LMO2COM_01(0.93)
         12                                    <-----------V$LMO2COM_01(0.93)
         13                                    <-----------V$MYOD_01(0.88)
         17--->V$AP1FJ_Q2(0.95)                 <---------V$AP4_Q6(0.99)
         20---->V$CREBP1_Q2(0.93)               <---------V$MYOD_Q6(0.96)
         21---->V$CREB_Q2(0.95)                                             Transcription start
         23-------->V$IK2_01(0.85)
         24                               <=========== E2F  (0.80)
MMCFOS_1   TAGGAAGTCCATCCATTCACAGCGCTTCTATAAAGGCGCCAGCTGAGGCGCCTACTACTC   480

          1               <-----------------V$CMYB_01(0.91)     -------...V$ER_Q6(0.86)
          2                <-----------V$LMO2COM_01(0.90)         <----...V$TCF11_01(0.87)
          3                 --------->V$MYOD_Q6(0.90)      -------->V$STAT_01(0.93)
          4                 --------->V$VMYB_01(0.89)      <--------V$STAT_01(0.89)
          5--------------V$CMYB_01(0.93)   -------->V$LMO2COM_02(0.93)
          6------>V$VMYB_02(0.89)      <-----------V$CAAT_01(0.85)
          7                 -------->V$VMYB_02(0.88)
          8                       -------------->V$EVI1_04(0.86)
          9                             ------------->V$GATA1_02(0.93)
         12                               <------------V$ZID_01(0.85)
         13                               <----------V$CP2_01(0.97)
         14                                 ---------->V$GATA_C(0.92)
         15                                     ----------------->V$CMYB_01(0.86)
         16                                          --------->V$CREL_01(0.91)
         24                                               <=========== E2F  (0.82)
MMCFOS_1   CAACCGCGACTGCAGCGAGCAACTGAGAAGACTGGATAGAGCCGGCGGTTCCGCGAACGA   540
```

Composite Modules (CM)

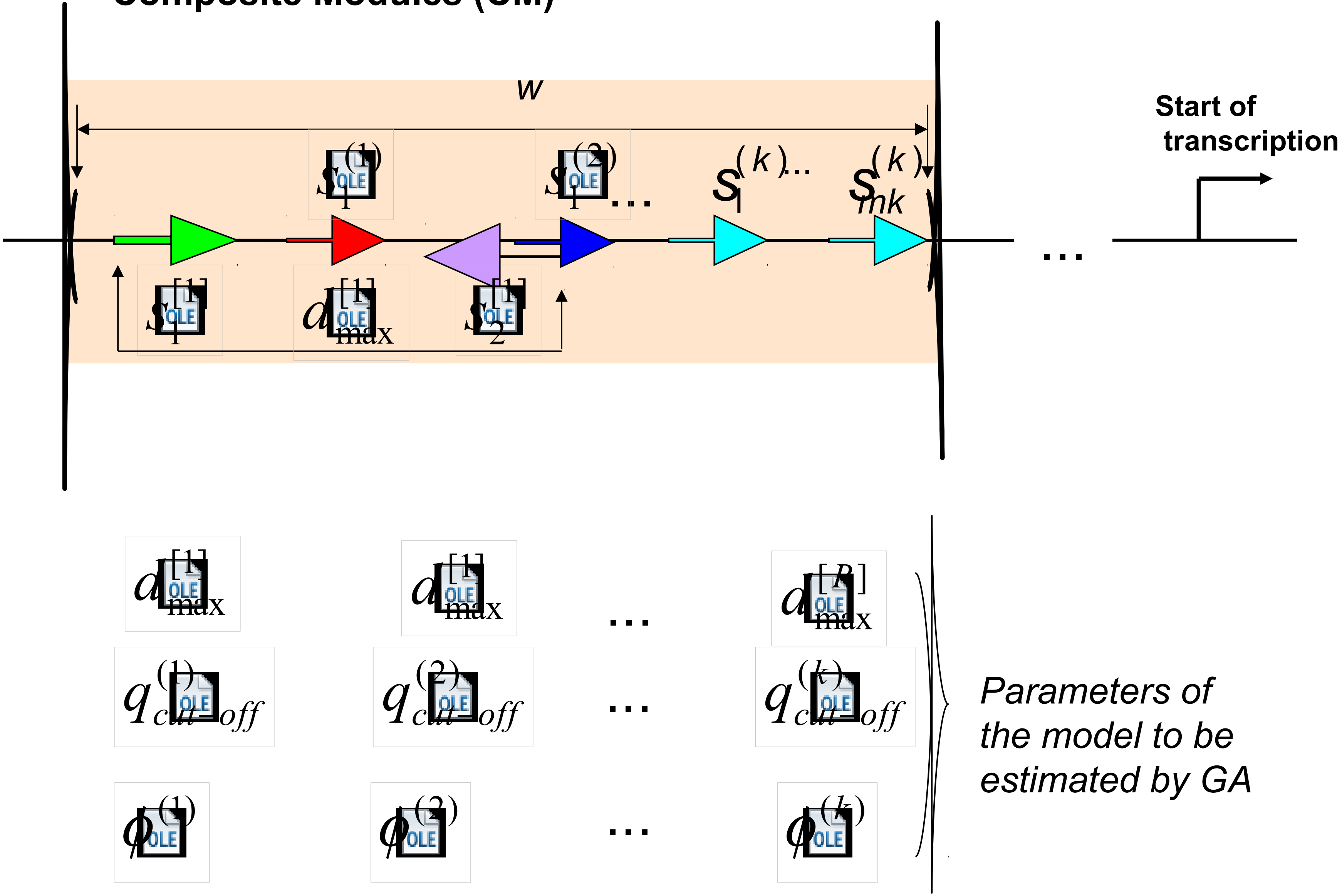


Composite Module (CM)

FIGURE 3.3. The human interferon-β enhanceosome. HMG represents HMGI/Y, a ubiquitous protein that binds cooperatively with the three activators. HMGI/Y both bends the DNA and contacts the activators. Each of the transcription factors shown is a member of a family of related activators.

(Mark Ptashne, Alexander Gann Genes and Signals, 2002)

Composite Modules (CM)

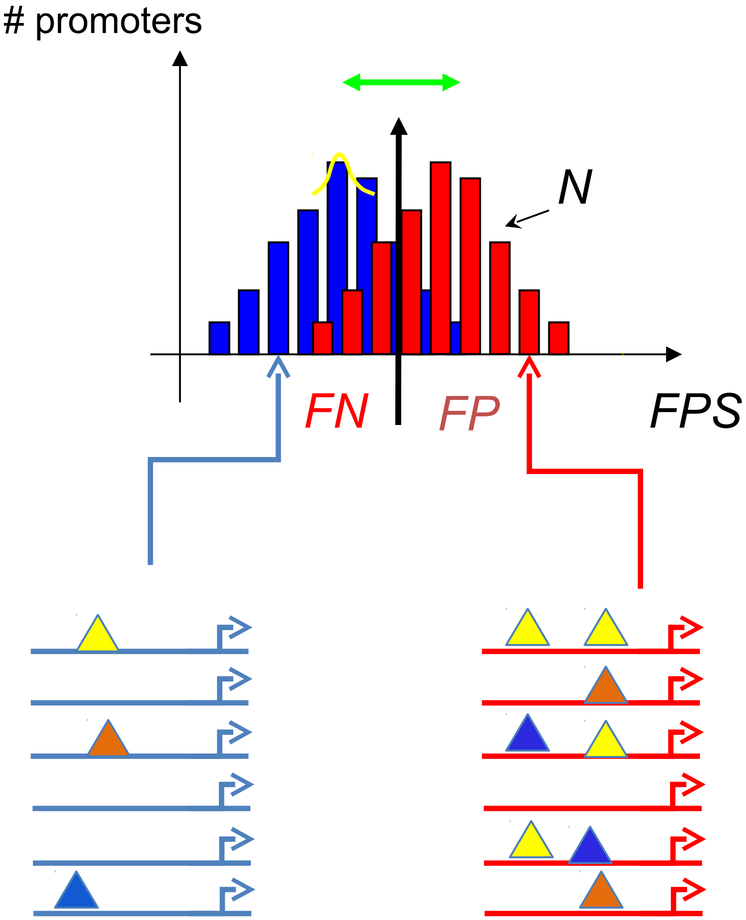


Parameters of the model to be estimated by GA

We created a genetic algorithm to find site combinations

Fitness function of the Genetic-Regression Algorithm (GRA)

$$F = \alpha \cdot R + \beta \cdot (1 - FN) + (1 - \beta) \cdot (1 - FP) + \gamma \cdot T + \delta \cdot N - \mu \cdot k$$



R – linear regression

FN – false negatives

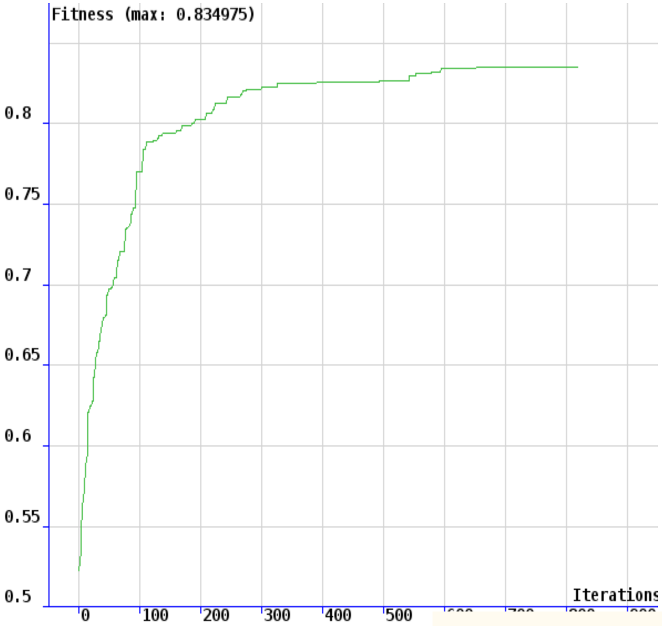
FP – false positives

T – T-test (difference between mean values)

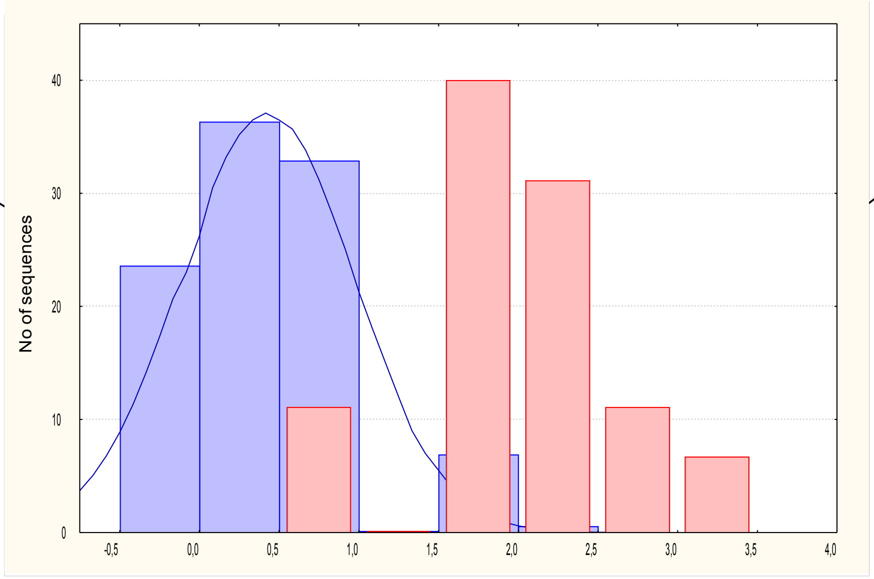
N – normal likeness

k – number of free parameters

Composite module in promoters of cell cycle-related genes



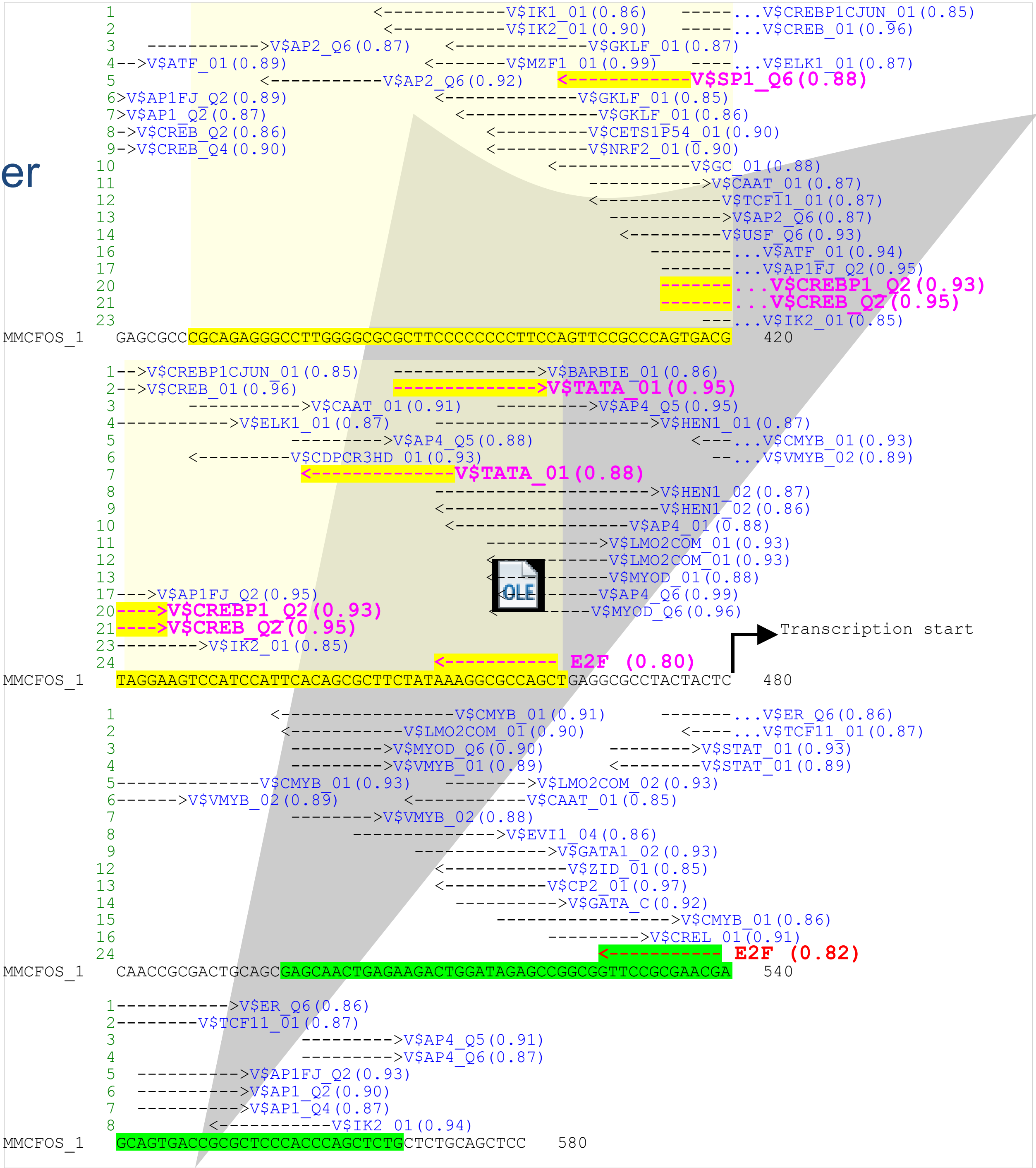
Weight: ϕ	$q_{cut-off}$	TF matrix
1.000000	0.840072	V$E2F_19
0.954483	0.737637	V$TATA_01
0.888064	0.939687	V$CREB_01
0.816179	0.941583	V$SP1_Q6
0.039746	0.839702	V$TAL1BETAE47_01



Background sequences

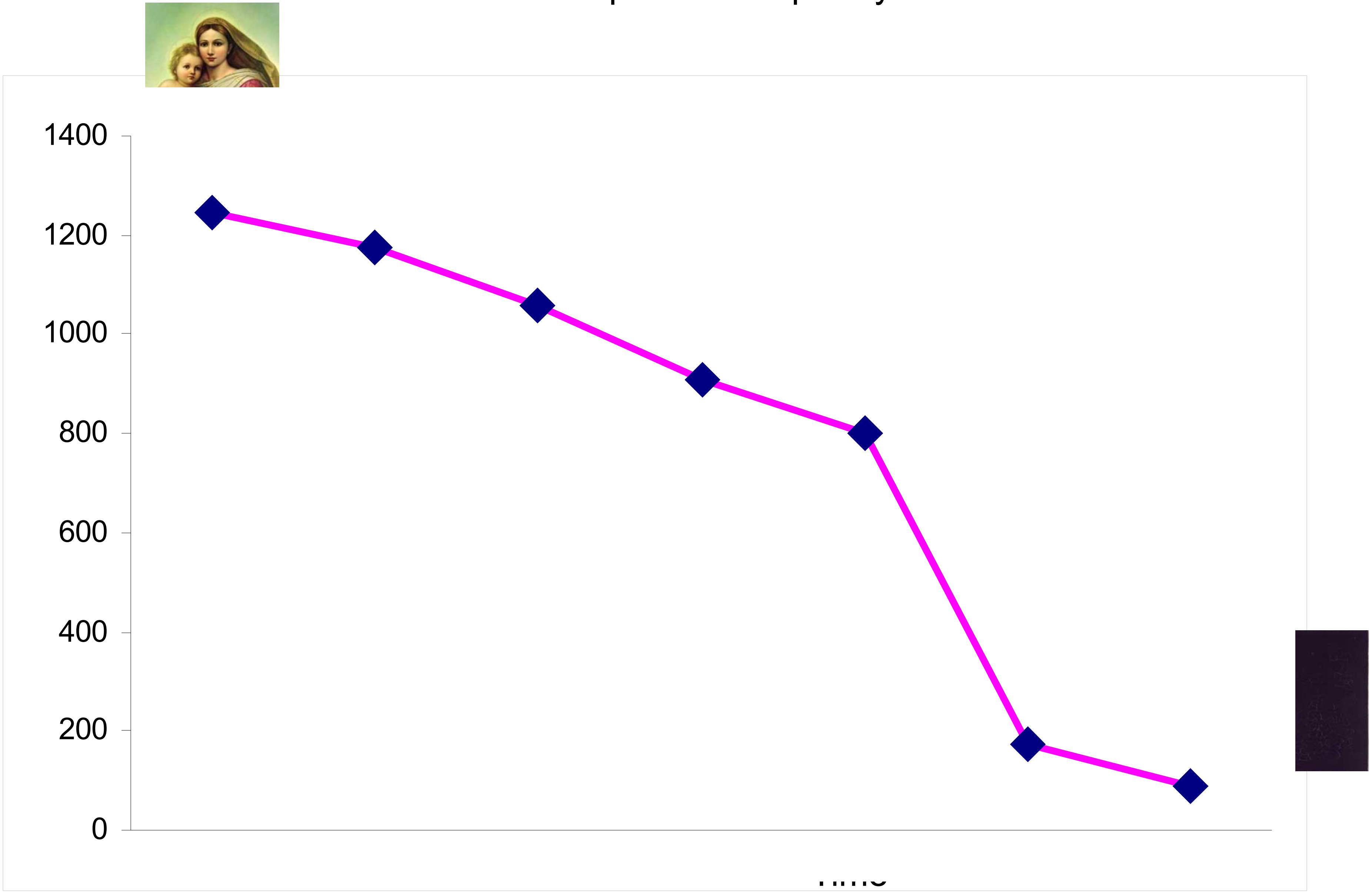
Cell cycle-related promoters

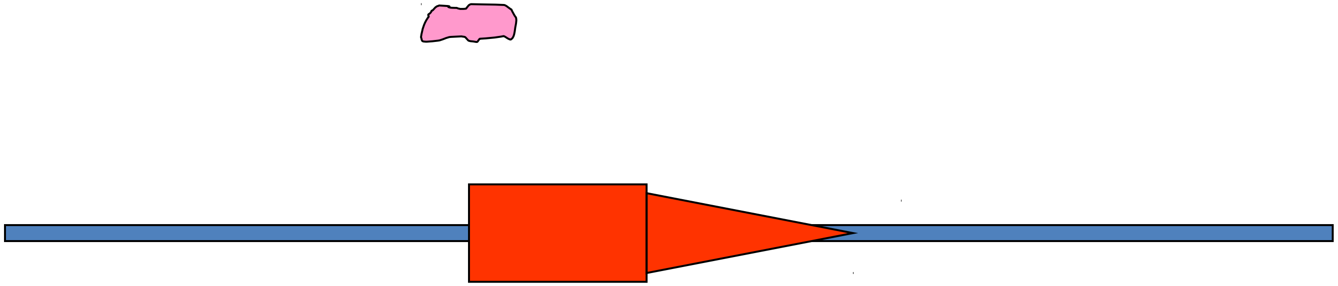
Mouse c-fos promoter

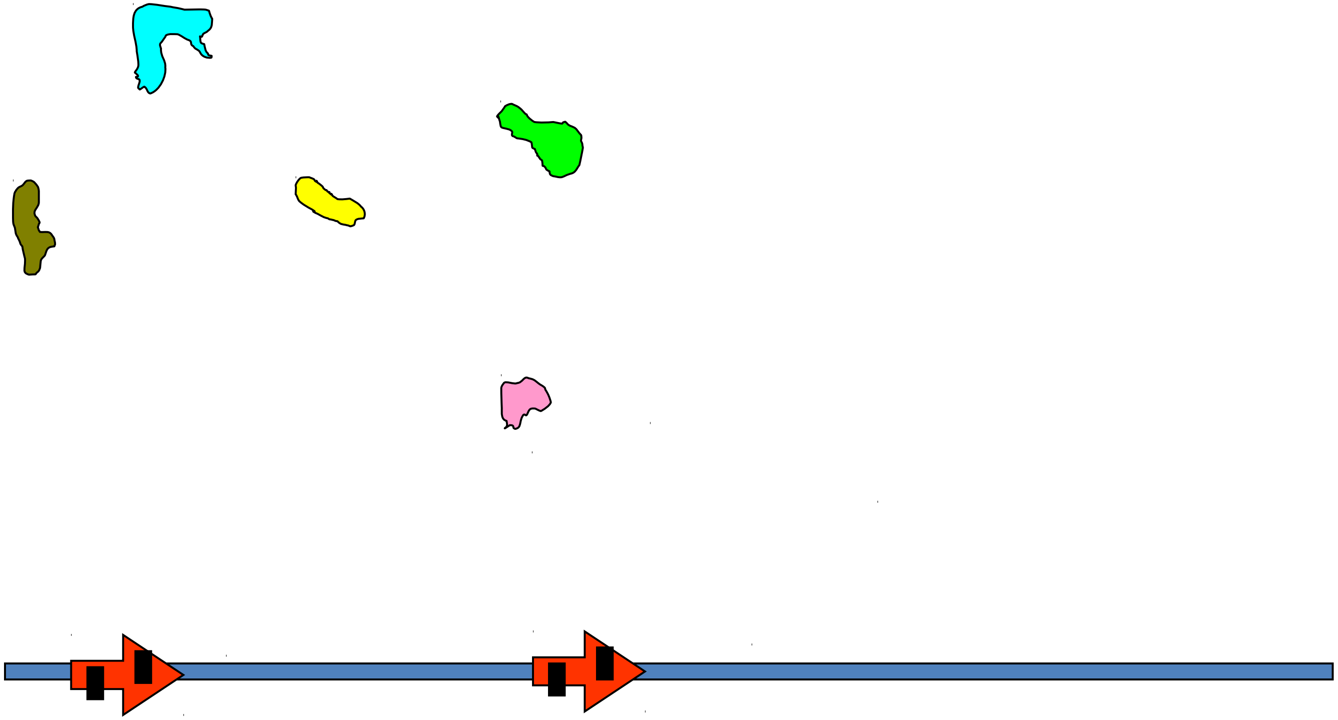


```
          1                         <------------V$IK1_01(0.86)   -----...V$CREBP1CJUN_01(0.85)
          2                          <-----------V$IK2_01(0.90)   -----...V$CREB_01(0.96)
          3   ----------->V$AP2_Q6(0.87)   <-------------V$GKLF_01(0.87)
          4-->V$ATF_01(0.89)             <-------V$MZF1_01(0.99)   ----...V$ELK1_01(0.87)
          5              <-----------V$AP2_Q6(0.92)   <------------V$SP1_Q6(0.88)
          6>V$AP1FJ_Q2(0.89)              <-------------V$GKLF_01(0.85)
          7>V$AP1_Q2(0.87)                  <-------------V$GKLF_01(0.86)
          8->V$CREB_Q2(0.86)                   <---------V$CETS1P54_01(0.90)
          9->V$CREB_Q4(0.90)                   <---------V$NRF2_01(0.90)
         10                                          <-------------V$GC_01(0.88)
         11                                              ----------->V$CAAT_01(0.87)
         12                                              <------------V$TCF11_01(0.87)
         13                                                ----------->V$AP2_Q6(0.87)
         14                                                 <---------V$USF_Q6(0.93)
         16                                                    --------...V$ATF_01(0.94)
         17                                                     -------...V$AP1FJ_Q2(0.95)
         20                                                     -------...V$CREBP1_Q2(0.93)
         21                                                     -------...V$CREB_Q2(0.95)
         23                                                         ---...V$IK2_01(0.85)
MMCFOS_1   GAGCGCCCGCAGAGGGCCTTGGGGCGCGCTTCCCCCCCCTTCCAGTTCCGCCCAGTGACG   420

          1-->V$CREBP1CJUN_01(0.85)   -------------->V$BARBIE_01(0.86)
          2-->V$CREB_01(0.96)         -------------->V$TATA_01(0.95)
          3       ----------->V$CAAT_01(0.91)   --------->V$AP4_Q5(0.95)
          4----------->V$ELK1_01(0.87)    --------------------->V$HEN1_01(0.87)
          5                 --------->V$AP4_Q5(0.88)               <---...V$CMYB_01(0.93)
          6       <---------V$CDPCR3HD_01(0.93)                      --...V$VMYB_02(0.89)
          7                  <--------------V$TATA_01(0.88)
          8                               --------------------->V$HEN1_02(0.87)
          9                               <---------------------V$HEN1_02(0.86)
         10                                <-----------------V$AP4_01(0.88)
         11                                    ----------->V$LMO2COM_01(0.93)
         12                                    <-----------V$LMO2COM_01(0.93)
         13                                    <-----------V$MYOD_01(0.88)
         17--->V$AP1FJ_Q2(0.95)                <----------V$AP4_Q6(0.99)
         20---->V$CREBP1_Q2(0.93)              <---------V$MYOD_Q6(0.96)
         21---->V$CREB_Q2(0.95)                                              Transcription start
         23-------->V$IK2_01(0.85)
         24                               <----------- E2F  (0.80)
MMCFOS_1   TAGGAAGTCCATCCATTCACAGCGCTTCTATAAAGGCGCCAGCTGAGGCGCCTACTACTC   480

          1               <-----------------V$CMYB_01(0.91)     -------...V$ER_Q6(0.86)
          2                <-----------V$LMO2COM_01(0.90)         <----...V$TCF11_01(0.87)
          3                 --------->V$MYOD_Q6(0.90)      -------->V$STAT_01(0.93)
          4                 --------->V$VMYB_01(0.89)      <--------V$STAT_01(0.89)
          5--------------V$CMYB_01(0.93)   -------->V$LMO2COM_02(0.93)
          6------>V$VMYB_02(0.89)      <-----------V$CAAT_01(0.85)
          7                 -------->V$VMYB_02(0.88)
          8                       -------------->V$EVI1_04(0.86)
          9                             ------------->V$GATA1_02(0.93)
         12                               <------------V$ZID_01(0.85)
         13                               <----------V$CP2_01(0.97)
         14                                 ---------->V$GATA_C(0.92)
         15                                     ---------------->V$CMYB_01(0.86)
         16                                          --------->V$CREL_01(0.91)
         24                                               <----------- E2F  (0.82)
MMCFOS_1   CAACCGCGACTGCAGCGAGCAACTGAGAAGACTGGATAGAGCCGGCGGTTCCGCGAACGA   540

          1----------->V$ER_Q6(0.86)
          2--------V$TCF11_01(0.87)
          3                  --------->V$AP4_Q5(0.91)
          4                  --------->V$AP4_Q6(0.87)
          5  ---------->V$AP1FJ_Q2(0.93)
          6  ---------->V$AP1_Q2(0.90)
          7  ---------->V$AP1_Q4(0.87)
          8         <-----------V$IK2_01(0.94)
MMCFOS_1   GCAGTGACCGCGCTCCCACCCAGCTCTGCTCTGCAGCTCC   580
```

Size of zip file = complexity

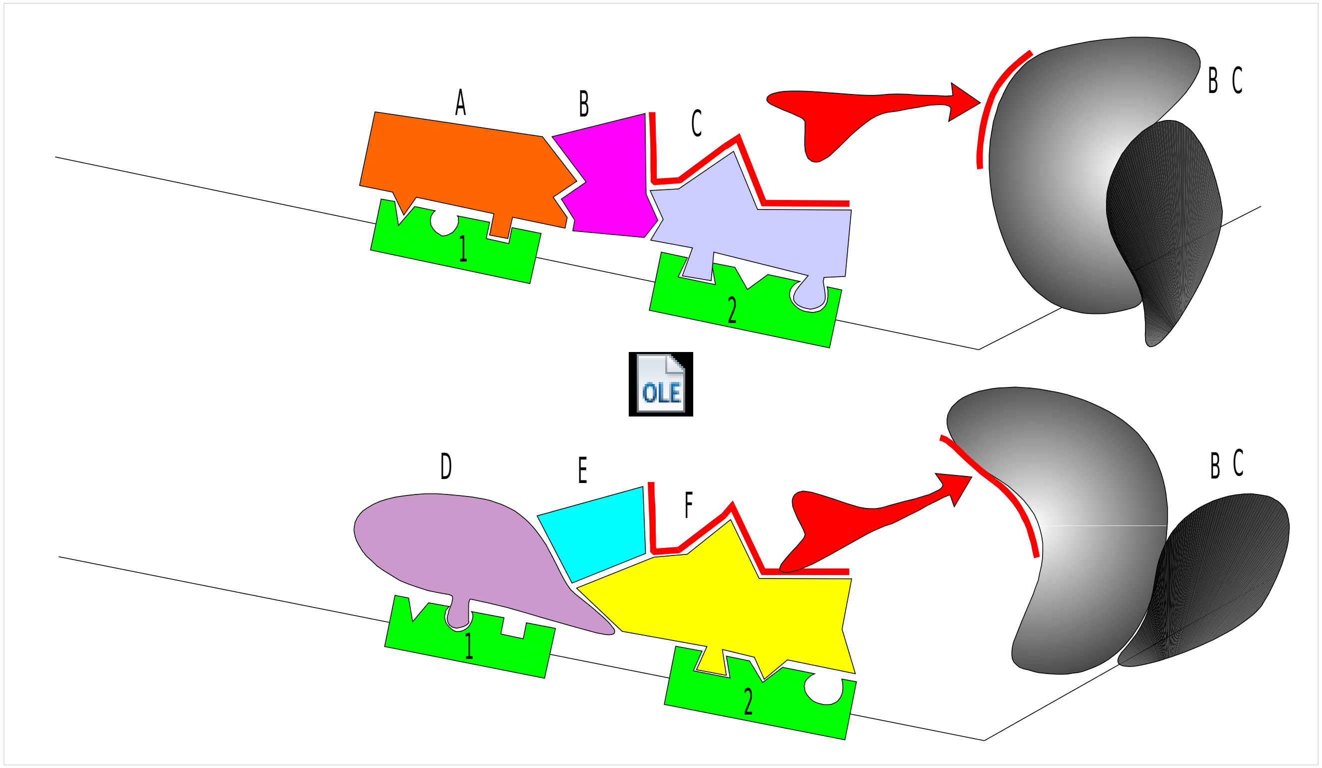






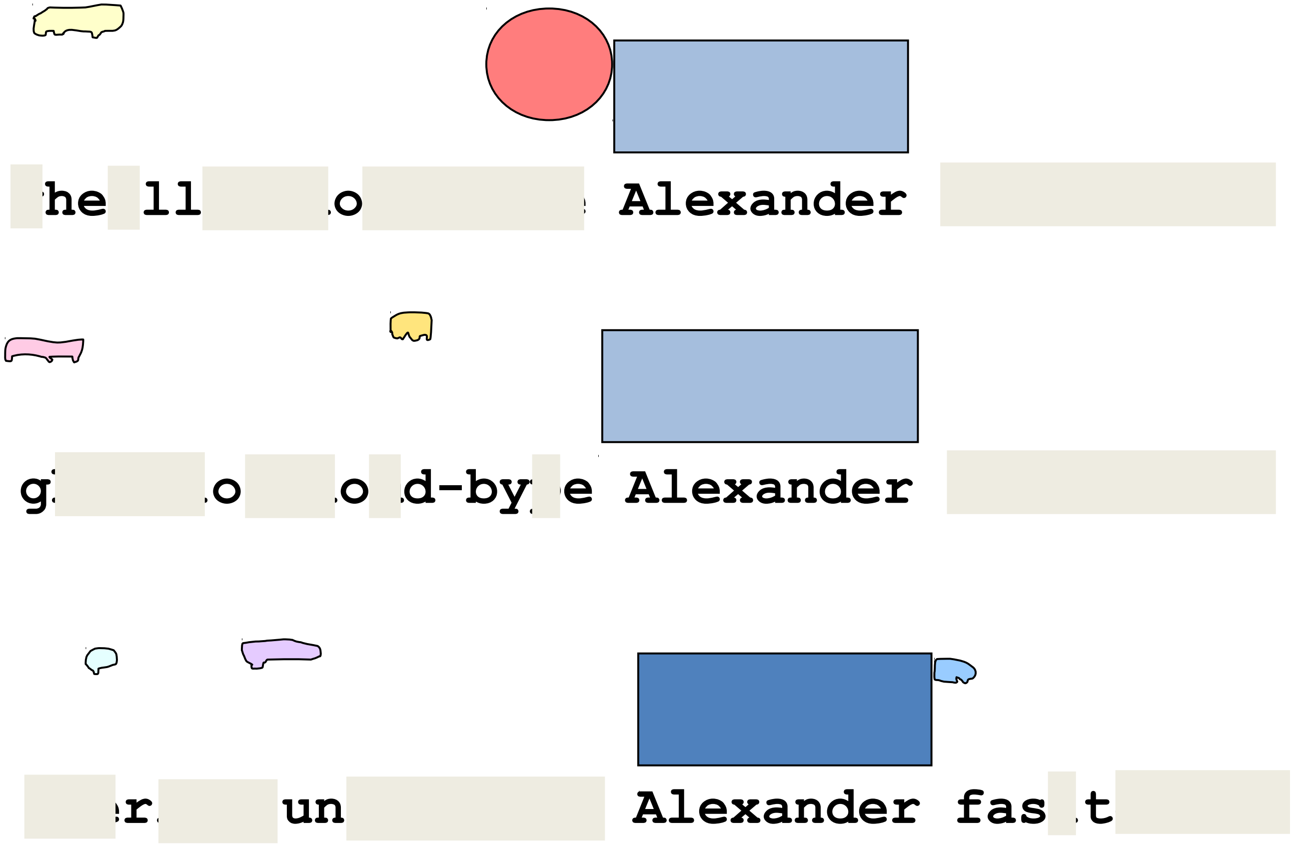
Fuzzy puzzle - multipurpose structure of the eukaryotic promoters

Mechanism of network plasticity



Gene regulatory code ?

`gherllojunomd-bype Alexander fasltoiwany`



1) `he ll o Alexander`

2) `g o o d-by e Alexander`

3) `er un Alexander fas t`

Even some messages which were not written

gherllojunomd-bype Alexander fasltoiwany