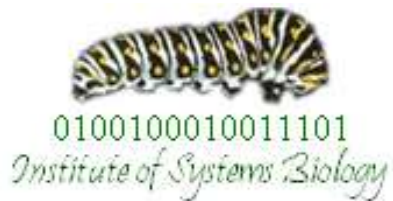
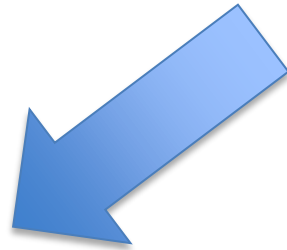


Моделирование шагающих паттернов в стволовых клетках

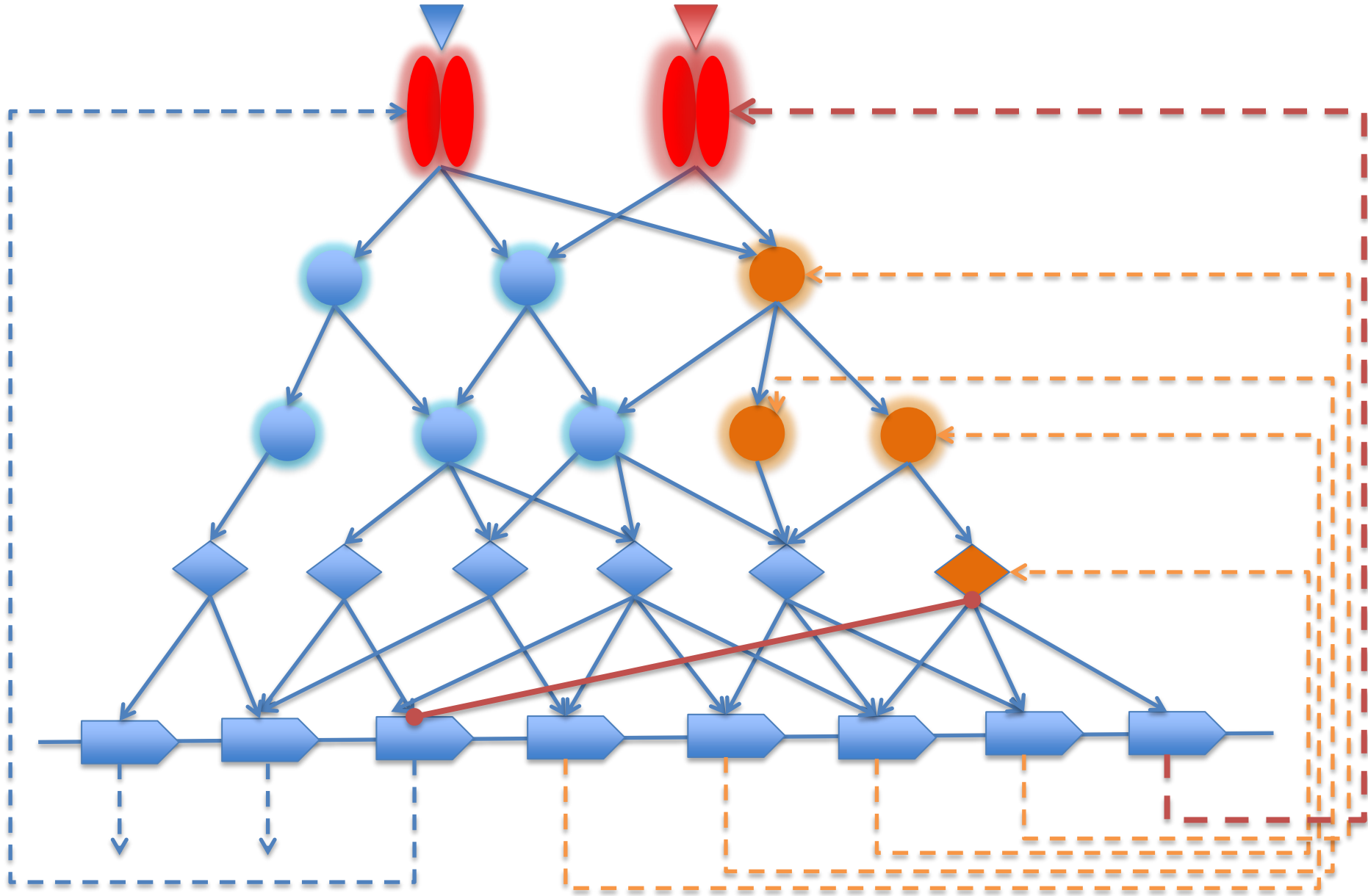
Кель А.Э

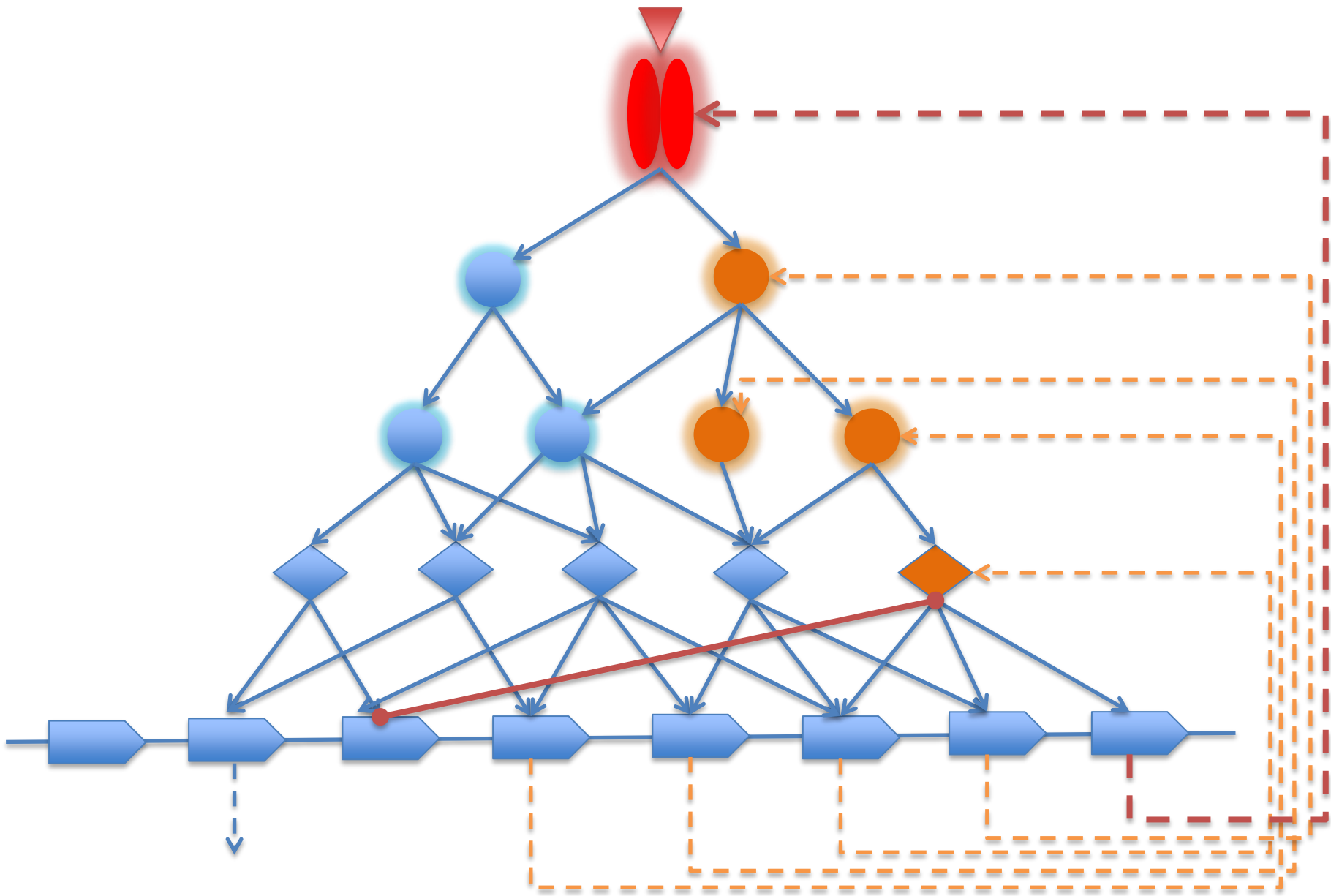


alexander.kel@genexplain.com

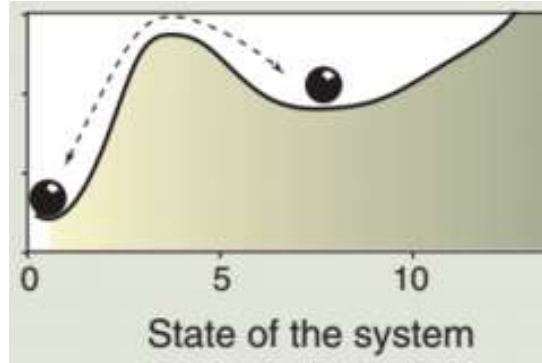


Walking pathways

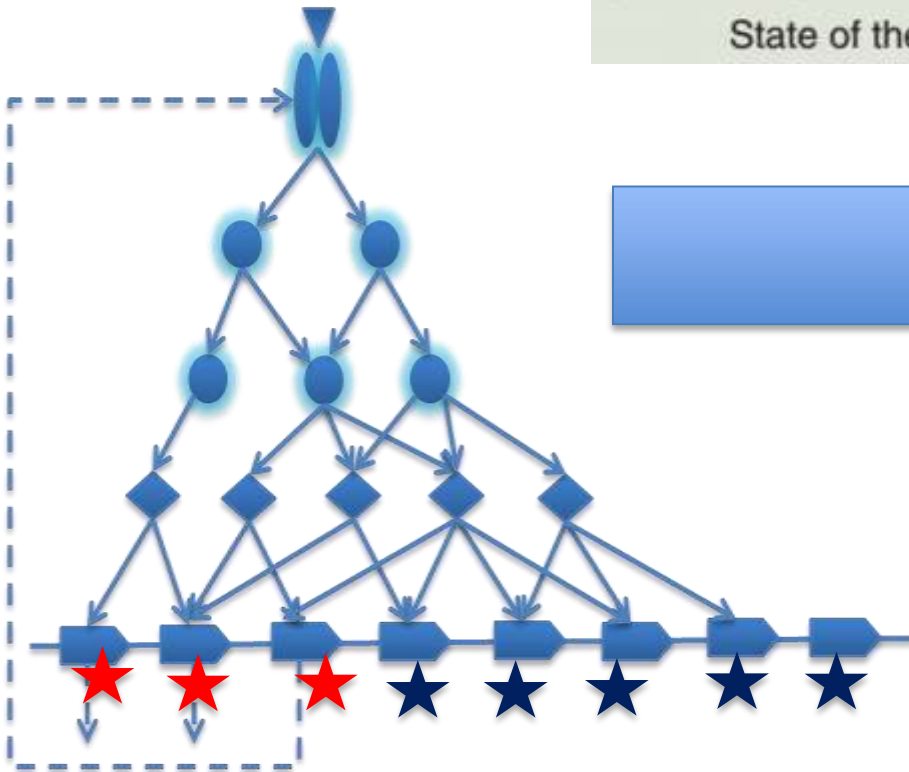




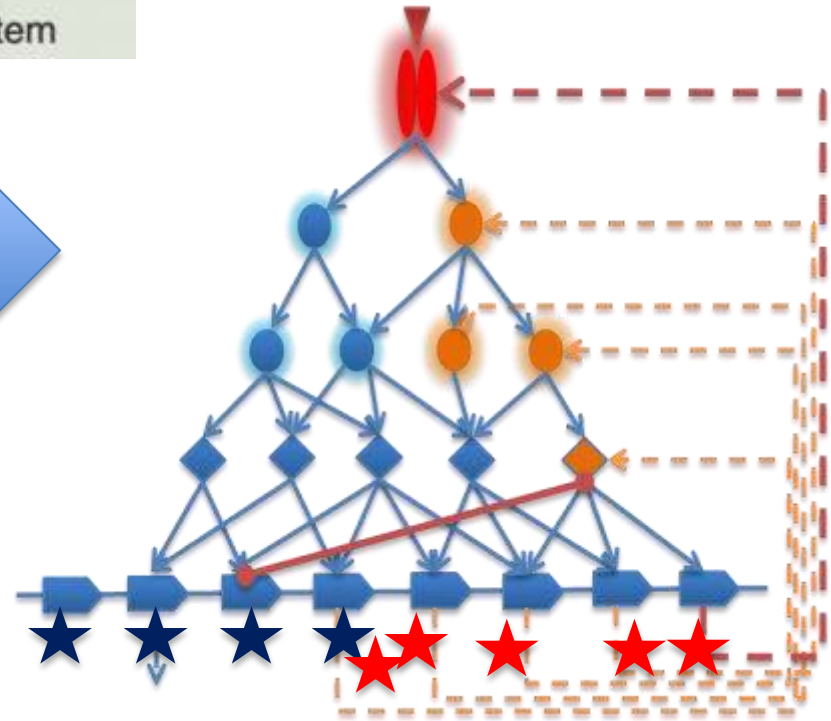
„Walking pathways“



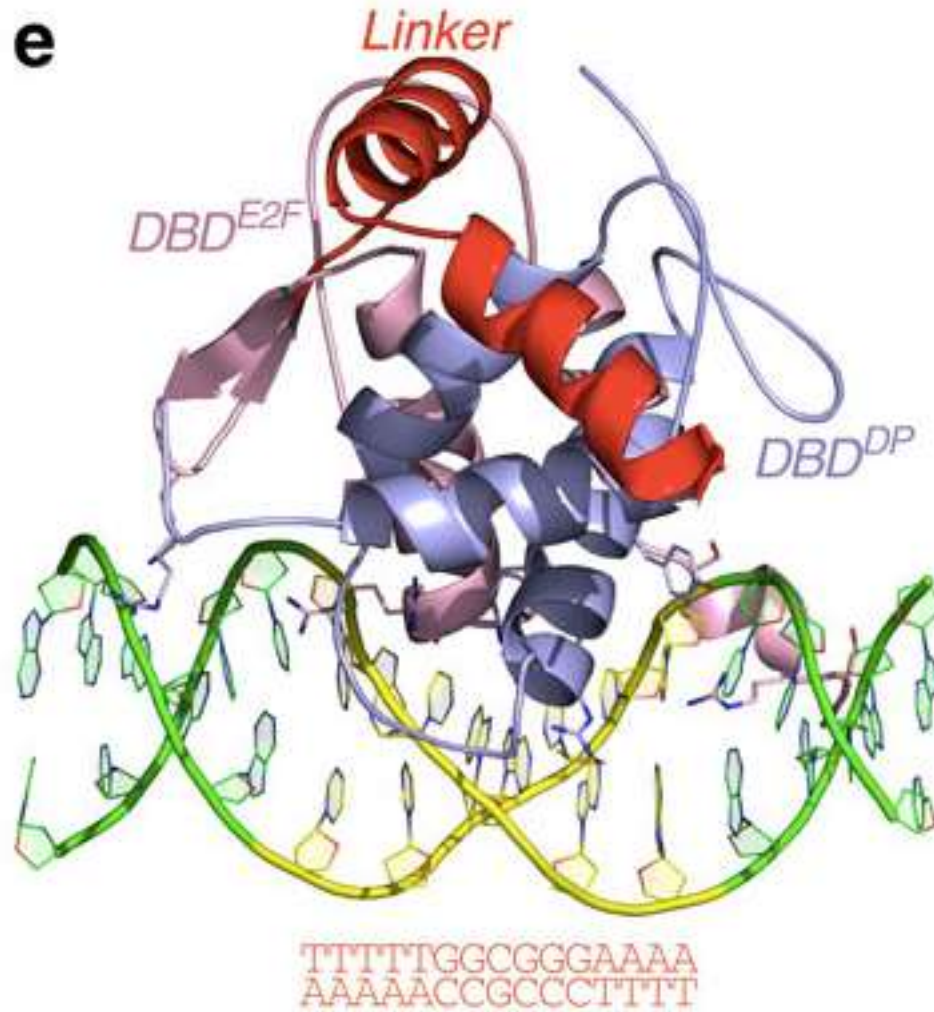
Healthy



Disease

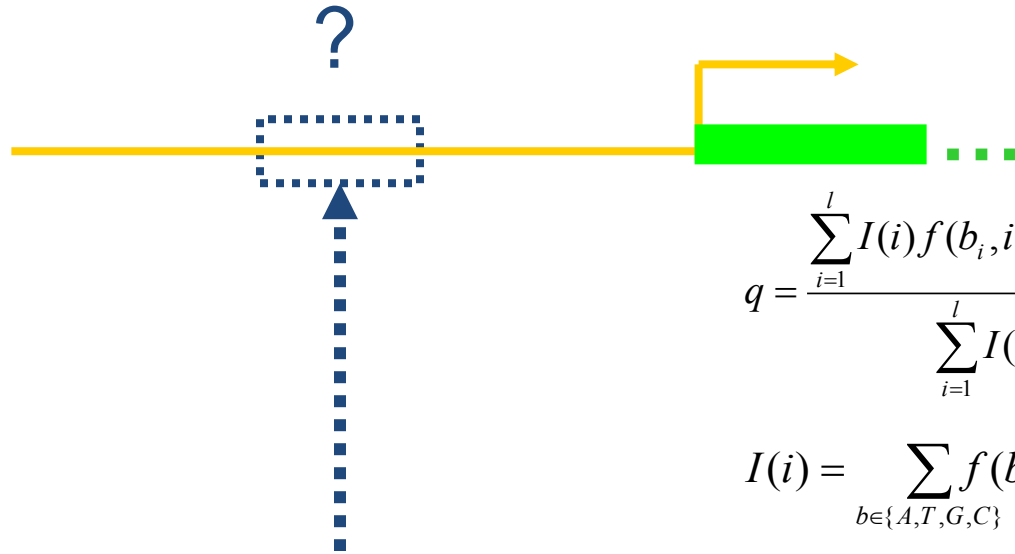
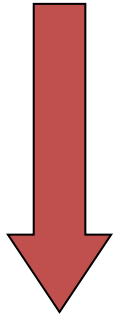


E2F-8 (human)



Search for new TF binding sites (with PWMs from TRANSFAC)

TRANSFAC®



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

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

Matrix length: 12

Logo



E2F

A	0.282	0.026	0.154	0.000	0.000	0.000	0.000	0.000	0.000	0.846	0.872	0.692	0.231
C	0.282	0.205	0.667	0.308	1.000	0.000	0.923	0.513	0.000	0.077	0.051	0.359	
G	0.103	0.359	0.154	0.667	0.000	1.000	0.077	0.410	0.154	0.051	0.205	0.179	
T	0.333	0.410	0.026	0.026	0.000	0.000	0.000	0.077	0.000	0.000	0.051	0.231	

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

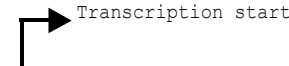
```

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$CTS1P54_01(0.90)
9->V$CREB_Q4(0.90)  <-----V$NRF2_01(0.90)
10         <-----V$GC_01(0.88)
11         ----->V$CAAT_01(0.87)
12         <-----V$TCF1_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)
MMCFOF_1  GAGCGCCCGCAGAGGGCCTTGGGGCGCGCTTCCCCCCTTCCAGTTCGCCCCAGTGCAG  420

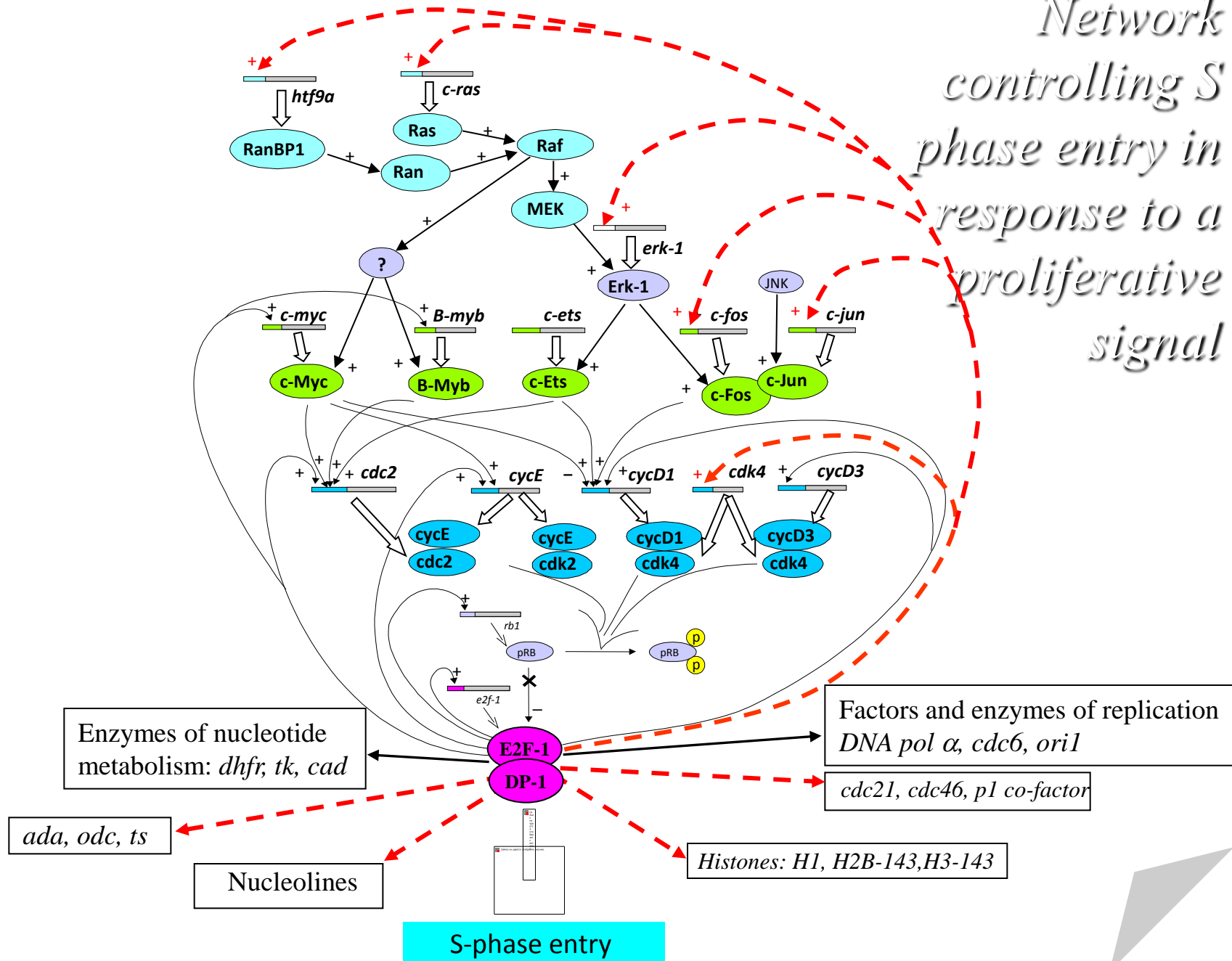
1-->V$CREBP1CJUN_01(0.85)  ----->V$BARBIE_01(0.86)
2-->V$CREB_01(0.96)  ----->V$TATA_01(0.95)
3          ----->V$CAAT_01(0.91)  ----->V$AP4_Q5(0.95)
4----->V$ELK1_01(0.87)  ----->V$HEN1_01(0.87)
5          ----->V$AP4_Q5(0.88)  <---...V$CMYB_01(0.93)
6          <-----V$CDPCR3HD_01(0.93)  ---...V$VMYB_02(0.89)
7          <-----V$TATA_01(0.88)
8          ----->V$HEN1_02(0.87)
9          <-----V$HEN1_02(0.86)
10         <-----V$AP4_Q1(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)
23----->V$IK2_01(0.85)
24         <===== E2F (0.80)
MMCFOF_1  TAGGAAGTCCATCCATTACAGCGCTTCTATAAAGCGCCAGCTGAGGCGCCTACTACT  480

1          <-----V$CMYB_01(0.91)  -----...V$ER_Q6(0.86)
2          <-----V$LMO2COM_01(0.90)  <---...V$TCF1_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)
MMCFOF_1  CAACCGGACTGCAGCGGCAACTGAGAAGACTGGATAGAGCCGGGTTCCCGAACGA  540

```



*Network
controlling S
phase entry in
response to a
proliferative
signal*



Enzymes of nucleotide metabolism: *dhfr*, *tk*, *cad*

Factors and enzymes of replication
DNA pol α , *cdc6*, *ori1*

cdc21, *cdc46*, *p1* co-factor

ada, *odc*, *ts*

Nucleolines

Histones: H1, H2B-143, H3-143

S-phase entry



Bifurcation analysis of the regulatory modules of the mammalian G₁/S transition

Maciej Swat^{1,*}, Alexander Kel^{2,3} and Hanspeter Herzel¹

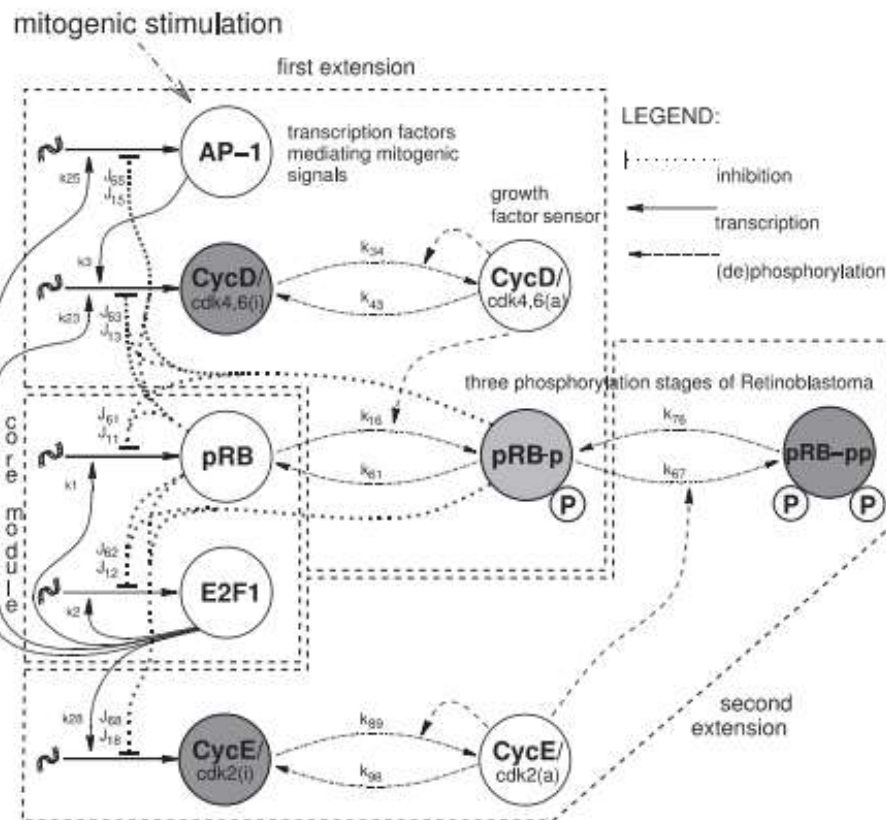
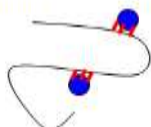
¹Institute for Theoretical Biology, Humboldt University Berlin, Invalidenstrasse 43, Berlin, D-10115, Germany, ²BIOBASE, Halchtersche Strasse 33, Wolfenbüttel, D-38304, Germany and ³Institute of Cytology and Genetics SB RAN, 10 Lavrentyev pr., Novosibirsk, 630090, Russia

Received on October 12, 2003; accepted

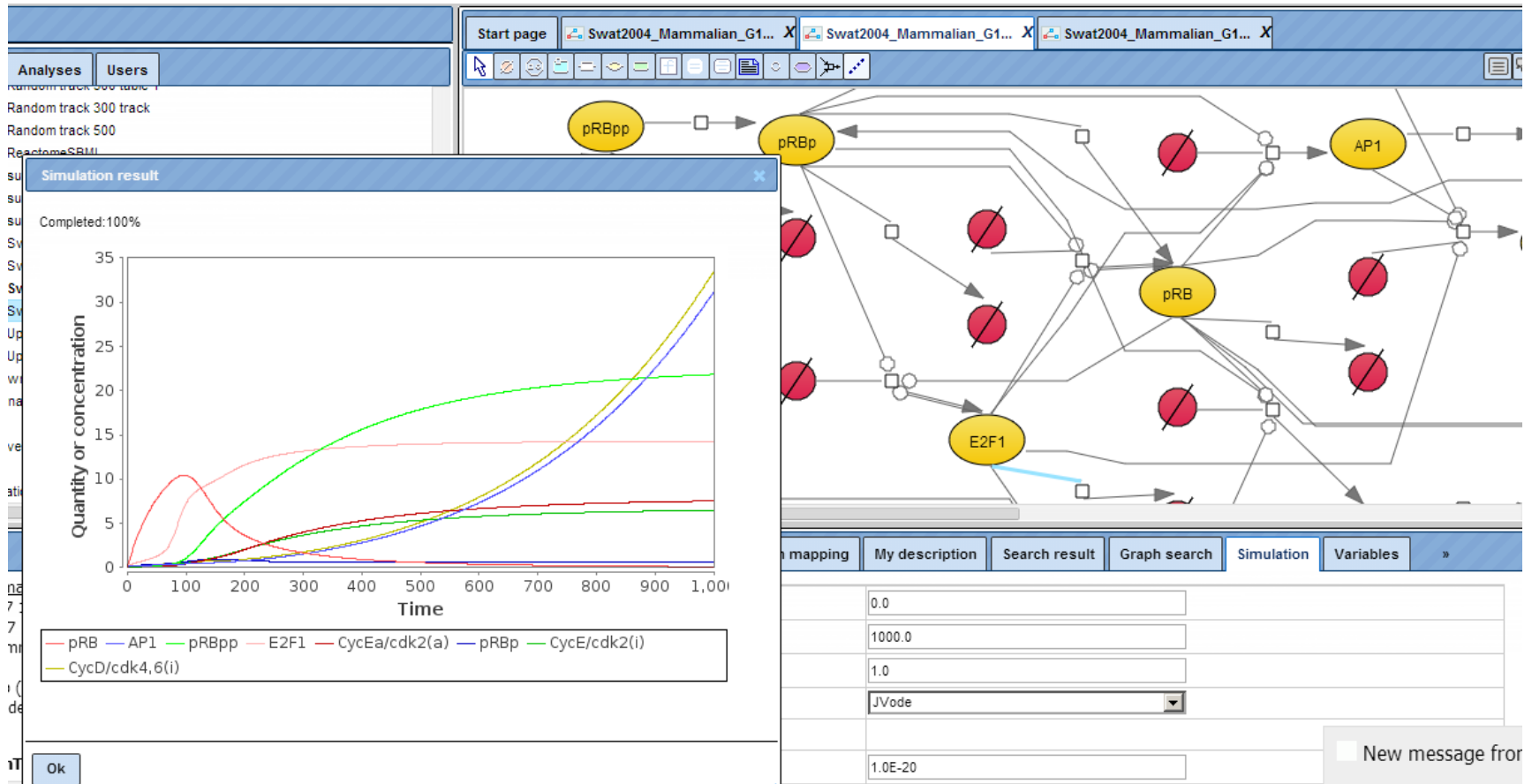
Computationally predicted E2F target gene confirmed by *in vivo* footprint

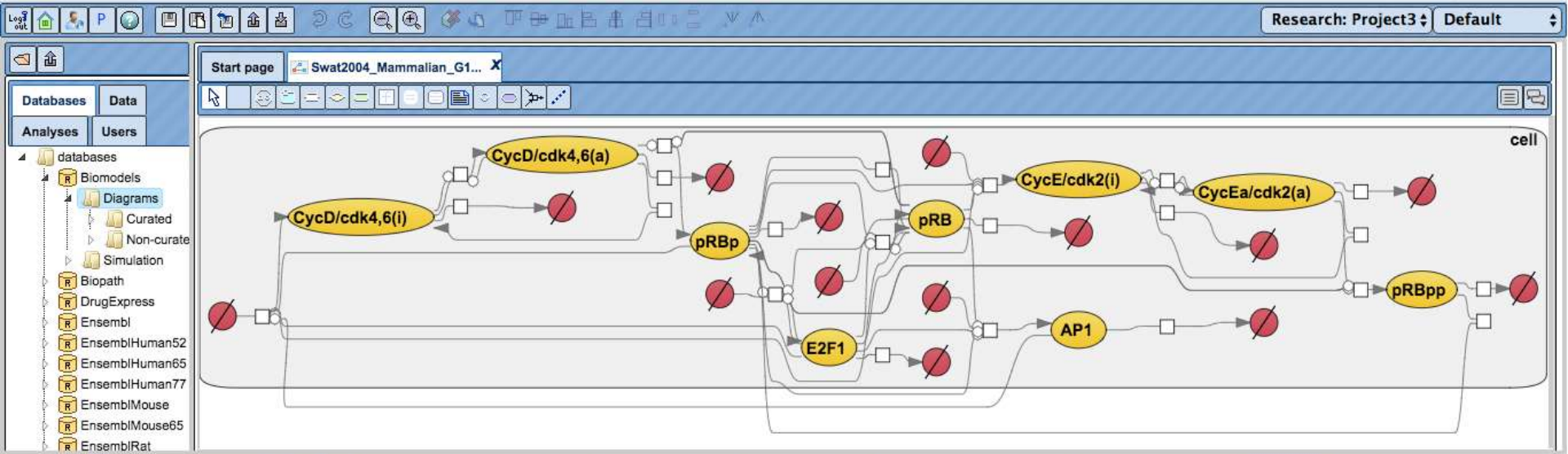
Gene	EMBL	Sequence of the potential sites	Position rel. start of transcription	Score -E
<i>c-fos</i> , H s	H3205	(-) gcCTTGGCGGTGTcc	-165 .. -176	0.92
		(-) ggGGTGGCGCCGGGgc	-91 .. -103	0.84
		(+) ccTCTGGCGCCACCGc	-90 .. -79	0.88
		(-) ccGGTGGCGCCAGAgg	-78 .. -89	0.83
<i>JunB</i> , H s	H3207341	(+) gcTATCGCGCCAGAgc	79 .. 90	0.89
		(-) tcTCTGGCGCCGATAgc	91 .. 80	0.91
		(-) ggGCTGGCGCCGGCGgg	169 .. 158	0.82

Chromatin crosslinking



Simulation engine in geneXplain platform





Search Info **Default**

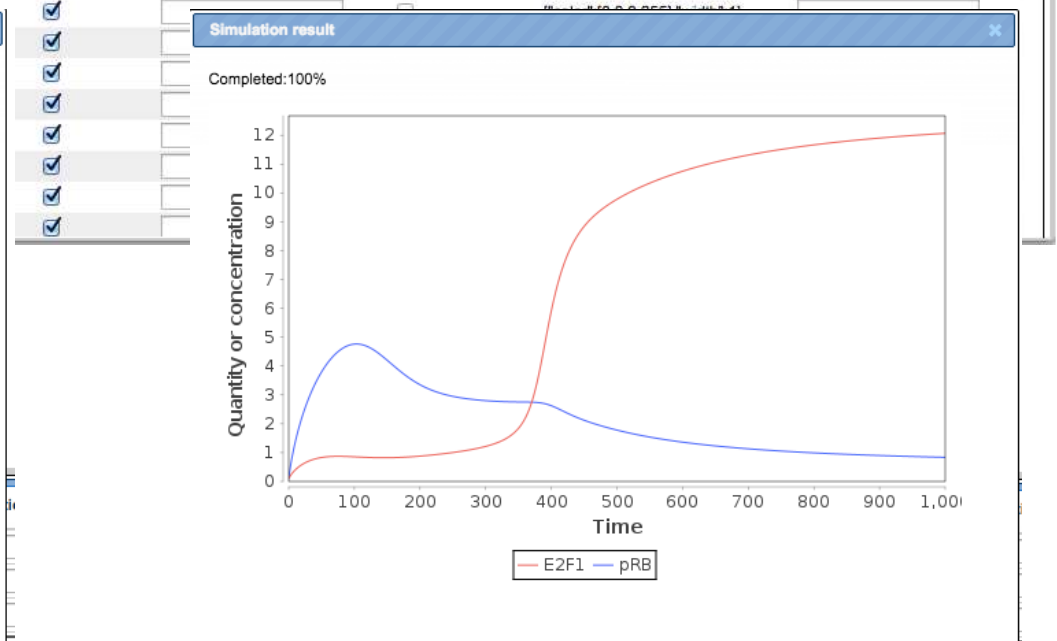
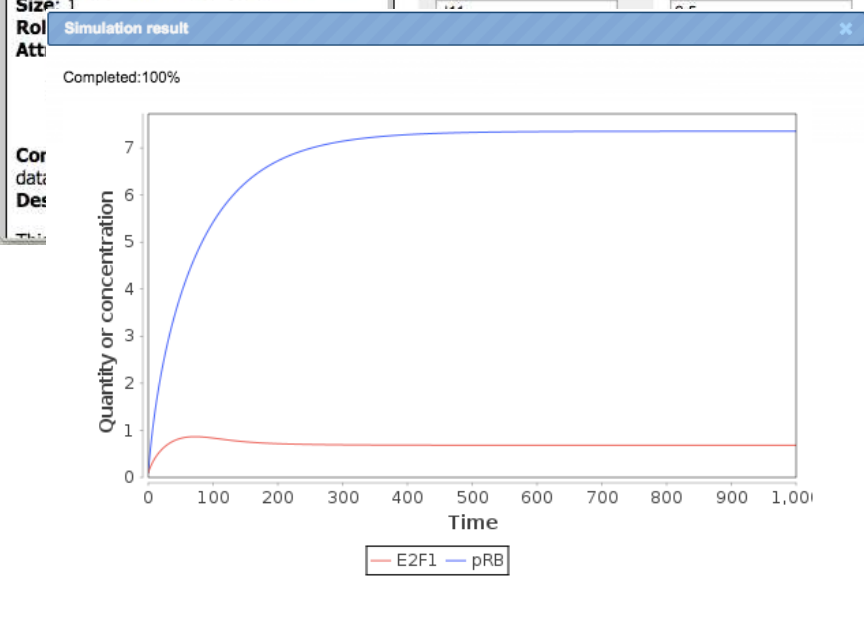
ID: BIOMD000000228
 Title: Swat2004_Mammalian_G1_S_Transition
 Size: 1
 Role: Simulation result
 Attachment:

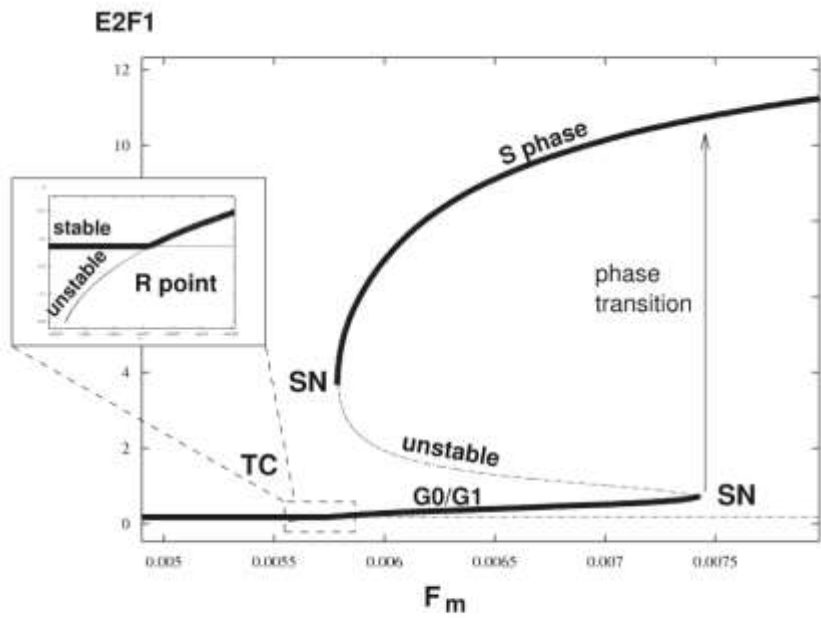
Overview Layout Expression mapping My description Search result Graph search Simulation Variables Parameters Script Clipboard Tasks

First Previous **1** Next Last

Name	Initial value	Constant	Units	Show in plot	Plot line spec	Comment
Fm	0.002	<input checked="" type="checkbox"/>		<input type="checkbox"/>	{color:[0,0,0,255],width:1}	<p xmlns="http://www.w3

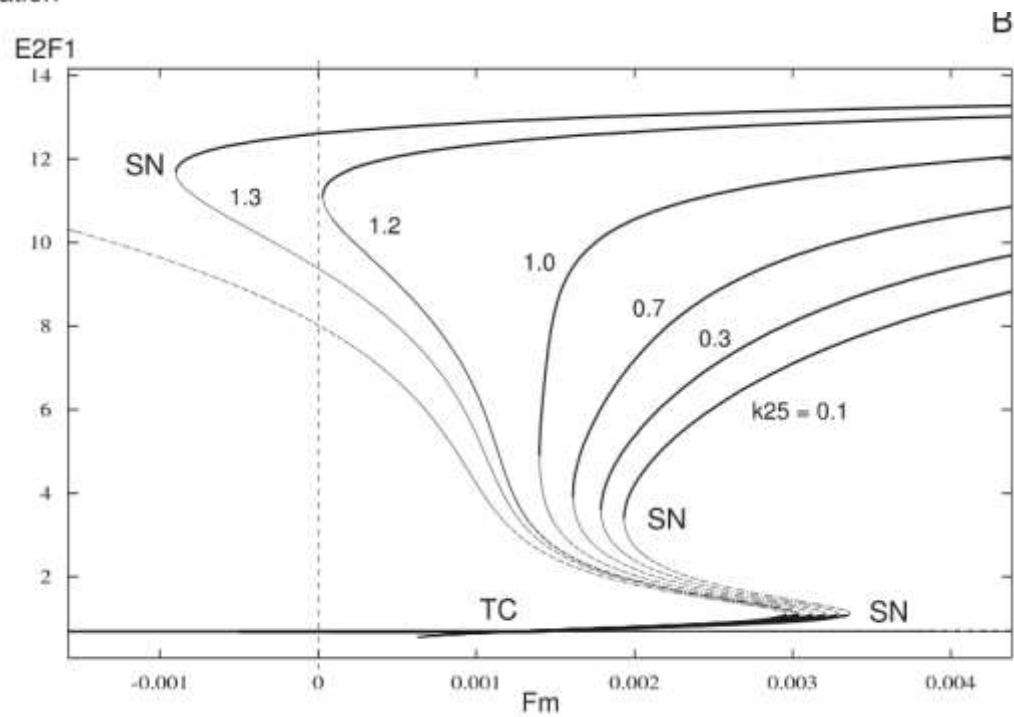
Show 50 entries





SN – saddle node bifurcation

TC – transcritical bifurcation



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

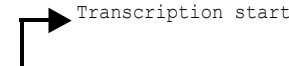
```

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$CTS1P54_01(0.90)
9->V$CREB_Q4(0.90)  <-----V$NRF2_01(0.90)
10         <-----V$GC_01(0.88)
11         ----->V$CAAT_01(0.87)
12         <-----V$TCF1_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)
MMCFO_1  GAGCGCCCGCAGAGGGCCTTGGGGCGCGCTTCCCCCCTTCCAGTTCGCCCGAGTGAGC  420

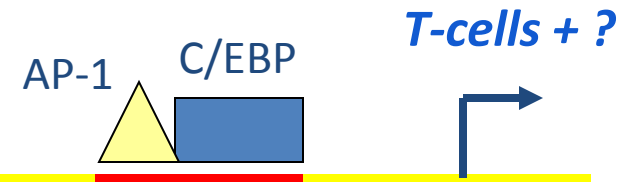
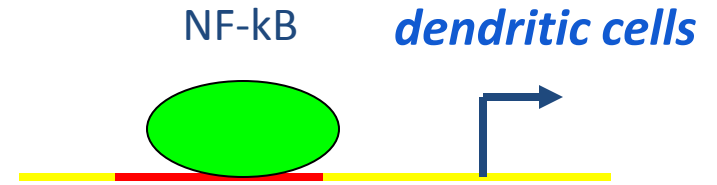
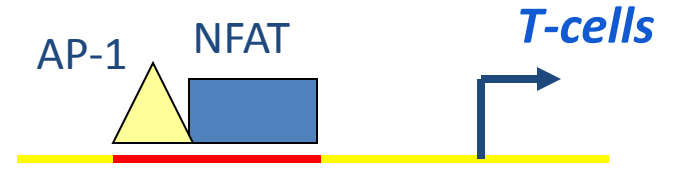
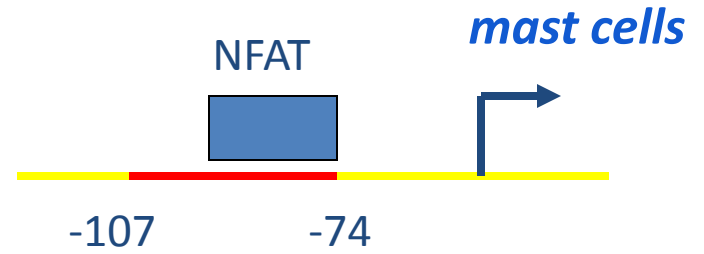
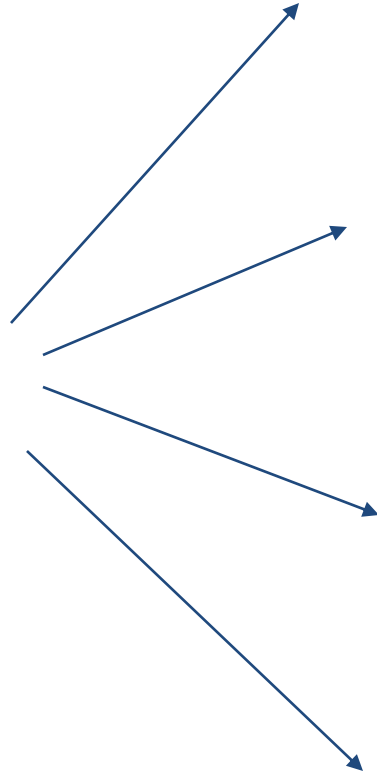
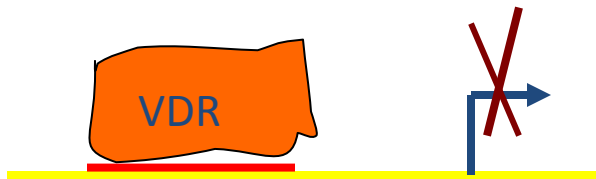
1-->V$CREBP1CJUN_01(0.85)  ----->V$BARBIE_01(0.86)
2-->V$CREB_01(0.96)  ----->V$TATA_01(0.95)
3          ----->V$CAAT_01(0.91)  ----->V$AP4_Q5(0.95)
4----->V$ELK1_01(0.87)  ----->V$HEN1_01(0.87)
5          ----->V$AP4_Q5(0.88)  <---...V$CMYB_01(0.93)
6          <-----V$DPCR3HD_01(0.93)  ---...V$VMYB_02(0.89)
7          <-----V$TATA_01(0.88)
8          ----->V$HEN1_02(0.87)
9          <-----V$HEN1_02(0.86)
10         <-----V$AP4_Q1(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)
23----->V$IK2_01(0.85)
24         <===== E2F (0.80)
MMCFO_1  TAGGAAGTCCATCCATTACAGCGCTTCTATAAAGCGCCAGCTGAGGCGCCTACTACT  480

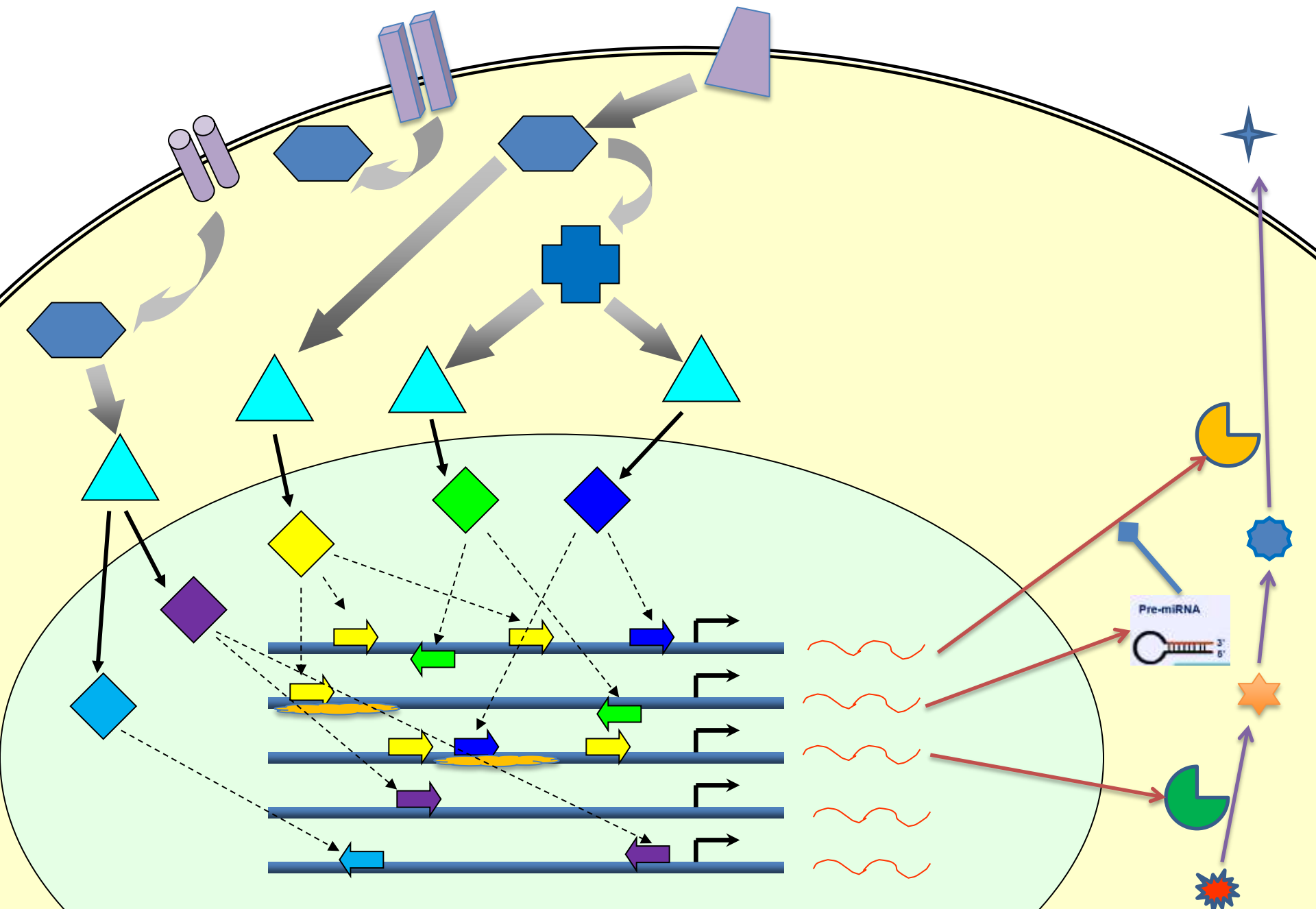
1          <-----V$CMYB_01(0.91)  -----...V$ER_Q6(0.86)
2          <-----V$LMO2COM_01(0.90)  <---...V$TCF1_01(0.87)
3          ----->V$MYOD_Q6(0.90)  ----->V$STAT_01(0.93)
4          ----->V$VMYB_01(0.89)  <-----V$STAT_01(0.89)
5----->V$CMYB_01(0.93)  ----->V$LMO2COM_02(0.93)
6----->V$VMYB_02(0.89)  <-----V$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)
MMCFO_1  CAACCGGACTGCAGCGAGCAACTGAGAAGACTGGATAGAGCCGGGTTCCCGAACGA  540

```



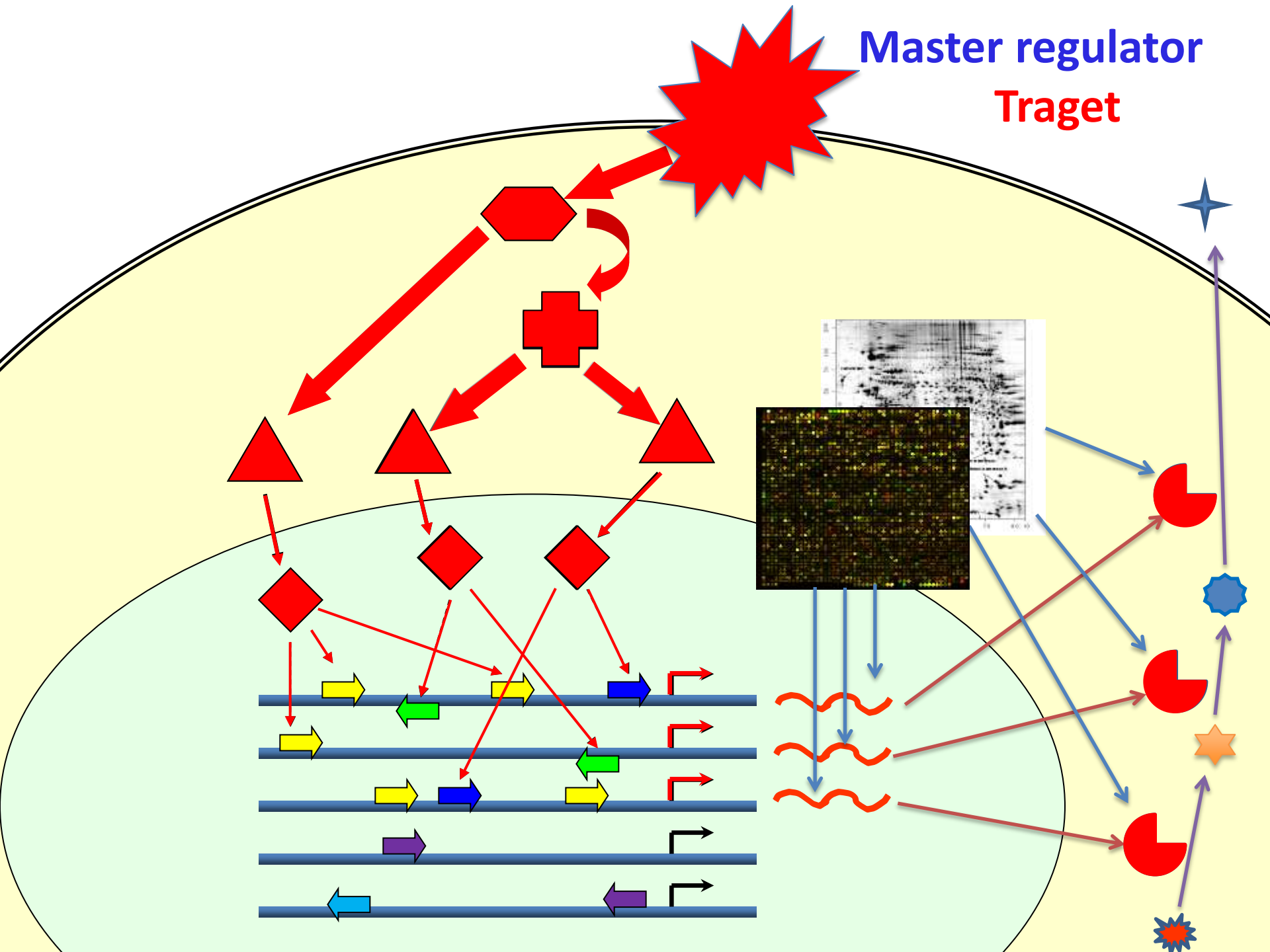
human TNF α promoter





Master regulator

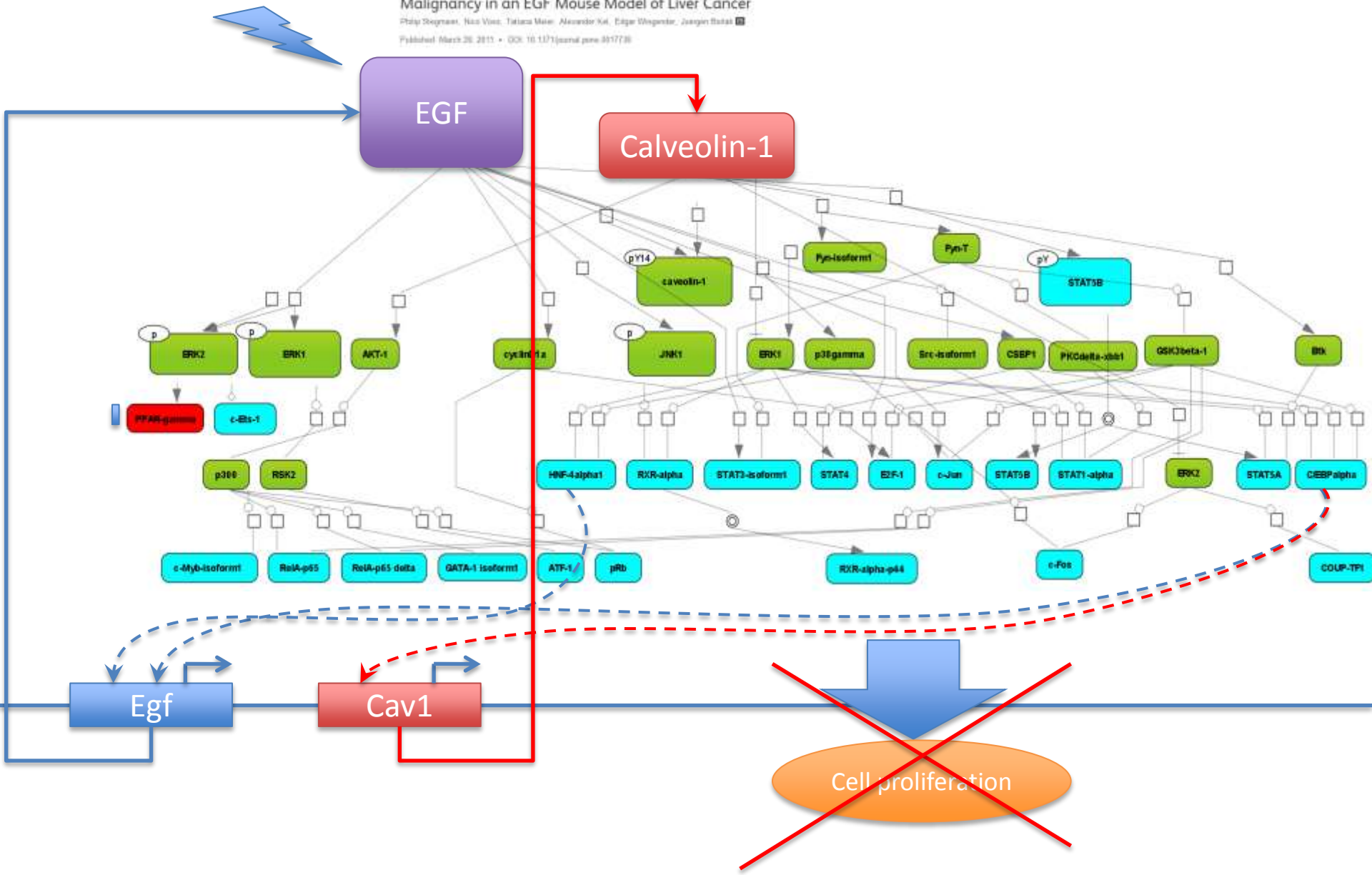
Traget



Advanced Computational Biology Methods Identify Molecular Switches for Malignancy in an EGF Mouse Model of Liver Cancer

Pshy Shyngnan, Nao Vito, Takara Masu, Alexander Kai, Edgar Wengeler, Jaesun Scharf

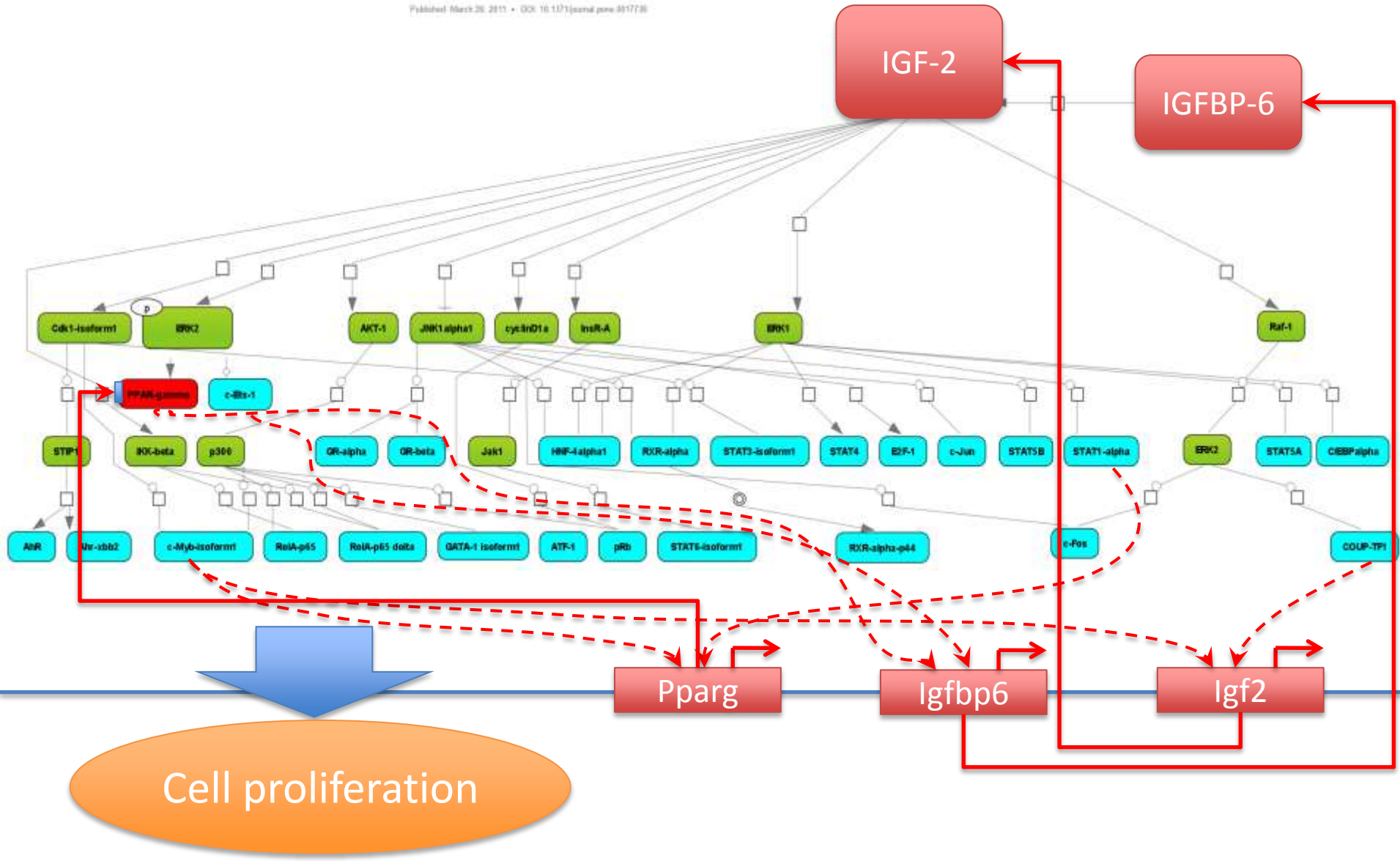
Published March 28, 2011 • DOI: 10.1371/journal.pone.0017730



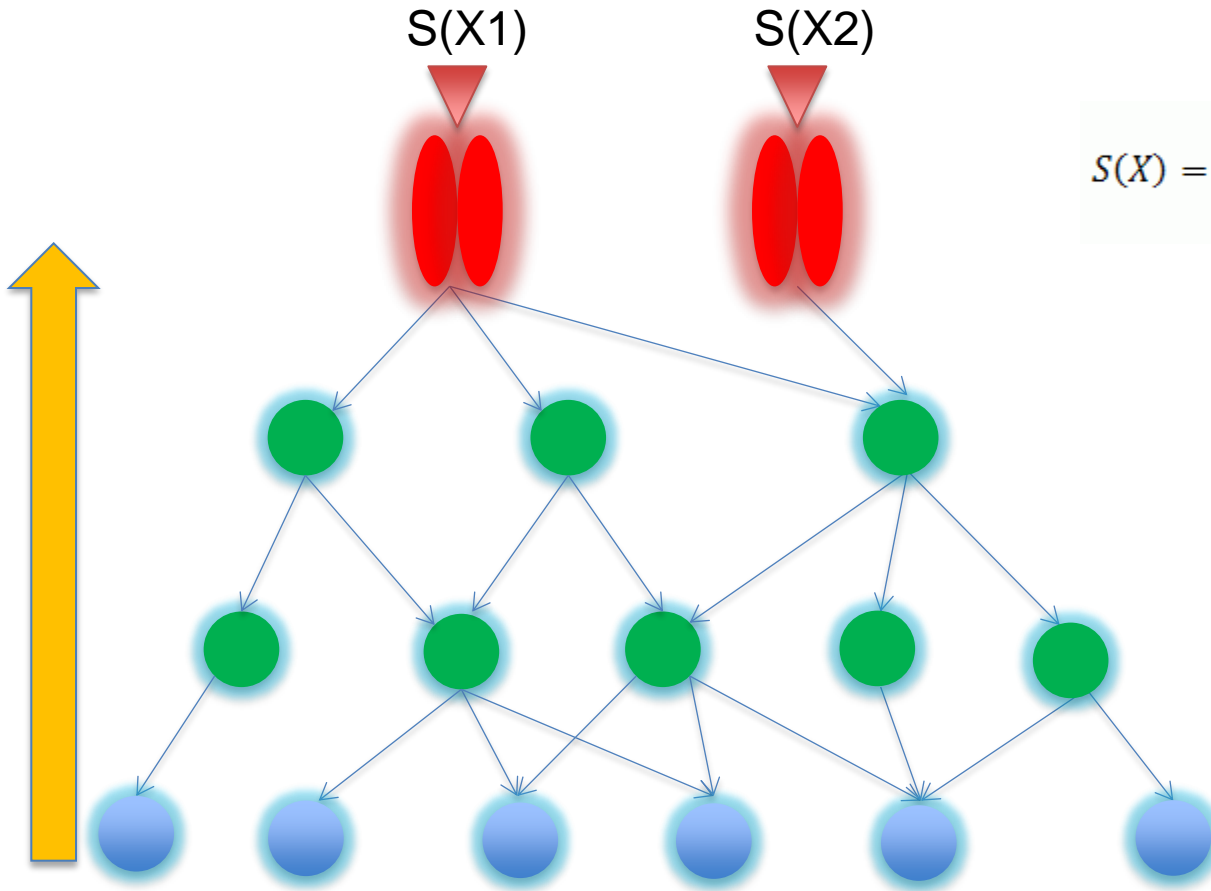
Advanced Computational Biology Methods Identify Molecular Switches for Malignancy in an EGF Mouse Model of Liver Cancer

Philip Steynson, Nico Vries, Tatjana Meier, Alexander Kai, Edgar Wingenler, Jürgen Basten

Published March 20, 2015 • DOI: 10.1371/journal.pone.0117730



Search for master regulators



$$S(X) = \sum_{r=1}^R \frac{M(X,r)}{M_{max}(r)} \cdot \frac{1}{1 + pN(X,r)/N_{max}(r)}$$

Where:

R - Max radius (input parameter)

p - Penalty (input parameter)

N(X,r) - total number of molecules reachable from key molecule X within the radius r.

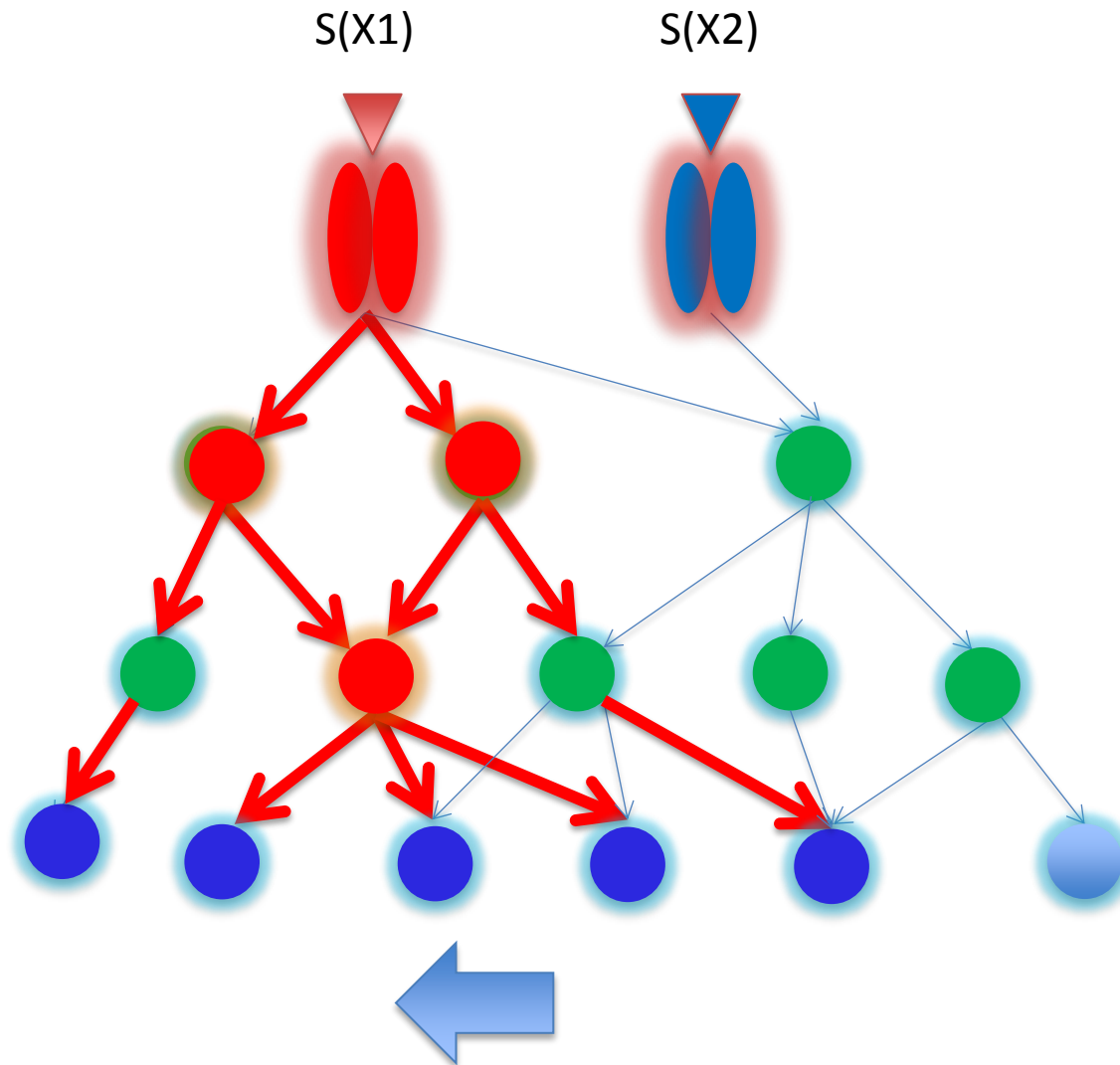
N_{max}(r) - maximal value of N(X,r) over all key molecules X found for this radius.

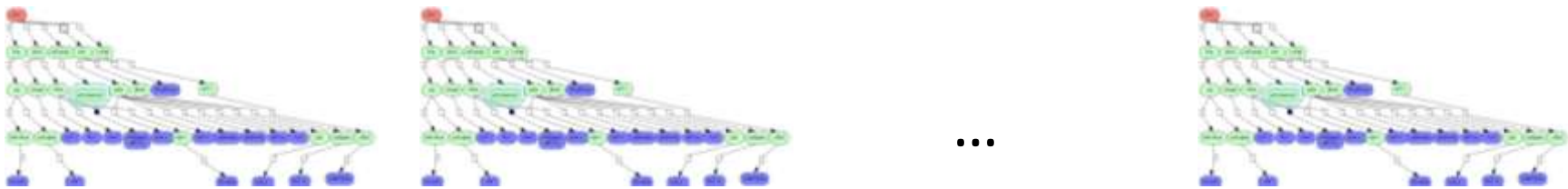
M(X,r) - sum of w(X) for all hits reachable from key molecule X within the radius r, where w(X) - weight of hit X.

M_{max}(r) - maximal value of M(X,r) over all key molecules X found for this radius.

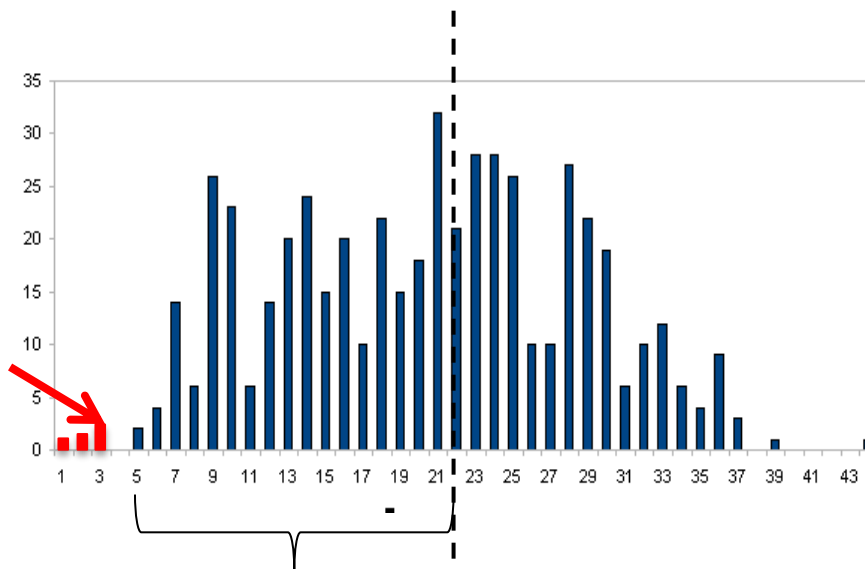
Kel, A., Voss, N., Jauregui, R., Kel-Margoulis, O. and Wingender, E.: Beyond microarrays: Find key transcription factors controlling signal transduction pathways *BMC Bioinformatics* 7(Suppl. 2), S13 (2006).

Use of context proteins (red nodes) to find where pathway walks.





Master molecule name	Maximal radius	Reached from set	Reachable total	Score	FDR	Z-Score	Ranks sum	Hits names
betaARK-1(h)	9.925	148	23946	0.57997	0	4.76362	39	14-3-3eta(h), AKT-3-isoform1(h), AKT-3-isoform2(h), AKT-3-p51(h), Amphiphysin-BIN1(h), (more)
PDGFRbeta(h)	9.795	154	24088	0.49885	0	4.72898	45	14-3-3eta(h), AKT-3-isoform1(h), AKT-3-isoform2(h), AKT-3-p51(h), Amphiphysin-BIN1(h), (more)
RPTPalphaisoform3(h)	9.82	151	24484	0.47905	0	4.38887	57	14-3-3eta(h), AKT-3-isoform1(h), AKT-3-isoform2(h), AKT-3-p51(h), Amphiphysin-BIN1(h), (more)
DDTDalpha(h)	9.82	151	24484	0.47905	0	4.35105	61	14-3-3eta(h), AKT-3-isoform1(h), AKT-3-isoform2(h), AKT-3-p51(h), Amphiphysin



Rank of the key node

$$Z = \frac{X - \mu}{\sigma}$$

Overview of the geneXplain platform

The screenshot displays the geneXplain platform interface. At the top, the browser address bar shows `genexplain-platform.com/bioulmweb/#`. The navigation bar includes a search field with the text "Research: RESOL" and a dropdown menu set to "Default".

On the left, a "Tree of data analysis" is visible, showing a hierarchical structure of data folders and files. The main dashboard area is titled "Start page" and contains several colored tiles for different analysis types: RNA-seq, Proteomics, Epigenomics, ChIP-seq, Sequence analysis, miRNA, Microarrays, Drug targets, Pathways, NGS, Genomic variants, Popular functions, Gene or protein list, and Complete list of workflows.

At the bottom, the "Track SNP_indels.vcf" section is shown. It includes a "Sequence collection" field with the value "databases/Ensembl/Sequences/chromosomes GRCh37", a "Site count" of 10113, and a "Result of analysis" field with the value "gatk_unified_genotyper". Below this, an "Info Box" provides parameters for the analysis, such as "Choose the source" and "BAM file".

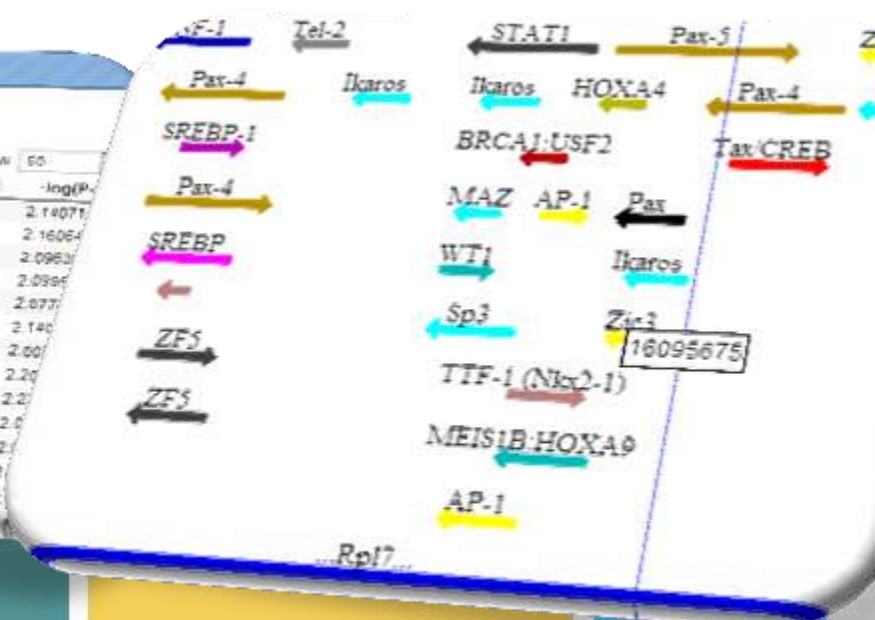
The "Operations field" at the bottom right shows a table of system operations. The table has columns for Type, Source, Start, End, and Status. The operations listed include "Add reactants" and "Venn diagrams".

Type	Source	Start	End	Status
analysis	Add reactants	2015.06.12 18:14:23	2015.06.12 18:15:11	Completed
analysis	Venn diagrams	2015.06.12 15:48:33	2015.06.12 15:48:47	Completed

geneXplain platform – a powerful tool to study molecular mechanisms of cellular processes using systems medicine approach

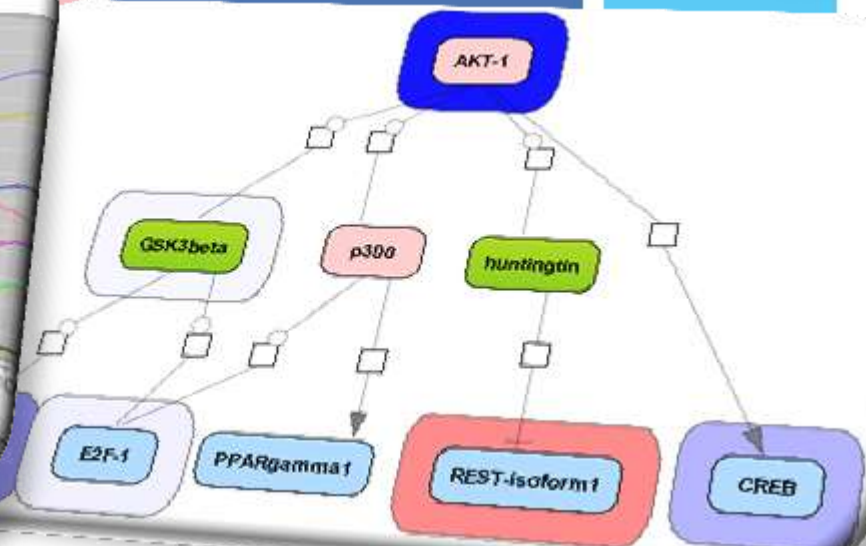
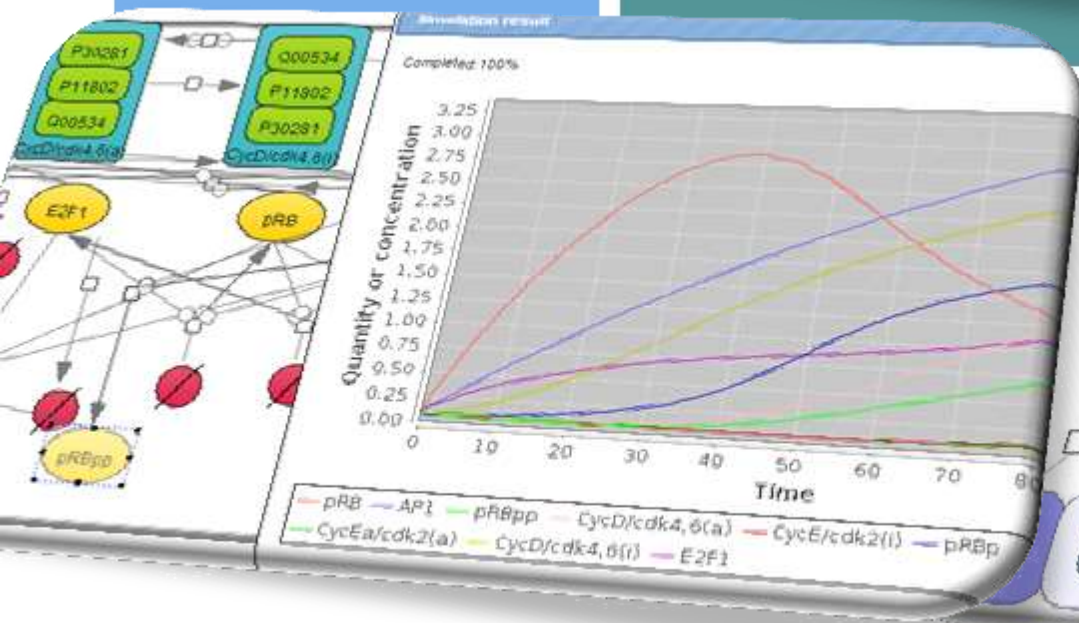
Ag_c-d-AMP_aged GS Upreg 1

ID	Gene symbol ID	ID	LogFoldChange	-log(P)
ENSGUSG00000000000	Mtck1	1.0499E7	3.26298	2.14077
ENSGUSG00000000000	Ahr5f1	1.0401E7	1.85450	2.16064
ENSGUSG00000000000	Ctla	1.0374E7	1.61515	2.09830
ENSGUSG00000000000	Nhg2	1.0378E7	1.68971	2.09960
ENSGUSG00000000000	Temp1	1.0598E7	1.7124	2.07700
ENSGUSG00000000000	Mars	1.0458E7	1.36993	2.14000
ENSGUSG00000000000	Fam50a	1.06E7	1.86801	2.00000
ENSGUSG00000000000	Bvrra	1.0470E7	1.5305	2.20000
ENSGUSG00000000000	Dexx	1.0444E7	1.3875	2.20000
ENSGUSG00000000000	L3kb7	1.0448E7	1.53795	2.20000
ENSGUSG00000000000	Mip2	1.0448E7	1.50815	2.20000
ENSGUSG00000000000	Rpl12s	1.0553E7	1.69186	2.20000
ENSGUSG00000000000	Map2	1.0367E7	1.38006	2.20000
ENSGUSG00000000000	Rpl7	1.0499E7	2.4001	2.20000



Microarrays

Drug targets



ChIP-Seq - Identify compo x

platform.genexplain.com/bioulweb/#de=analyses/Workflows/TRANSFAC/ChIP-Seq - Identify composite modules on peaks (TRANSFAC(R))

Research: 4Present

Start page ChIP-Seq - Identify composi... X

Databases Data
Analyses Users

data

- Examples
 - Atherosclerosis (
 - Brain Tumor GSE
 - Breast Cancer G
 - E2F1 binding reg
- Data
 - Gene cla
 - GSE141
 - GSM558
 - GSM558
 - GSM558
 - GSM558
 - GSM558
 - GSM558
 - Normaliz
- Journal
- Optimization
- Response to BMI
- SMAD1,5 binding
- Sample data
- Scripts
- Projects

Results folder

Input Yes track

Site search on track

out all

Sequence source

Input No track

Site search on track

out2 all

Profile

Maximal number of pairs

Number of iterations

Minimal number of pairs

Site search result optimization

out

out2

Construct composite modules c tracks

modules

mod

Site optimization summary

Search Info Default Overview My description Graph search Script Clipboard Tasks Workflow History

Master regulators in networks

Firefox

geneXplain 2.4.1 web edition

Mitotic regulation of the human ana...

cdg627 6598..6609 - cdg627.pdf

О компании | DevelopmentOnTheEd...

platform.genexplain.com/bioulweb/#

Conduit Search

Research: PNaS13

Start page

- [Normalize data](#)
- [Detect differentially expressed genes](#)
- [Discover functional enrichment](#)
- [Identify master regulators in networks](#)
- [Analyze regulatory genome regions](#)
- [Find potential targets](#)
- [Analyze ChIP-Seq data](#)
- [Functional analysis of genomic variations](#)

[Create your own workflow](#)

[About geneXplain](#)

Search Info Default

My description Graph search Script Clipboard Tasks

ID: Data
Size: 4
Size on disk: 7.3Mb (7,606,235 bytes)
Complete name: data/Projects/PNaS13/Data

Put your comment here - press Edit button above

Let's save the beauty!





Transcriptome analysis - Illumina HumanRef-8 v3.0 Gene Expression BeadChips (containing more than 24000 probes from RefSeq database).

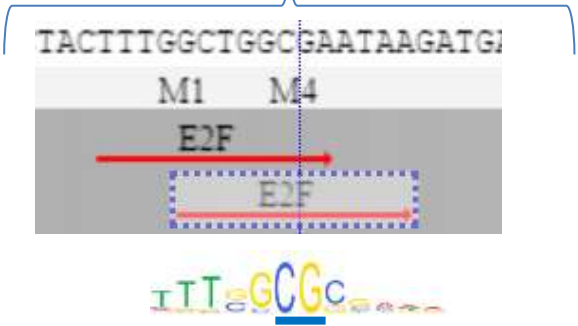
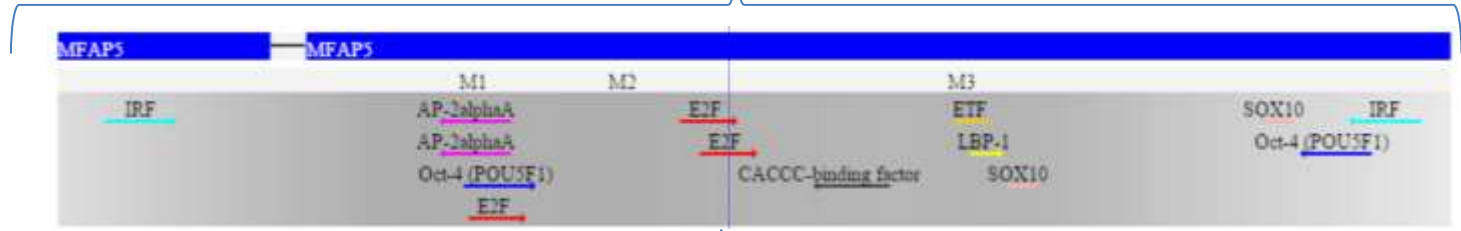
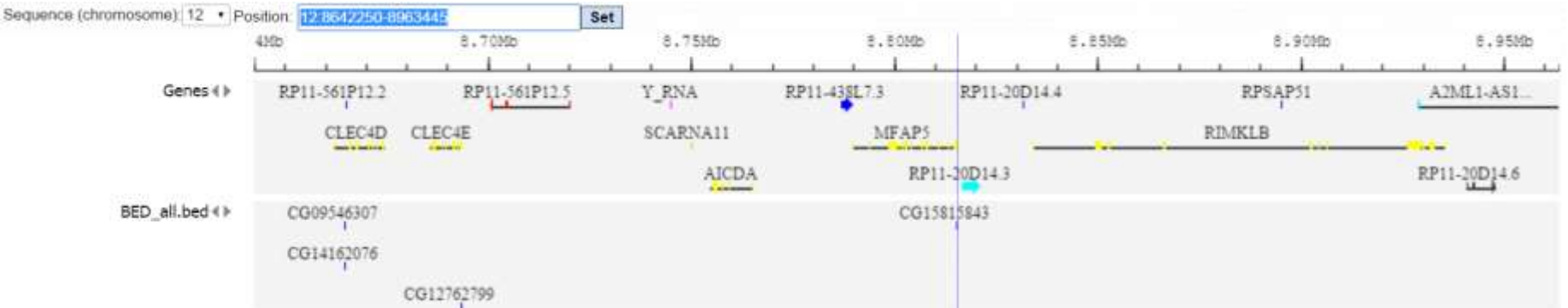
The primary data analysis - Illumina GenomeStudio software, Gene Expression Module (the raw intensities data were quantile-normalized). DiffScore cut-off was ± 20 (that corresponds to $p < 0.01$). **We have identified 65 genes** (39 of them - up-regulated, and 26 - down), **whose expression level is significantly different in varicose and healthy veins.**

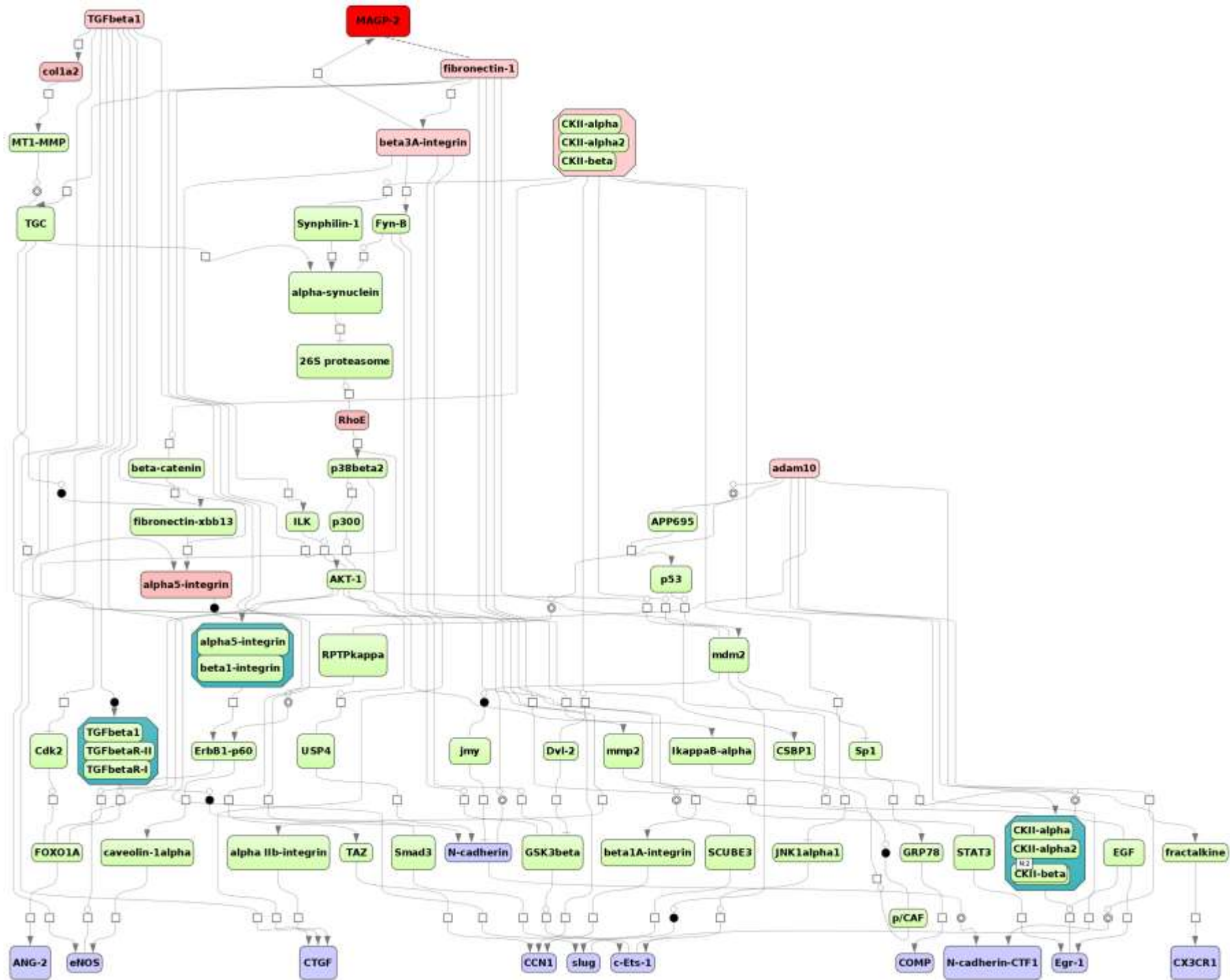


Methylation events - Illumina Infinium HumanMethylation27 BeadChips (27578 CpG measurements spanning 14495 genes).

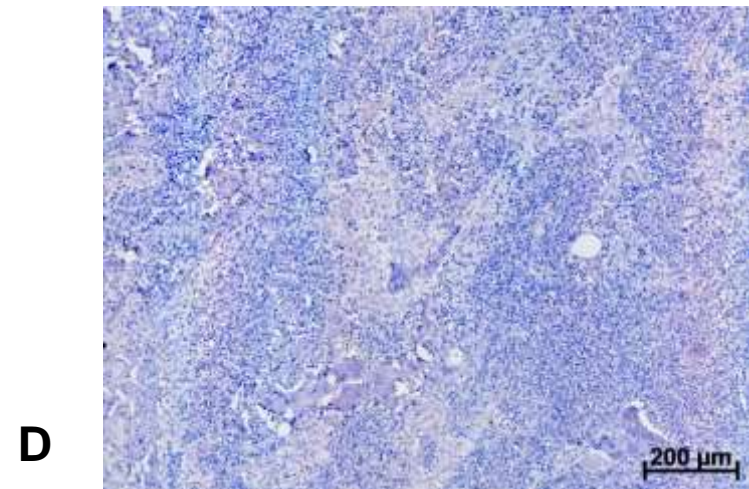
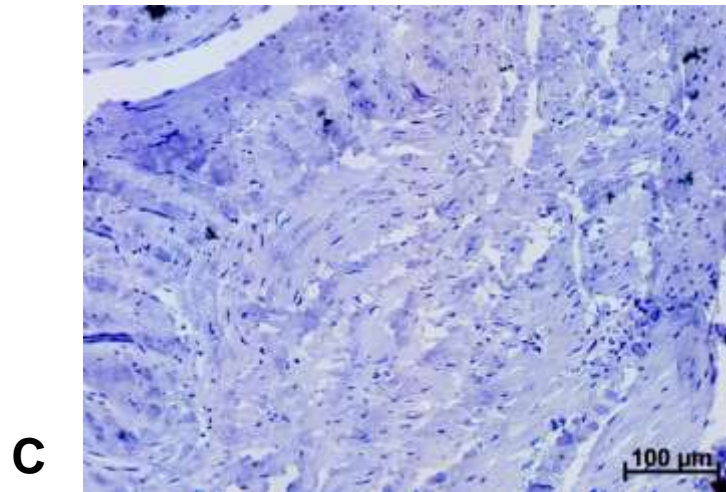
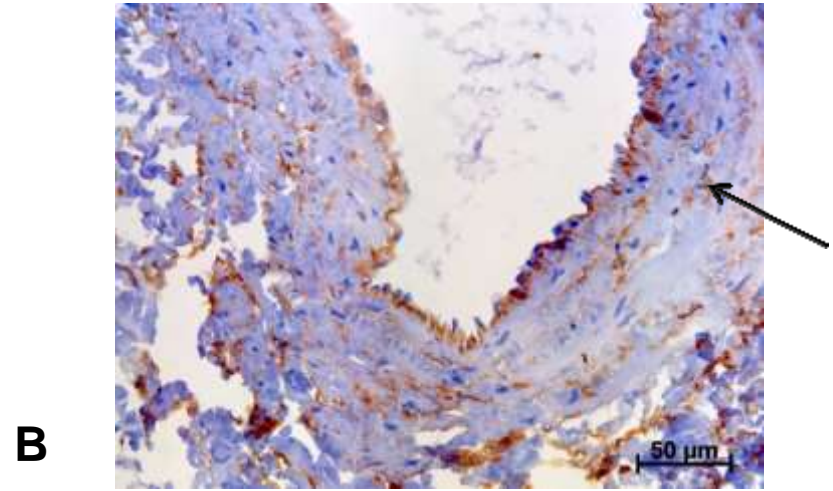
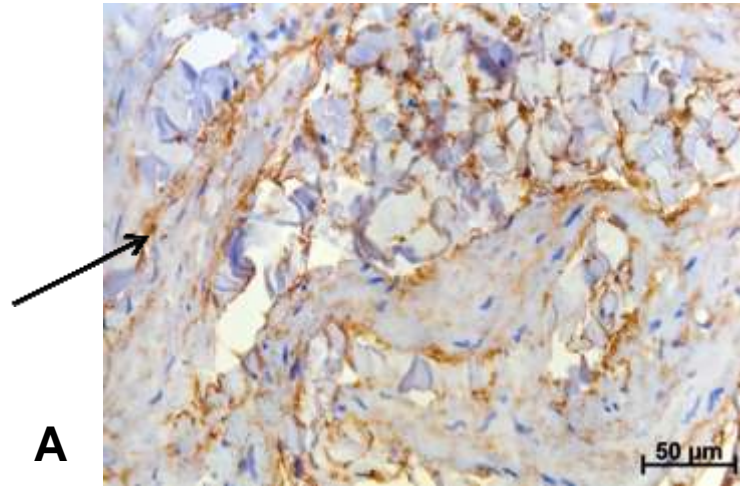
The primary data analysis - Illumina GenomeStudio software, Methylation Module (the raw intensities data were quantile-normalized). DiffScore cut-off was ± 20 (that corresponds to $p < 0.01$). **We have identified 288 genes** (227 of them - hypermethylated, and 61 - hypomethylated). **Then gene expression data were integrated with methylation project.**

Varicose.DiffScore ▼	Pearson ▲	Gene symbol ▲	Pearson_Negative055 Up_TSS1200: Schematic ▲
19.99524	-0.68412	MFAP5	
16.99033	-0.5834	SMS	
15.66516	-0.65709	RIOK2	
13.27985	-0.71742	SOCS2	
10.42101	-0.5856	C10orf32	
7.89341	-0.61673	CFL2	
7.75691	-0.68584	ATP6V0A2	
7.45296	-0.56828	SUMO2	
6.5679	-0.88606	MEGF11	
6.52747	-0.6005	OR3A1	
6.19046	-0.62481	TXNDC12	
6.00776	-0.56719	MTHFS	
5.91968	-0.70275	FANCL	
5.60635	-0.72475	TKT	
5.59807	-0.5501	PAFAH2	



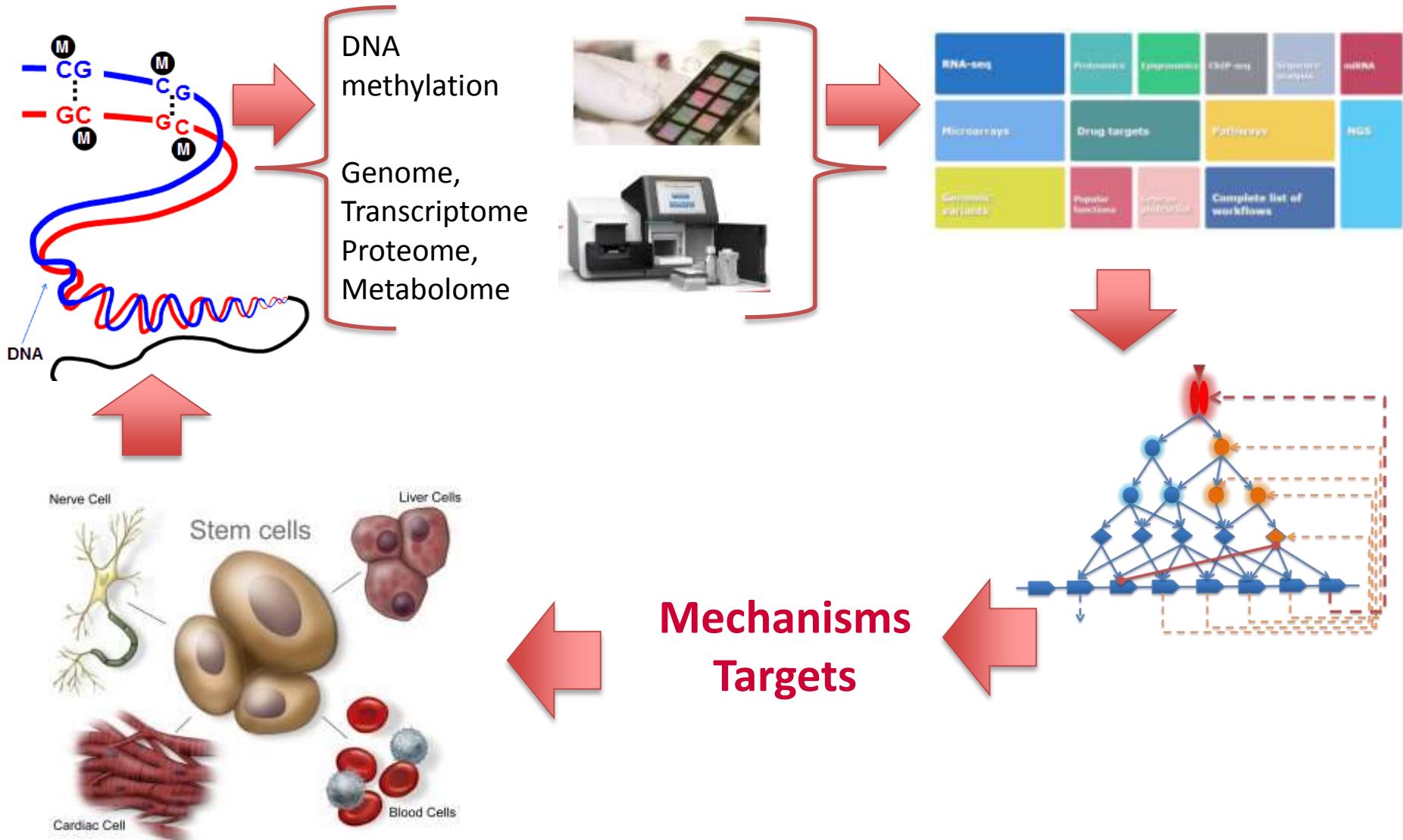


Immunohistochemical staining of MFAP5 for varicose and healthy human veins



Expression of MFAP5 in varicose and normal/healthy veins. The MFAP5-positive staining in varicose veins (A) was increased compared with that in normal/healthy veins (B). There was no staining in varicose veins without primary (MFAP5) anti-body (C). No specific staining was observed in lymph nodes (D) used as negative control.

Application of geneXplain platform to stem cell research



Scope: Format: Amount: GEO accession:

Series GSE74151 [Query DataSets for GSE74151](#)

Status Public on Oct 30, 2015
Title Expression data from three types of spermatogonial stem cells.
Organism [Mus musculus](#)
Experiment type Expression profiling by array
Summary Multipotent spermatogonial stem cells (mSSCs) derived from SSCs are a potential new source of individualized pluripotent cells in regenerate medicine such as ESCs. We hypothesized that the culture-induced reprogramming of SSCs was mediated by a mechanism different from that of iPS, and was due to up-regulation of specific pluripotency-related genes during cultivation. Through a comparative analysis of expression profile data, we try to find cell reprogramming candidate factors from mouse spermatogonial stem cells. We used microarrays to analyze the gene expression profiles of culture-induced reprogramming converting unipotent spermatogonial stem cells to pluripotent spermatogonial stem cells.

Overall design Three types of spermatogonial stem cells were mechanically collected according to morphological criteria for RNA extraction and hybridization on Affymetrix microarrays.

Contributor(s) [Lee DR, Lee I](#)
Citation missing *Has this study been published? Please [login](#) to update or [notify GEO](#).*
Submission date Oct 19, 2015
Last update date Nov 24, 2015
Contact name Insuk Lee
E-mail insuksysbio@gmail.com
Organization name Yonsei University
Department Biotechnology
Lab Network biology laboratory
Street address 50 Yonsei-ro, Seodaemun-gu
City Seoul
ZIP/Postal code 120-749
Country South Korea

- Platforms (1)** [GPL1261 \[Mouse430_2\] Affymetrix Mouse Genome 4](#)
- Samples (6)**
[GSM1911697 SSC_rep1](#)
[GSM1911698 SSC_rep2](#)
[GSM1911699 iSSC_rep1](#)
[GSM1911700 iSSC_rep2](#)
[GSM1911701 mSSC_rep1](#)
[GSM1911702 mSSC_rep2](#)

This SubSeries is part of SuperSeries:
[GSE74156](#) An integrated systems biology approach identifies positive cell pluripotency regulatory factor

Relations
 BioProject [PRJNA299186](#)

Analyze with GEO2R

Download family

- [SOFT formatted family file\(s\)](#)
- [MINiML formatted family file\(s\)](#)
- [Series Matrix File\(s\)](#)

Supplementary file	Size	Download
GSE74151_RAW.tar	21.8 Mb	(http)(custom)
<i>Raw data provided as supplementary file</i>		
<i>Processed data provided as supplementary file</i>		

Example of applying geneXplain platform In stem cell research

Step 1: Identification of differentially expressed genes

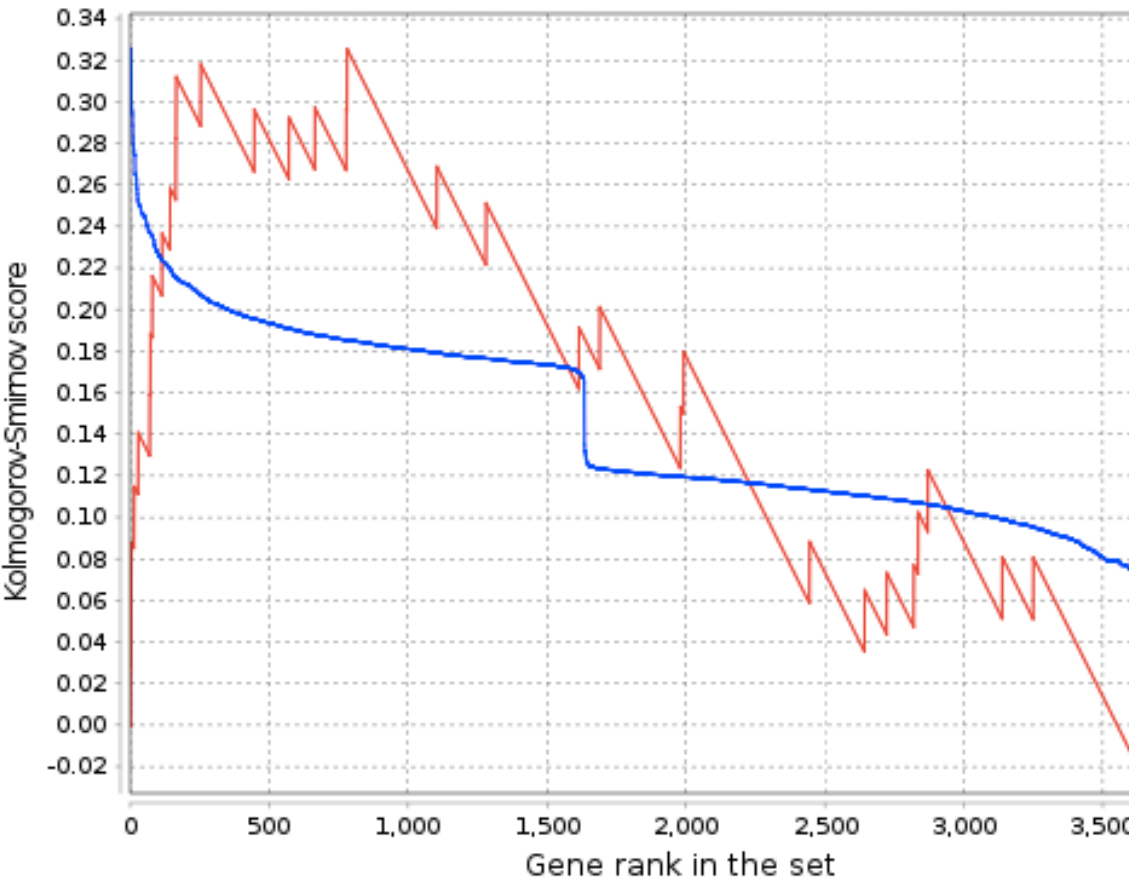
ID	Gene description	Gene symbol	Affymetrix ID	Condition_3 FC
ENSMUSG00000032494	teratocarcinoma-derived growth factor 1	TdGF1	1450989_at	8.67469
ENSMUSG00000012396	Nanog homeobox	Nanog	1429388_at	8.6699
ENSMUSG00000067860	zinc finger protein of the cerebellum 3	Zic3	1423424_at	7.96982
ENSMUSG00000050917	fibroblast growth factor 4	Fgf4	1420085_at	7.50462
ENSMUSG00000047146	tet methylcytosine dioxygenase 1	Tet1	1455425_at	7.43141
ENSMUSG00000025056	nuclear receptor subfamily 0, group B, member 1	Nr0b1	1417760_at	7.23791
ENSMUSG00000095180	reproductive homeobox 5	Rhox5	1423429_at	7.19038
ENSMUSG00000087166	LINE-1 type transposase domain containing 1	L1td1	1457314_at	7.11816
ENSMUSG00000074303	predicted gene 10664	Gm10664	1436287_at	6.87779
ENSMUSG00000079471	predicted gene 7325	Gm7325	1456242_at	6.60537
ENSMUSG00000043629	RIKEN cDNA 1700019D03 gene	1700019D03Rik	1430368_s_at	6.43018
ENSMUSG00000005892	thyrotropin releasing hormone	Trh	1418756_at	6.42415
ENSMUSG00000027702	leucine rich repeat containing 34	Lrrc34	1429366_at	6.32934
ENSMUSG00000000730	DNA (cytosine-5-)-methyltransferase 3-like	Dnmt3l	1425035_s_at	6.29559
ENSMUSG00000021255	estrogen related receptor, beta	Esrrb	1436926_at	6.19174
ENSMUSG00000030669	calcitonin/calcitonin-related polypeptide, alpha	Calca	1452004_at	6.1866
ENSMUSG00000021848	orthodenticle homolog 2 (Drosophila)	Otx2	1425926_a_at	6.17099
ENSMUSG000000031297	solute carrier family 7 (cationic amino acid transporter, y+ system), member 3	Slc7a3	1417022_at	5.73722
ENSMUSG00000066652	left-right determination factor 2	Lefty2	1436227_at	5.73566
ENSMUSG00000021953	L-threonine dehydrogenase	Tdh	1449064_at	5.71968
ENSMUSG000000081037	predicted gene 13929	Gm13929	1449064_at	5.71968
ENSMUSG000000035186	ubiquitin D	Ubd	1419762_at	5.63106
ENSMUSG00000029695	aminoadipate-semialdehyde synthase	Aass	1423523_at	5.62094
ENSMUSG000000031139	mcf.2 transforming sequence	Mcf2	1419021_at	5.4658
ENSMUSG000000054753	expressed sequence AU018091	AU018091	1435154_at	5.37406
ENSMUSG00000044393	desmoglein 2	Dsg2	1439476_at	5.27011
ENSMUSG000000027500	stathmin-like 2	Stmn2	1423281_at	5.16293
ENSMUSG000000053062	junction adhesion molecule 2	Jam2	1431417_at	5.09923
ENSMUSG000000037171	nodal	Nodal	1422057_at	5.06873
ENSMUSG00000077886	microRNA 295	Mir295	1444292_at	5.01388
ENSMUSG00000077903	microRNA 294	Mir294	1444292_at	5.01388
ENSMUSG00000078035	microRNA 293	Mir293	1444292_at	5.01388
ENSMUSG00000070708	gametocyte specific factor 1-like	Gtsf1l	1418994_at	4.99413
ENSMUSG00000042414	PR domain containing 14	Prdm14	1444390_at	4.98854
ENSMUSG00000028773	fatty acid binding protein 3, muscle and heart	Fabp3	1416023_at	4.98846
ENSMUSG00000056366	fatty acid binding protein 3, muscle and heart, pseudogene 1	Fabp3-ps1	1416023_at	4.98846
ENSMUSG000000032346	oocyte expressed protein	Ooep	1460471_at	4.9546
ENSMUSG000000028696	IAP promoted placental gene	Ipp	1421771_a_at	4.94328

Step 2: GSEA (Gene Set Enrichment Analysis) of Up and Dn regulated genes – Identification of **activated functions/systems in the cells**

ID	Title	Group size	Expected hits	Nominal P-value	ES	Rank at max	NES	FDR	Number of hits	Plot	Hit names
GO:0000087	M phase of mitotic cell cycle	197	30.88841	0	0.53871	1392	3.64071	0	31	View	Ahctf1, Birc5, Cdc25a, Cdk1, Cenpk, (more)
GO:0000279	M phase	198	31.0452	0	0.53871	1392	3.54453	0	31	View	Ahctf1, Birc5, Cdc25a, Cdk1, Cenpk, (more)
GO:0000278	mitotic cell cycle	364	57.073	0	0.36121	1621	3.37359	0	65	View	Ahctf1, Birc5, Ccna2, Ccnd1, Ccne1, (more)
GO:0000236	mitotic prometaphase	100	15.6794	0	0.65379	1318	3.3642	0	18	View	Ahctf1, Birc5, Cdk1, Cenpk, Csnk2a1, (more)
GO:1901605	alpha-amino acid metabolic process	222	34.80826	0	0.3259	1610	3.27387	0	71	View	Aass, Adi1, Aldh18a1, Aldh1l1, Aldh5a1, (more)
GO:0006520	cellular amino acid metabolic process	432	67.73499	0	0.2508	1641	3.1895	0	121	View	Aasdhppt, Aass, Abat, Acat1, Adi1, (more)
GO:0006396	RNA processing	655	102.70004	0	0.26813	1642	3.12109	6.669E-5	102	View	Adad1, Adar, Adat2, Ago2 (ENSMUSG00000036), (more)
GO:0046395	carboxylic acid catabolic process	200	31.35879	0	0.34628	1610	3.0486	1.5795E-4	56	View	Aass, Abat, Abcd2, Abhd10, Acads, (more)
GO:0010467	gene expression	657	103.01363	0	0.25648	2085	3.04602	1.539E-4	106	View	Adar, Akt1, Alyref, Anp, Brf1, (more) 1190003.115Rik
GO:0003407	neural retina development	41	7.23144	0	0.71322	548	2.30586	0.00845	7	View	Esrrb, Gli2, Isl1, Pou4f2, Prdm1, Tulp1, Zic3
GO:0046395	carboxylic acid catabolic process	136	23.9872	0	0.34627	1345	2.30419	0.00841	30	View	Aass, Abcd2, Acad9, Acads, Acadvl, (more)
GO:0016071	mRNA metabolic process	199	35.09892	0	0.31966	1661	2.27777	0.00978	37	View	Adar, Apobec2, Auh, Bag4, Cdk12, (more)
GO:0048863	stem cell differentiation	184	32.45327	0.001	0.25914	547	2.27644	0.00978	60	View	2410018M08Rik, Adipor1, Akt1, Arg1, Bcas3, (more)
GO:0048864	stem cell development	95	16.75577	0.001	0.3564	1320	2.2634	0.01065	27	View	Akt1, Arg1, Chd1, Eph4, Etv5, (more)
GO:0009310	amine catabolic process	83	14.63925	0	0.45796	1345	2.24409	0.01205	16	View	Aass, Aldh1l1, Aldh5a1, Arg2, Bcat2, (more)
GO:0019827	stem cell maintenance	85	14.992	0	0.38871	1003	2.24385	0.01204	23	View	Akt1, Chd1, Eph4, Etv5, Gdf3, (more)

Aticated system: "Stem cell maintenance"

Image x



— Condition_3 FC — GO:0019827 stem cell maintenance

Condition_3 FC	Gene symbol
8.6699	Nanog
7.50462	Fgf4
7.43141	Tet1
6.19174	Esrrb
5.06873	Nodal
4.35994	Dppa2,Gm9788
4.33451	Tfap2c
4.26574	Sox2
3.70146	Tbx3
3.49861	Pou5f1
3.29016	Cdc73
3.29016	Cdc73
2.87362	Mtf2
2.32343	Tcf7l1
2.09654	Dll1
1.95014	Mcp1
1.82427	Sfrp1
1.82229	Erdr1,Gm21887
1.52504	Smc1a

TF site enrichment analysis

- **Use cases:**
 - **Find transcription factors for co-regulated gene set**
 - **Identify motifs “co-enriched” in ChIP-seq regions**
- Which binding sites are significantly enriched in the Yes sequences?
- Site optimization tool
- Automatic threshold optimization
- One-sided binomial test for significant enrichment of sites

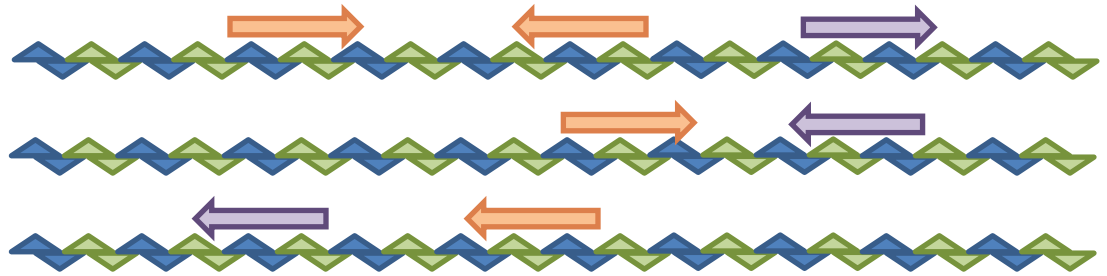
$$P(k \geq n) = \sum_{k=n}^N \binom{N}{k} \cdot p^k \cdot (1-p)^{N-k}$$

$$p = \frac{\#Seq_{Yes}}{\#Seq_{Yes} + \#Seq_{No}}$$

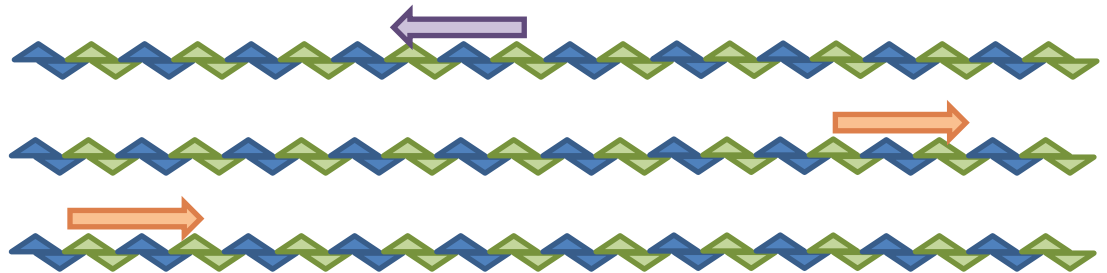
$$n = \#Sites_{Yes}$$

$$N = \#Sites_{Yes} + \#Sites_{No}$$

Yes sequences



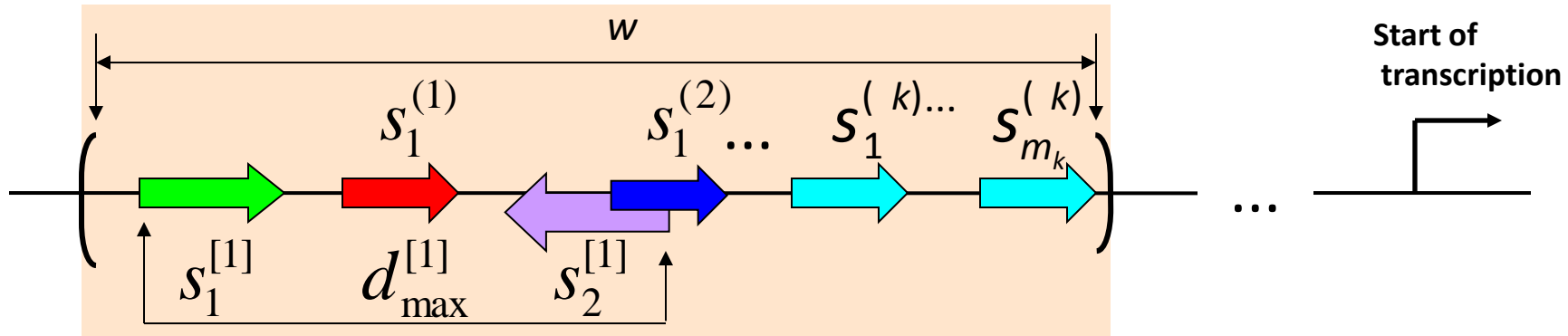
No sequences



Results of TF site enrichment analysis for promoters of up-regulated genes of “stem cell maintenance”

ID	Yes density per 1000bp	No density per 1000bp	Yes-No ratio	Model cutoff	P-value
<u>V\$NANOG_01</u>	0.05051	0.00193	26.22222	0.9839	0.01435
V\$ZNF143_03	0.05051	0.00193	26.22222	0.9561	0.01435
V\$CPHX_01	0.10101	0.00578	17.48148	0.979	7.4128E-4
V\$HDX_01	0.07576	0.0077	9.83333	0.9915	0.01002
V\$BBX_03	0.25253	0.06163	4.09722	0.9223	5.7023E-4
<u>V\$POU2F1_Q6</u>	0.15152	0.04237	3.57576	0.9894	0.01229
V\$PAX_Q6	0.22727	0.07319	3.10526	0.9316	0.00521
V\$GCM2_01	0.35354	0.12519	2.82393	0.9249	0.00124
V\$HMGY_Q3	0.35354	0.15408	2.29444	0.9461	0.00643
V\$MAZR_01	0.60606	0.30239	2.00425	0.9651	0.00233
V\$GEN_INI_B	0.37879	0.19068	1.98653	0.9949	0.01494
V\$ZFP161_04	1.13636	0.78005	1.45679	0.8306	0.01322
V\$NF1A_Q6_01	1.26263	0.87635	1.44078	0.9932	0.0113
<u>V\$P53_Q3</u>	1.69192	1.17874	1.43537	0.9514	0.00428
V\$GLI_Q3	1.66667	1.22304	1.36273	0.9363	0.01229
V\$SRY_Q6	2.12121	1.56972	1.35133	1	0.00655
V\$PBX_Q3	2.62626	2.04738	1.28274	0.839	0.01058
V\$FAC1_01	4.11616	3.25116	1.26606	0.8738	0.00293
V\$P53_04	8.10606	6.6198	1.22452	0.765	3.9839E-4
V\$AP2ALPHA_03	6.91919	6.00924	1.15142	0.7523	0.01495

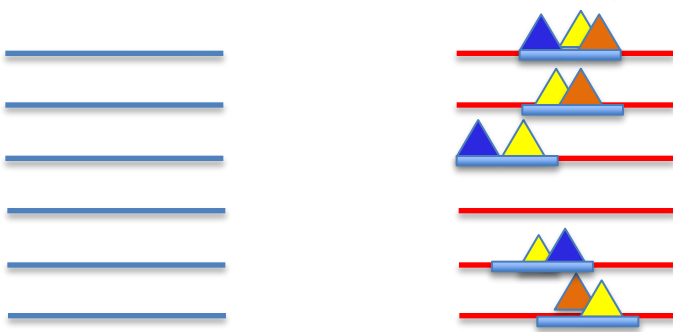
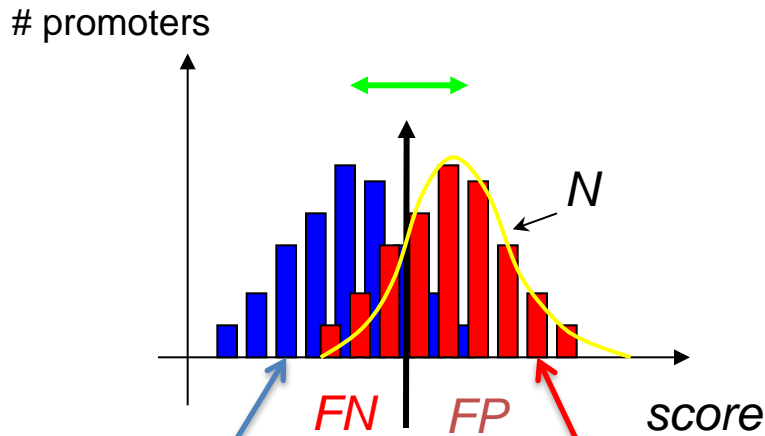
Строится модель энхансера специфического для стволовых клеток из комбинаций сайтов связывания транскрипционных факторов



$d_{\max}^{[1]}$	$d_{\max}^{[1]}$...	$d_{\max}^{[R]}$	} Параметры модели оптимизируются генетическим алгоритмом
$q_{cut-off}^{(1)}$	$q_{cut-off}^{(2)}$...	$q_{cut-off}^{(k)}$	
$\phi^{(1)}$	$\phi^{(2)}$...	$\phi^{(k)}$	

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

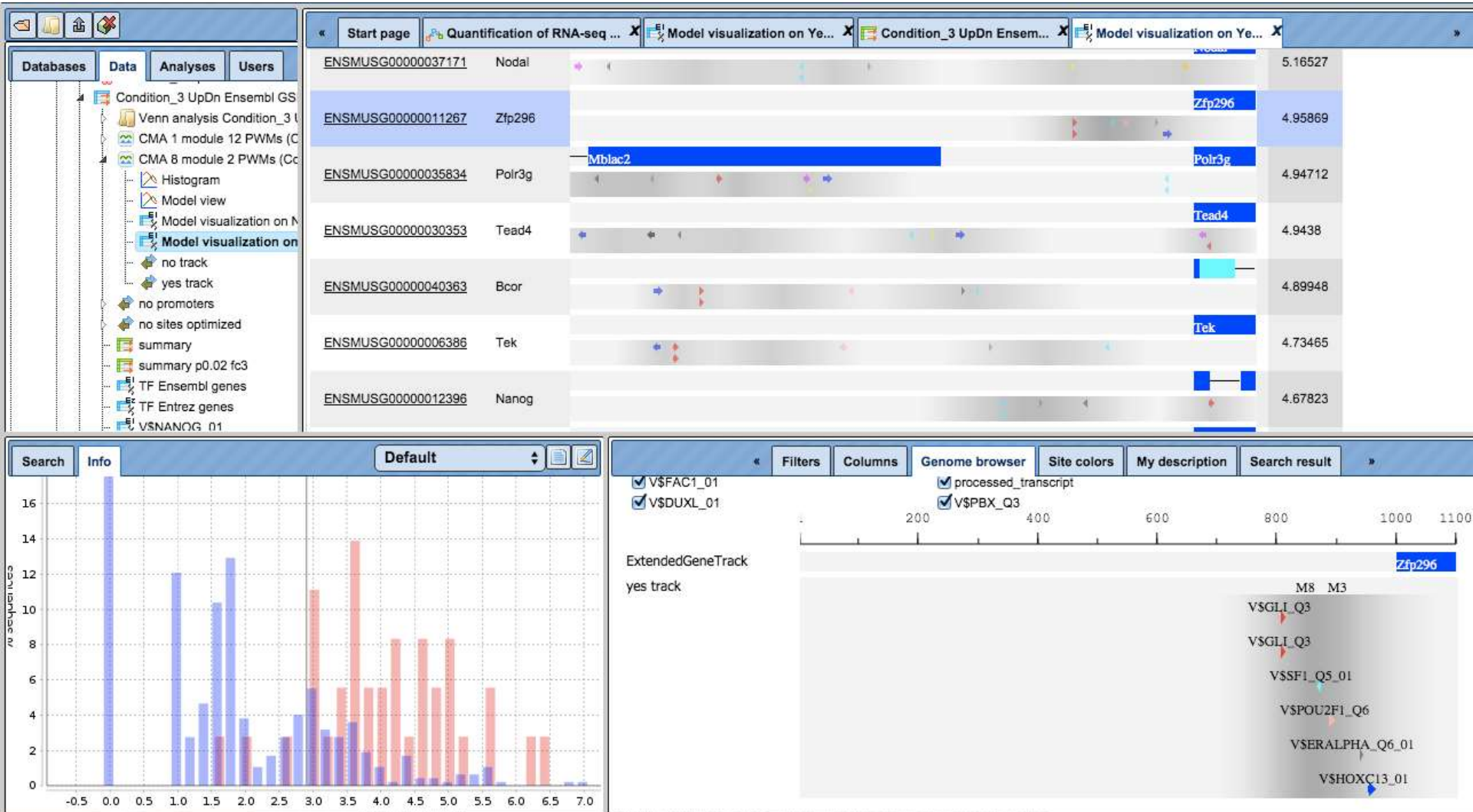
FP – false positives

T – T-test (difference between mean values)

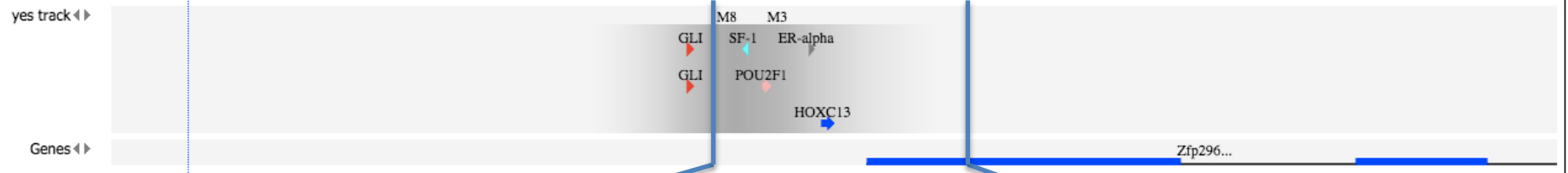
N – normal likeness

k – number of free parameters

Results of analysis Composite modules (combinations of TF binding sites) in promoters of up-regulated genes of “stem cell maintenance”

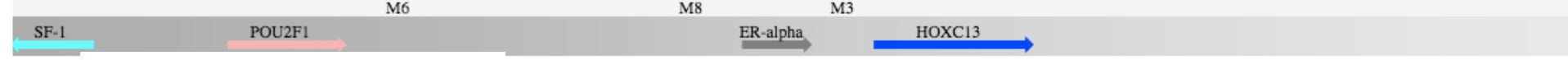


(chromosome): 7 Position: 7:19576468-19578035 Set



7152 19577180 19577200 19577220 19577240 19577260 19577280 19577300

:caaggcctgtggaggagggcacaatttgcatacaggtcooctgggcagcgggctgggagggaaggtcgagtttgacctccgagactggccccgtaaagacccccagtgggcggagacaaggttaggaccagcacgtgtaaaaagttcggc



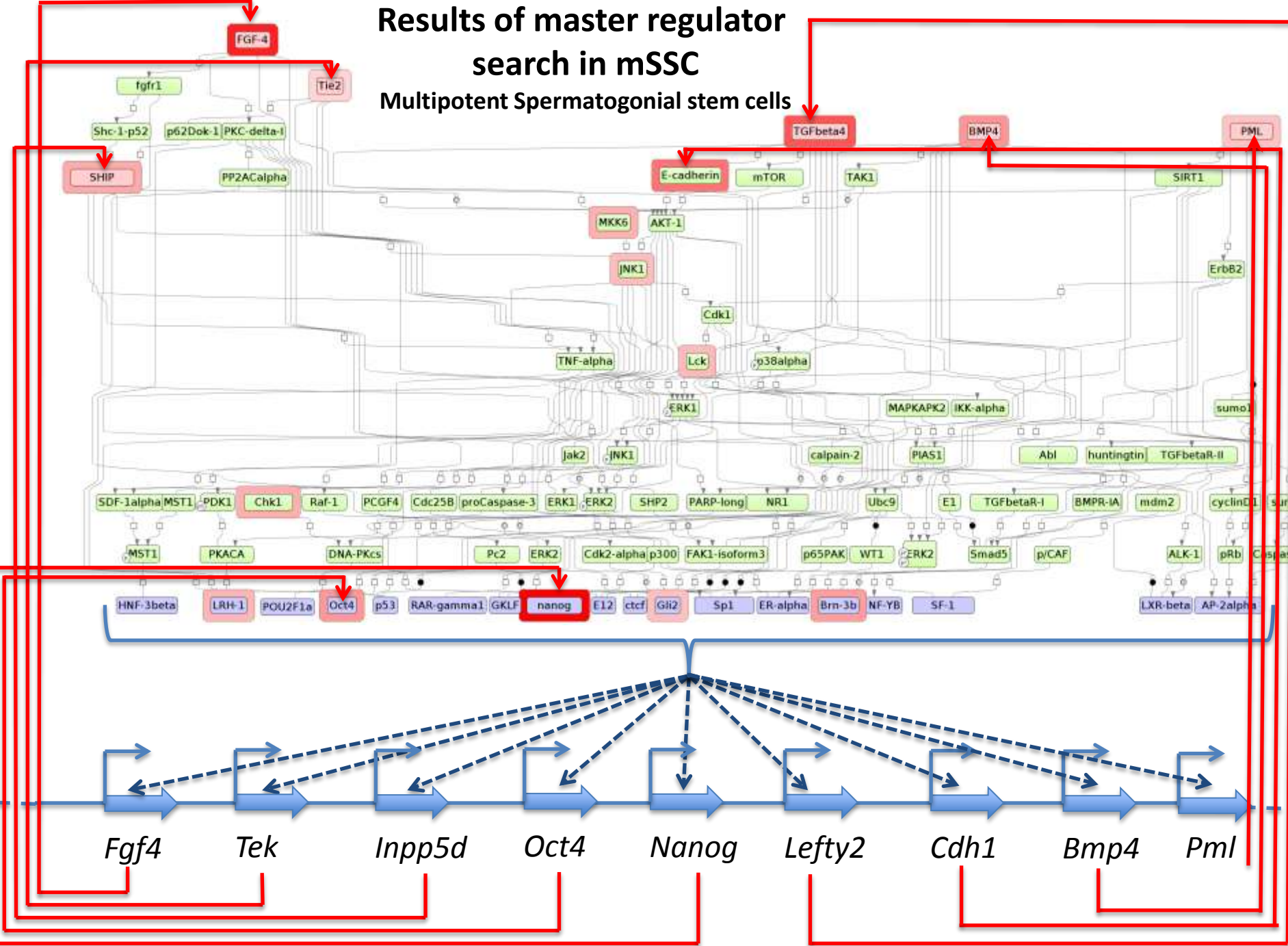
Oct-4



Zfp296... 19 57

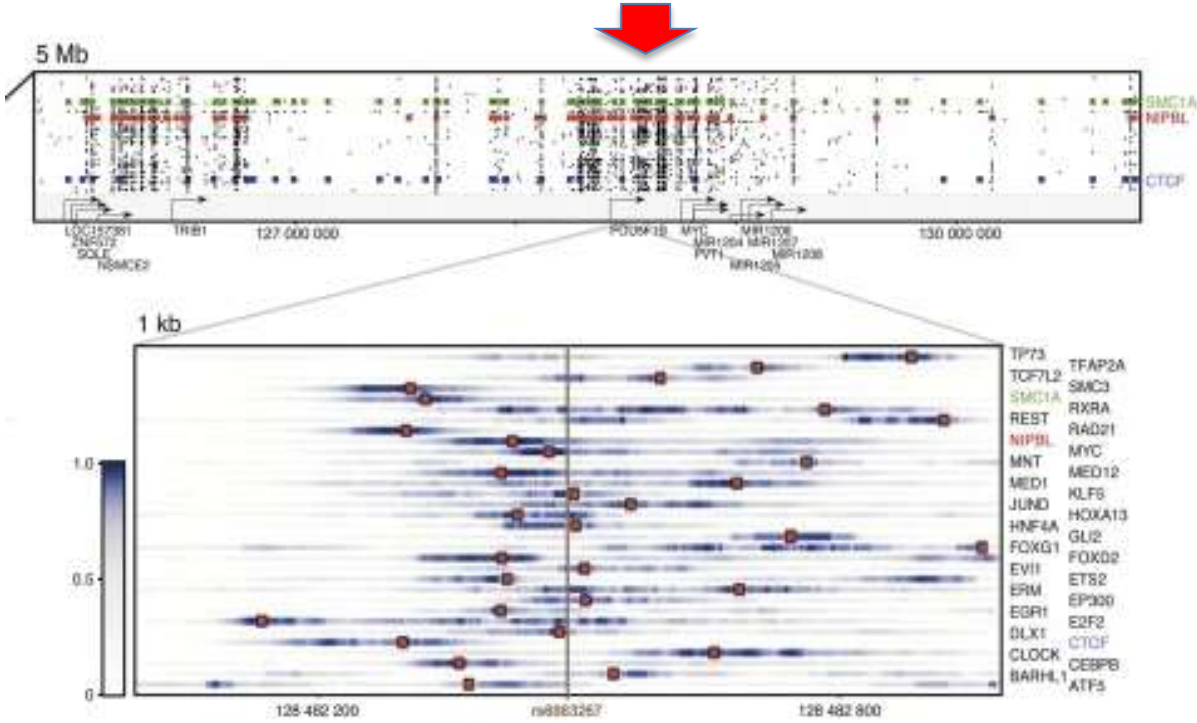
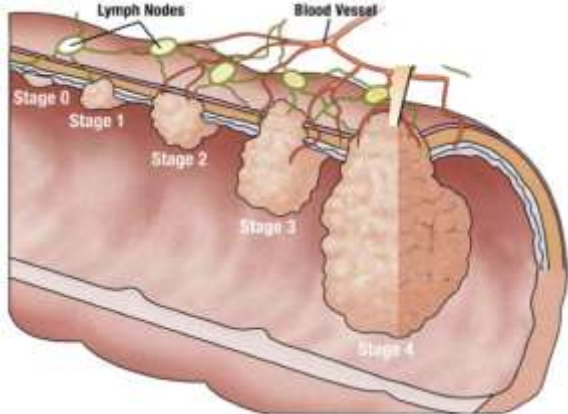
Results of master regulator search in mSSC

Multipotent Spermatogonial stem cells



Colon cancer

Super-enhancer



b

d

Sur et al., Science 2012



In this work, we generated mice deficient in *Myc-335*, a putative *MYC* regulatory element that contains rs6983267, a SNP accounting for more human cancer-related morbidity than any other genetic variant or mutation. In *Myc-335* null mice, *Myc* transcripts were expressed in the intestinal crypts in a pattern similar to that in wild-type mice but at modestly reduced levels. The mutant mice displayed no overt phenotype but were markedly resistant to intestinal tumorigenesis induced by the *APC^{min}* mutation. These results highlight the fact that although a disease-associated polymorphism typically has a relatively modest effect size, the element that it affects can be critically important for the underlying pathological process. The finding also indicates that normal growth control and pathological growth induced by cancer can utilize different mechanisms.

High-altitude adaptation in humans

From Wikipedia, the free encyclopedia

High-altitude adaptation in humans is an instance of **evolutionary modification** in human populations in **Tibet**, the **Andes** and **Ethiopia**, who have acquired the **ability to survive at extremely high altitudes**. The phrase is used to signify irreversible, **long-term physiological responses** to high-altitude environments, associated with heritable **behavioural** and **genetic changes**. While the rest of human population would suffer serious health consequences, these native inhabitants thrive well in the highest parts of the world. These people have undergone extensive physiological and genetic changes, particularly in the regulatory systems of **respiration** and **circulation**, when compared to the general lowland population.^{[1][2]} This special **adaptation** is now recognised as a clear example of **natural selection** in action.^[3] In fact, the adaptation account of the Tibetans has become the fastest case of **human evolution** in the scientific record, as it is estimated to have occurred in less than 3,000 years.^{[4][5][6]}



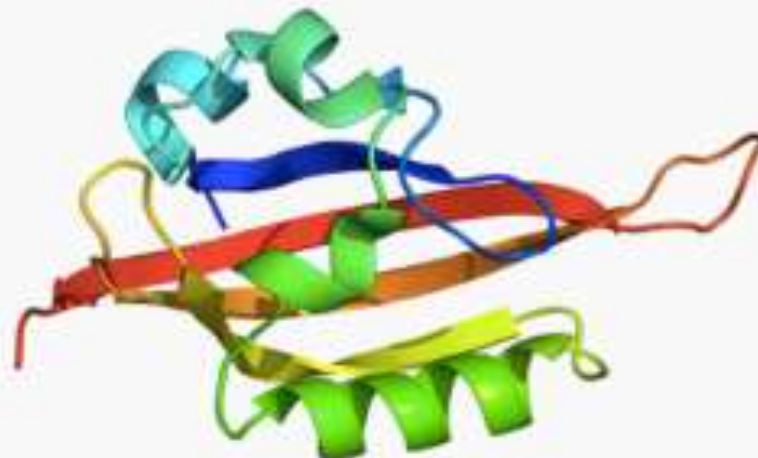
EPAS1

From Wikipedia, the free encyclopedia

Endothelial PAS domain-containing protein 1 (**EPAS1**, also known as hypoxia-inducible factor-2alpha (HIF-2alpha)) is a [protein](#) that in humans is encoded by the *EPAS1* gene. It is a type of [hypoxia-inducible factor](#), a group of [transcription factors](#) involved in body response to oxygen level.^{[1][2][3][4]} The gene is active under low oxygen condition called [hypoxia](#). It is also important in the development of the heart, and maintaining [catecholamine](#) balance required for protection of the heart. Mutation often leads to neuroendocrine tumors.

However, a special version (allele) of *EPAS1* produces EPAS1 which is responsible for [high-altitude adaptation in humans](#).^{[5][6]} It is known that the variant gene confers increased athletic performance in some people, and hence, is dubbed the "super athlete gene".^[7]

Endothelial PAS domain protein 1



A 3.4-kb Copy-Number Deletion near EPAS1 Is Significantly Enriched in High-Altitude Tibetans but Absent from the Denisovan Sequence.

Lou H¹, Lu Y¹, Lu D¹, Fu R¹, Wang X², Feng Q¹, Wu S¹, Yang Y³, Li S³, Kang L⁴, Guan Y⁵, Hoh BP⁶, Chung YJ⁷, Jin L³, Su B⁸, Xu S⁹.

⊕ Author information

Abstract

Tibetan high-altitude adaptation (HAA) has been studied extensively, and many candidate genes have been reported. Subsequent efforts targeting HAA functional variants, however, have not been that successful (e.g., no functional variant has been suggested for the top candidate HAA gene, EPAS1). With WinXPCNVer, a method developed in this study, we detected in microarray data a Tibetan-enriched deletion (TED) carried by 90% of Tibetans; 50% were homozygous for the deletion, whereas only 3% carried the TED and 0% carried the homozygous deletion in 2,792 worldwide samples ($p < 10^{-15}$). We employed long PCR and Sanger sequencing technologies to determine the exact copy number and breakpoints of the TED in 70 additional Tibetan and 182 diverse samples. The TED had identical boundaries (chr2: 46,694,276-46,697,683; hg19) and was 80 kb downstream of EPAS1. Notably, the TED was in strong linkage disequilibrium (LD; $r(2) = 0.8$) with EPAS1 variants associated with reduced blood concentrations of hemoglobin. It was also in complete LD with the 5-SNP motif, which was suspected to be introgressed from Denisovans, but the deletion itself was absent from the Denisovan sequence. Correspondingly, we detected that footprints of positive selection for the TED occurred 12,803 (95% confidence interval = 12,075-14,725) years ago. We further whole-genome deep sequenced ($>60\times$) seven Tibetans and verified the TED but failed to identify any other copy-number variations with comparable patterns, giving this TED top priority for further study. We speculate that the specific patterns of the TED resulted from its own functionality in HAA of Tibetans or LD with a functional variant of EPAS1.

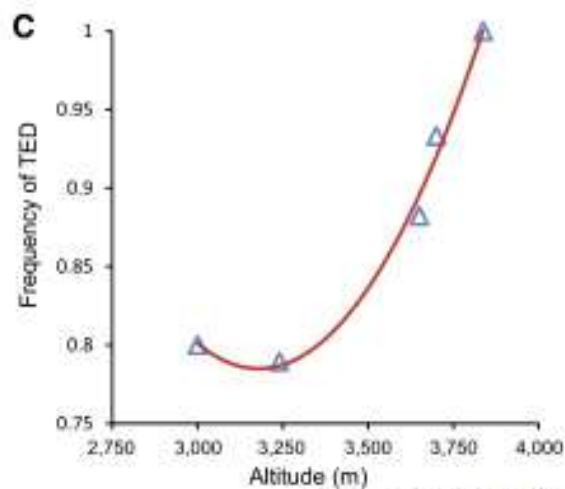
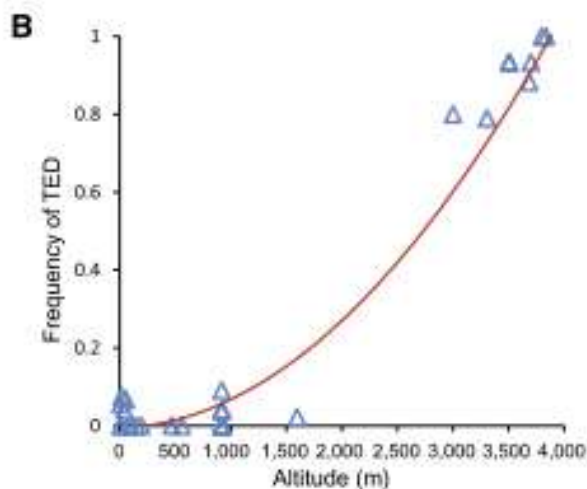
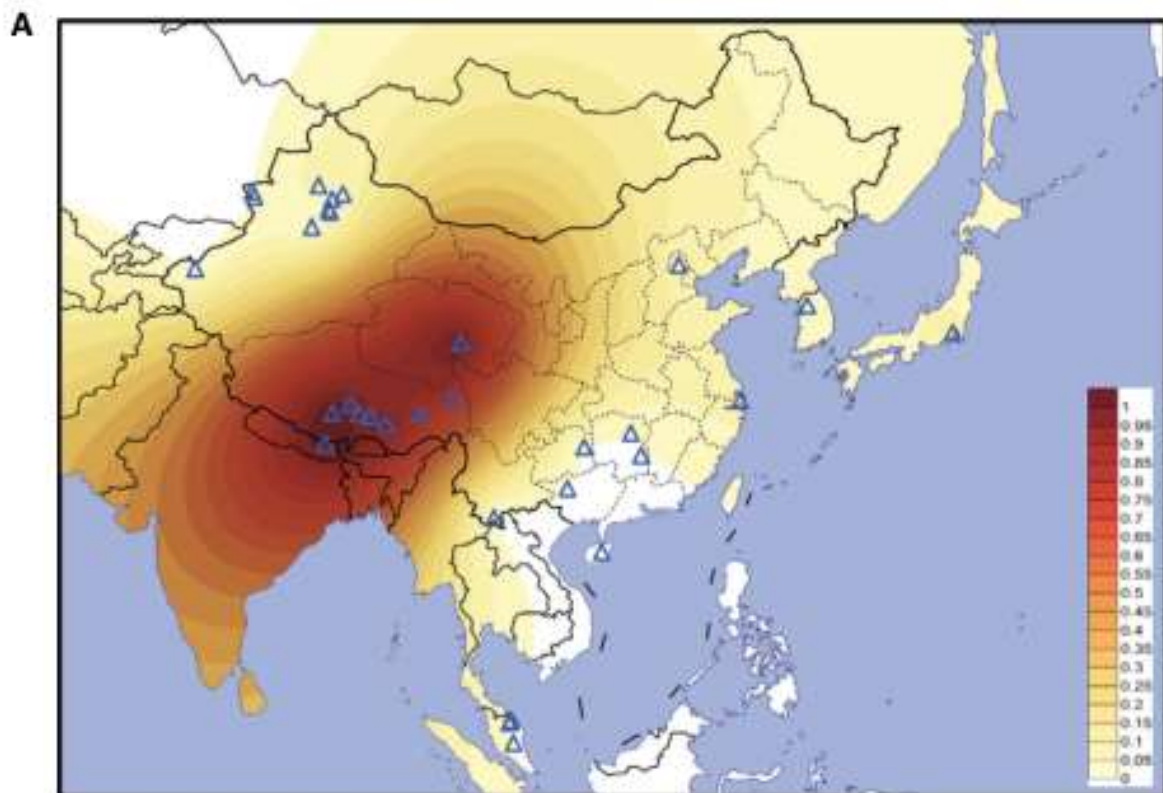


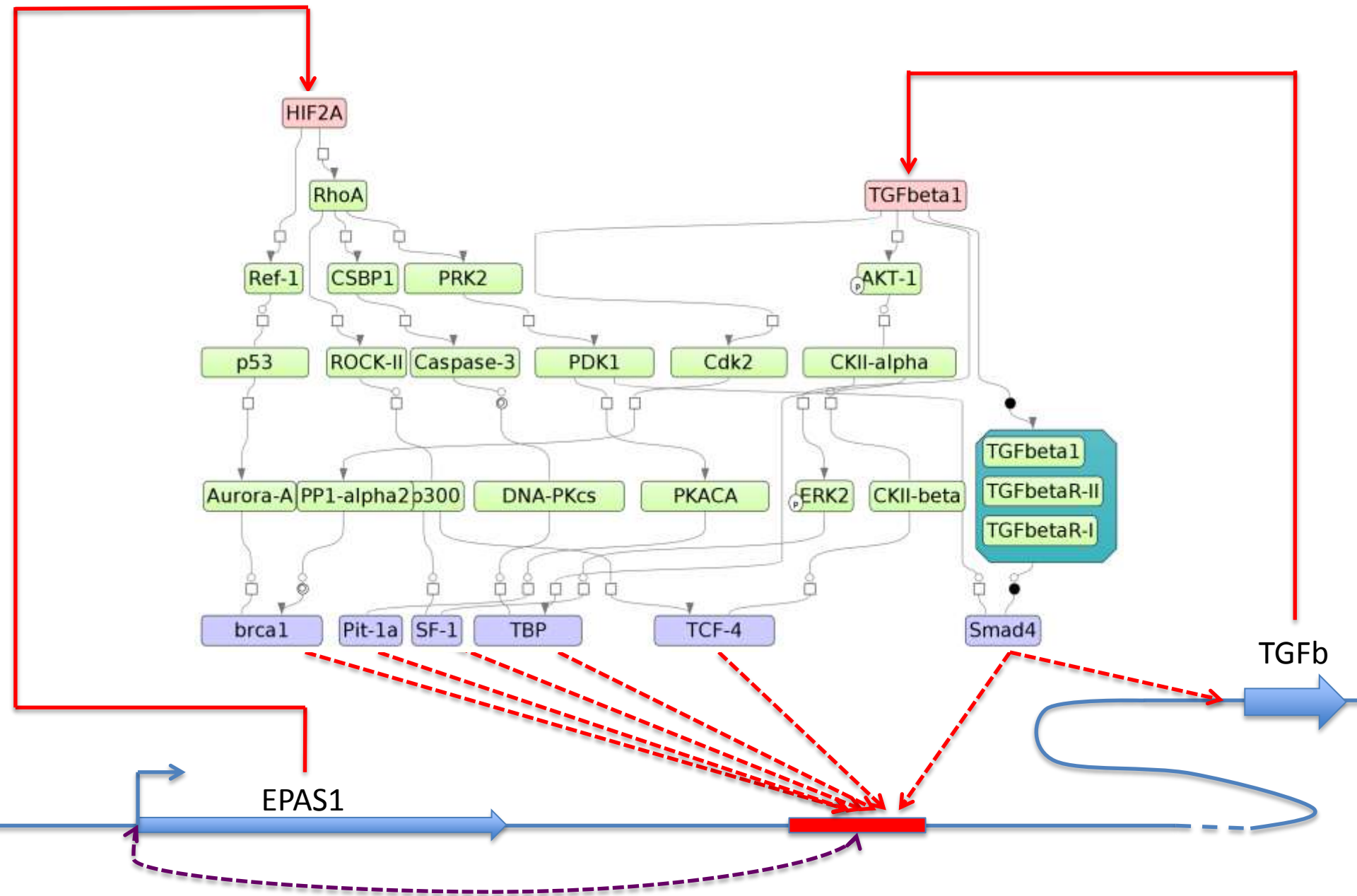
Figure 2. Distribution of the TED Frequency among Populations and Its Correlation with Altitude

(A) Distribution of deletion frequency in Asian populations. Colors from yellow to red indicate the frequency from low to high, respectively. Each blue triangle represents a sampled population.

(B) Deletion frequency correlated ($R^2 = 0.958$) with altitude in Asian populations (population information is listed in Table 2).

(C) Deletion frequency correlated ($R^2 = 0.989$) with altitude in five Tibetan sub-groups (Lhasa, Nyingchi, Qamdo, Shannan, and Shigatse).

(Figure S6). We inferred the haplotype of the TED with highly differentiated SNPs ($F_{ST} > 0.5$). Interestingly, one dominant haplotype occupied 88% of all the haplotypes with the deletion in Tibetans, and the deletion haplotype in the two Sherpa individuals was identical to the dominant one in the Tibetans (Figure S7). In addition, we investigated the relationship between TED frequency and the altitude of locations where Tibetan individuals live. Interestingly, we observed a strong correlation between the deletion frequency and the altitude in Asian populations ($R^2 = 0.958$) and





Bioinformatics SME in Lower Saxony

- Original world's biggest biological data
TRANSFAC®,
TRANSPATH (30 years of development)
- Unique approach for automatic discovery of
drug targets and biomarkers
- 1 mln. lines of code of geneXplain software
platform covering all bioinformatics fields.
- >1200 world-wide users including Big
Pharmas, red and green biotech and
academia.
- 12 employees in Germany and India and
cooperation with Russian software
development company (50 employees).



EU and German grants

