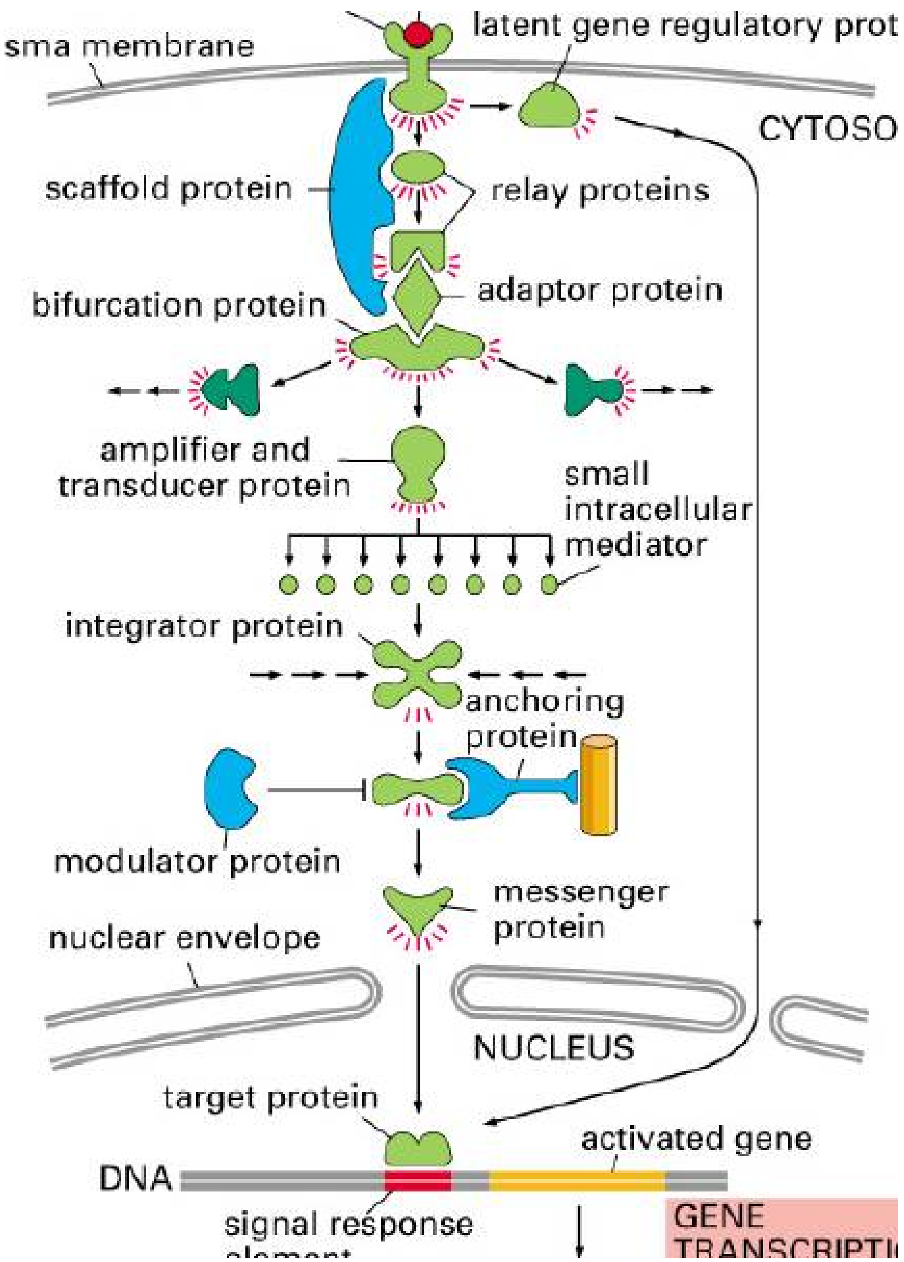
Integrative analyses of ErbB receptor signaling and transcriptional network

Takashi Nakakuki and Mariko Hatakeyama
Cellular Systems Biology Team
RIKEN Genomic Sciences Center

FEBS SysBio2007

Methodology

General structure of signal flow



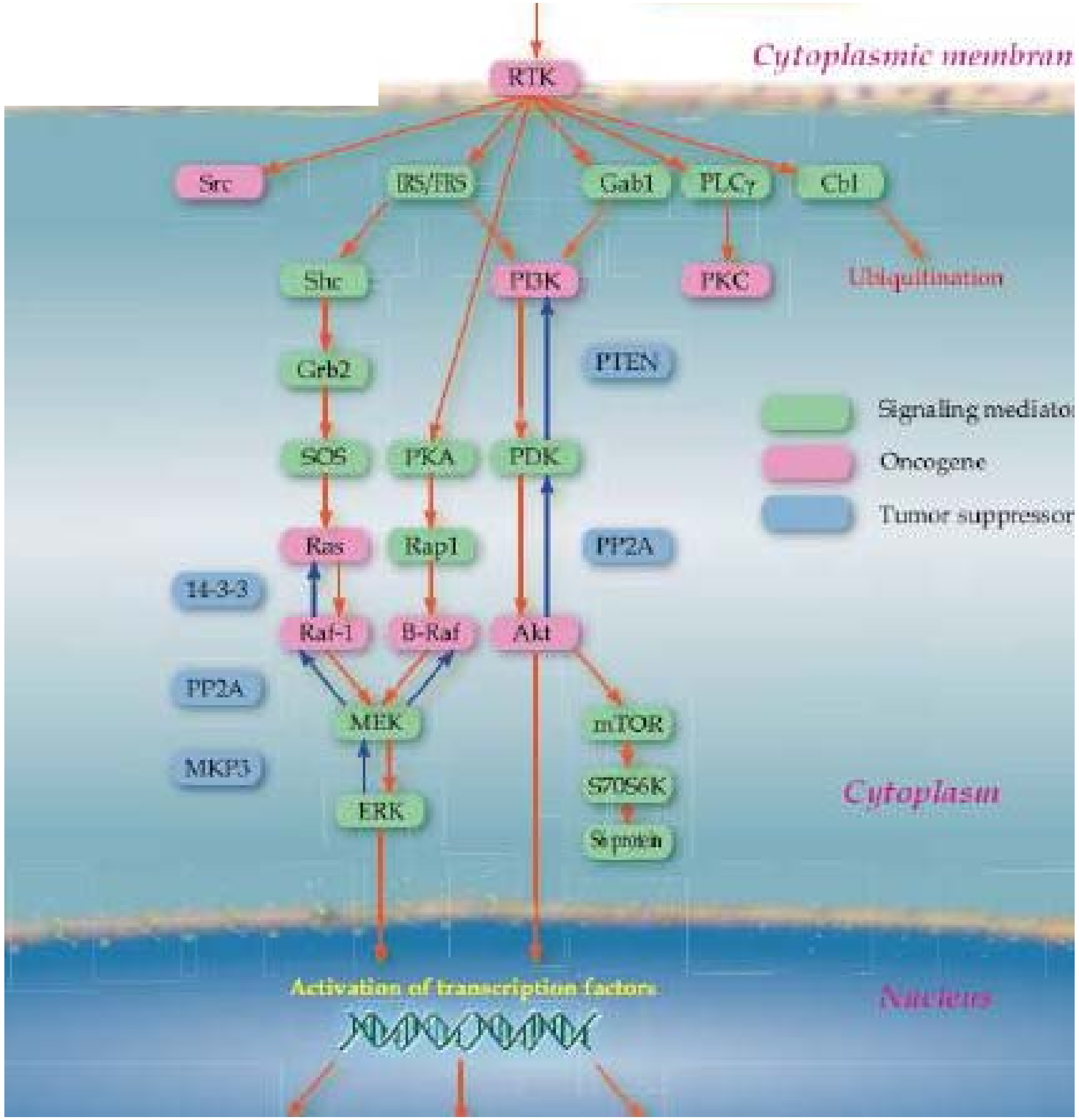
cell fate control
for survival

--> differentiation
--> proliferation
--> apoptosis

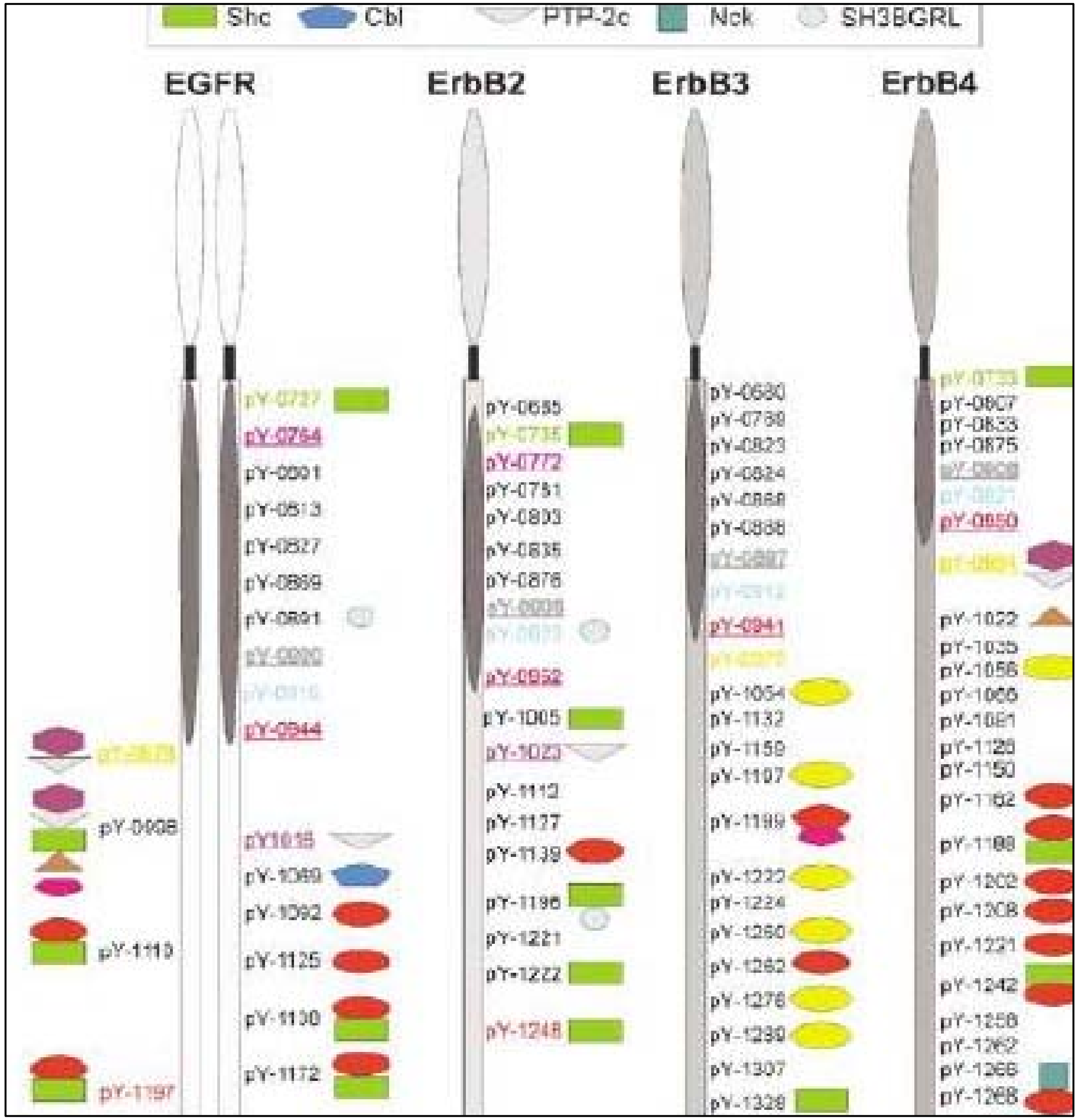
distinct cell fate

What makes
such a decision?

Signal transduction pathways are mostly overlapped regardless of different ligand or cell types!

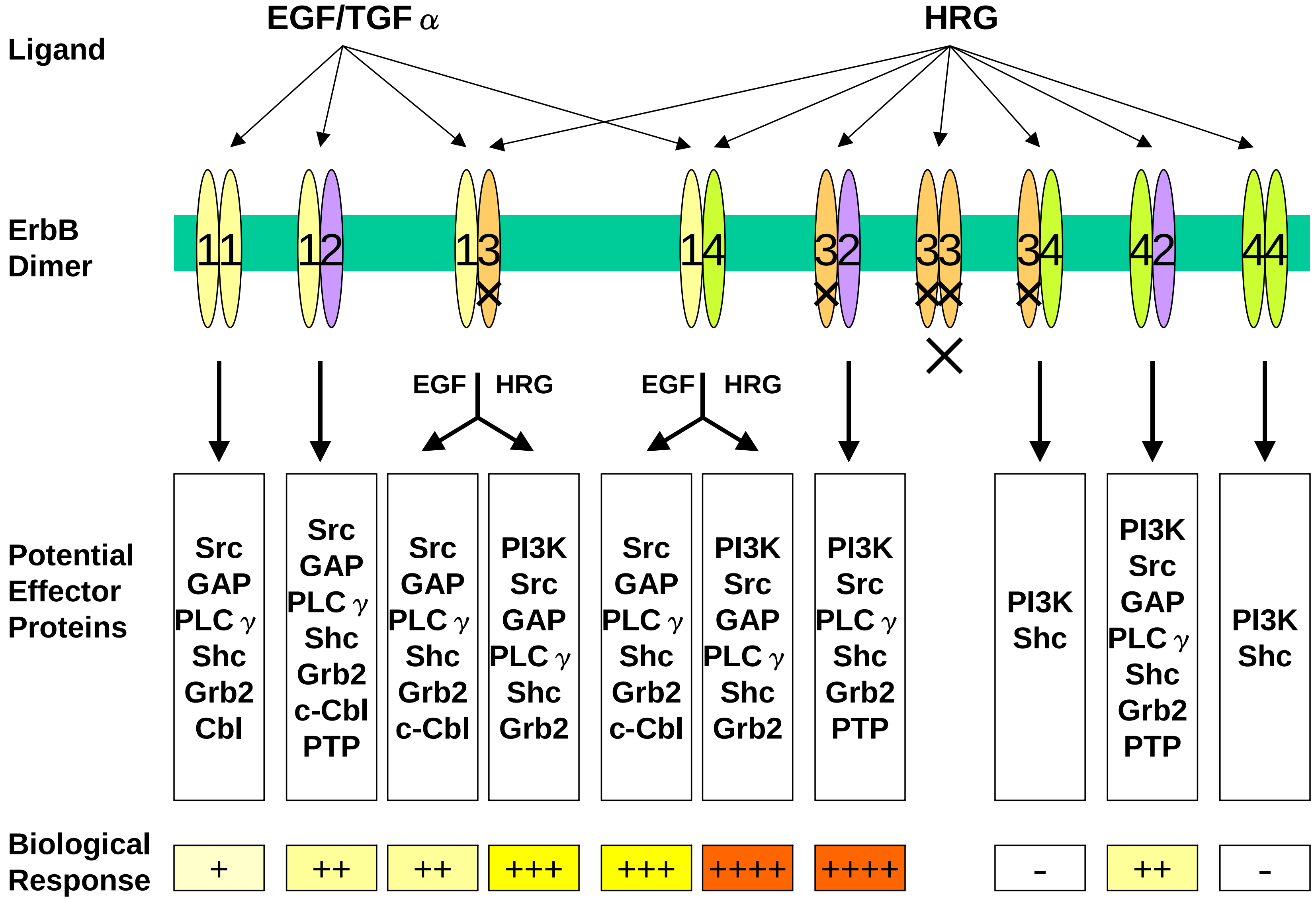


Furthermore, primary binding proteins for ErbBs are almost same !



Schulze, et al. Mol Syst Biol. 2005

However, ErbB signaling network is diverse



Ligand: EGF/TGF α, HRG

ErbB Dimer	11	12	13	13	14	14	32	33	34	42	44
			EGF	HRG	EGF	HRG					
Potential Effector Proteins	Src GAP PLC γ Shc Grb2 Cbl	Src GAP PLC γ Shc Grb2 c-Cbl PTP	Src GAP PLC γ Shc Grb2 c-Cbl	PI3K Src GAP PLC γ Shc Grb2	Src GAP PLC γ Shc Grb2 c-Cbl	PI3K Src GAP PLC γ Shc Grb2	PI3K Src PLC γ Shc Grb2 PTP		PI3K Shc	PI3K Src GAP PLC γ Shc Grb2 PTP	PI3K Shc
Biological Response	+	++	++	+++	+++	++++	++++		-	++	-

Modified from Alroy & Yarden, 1997.

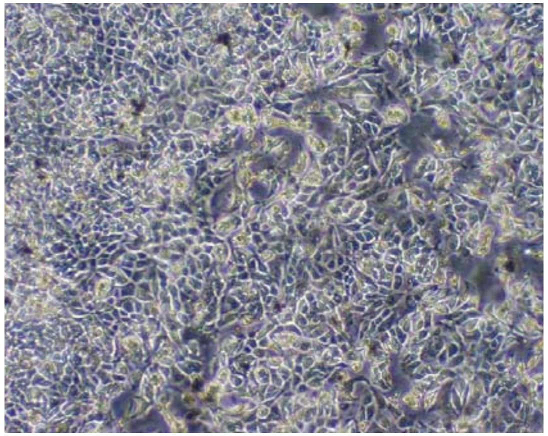
How do cells induce specificity using overlapping signaling cassettes?

- **Cell specificity**
  different ErbB receptor expression (single or co-expression),
  = distinct cellular transformation outcomes
  - prediction of mechanism using kinetic model
  - identification of pathway structure and key regulators

- **Ligand specificity**
  different ligand induces distinct cell fate
  - identification of mechanism
  - quantitative transcriptome analysis
  - relationship with upstream signaling

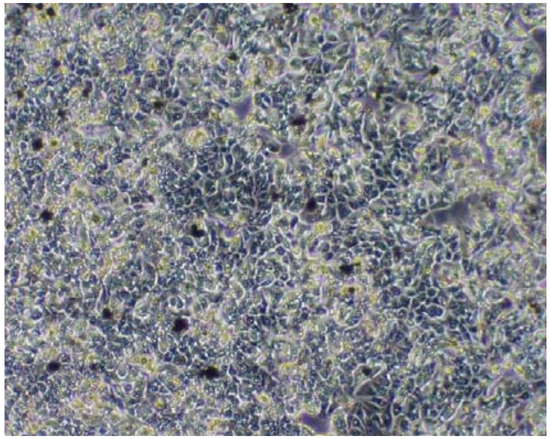
Different responses to the ligands.

Control
(human breast cancer MCF-7 cells)

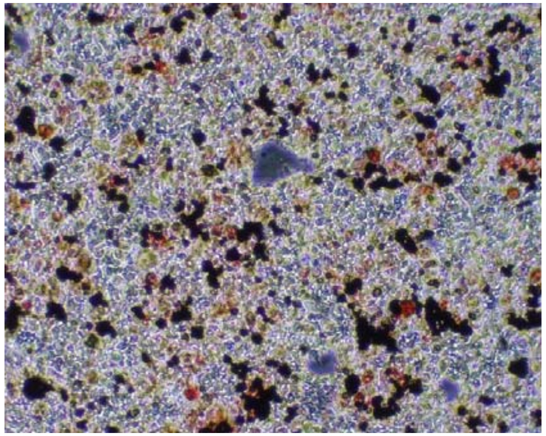


EGF

HRG



proliferation



differentiation (lipid accumulation)

Ligands and ErbB receptors in our study

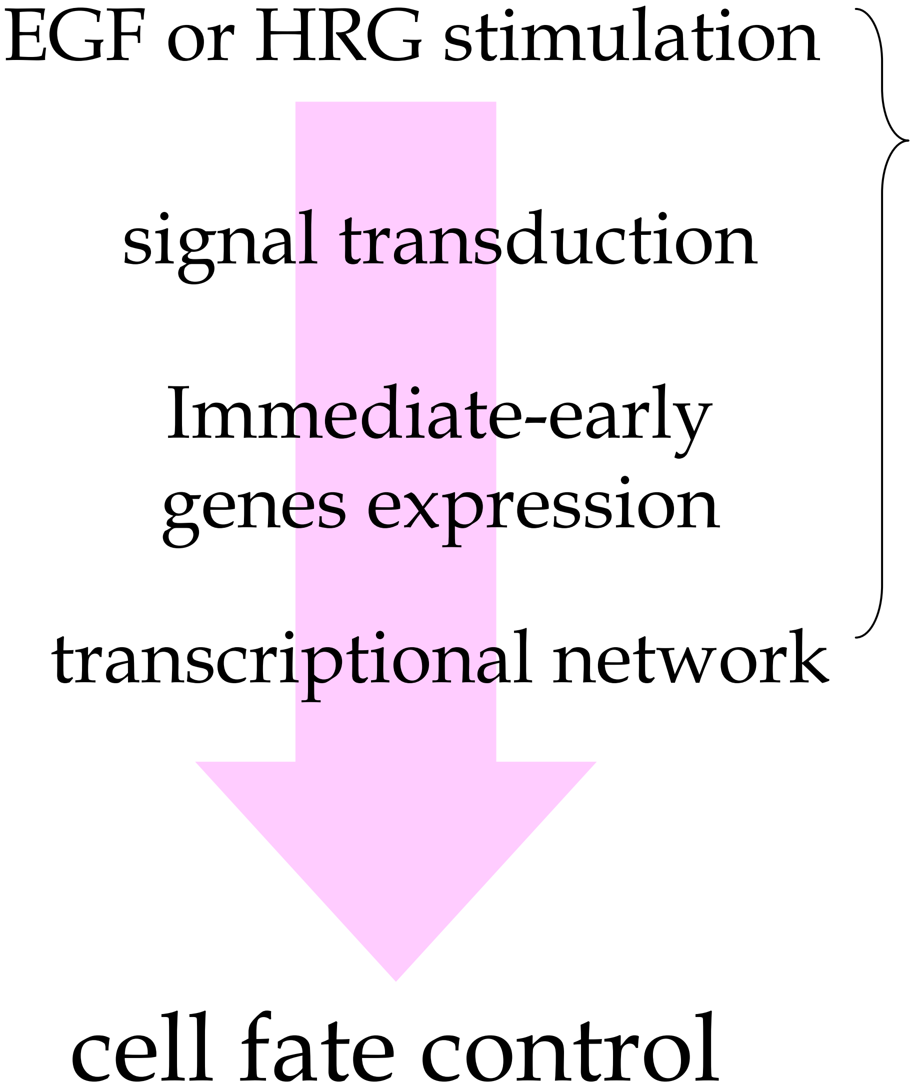
EGF …

- is a high affinity ligand for EGFR/ErbB1 receptor.
- down-regulates EGFR activation through Cbl ubiquitination.

HRG …

- is a high affinity ligand for ErbB3/4 receptor and trans-activate ErbB2.
- ErbB2/ErbB3 heterodimer forms the most potent oncogenic unit (ErbB2 activates MAPK & ErbB3 activates PI3K-Akt).

Flow of cell fate control



EGF or HRG stimulation

signal transduction

Immediate-early genes expression

transcriptional network

cell fate control

Short term protein phosphorylation and gene expression were investigated.

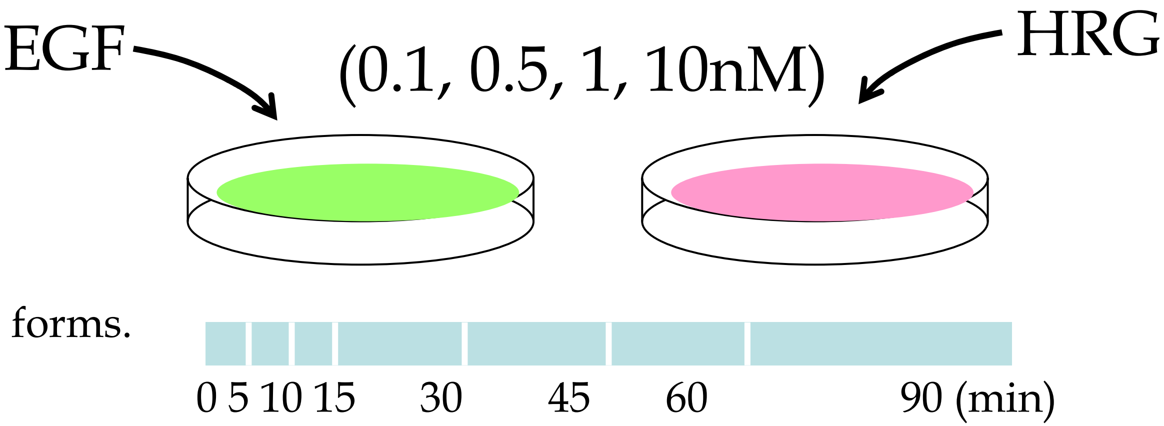
8 time points (0-90min)
4 dosages (0.1-10nM)

Identification of IE genes may be directly affected by signal transduction.

Protein assay

Western blot

Measured species:
EGRF, ErbB2,
and their phosphorylated forms.



Bio-Plex suspension array system
(Bio-Rad Laboratories, Inc)

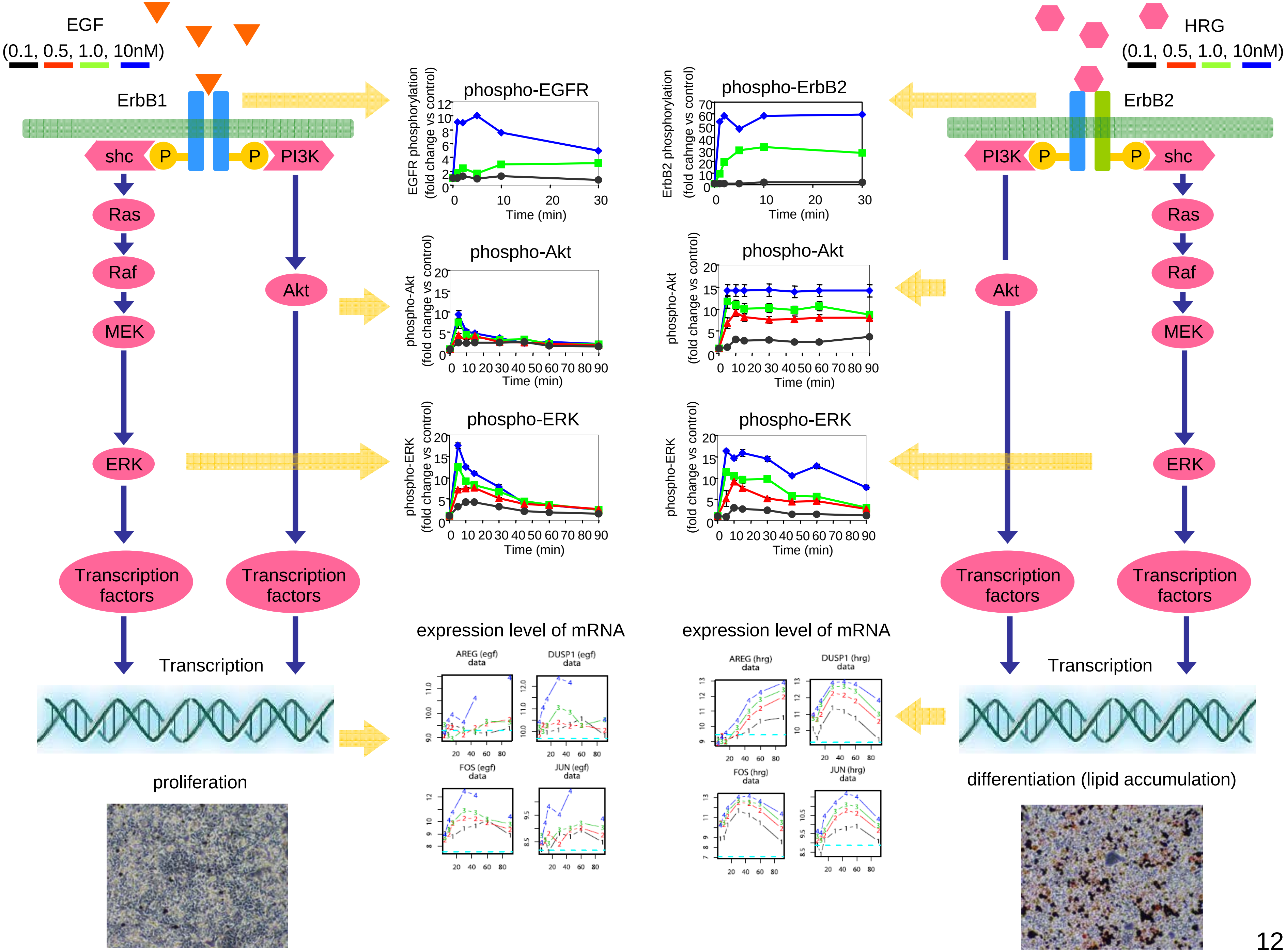
➡ multiplex assay (96 well)

easy to normalize raw data

Measured species:
ERK, Akt, and their phosphorylated forms.

Gene expression

GeneChip (Affymetrix)
Human Genome U133A 2.0 Array
(22,000 probe sets ~ 18,400 transcripts ~ 14,500 genes)



EGF

(0.1, 0.5, 1.0, 10nM)

HRG

(0.1, 0.5, 1.0, 10nM)

ErbB1

ErbB2

shc

P

PI3K

Ras

Raf

MEK

ERK

Akt

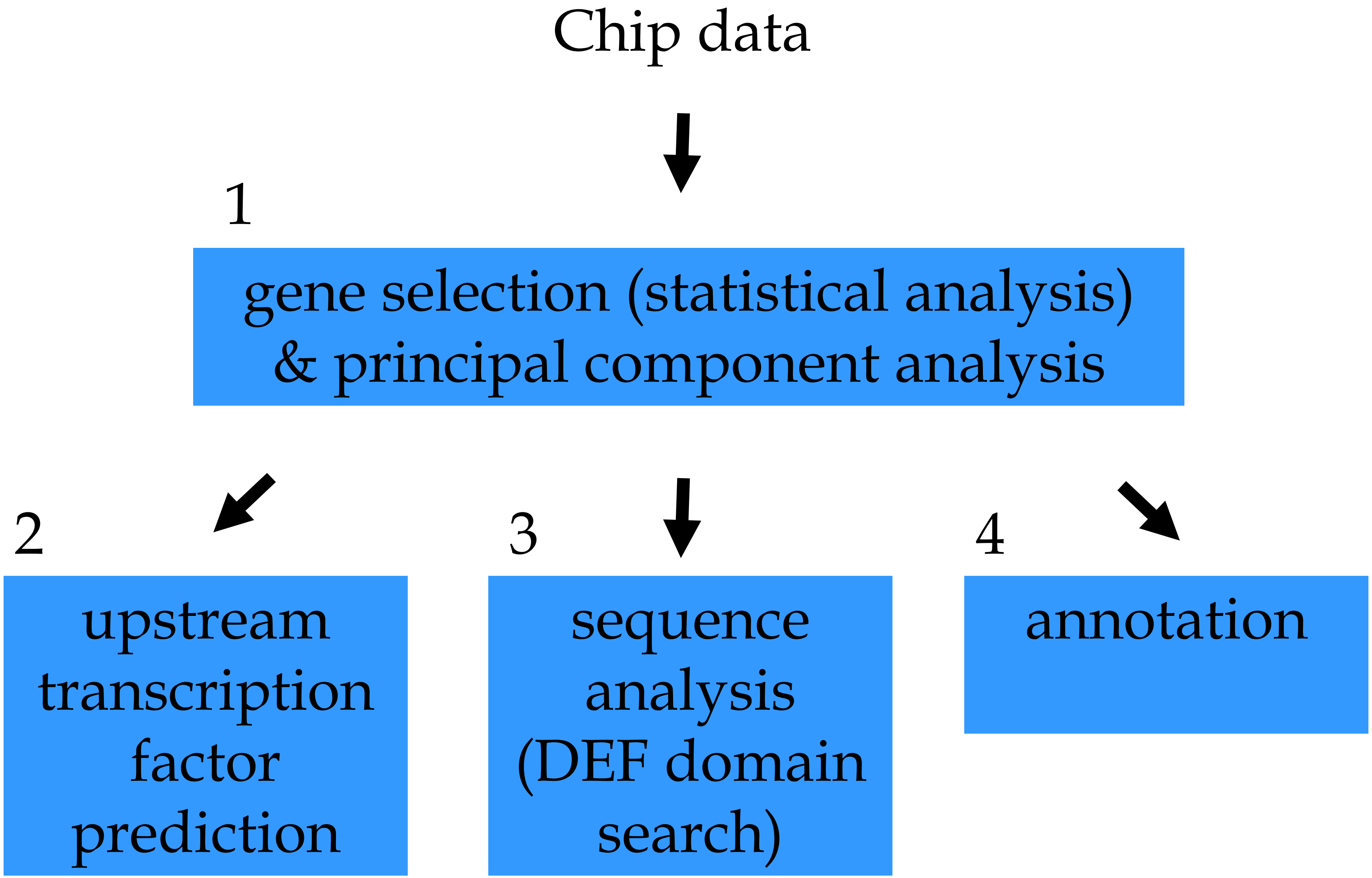
Transcription factors

Transcription

proliferation

differentiation (lipid accumulation)

Flow chart of gene expression analysis in this study



Chip data

1

gene selection (statistical analysis) & principal component analysis

2

upstream transcription factor prediction

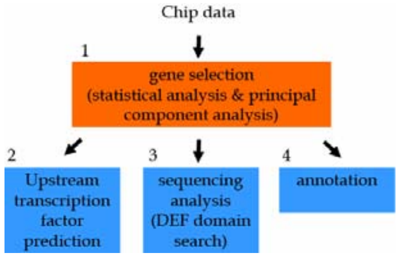
3

sequence analysis (DEF domain search)

4

annotation

1. Gene selection (Statistical analysis)



Chip data

1 gene selection (statistical analysis & principal component analysis)

2 Upstream transcription factor prediction

3 sequencing analysis (DEF domain search)

4 annotation

Wave shaping by Multiplicative decomposition model

- Log-transformed gene expression level at time point *i* and dose level *j* is modeled as following equation.

$$X(i,j) = A(i)B(j) + C(j) + error \quad X(0,0) = C(0) + error$$

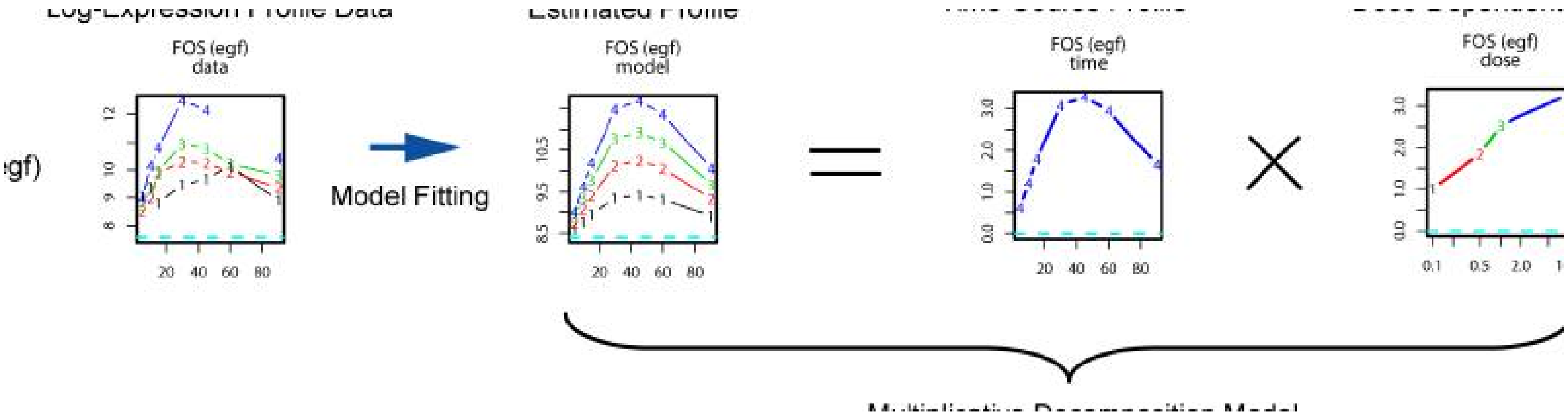
- where,
- $A(i)$ is time-course profile
- $B(j)$ is dose-dependent profile
- $C(j)$ is baseline expression level depending on dose level j.
- Three component $A(i)$, $B(j)$ and $C(j)$ are estimated by minimizing following quantity.

$$RSS = \sum_{i=1}^{I}\sum_{j=1}^{J} w_x(i,j)\left(X(i,j) - A(i)B(j) + C(j)\right)^2 + w_X(0,0)\left(X(0,0) - C(0)\right)^2$$

- with constraint $\left(B(1)+\cdots B(J)\right)/J$
- Numerical optimization was done by *optim* function implemented in R programming language with multiple initial values.

Shimodaira, H, TITEC

Multiplicative decomposition of a gene expression profile into time-course and dose-dependent profiles



Tani Y, et al. *Proc. Genome Inform* P058, 2005.

Gene selection procedure

- Discard the gene if $S < \log_2(1.5)$ where $S = \max_{i=1}^{I} | A(i)B(J) |$
- Discard the gene If $\max_{i=1}^{I} \max_{j=1}^{J} X(i, j) < \log_2(80)$
- Calculate the z-statistic $z = \dfrac{S - \log_2(1.5)}{\sigma_S}$

Then, calculate *p*-value of the normal approximation test

$$p = 1 - \Phi(z)$$

where Φ is the standard normal distribution function.
Select the gene if $p < 0.0001$.

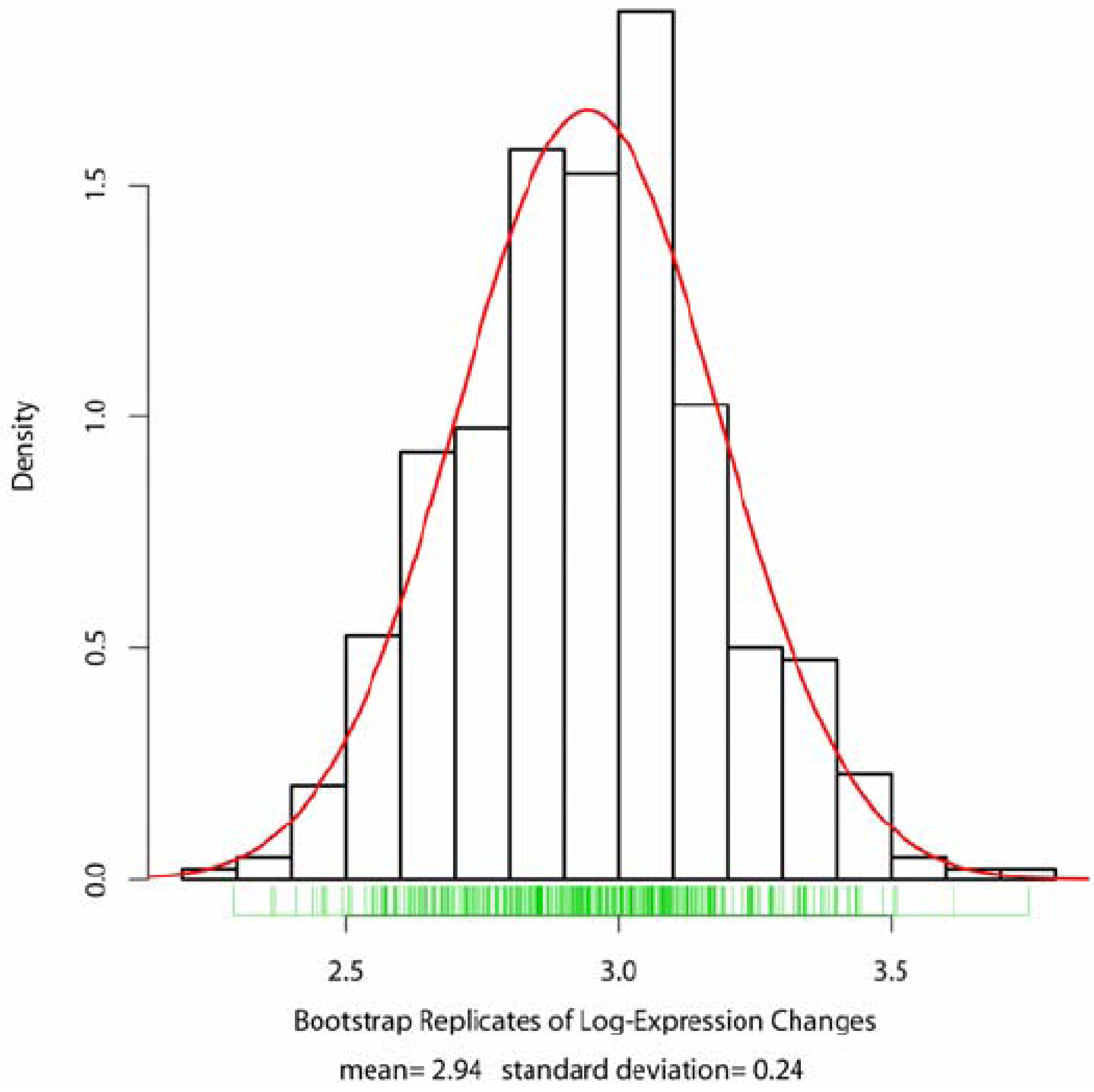
- σ_S is calculated using the resampling residuals method, a version of the bootstrap resampling method for regression analysis.

Tani, Y., Kamimura, T., Nagashima, T., Ide, K., Hatakeyama, M. & Shimodaira,
ltiplicative Decomposition of Time- and Dose-Dependent Gene Expression Changes. *Pr*
h International Conference on Genome Informatics (GIW 2005), Poster Abstract P(
05).

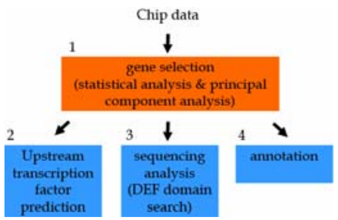
Mandel, J. Analysis of Two-Way Layouts. Chapman & Hall/CRC (1994).

Efron, B. and Tibshirani, R. An Introduction to the Bootstrap. Chapman & Hall/CRC (199

400 bootstrap replicates S^* for the FOS gene



Principal component analysis



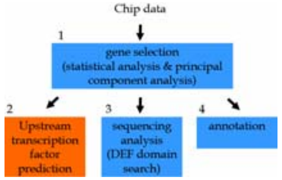
For given multiplicative decomposition model:

$$X(i,j)=A(i)B(j)+C(j)+error \quad X(0,0)=C(0)+error$$

<PCA for **time-course** profiles>
i) A(i) is a 7 dimensional vector. Put the vectors in the 7 dimensional space, and Find principal component.

<PCA for **dose-dependent** profiles>
i) B(i) is a 4 dimensional vector. Put the vectors in the 4 dimensional space, and Find principal component.

2. TF analysis (Find specific upstream TFs)



In-house tool

selected genes
gene1
gene2
…
geneN

➡

predicted TFs (provided by UCSC genome browser)
predicted TFs for gene1
predicted TFs for gene2
…
predicted TFs for geneN

➡

TF_1
TF_2
…
TF_M

For TF_x (x=1,2,…,M), Fisher's Exact Test was performed.

	selected genes (N)	others (22277-N)
bind	number of genes	number of genes
not bind	number of genes	number of genes

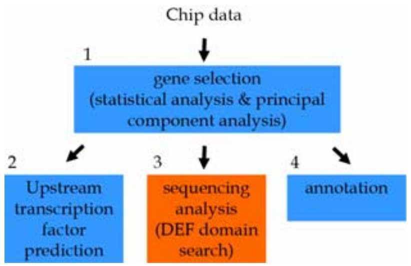
We can use "fisher.test" on R Language to get p-value.

Public tool

⬍ combinational analysis

Utilize web service like "MOTIF Search". (http://motif.genome.jp/)

3. Sequence analysis for DEF domain search (FXFP & FXYP)

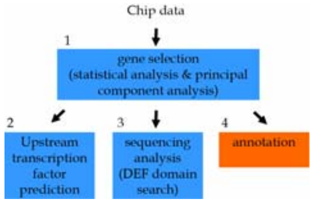


For selected genes,
i) retrieve amino-acid sequence information

ii) select genes that have a target motif
(In this study, this selection was just done with perl script.)

No	Symbol	Probe Set ID	Gene ID	Protein acc.	Seq. len.	Motif			Alignment		
						Start	End	Seq.	Start	End	Seq. + gap
1	CSNK1D	207945_s_at	1453	Q6ZNS0	128	99	102	FSYP	79	122	-ATGGFLKMPPPKYVLWSYFFFSYPQEHVSSVCFHSTGSATIHMA--
2	DST	204455_at	667	NP_001714	2649	2319	2322	**FLFP**	2299	2342	-QALYYSELLRMCVFDVESQCFLFPFGERNISNLNVKKTHRISVV--
	DST	204455_at	667	NP_065121	3062	2555	2558	**FPFP**	2535	2578	-HLKLLPGKNTRDSFKLINSQFPFPQITNNEELNQKGSLKKATVT--
	DST	204455_at	667	Q03001	3214	2884	2887	**FLFP**	2864	2907	-QALYYSELLRMCVFDVESQCFLFPFGERNISNLNVKKTHRISVV--
3	**DUSP1**	201041_s_at	1843	P28562	367	339	342	**FNFP**	319	362	AEAGSPAMAVLDRGTSTTT-VFNFPVSIPVHSTNSALSYLQSPIT--
4	DUSP4	204015_s_at	1846	NP_476499	303	275	278	**FSFP**	255	298	-PSGPLRERGKTPATPTSQFVFSFPVSVGVHSAPSSLPYLHSPIT--
	DUSP4	204015_s_at	1846	NP_001385	394	366	369	**FSFP**	346	389	-PSGPLRERGKTPATPTSQFVFSFPVSVGVHSAPSSLPYLHSPIT--
	DUSP4	204015_s_at	1846	Q13649	411	383	386	**FSFP**	363	406	-PSGPLRERGKTPATPTSQFVFSFPVSVGVHSAPSSLPYLHSPIT--
5	**F2RL1**	213506_at	2150	P55085	397	251	254	**FLFP**	231	274	-PEQLLVGDMFNYFLSLAIGVFLFPAFLTASAYVLMIRMLRSSAM--
6	F3	204363_at	2152	Q86WH3	145	15	18	FSYP	1	38	-------EIVKDVKQTYLARVFSYPAGNVESTGSAGEPLYENSPE--
	F3	204363_at	2152	Q86SE7	220	90	93	FSYP	70	113	-ECDLTDEIVKDVKQTYLARVFSYPAGNVESTGSAGEPLYENSPE--
	F3	204363_at	2152	P13726	295	108	111	FSYP	88	131	-ECDLTDEIVKDVKQTYLARVFSYPAGNVESTGSAGEPLYENSPE--
7	**FOS**	209189_at	2353	P01100	380	272	275	**FLFP**	252	295	-SVEPVKSISSMELKTEPFDDFLFPASSRPSGSETARSVPDMDLS--
	FOS	209189_at	2353	P01100	380	343	346	FTYP	323	366	-LCTPVVTCTPSCTAYTSSFVFTYPEADSFPSCAAAHRKGSSSNE--
8	**FOSL1**	204420_at	8061	NP_005429	271	235	238	FTYP	215	258	-LHTPTLMTTPSLTPFTPSLVFTYPSTP--EPCASAHRKSSSSSGDP
9	GATA2	209710_at	2624	NP_116027	480	171	174	**FGFP**	151	194	-GSGSSVASLTPTAAHSGSHLFGFPPTPPKEVSPDPSTTGAASPA--
10	**JMJD3**	213146_at	23135	O15054	1682	671	674	**FDFP**	651	694	-KAPQPVPPGVGELPARGPRLFDFPPTPLEDQFEEPAEFKILPDG--
11	KLF2	219371_s_at	10365	Q8IUN4	224	73	76	FYYP	53	96	-LDGLGAEAAPEPPPPPPPPAFYYPEPGAPPPYSAPAGGLVSELL--
12	**MYC**	202431_s_at	4609	P01106	439	195	198	FPYP	175	218	-SLYLQDLSAAASECIDPSVVFPYPLNDSSSPKSCASQDSSAFSP--
	MYC	202431_s_at	4609	NP_002458	454	210	213	FPYP	190	233	-SLYLQDLSAAASECIDPSVVFPYPLNDSSSPKSCASQDSSAFSP--
13	NDEL1	208093_s_at	81565	Q6ZW10	148	8	11	**FTFP**	1	31	-------MCAGVH-------VFTFPPAETSDAVKPSHIKQYTLTG--
14	OSR2	213568_at	116039	Q8N2R0	312	83	86	**FPFP**	63	106	-EITRSTITEMAAAQGLVDARFPFPALPFTTHLFHPKQGAIAHVL--
15	PGAP1	220576_at	80055	Q75T13	922	43	46	FEYP	23	66	-LGLWDVFFGFEENKCSMSYMFEYPEYQKIELPKKLAKRYPAYEL--
	PGAP1	220576_at	80055	Q6AW92	748	196	199	**FTFP**	176	219	-VLVKVSKWTYVAYNESEKIYFTFPLEN--HRKIYTHVYCQSAMLDT
	PGAP1	220576_at	80055	Q75T13	922	370	373	**FTFP**	350	393	-VLVKVSKWTYVAYNESEKIYFTFPLEN--HRKIYTHVYCQSTMLDT
16	PPP1R15A	37028_at	23645	Q6IA96	674	177	180	FSYP	157	200	-NPGEEKAEEEGVAEEEGVNKFSYPPSHRECCPAVEEEDDEEAVK--
17	PPP1R3D	204554_at	5509	Q86X09	299	238	241	**FGFP**	218	261	-EAVARWRGPAGPEGTEDVFTFGFPVPPFLLELGSRVHFAVRYQV--
18	RARA	203749_s_at	5914	P10276	462	26	29	**FFFP**	6	49	-SSCPTPGGGHLNGYPVPPYAFFFPPMLGGLSPPGALTTLQHQLP--
19	TSC22D2	210953_at	9819	O75157	780	357	360	FAYP	337	380	-SLPPQPGPAVGAPAAQQPQQFAYPQPQIPPGHLLPVQPSGQSEY--
20	VIL2	208621_s_at	7430	Q8IXG9	167	59	62	FNYP	39	82	-TVREILTSPSCWQYAVLLNRFNYPFELEKDLHLKGYHTLSQGSL--
21	ZNF287	220055_at	57336	NP_065704	754	48	51	FPYP	28	71	EILTSRFLRDTETCRQNFR-NFPYPDLAGPRKALSQLRELCLKWL--
22	ZNF307	213625_at	387032	NP_061983	545	59	62	FRYP	39	82	VRAPCSPARGPERSRQRFR-GFRYPEAAGPREALSRLRELCGQWL--

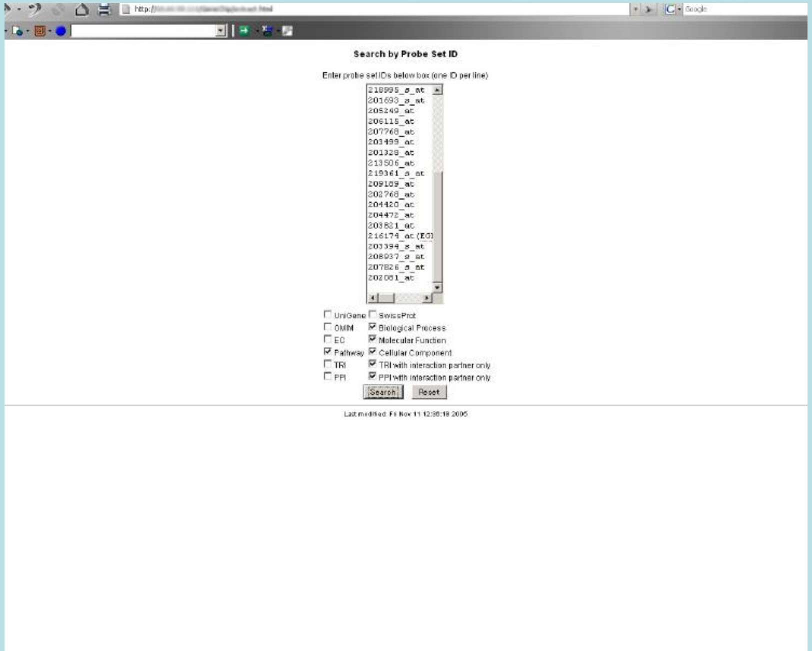
4. Annotation



For selected genes, annotation information was added with …

In-house tool

We developed annotation support tool.



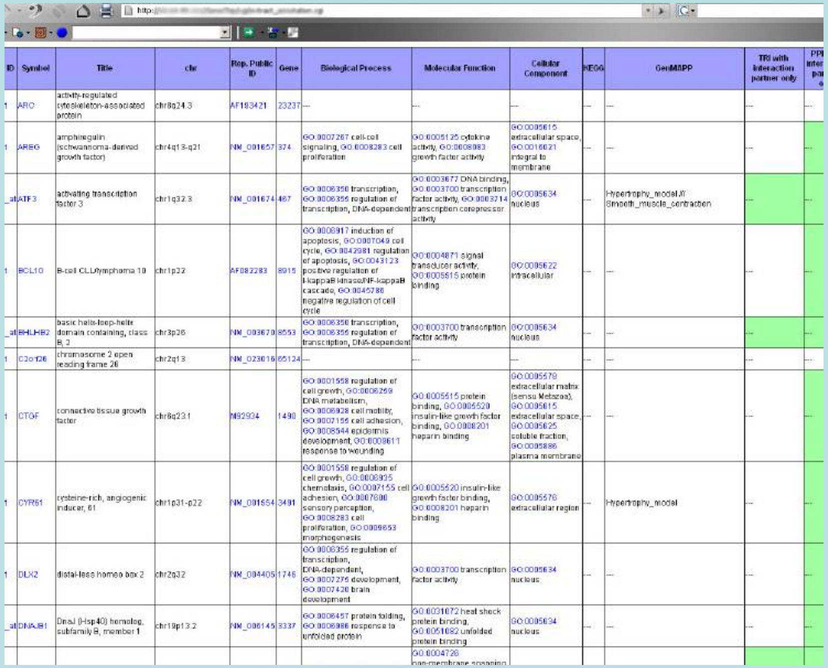
Search by Probe Set ID

Enter probe set IDs below box (one ID per line)

```
218995_s_at
201693_s_at
205249_at
206115_at
207768_at
203499_at
201328_at
213506_at
219361_s_at
209189_at
202768_at
204420_at
204472_at
203821_at
216174_at (EG)
203394_s_at
208937_s_at
207826_s_at
202081_at
```

☐ UniGene ☐ SwissProt
☐ OMIM ☑ Biological Process
☐ EC ☑ Molecular Function
☑ Pathway ☑ Cellular Component
☐ TRI ☑ TRI with interaction partner only
☐ PPI ☑ PPI with interaction partner only

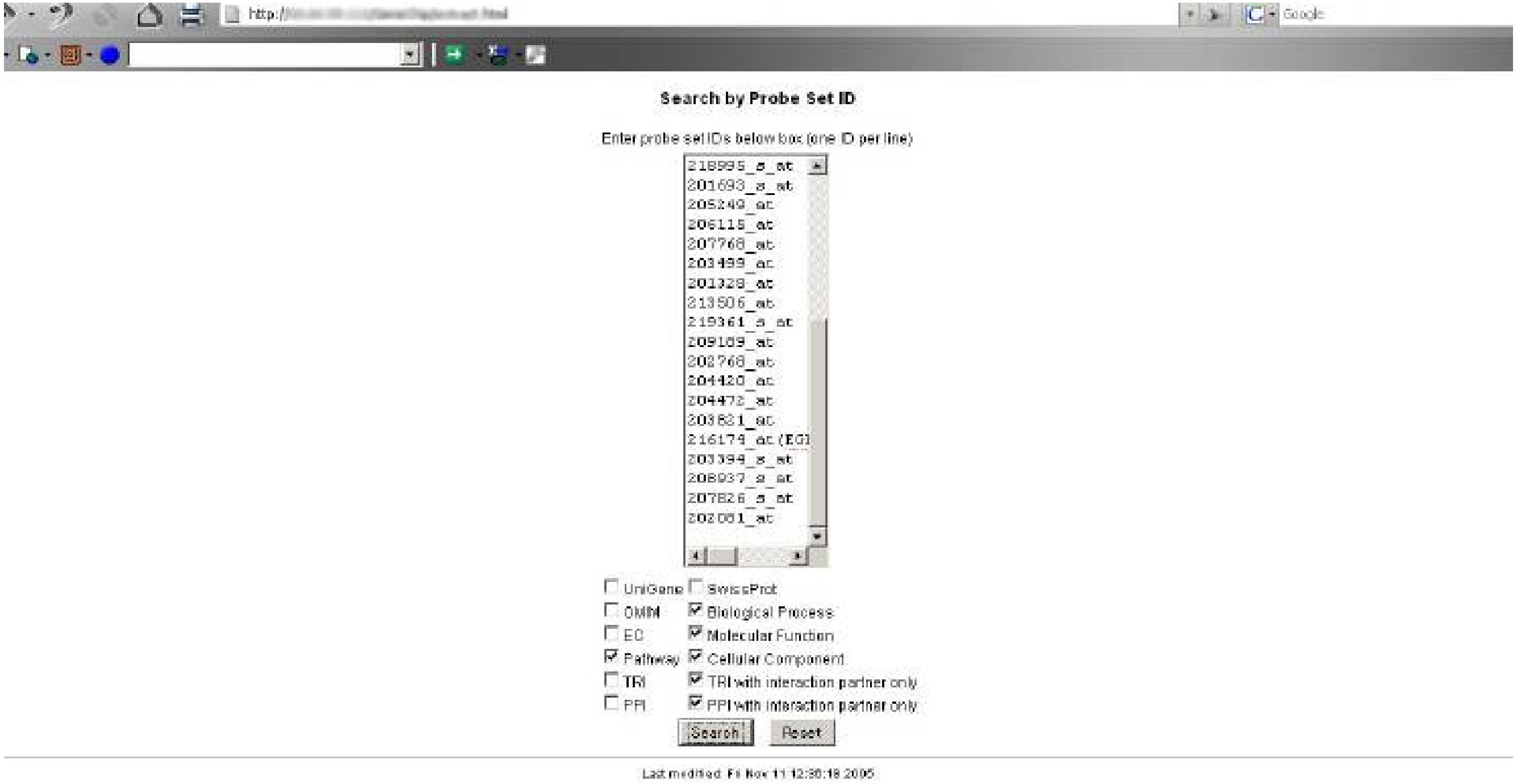
Last modified: Fri Nov 11 12:38:18 2005



ID	Symbol	Title	chr	Rep. Public ID	Gene	Biological Process	Molecular Function	Cellular Component	KEGG	GenMAPP	TRI with interaction partner only	PPI
1	ARC	activity-regulated cytoskeleton-associated protein	chr8q24.3	AF193421	23237	---	---	---	---	---	---	---
1	AREG	amphiregulin (schwannoma-derived growth factor)	chr4q13-q21	NM_001657	374	GO:0007267 cell-cell signaling, GO:0008283 cell proliferation	GO:0005125 cytokine activity, GO:0008083 growth factor activity	GO:0005615 extracellular space, GO:0016021 integral to membrane	---	---	---	---
_at	ATF3	activating transcription factor 3	chr1q32.3	NM_001674	467	GO:0006350 transcription, GO:0006355 regulation of transcription, DNA-dependent	GO:0003677 DNA binding, GO:0003700 transcription factor activity, GO:0003714 transcription corepressor activity	GO:0005634 nucleus	---	Hypertrophy_model.M Smooth_muscle_contraction	---	---
1	BCL10	B-cell CLL/lymphoma 10	chr1p22	AF082283	8915	GO:0006917 induction of apoptosis, GO:0007049 cell cycle, GO:0042981 regulation of apoptosis, GO:0043123 positive regulation of I-kappaB kinase/NF-kappaB cascade, GO:0045786 negative regulation of cell cycle	GO:0004871 signal transducer activity, GO:0005515 protein binding	GO:0005622 intracellular	---	---	---	---
_at	BHLHB2	basic helix-loop-helix domain containing, class B, 2	chr3p26	NM_003670	8553	GO:0006350 transcription, GO:0006355 regulation of transcription, DNA-dependent	GO:0003700 transcription factor activity	GO:0005634 nucleus	---	---	---	---
1	C2orf26	chromosome 2 open reading frame 26	chr2q13	NM_023016	65124	---	---	---	---	---	---	---
1	CTGF	connective tissue growth factor	chr8q23.1	M92934	1490	GO:0001558 regulation of cell growth, GO:0006259 DNA metabolism, GO:0006928 cell motility, GO:0007155 cell adhesion, GO:0008544 epidermis development, GO:0009611 response to wounding	GO:0005515 protein binding, GO:0005520 insulin-like growth factor binding, GO:0008201 heparin binding	GO:0005578 extracellular matrix (sensu Metazoa), GO:0005615 extracellular space, GO:0005625 soluble fraction, GO:0005886 plasma membrane	---	---	---	---
1	CYR61	cysteine-rich, angiogenic inducer, 61	chr1p31-p22	NM_001554	3491	GO:0001558 regulation of cell growth, GO:0006935 chemotaxis, GO:0007155 cell adhesion, GO:0007600 sensory perception, GO:0008283 cell proliferation, GO:0009653 morphogenesis	GO:0005520 insulin-like growth factor binding, GO:0008201 heparin binding	GO:0005576 extracellular region	---	Hypertrophy_model	---	---
1	DLX2	distal-less homeo box 2	chr2q32	NM_004405	1746	GO:0006355 regulation of transcription, DNA-dependent, GO:0007275 development, GO:0007420 brain development	GO:0003700 transcription factor activity	GO:0005634 nucleus	---	---	---	---
_at	DNAJB1	DnaJ (Hsp40) homolog, subfamily B, member 1	chr19p13.2	NM_006145	3337	GO:0006457 protein folding, GO:0006986 response to unfolded protein	GO:0031072 heat shock protein binding, GO:0051082 unfolded protein binding	GO:0005634 nucleus	---	---	---	---
							GO:0004726					

Public tool

literature study with PubMed



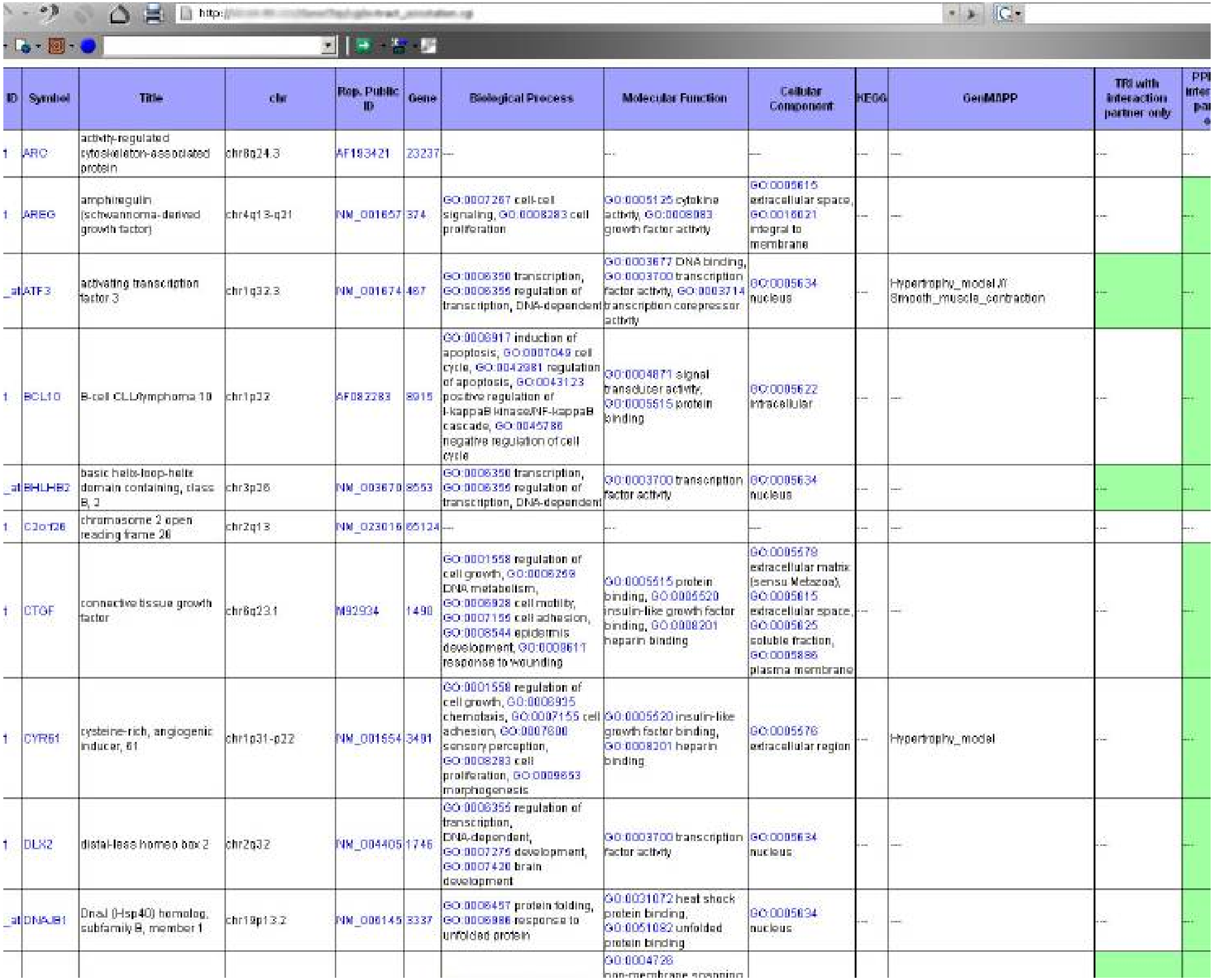
Search by Probe Set ID

Enter probe set IDs below box (one ID per line)

```
218995_s_at
201693_s_at
205249_at
206115_at
207768_at
203499_at
201328_at
213506_at
219361_s_at
209189_at
202768_at
204420_at
204472_at
203821_at
216174_at (EGI
203394_s_at
208937_s_at
207826_s_at
202081_at
```

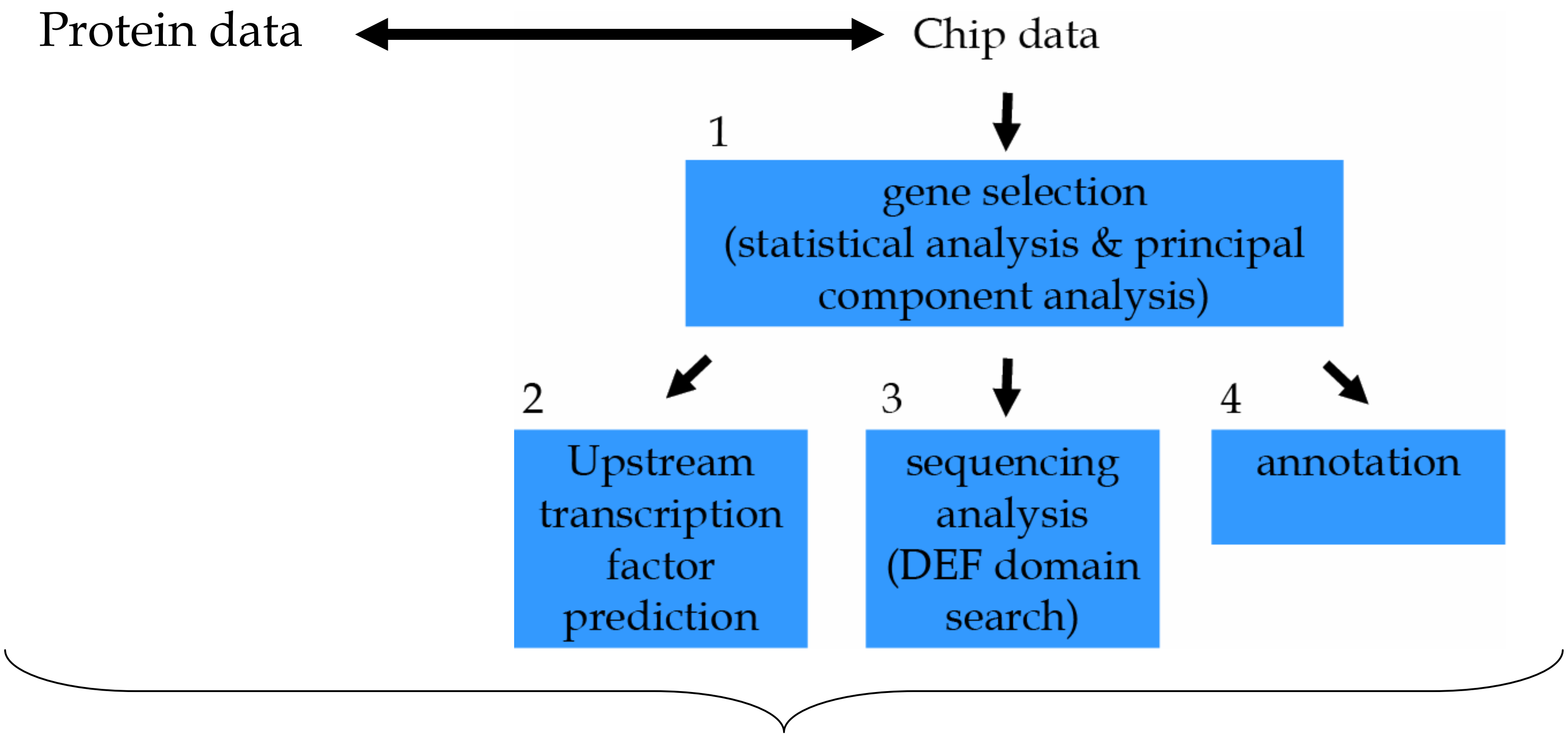
☐ UniGene ☐ SwissProt
☐ OMIM ☑ Biological Process
☐ EC ☑ Molecular Function
☑ Pathway ☑ Cellular Component
☐ TRI ☑ TRI with interaction partner only
☐ PPI ☑ PPI with interaction partner only

Last modified: Fri Nov 11 12:38:18 2005



ID	Symbol	Title	chr	Rep. Public ID	Gene	Biological Process	Molecular Function	Cellular Component	KEGG	GenMAPP	TRI with interaction partner only	PPI inter pa
1	ARC	activity-regulated cytoskeleton-associated protein	chr8q24.3	AF193421	23237	---	---	---	---	---	---	---
1	AREG	amphiregulin (schwannoma-derived growth factor)	chr4q13-q21	NM_001657	374	GO:0007267 cell-cell signaling, GO:0008283 cell proliferation	GO:0005125 cytokine activity, GO:0008083 growth factor activity	GO:0005615 extracellular space, GO:0016021 integral to membrane	---	---	---	---
_at	ATF3	activating transcription factor 3	chr1q32.3	NM_001674	467	GO:0006350 transcription, GO:0006355 regulation of transcription, DNA-dependent	GO:0003677 DNA binding, GO:0003700 transcription factor activity, GO:0003714 transcription corepressor activity	GO:0005634 nucleus	---	Hypertrophy_model /// Smooth_muscle_contraction	---	---
1	BCL10	B-cell CLL/lymphoma 10	chr1p22	AF082283	8915	GO:0006917 induction of apoptosis, GO:0007049 cell cycle, GO:0042981 regulation of apoptosis, GO:0043123 positive regulation of I-kappaB kinase/NF-kappaB cascade, GO:0045786 negative regulation of cell cycle	GO:0004871 signal transducer activity, GO:0005515 protein binding	GO:0005622 intracellular	---	---	---	---
_at	BHLHB2	basic helix-loop-helix domain containing, class B, 2	chr3p26	NM_003670	8553	GO:0006350 transcription, GO:0006355 regulation of transcription, DNA-dependent	GO:0003700 transcription factor activity	GO:0005634 nucleus	---	---	---	---
1	C2orf26	chromosome 2 open reading frame 26	chr2q13	NM_023016	65124	---	---	---	---	---	---	---
1	CTGF	connective tissue growth factor	chr6q23.1	M92934	1490	GO:0001558 regulation of cell growth, GO:0006259 DNA metabolism, GO:0006928 cell motility, GO:0007155 cell adhesion, GO:0008544 epidermis development, GO:0009611 response to wounding	GO:0005515 protein binding, GO:0005520 insulin-like growth factor binding, GO:0008201 heparin binding	GO:0005578 extracellular matrix (sensu Metazoa), GO:0005615 extracellular space, GO:0005625 soluble fraction, GO:0005886 plasma membrane	---	---	---	---
1	CYR61	cysteine-rich, angiogenic inducer, 61	chr1p31-p22	NM_001554	3491	GO:0001558 regulation of cell growth, GO:0006935 chemotaxis, GO:0007155 cell adhesion, GO:0007600 sensory perception, GO:0008283 cell proliferation, GO:0009653 morphogenesis	GO:0005520 insulin-like growth factor binding, GO:0008201 heparin binding	GO:0005576 extracellular region	---	Hypertrophy_model	---	---
1	DLX2	distal-less homeo box 2	chr2q32	NM_004405	1746	GO:0006355 regulation of transcription, DNA-dependent, GO:0007275 development, GO:0007420 brain development	GO:0003700 transcription factor activity	GO:0005634 nucleus	---	---	---	---
_at	DNAJB1	DnaJ (Hsp40) homolog, subfamily B, member 1	chr19p13.2	NM_006145	3337	GO:0006457 protein folding, GO:0006986 response to unfolded protein	GO:0031072 heat shock protein binding, GO:0051082 unfolded protein binding	GO:0005634 nucleus	---	---	---	---
							GO:0004726 non-membrane spanning					

Flow chart of analysis in this study



Protein data ↔ Chip data

1 gene selection (statistical analysis & principal component analysis)

2 Upstream transcription factor prediction

3 sequencing analysis (DEF domain search)

4 annotation

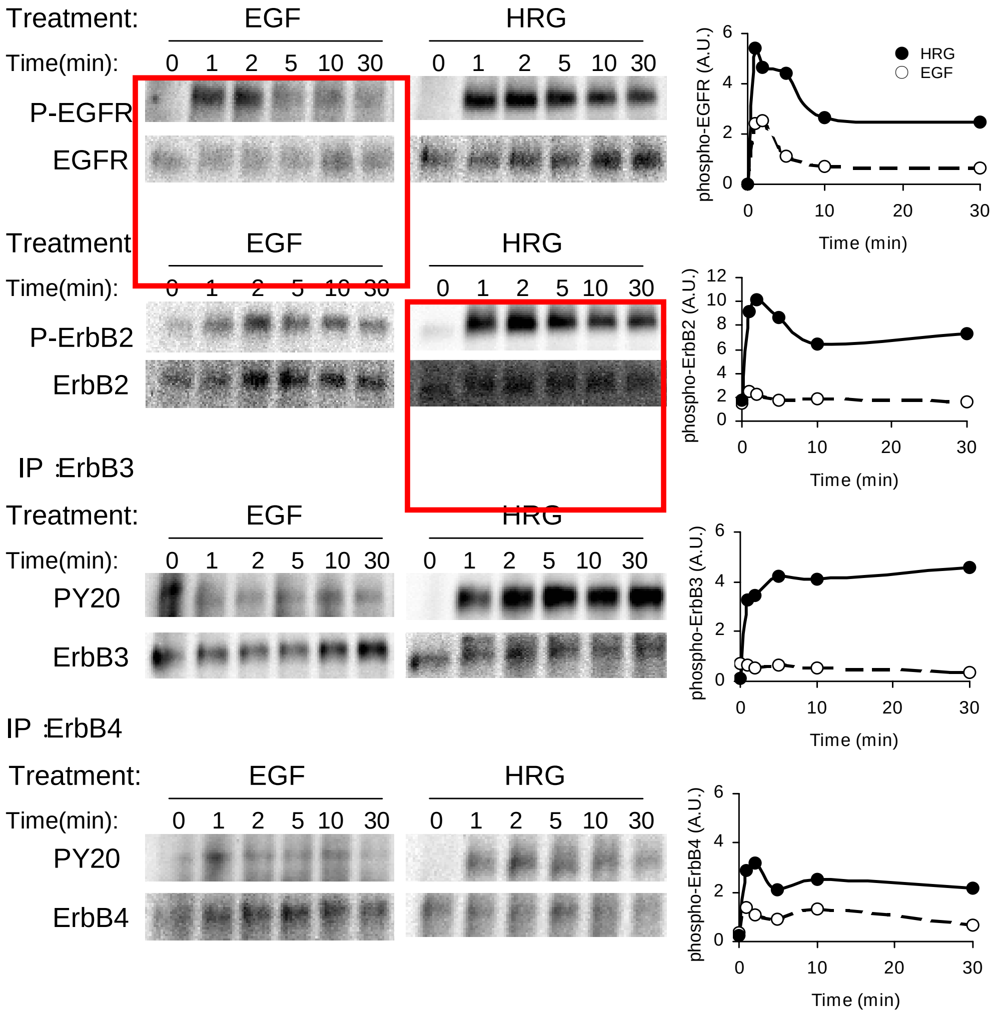
Integrative analysis : hypothesis for cell fate control

Thank you for your attention

to be continued later

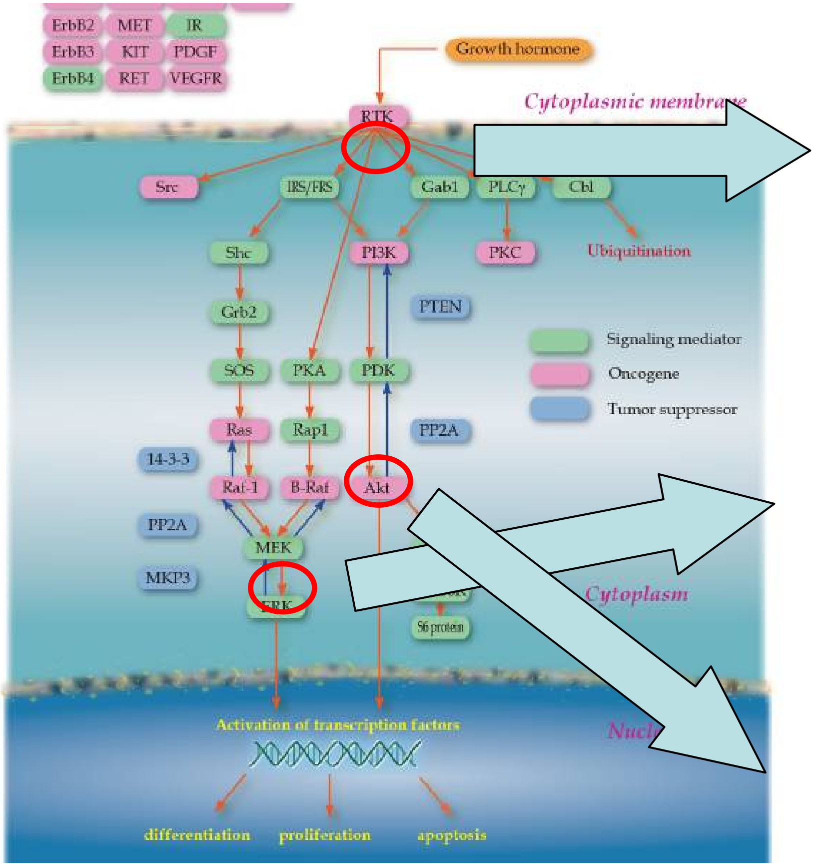
Results

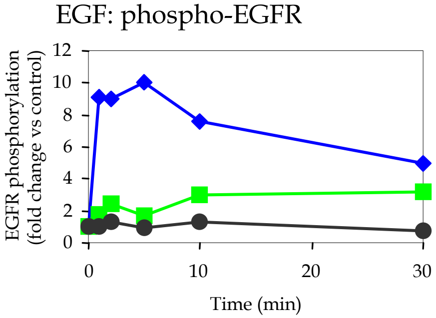
ErbB receptor phosphorylation

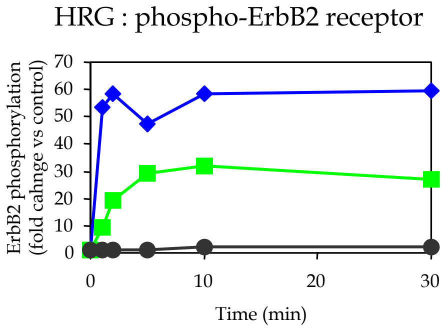


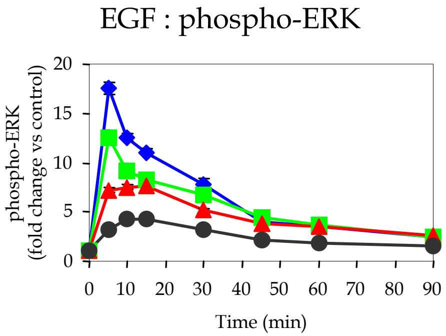
EGF and HRG preferentially activate EGFR and ErbB2, respectively.

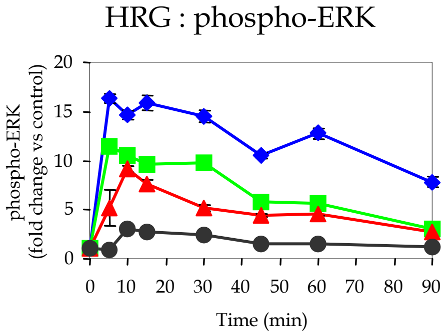
Dose-dependent activation of kinases in MCF-7 cells

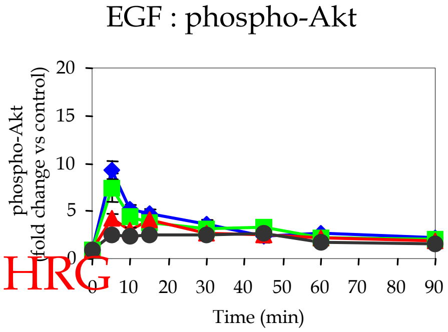


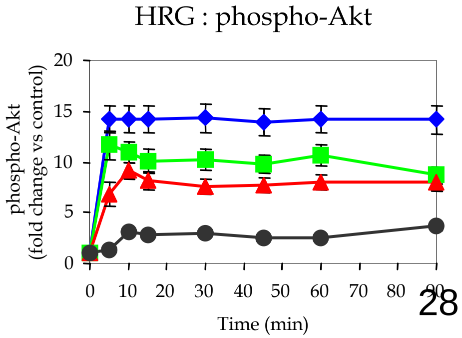

EGF: phospho-EGFR


HRG : phospho-ErbB2 receptor


EGF : phospho-ERK


HRG : phospho-ERK

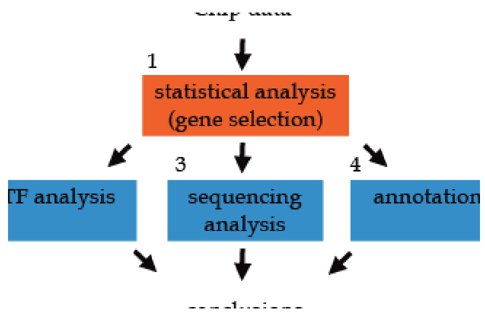

EGF : phospho-Akt

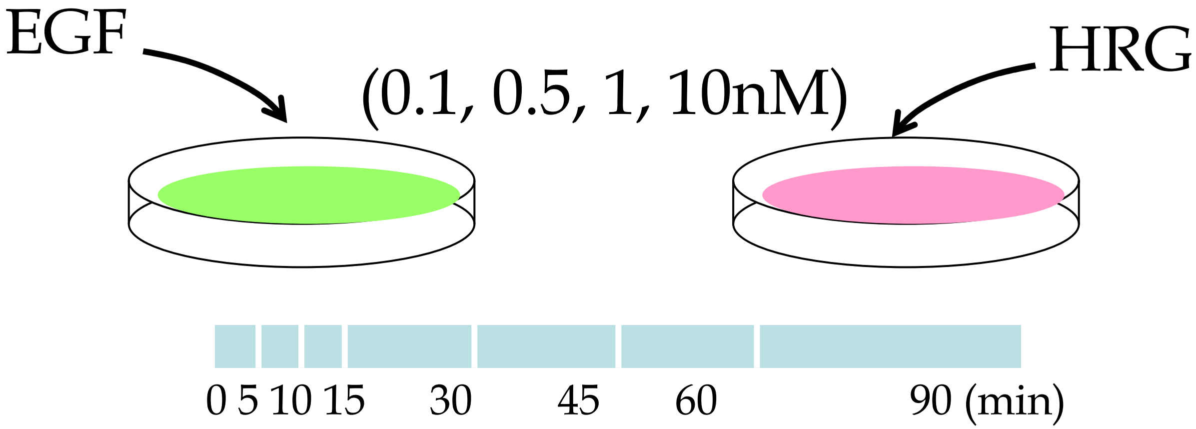

HRG : phospho-Akt

EGF causes transient ERK, Akt activation, while HRG causes prolonged activation.

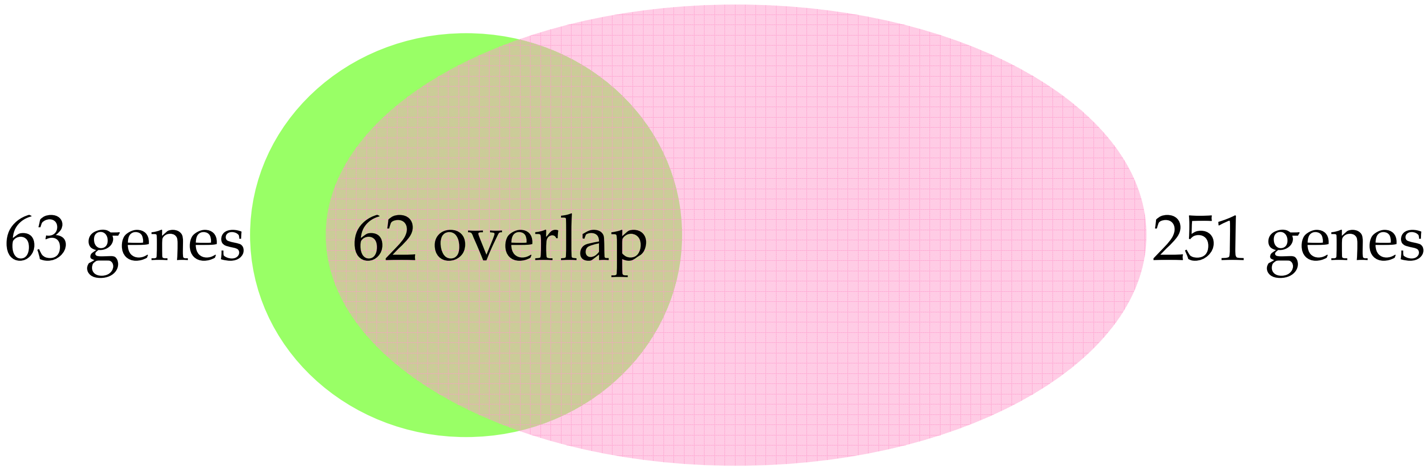
* pERK level, EGF=HRG, pAkt level, EGF<HRG

EGF, HRG-triggered early transcription



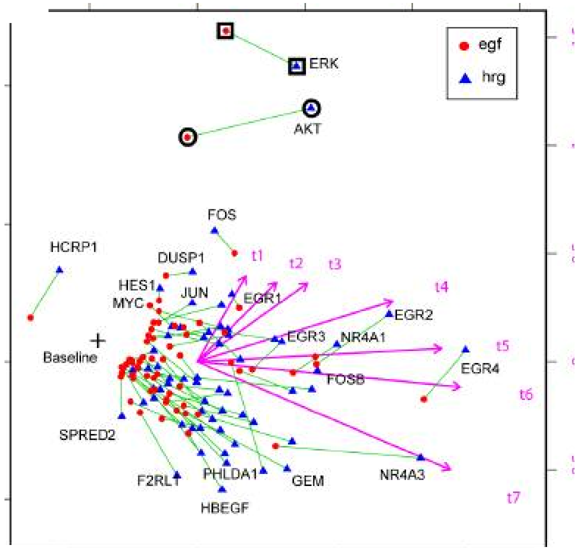


Significant change (>1.5 fold change, $p<0.0001$)

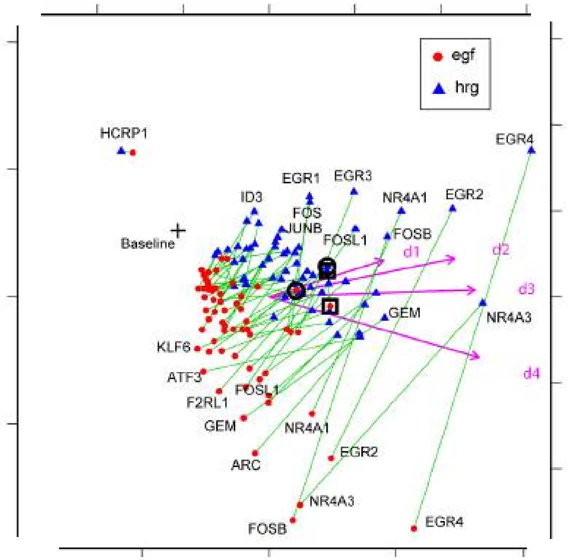


PCA of time-course and dose-dependent profiles (common 62 genes)

PCA of time-course profile

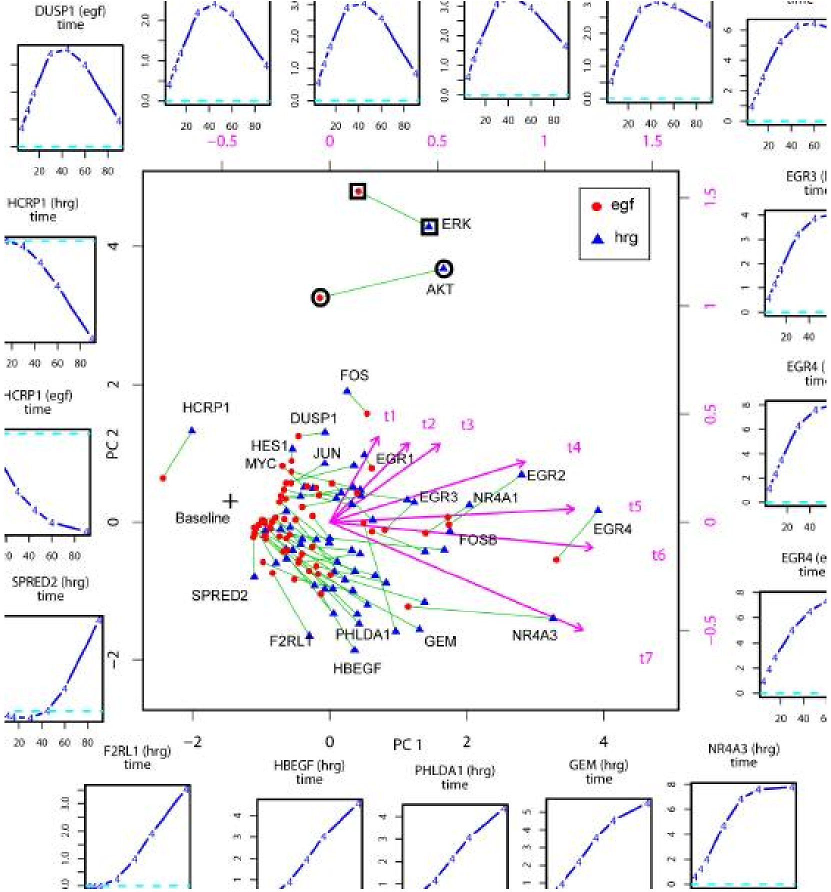


PCA of dose-dependent profile

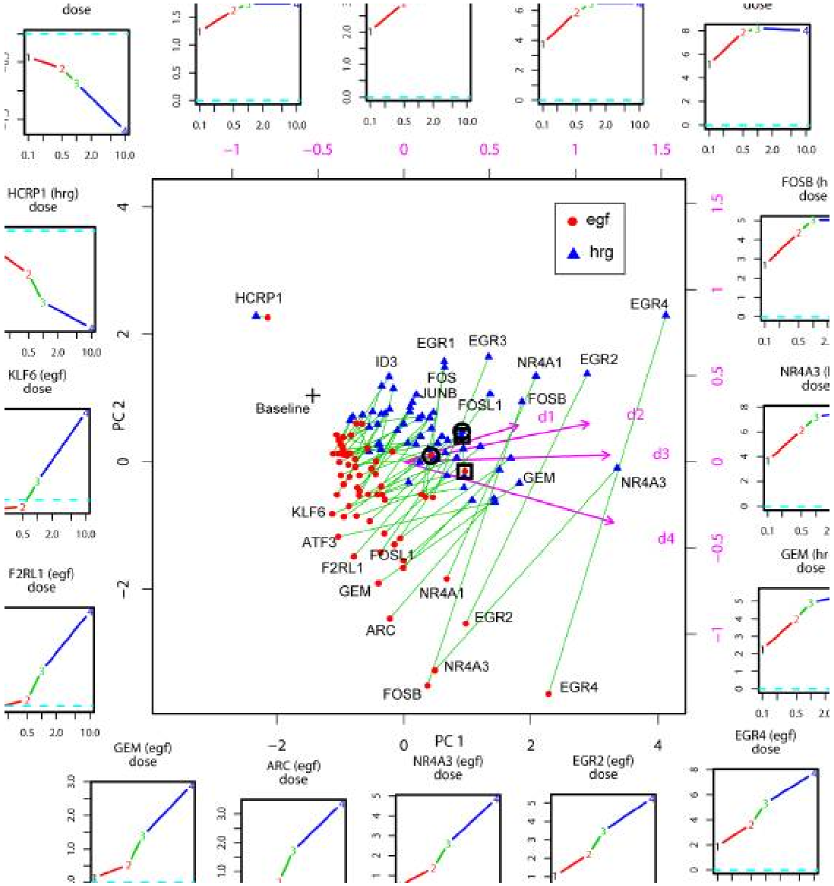


PCA of time-course and dose-dependent profiles (common 62 genes)

PCA of time-course profile



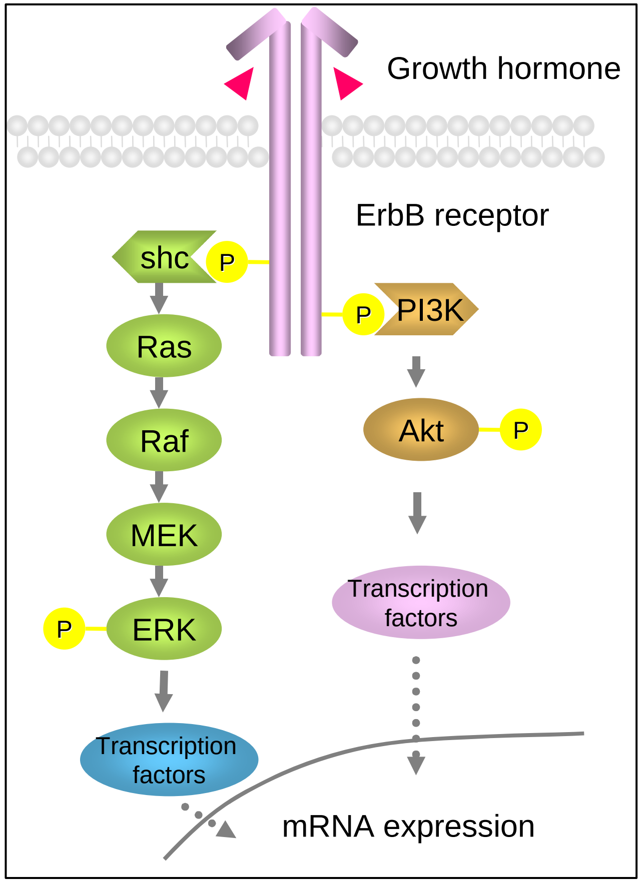
PCA of dose-dependent profile

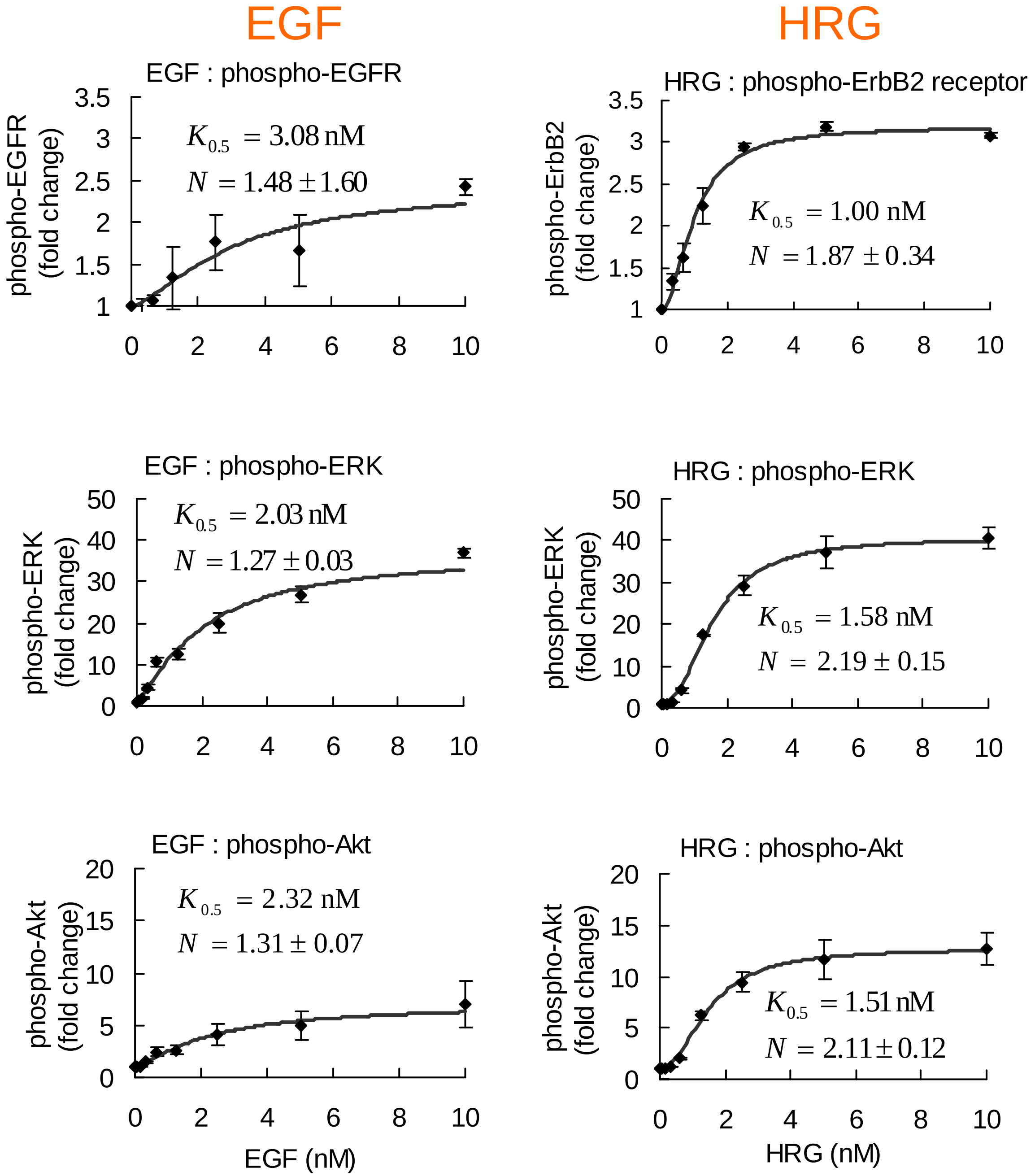


-- EGF-induced genes are closer to baseline (black cross) compared to HRG-induced gene (i.e., smaller expression average)

-- EGF and HRG regulated gene expression at the same time, while their effect of dosage was very different.

Ligand sensitivity of the receptors



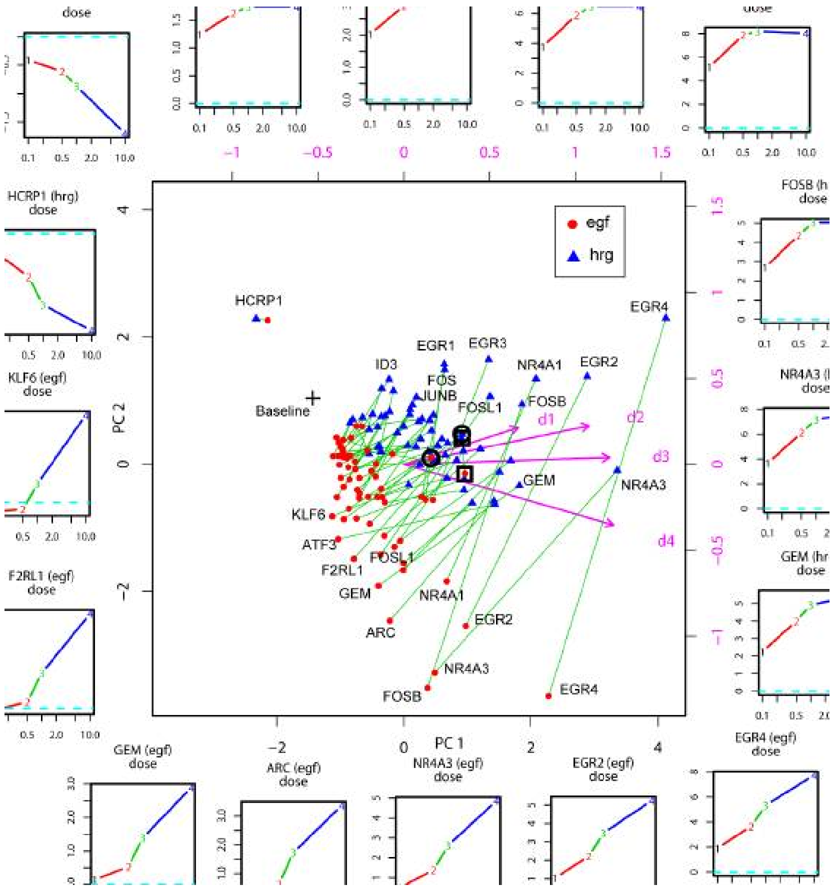


$K_{0.5}$: the concentration of [*EGF*] at which [*pERK*] is half-maximal

N : Hill coefficient.

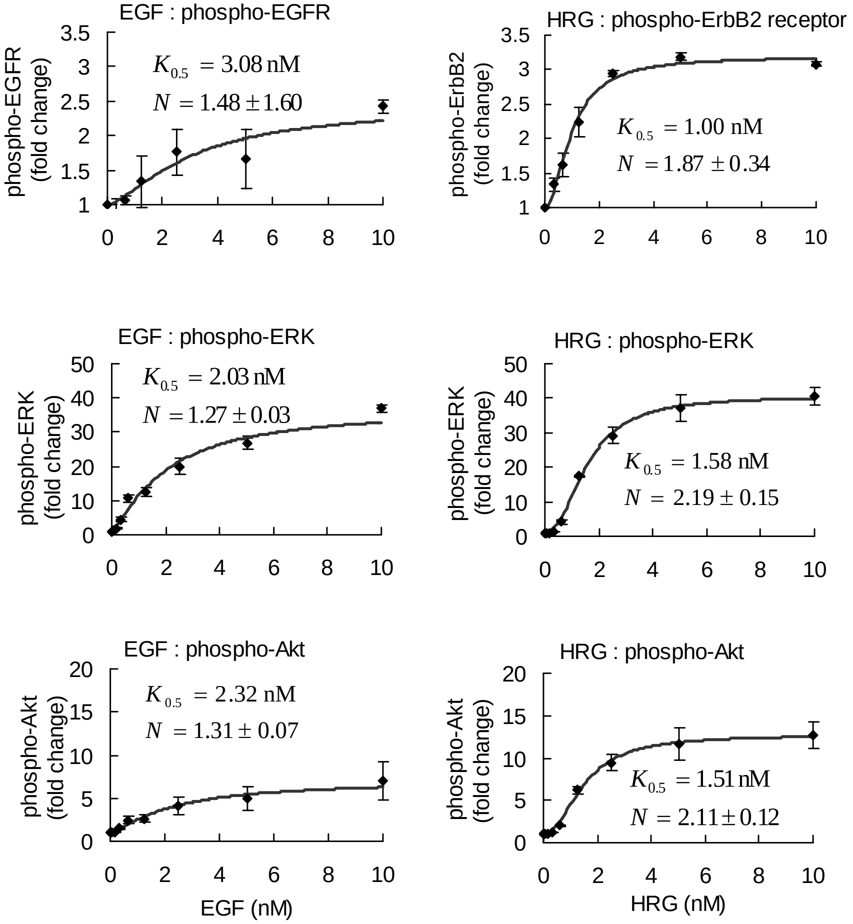
Ligand dose-dependent property of early transcription might be determined by the membrane receptor

PCA of dose-dependent profile



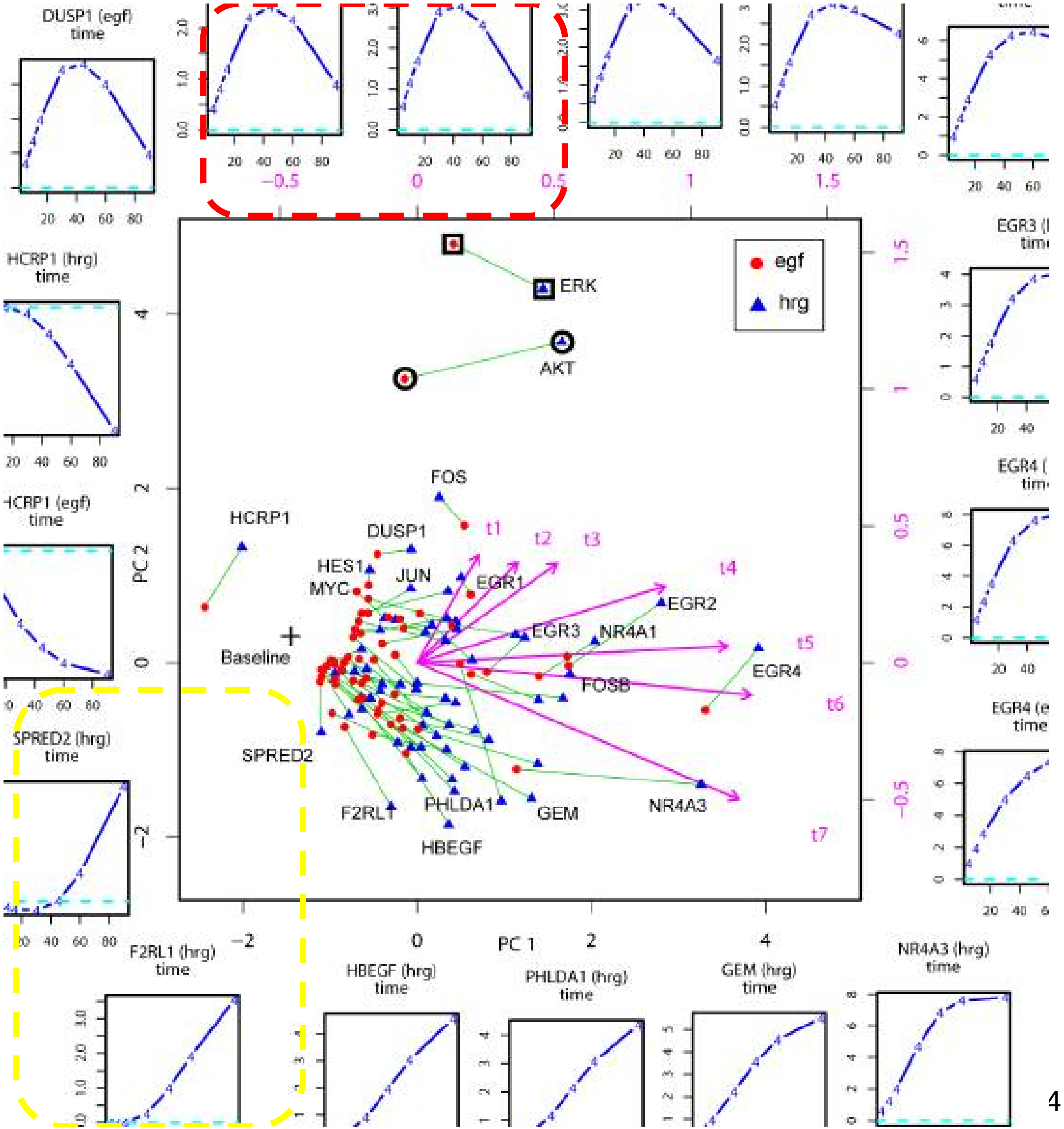
EGF

HRG

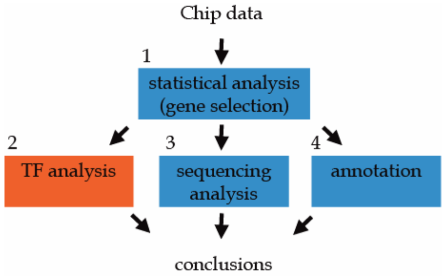


Time decomposition

Different kinetics
=
Different regulation



Transcription factor binding site



Chip data

1 statistical analysis (gene selection)

2 TF analysis

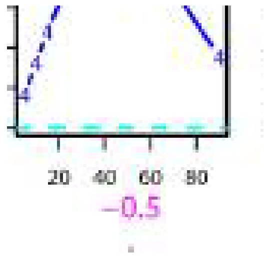
3 sequencing analysis

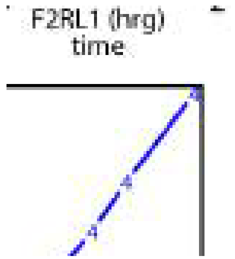
4 annotation

conclusions

Quick genes (23)

Slow genes (2)



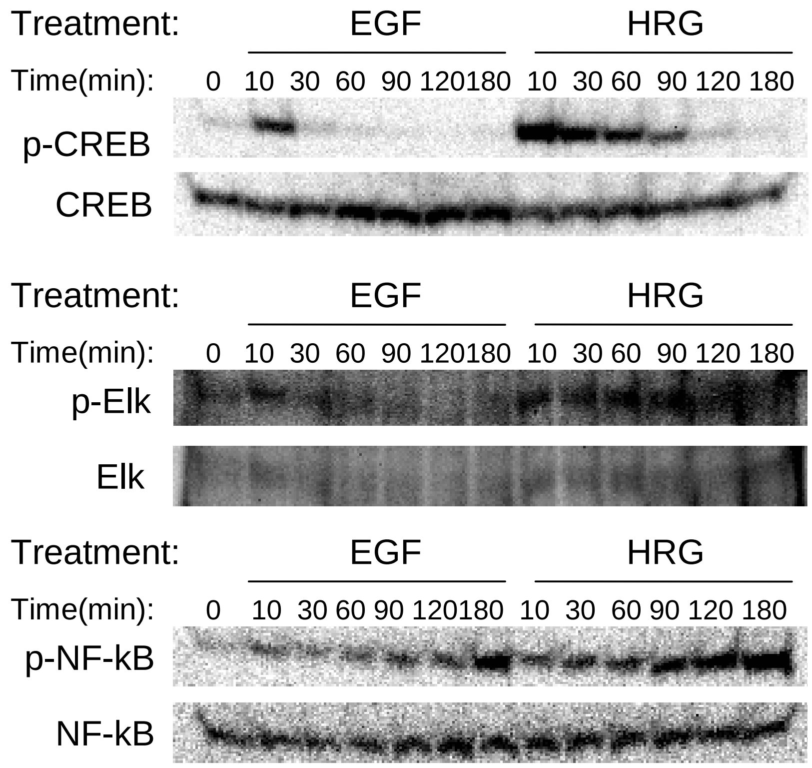


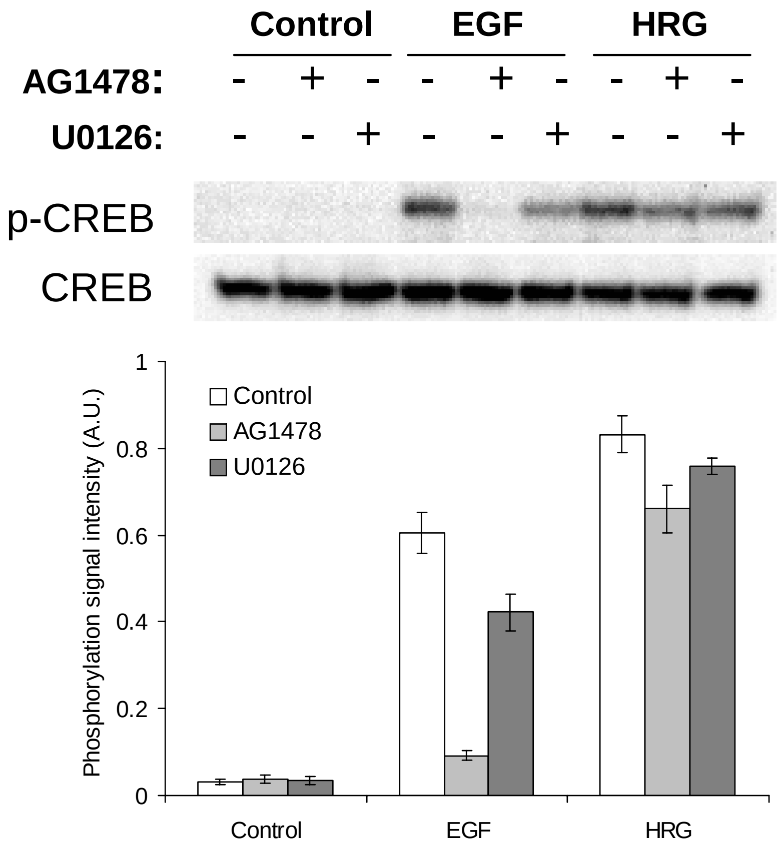
sequence survey – up to 2000bp upstream of 5′

Transcription factors	Quick genes (/23)	Slow genes (/2)
CREB	14	0
Myb	13	0
Elk-1	15	0
NF-kappaB	16	0

Only CREB was significant ($p<0.05$)

Time-selective activation of transcription factors

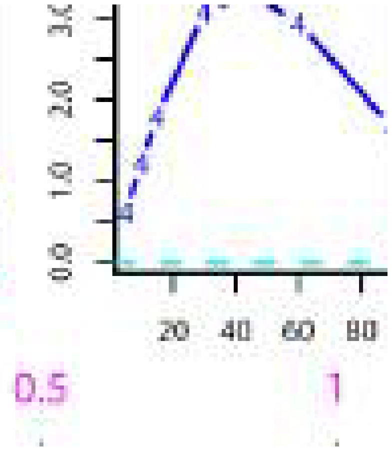


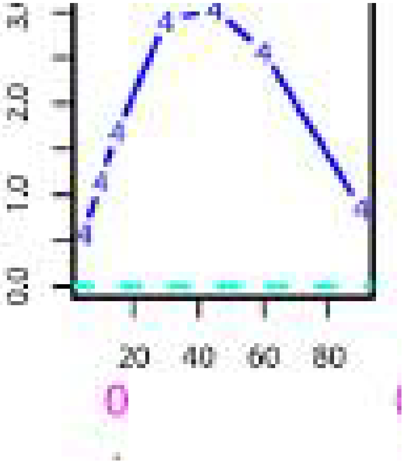


Effect of ligand for protein induction

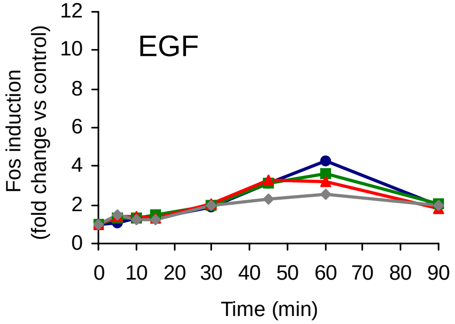
EGF HRG

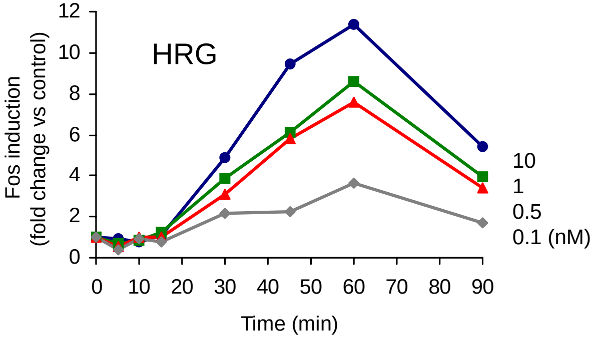
FOS transcript



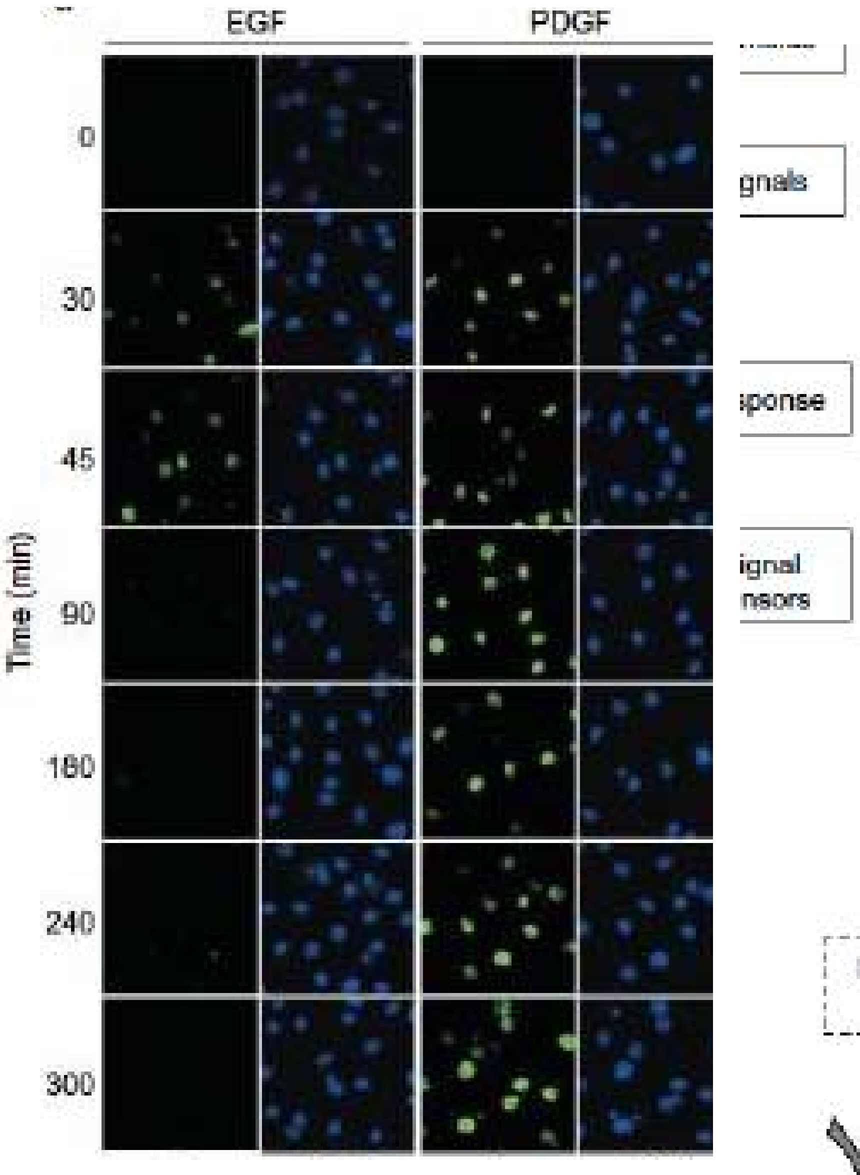


c-Fos protein

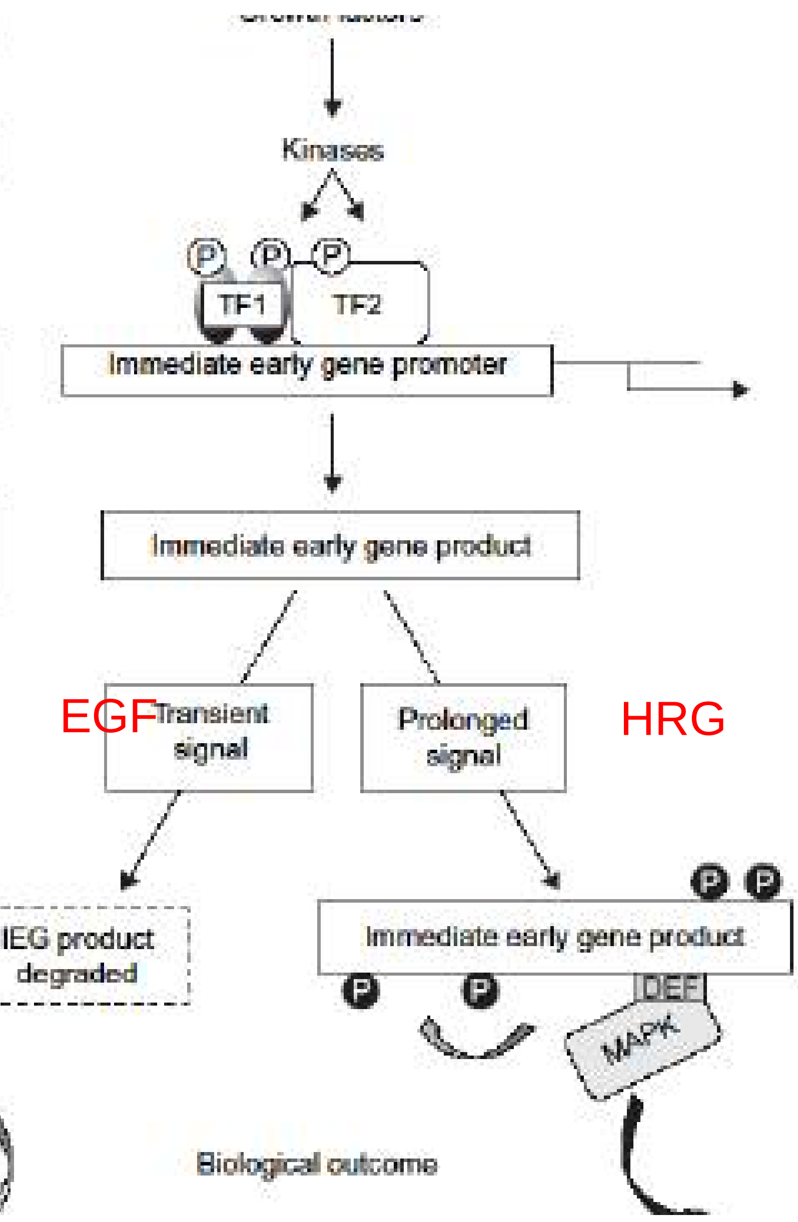




Prolonged ERK activity stabilizes proteins



Swiss 3T3 cells



EGF

HRG

Murphy LO, et al. *Nat Cell Biol.* 4:556-64, 2002

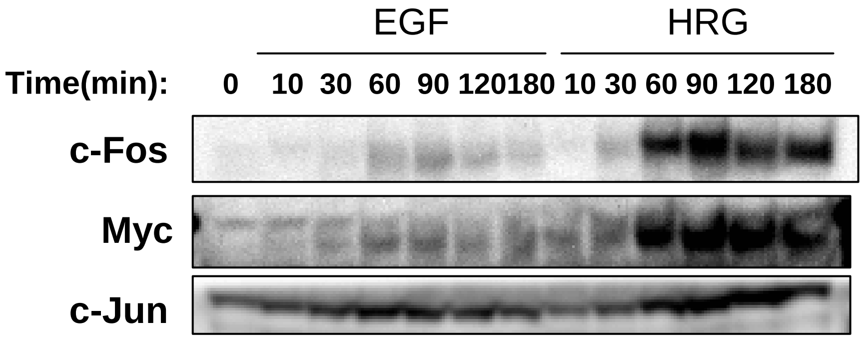
DEF domain and protein synthesis

le 1. DEF domains identified in immediate early gene products

os	^{333}GPMVTELEPLCTP-VVTCTPSCTTYTSSFV**FTYP**EEADS
-1	^{211}GP-VLEPEALHTPTLMT-TPSLTPFTPSLV**FTYP**STPEP
-2	^{188}GGFYGE-EPLHTP-IVVTSTPAITPGTSNLV**FTYP**SVLEQ
B	^{283}HSEVQV--LGDPFPVV-SPS--YTSSFVLTCPEVSAF
D	^{87}LLASPDLGLLKLASPELERLIIQS-NGLVTTTPTST-Q**FLYP**KV
B	^{71}GQGSDTGASLKLASSELERLIVPNSNGVITTTPTPPGQY**FYP**RG
un	^{58}LLTSPDVGLLKLASPELERLIIQSSNGHITTTPTPT-QFLCPKN
yc	^{56}LPTPPLSPSRRSGLCSPSYV//^{161}LTA-AASECIDPSVV**FPYP**LND
yc	^{77}AQSPGAGAASPAGRGHGGAAGA//^{118}AHPAAECVDPAVV**FPFP**VNK
-1	^{164}QSPPLSCAVPSNDSSPIYSAAPTFPTPNTD//^{247}PMIPD**YLFP**QQ
r1	^{447}PRGGPQPLPPAPTSVPPAAFPAPLVTPMVALPN**YLFP**TPSSY

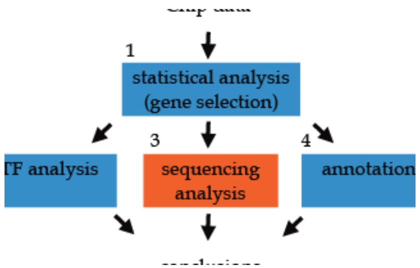
domains are in bold and numbers indicate amino acid position. Sequences
rom rat (c-Fos, Fra-1 and Fra-2) mouse (FosB, JunD, c-Jun, c-Myc and Egr-

Murphy LO, et al. *Nat Cell Biol.* 4:556-64, 2002.



EGF << HRG

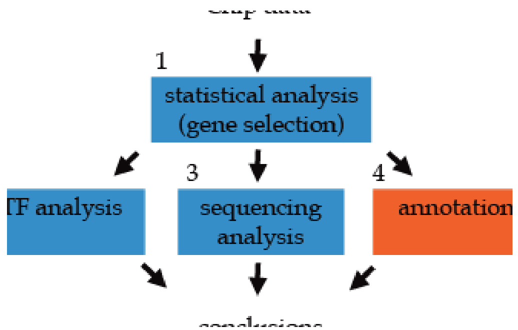
Ligand-induced transcripts with DEF domains

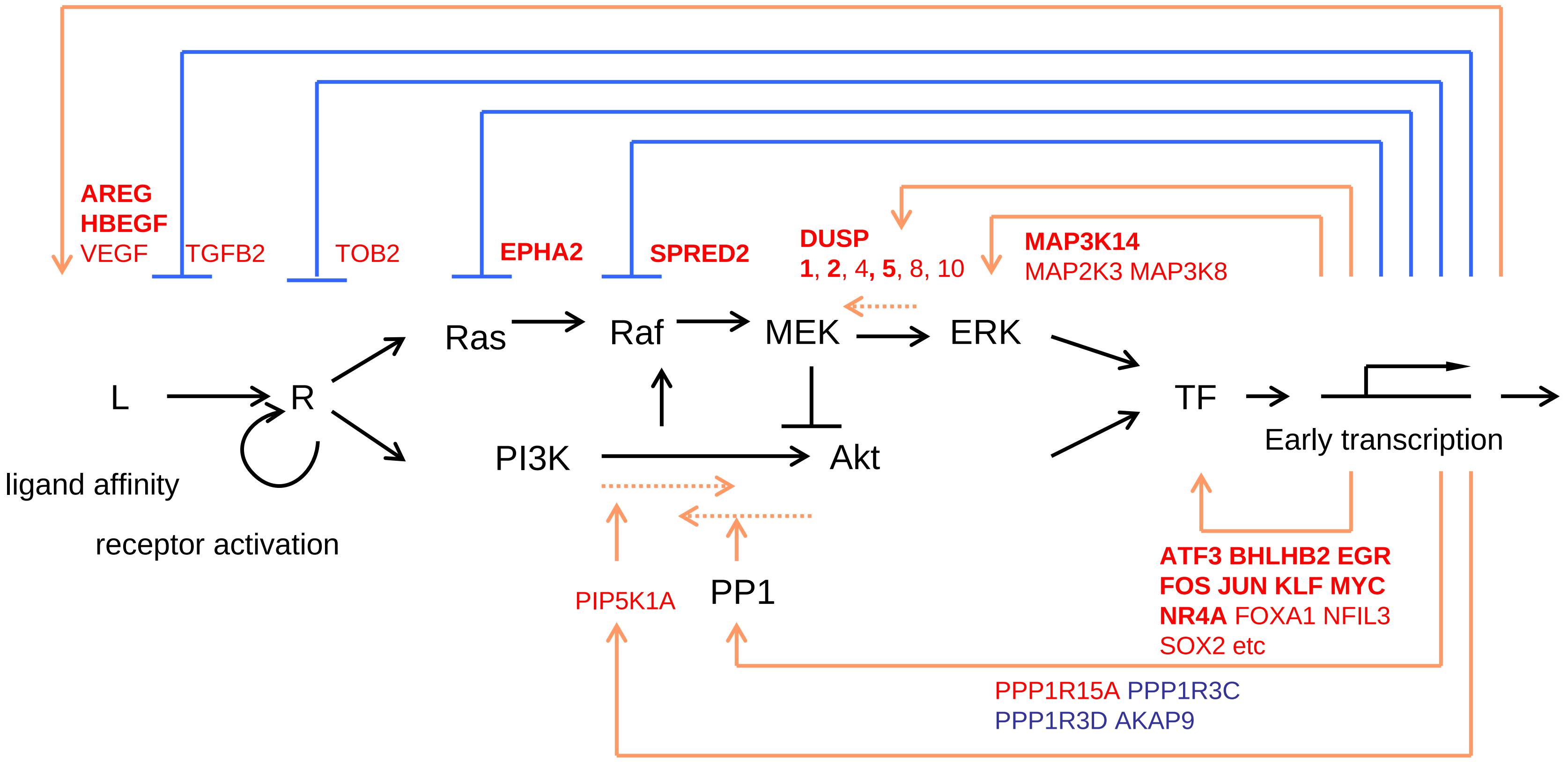


No	Symbol	Probe Set ID	Gene ID	Protein acc.	Seq. len.	Motif			Alignment		
						Start	End	Seq.	Start	End	Seq. + gap
1	CSNK1D	207945_s_at	1453	Q6ZNS0	128	99	102	FSYP	79	122	-ATGGFLKMPPPKYVLWSYFFFSYPQEHVSSVCFHSTGSATIHMA--
2	DST	204455_at	667	NP_001714	2649	2319	2322	**FLFP**	2299	2342	-QALYYSELLRMCVFDVESQCFLFPFGERNISNLNVKKTHRISVV--
	DST	204455_at	667	NP_065121	3062	2555	2558	**FPFP**	2535	2578	-HLKLLPGKNTRDSFKLINSQFPFPQITNNEELNQKGSLKKATVT--
	DST	204455_at	667	Q03001	3214	2884	2887	**FLFP**	2864	2907	-QALYYSELLRMCVFDVESQCFLFPFGERNISNLNVKKTHRISVV--
3	**DUSP1**	201041_s_at	1843	P28562	367	339	342	**FNFP**	319	362	AEAGSPAMAVLDRGTSTTT-VFNFPVSIPVHSTNSALSYLQSPIT--
4	DUSP4	204015_s_at	1846	NP_476499	303	275	278	**FSFP**	255	298	-PSGPLRERGKTPATPTSQFVFSFPVSVGVHSAPSSLPYLHSPIT--
	DUSP4	204015_s_at	1846	NP_001385	394	366	369	**FSFP**	346	389	-PSGPLRERGKTPATPTSQFVFSFPVSVGVHSAPSSLPYLHSPIT--
	DUSP4	204015_s_at	1846	Q13649	411	383	386	**FSFP**	363	406	-PSGPLRERGKTPATPTSQFVFSFPVSVGVHSAPSSLPYLHSPIT--
5	**F2RL1**	213506_at	2150	P55085	397	251	254	**FLFP**	231	274	-PEQLLVGDMFNYFLSLAIGVFLFPAFLTASAYVLMIRMLRSSAM--
6	F3	204363_at	2152	Q86WH3	145	15	18	FSYP	1	38	-------EIVKDVKQTYLARVFSYPAGNVESTGSAGEPLYENSPE--
	F3	204363_at	2152	Q86SE7	220	90	93	FSYP	70	113	-ECDLTDEIVKDVKQTYLARVFSYPAGNVESTGSAGEPLYENSPE--
	F3	204363_at	2152	P13726	295	108	111	FSYP	88	131	-ECDLTDEIVKDVKQTYLARVFSYPAGNVESTGSAGEPLYENSPE--
7	**FOS**	209189_at	2353	P01100	380	272	275	**FLFP**	252	295	-SVEPVKSISSMELKTEPFDDFLFPASSRPSGSETARSVPDMDLS--
	FOS	209189_at	2353	P01100	380	343	346	FTYP	323	366	-LCTPVVTCTPSCTAYTSSFVFTYPEADSFPSCAAAHRKGSSSNE--
8	**FOSL1**	204420_at	8061	NP_005429	271	235	238	FTYP	215	258	-LHTPTLMTTPSLTPFTPSLVFTYPSTP--EPCASAHRKSSSSSGDP
9	GATA2	209710_at	2624	NP_116027	480	171	174	**FGFP**	151	194	-GSGSSVASLTPTAAHSGSHLFGFPPTPPKEVSPDPSTTGAASPA--
10	**JMJD3**	213146_at	23135	O15054	1682	671	674	**FDFP**	651	694	-KAPQPVPPGVGELPARGPRLFDFPPTPLEDQFEEPAEFKILPDG--
11	KLF2	219371_s_at	10365	Q8IUN4	224	73	76	FYYP	53	96	-LDGLGAEAAPEPPPPPPPPAFYYPEPGAPPPYSAPAGGLVSELL--
12	**MYC**	202431_s_at	4609	P01106	439	195	198	FPYP	175	218	-SLYLQDLSAAASECIDPSVVFPYPLNDSSSPKSCASQDSSAFSP--
	MYC	202431_s_at	4609	NP_002458	454	210	213	FPYP	190	233	-SLYLQDLSAAASECIDPSVVFPYPLNDSSSPKSCASQDSSAFSP--
13	NDEL1	208093_s_at	81565	Q6ZW10	148	8	11	**FTFP**	1	31	-------MCAGVH-------VFTFPPAETSDAVKPSHIKQYTLTG--
14	OSR2	213568_at	116039	Q8N2R0	312	83	86	**FPFP**	63	106	-EITRSTITEMAAAQGLVDARFPFPALPFTTHLFHPKQGAIAHVL--
15	PGAP1	220576_at	80055	Q75T13	922	43	46	FEYP	23	66	-LGLWDVFFGFEENKCSMSYMFEYPEYQKIELPKKLAKRYPAYEL--
	PGAP1	220576_at	80055	Q6AW92	748	196	199	**FTFP**	176	219	-VLVKVSKWTYVAYNESEKIYFTFPLEN--HRKIYTHVYCQSAMLDT
	PGAP1	220576_at	80055	Q75T13	922	370	373	**FTFP**	350	393	-VLVKVSKWTYVAYNESEKIYFTFPLEN--HRKIYTHVYCQSTMLDT
16	PPP1R15A	37028_at	23645	Q6IA96	674	177	180	FSYP	157	200	-NPGEEKAEEEGVAEEEGVNKFSYPPSHRECCPAVEEEDDEEAVK--
17	PPP1R3D	204554_at	5509	Q86X09	299	238	241	**FGFP**	218	261	-EAVARWRGPAGPEGTEDVFTFGFPVPPFLLELGSRVHFAVRYQV--
18	RARA	203749_s_at	5914	P10276	462	26	29	**FFFP**	6	49	-SSCPTPGGGHLNGYPVPPYAFFFPPMLGGLSPPGALTTLQHQLP--
19	TSC22D2	210953_at	9819	O75157	780	357	360	FAYP	337	380	-SLPPQPGPAVGAPAAQQPQQFAYPQPQIPPGHLLPVQPSGQSEY--
20	VIL2	208621_s_at	7430	Q8IXG9	167	59	62	FNYP	39	82	-TVREILTSPSCWQYAVLLNRFNYPFELEKDLHLKGYHTLSQGSL--
21	ZNF287	220055_at	57336	NP_065704	754	48	51	FPYP	28	71	EILTSRFLRDTETCRQNFR-NFPYPDLAGPRKALSQLRELCLKWL--
22	ZNF307	213625_at	387032	NP_061983	545	59	62	FRYP	39	82	VRAPCSPARGPERSRQRFR-GFRYPEAAGPREALSRLRELCGQWL--

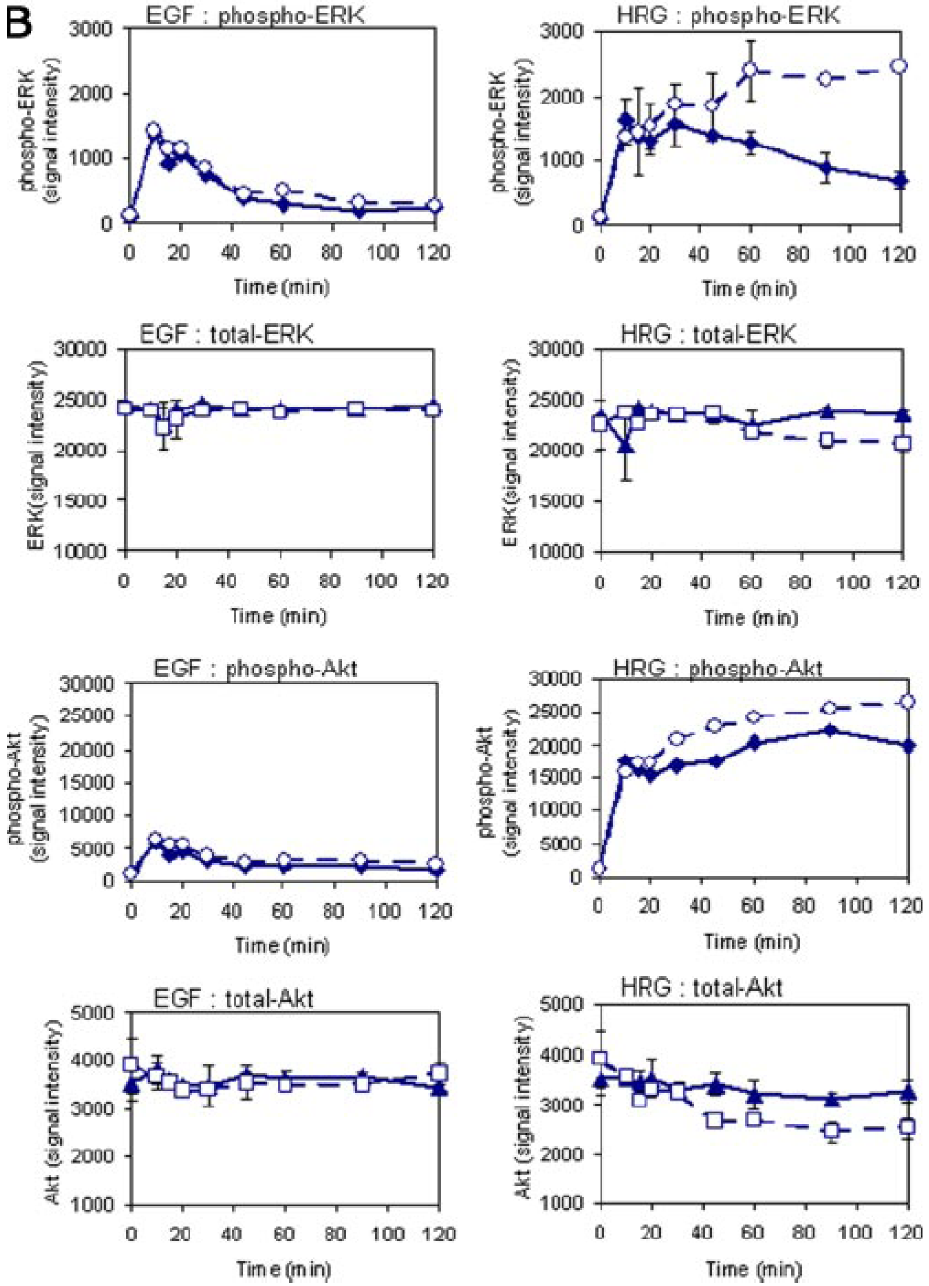
Bold letters, found in both EGF and HRG
Plain letters, only found in HRG

Transcriptional feedback to signaling pathway (EGF-, HRG-induced genes)



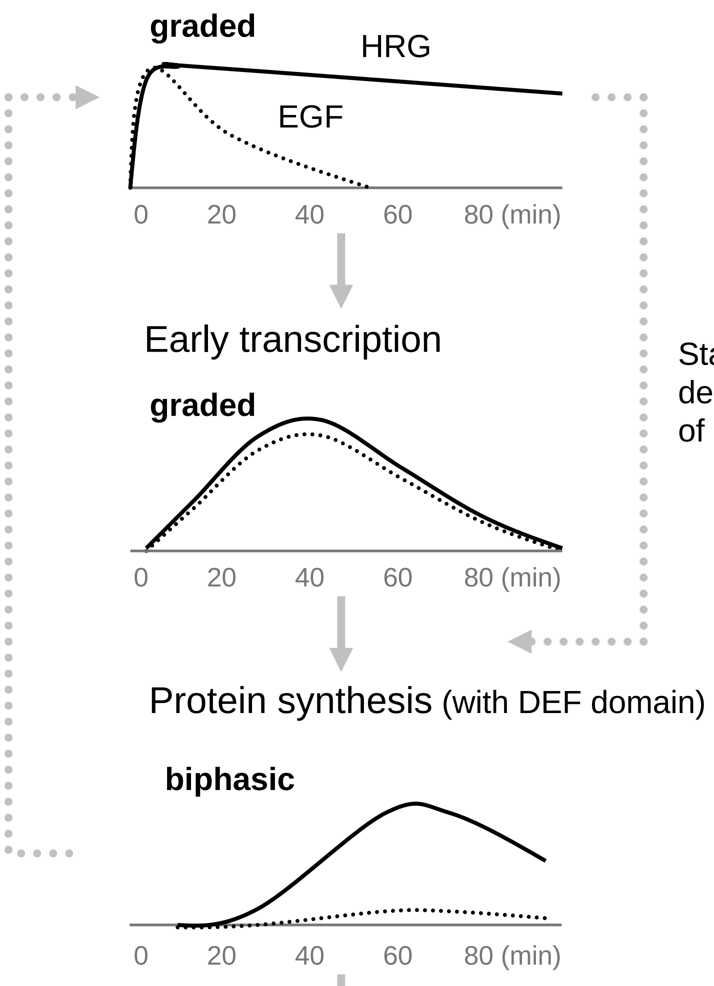


ErbB signaling kinetics might be regulated by early transcription products through negative and positive feedback


B
EGF : phospho-ERK
phospho-ERK (signal intensity)
0
1000
2000
3000
Time (min)
0 20 40 60 80 100 120
HRG : phospho-ERK
phospho-ERK (signal intensity)
0
1000
2000
3000
Time (min)
0 20 40 60 80 100 120
EGF : total-ERK
ERK(signal intensity)
10000
15000
20000
25000
30000
Time (min)
0 20 40 60 80 100 120
HRG : total-ERK
ERK(signal intensity)
10000
15000
20000
25000
30000
Time (min)
0 20 40 60 80 100 120
EGF : phospho-Akt
phospho-Akt (signal intensity)
0
5000
10000
15000
20000
25000
30000
Time (min)
0 20 40 60 80 100 120
HRG : phospho-Akt
phospho-Akt (signal intensity)
0
5000
10000
15000
20000
25000
30000
Time (min)
0 20 40 60 80 100 120
EGF : total-Akt
Akt (signal intensity)
1000
2000
3000
4000
5000
Time (min)
0 20 40 60 80 100 120
HRG : total-Akt
Akt (signal intensity)
1000
2000
3000
4000
5000
Time (min)
0 20 40 60 80 100 120

Conclusions

Intracellular signaling



graded

Early transcription

graded

Protein synthesis (with DEF domain)

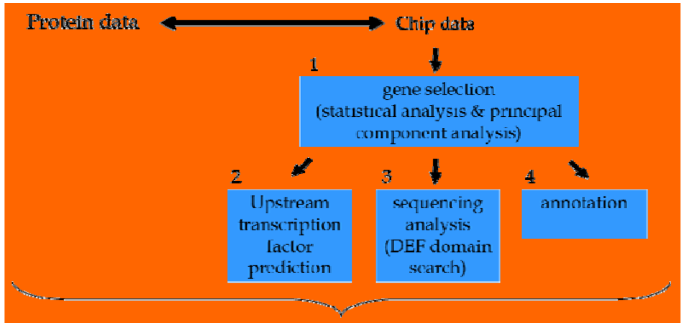
biphasic

Late transcription

Cell determination

Elongation
and
stabilization
of the signal

Stabilization or
degradation
of proteins



Protein data ↔ Chip data

1 gene selection (statistical analysis & principal component analysis)

2 Upstream transcription factor prediction

3 sequencing analysis (DEF domain search)

4 annotation

ErbB receptor signaling for cell fate control

Receptor

Define ligand sensitivity and duration of first signal

Early transcription = Quantitative control

Convert signal (duration and amplitude) to expression amplitude
Induce ligand-dependent biphasic induction of transcription factors
(coordination with upstream signaling)
=this part should be quantitatively examined using model & exp.

Late transcription = Qualitative control

Ligand-specific transcriptional network

Cell determination

Acknowledgements

Cellular Systems Biology Team, RIKEN GSC Members

Team leader

Mariko Hatakeyama

Bioinformatics analysis

Takeshi Nagashima

Experiments

Taishi Tsubouchi

Kaori Ide

Kaoru Takahashi

Noriko Yumoto

Kinetic modeling

Takashi Nakakuki

Collaborators

qRT-PCR

Harukazu Suzuki & Yoshihide Hayashizaki, RIKEN GSC

Statistics/Transcriptome analysis

Hidetoshi Shimodaira, TITEC

Takao Endo & Tetsuro Toyoda, RIKEN GSC

(For kinetic modeling)

Parameter Estimation

Shuhei Kimura, Tottori Univ.

Modeling

Takashi Naka, Kyusyu Sangyo Univ.

Boris Kholodenko, Thomas Jefferson Univ.

Thank you for your attention