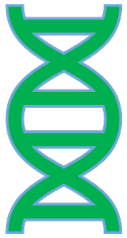
Gene regulatory network modeling

PHUSE SDE Heidelberg - 01-Dec-2022

www.mainanalytics.de

Agenda



Gene regulation

Gene expression

Gene regulation

GRN

Boolean networks

Propositional logic

Transition functions

Computational effort (complexity)

Construction Boolean Networks

Exciting insights

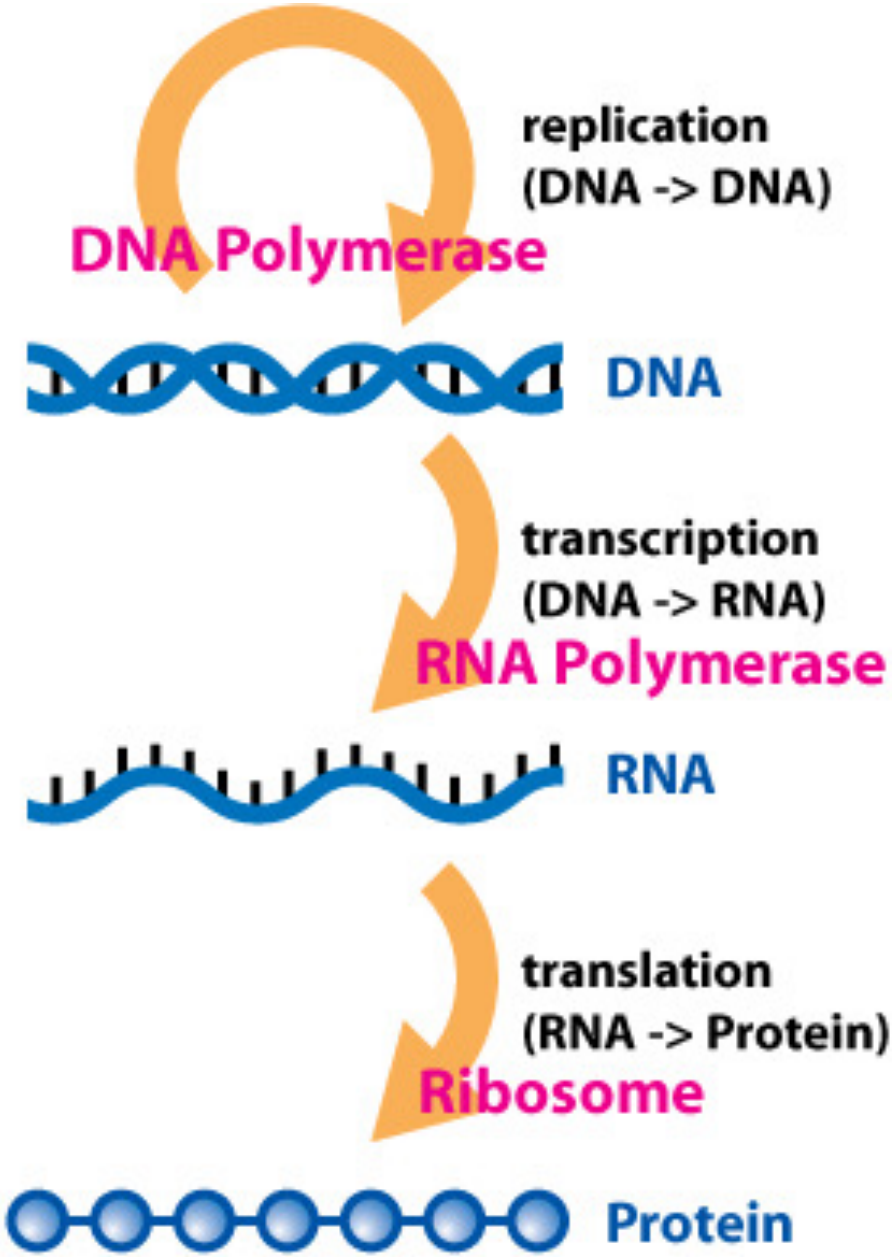
Studying connectivity

Studying regulatory functions

Living at the edge of chaos

Population based models

Gene expression



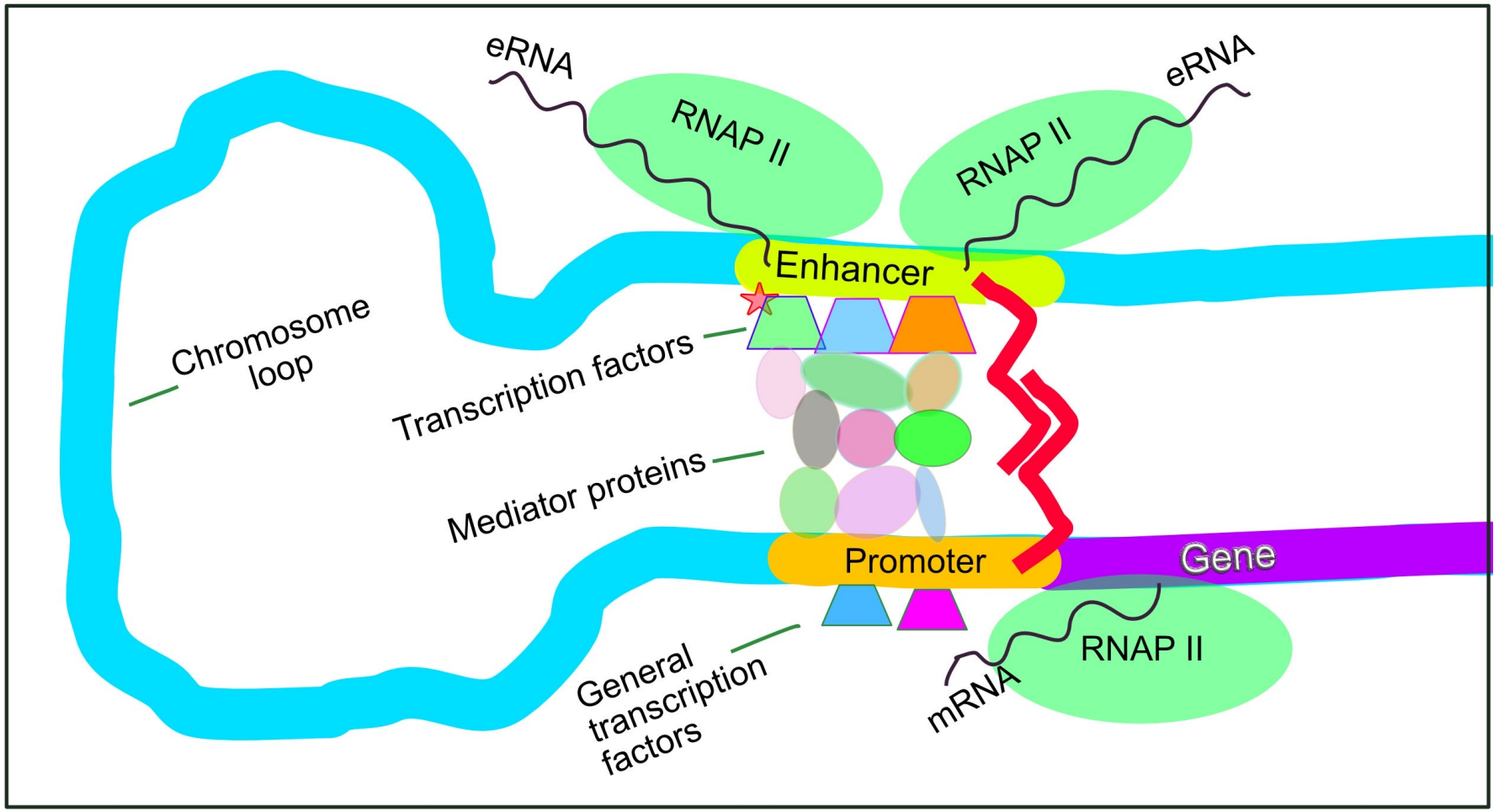
The Human genome

~ 3 000 000 000 base pairs

2bit per base pair
⇒ 750mb data

~ 20.300 genes (protein-coding)
~ 70,000 splice variants
~ 100 proteoforms per gene

Regulation of transcription



mediators :

complex of ~ 26 different proteins

enhancer RNAs (eRNAs)

relatively long non-coding RNA molecules (50-2000 nucleotides)

cis-regulatory modules (CRMs)

non-coding DNA -> regulates transcription of neighbouring genes

trans-acting factor

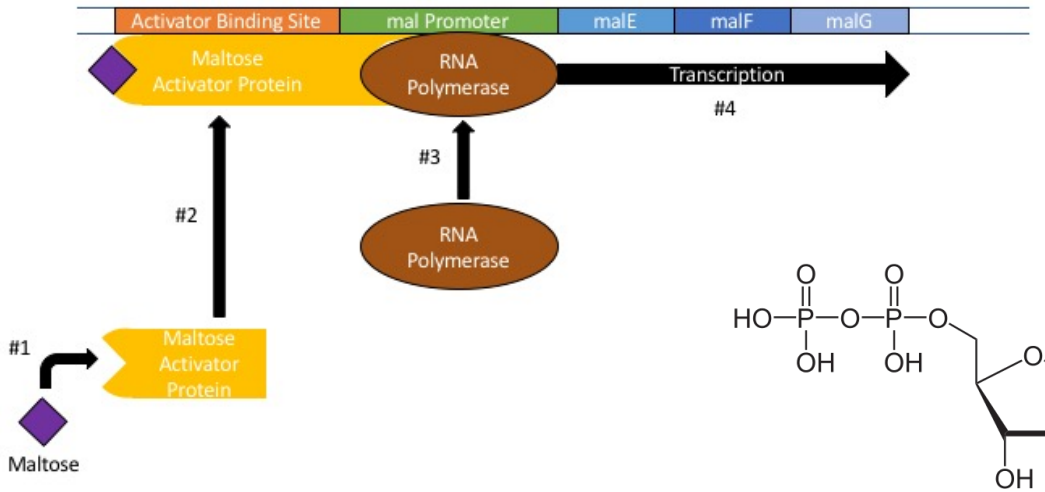
regulatory protein, binds to DNA.

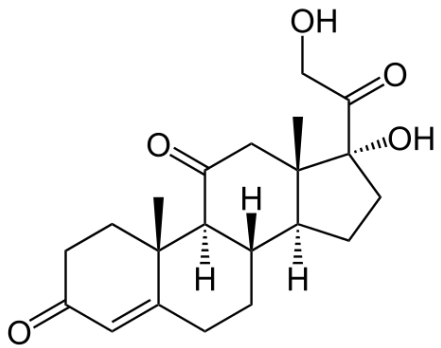
~ **1,600** different transcription factors in a human cell

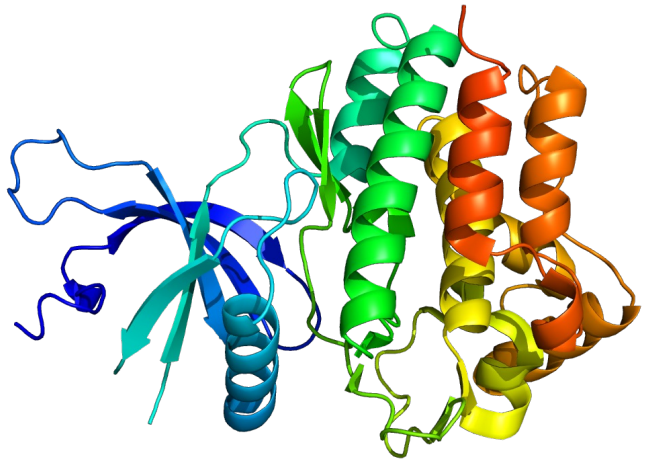
Non TFs that modifies gene expression

- Proteins
  - coactivators, chromatin remodelers, histone acetyltransferases, histone deacetylases, **kinases,** methylases
- RNA
  - eRNA (enhancer RNA), antisense RNA
- Metabolic products
  - maltose, ADP, …
- Hormones
  - steroids, L-Thyroxin, …
- DNA methylation
- Histones
- Protein phosphorylation

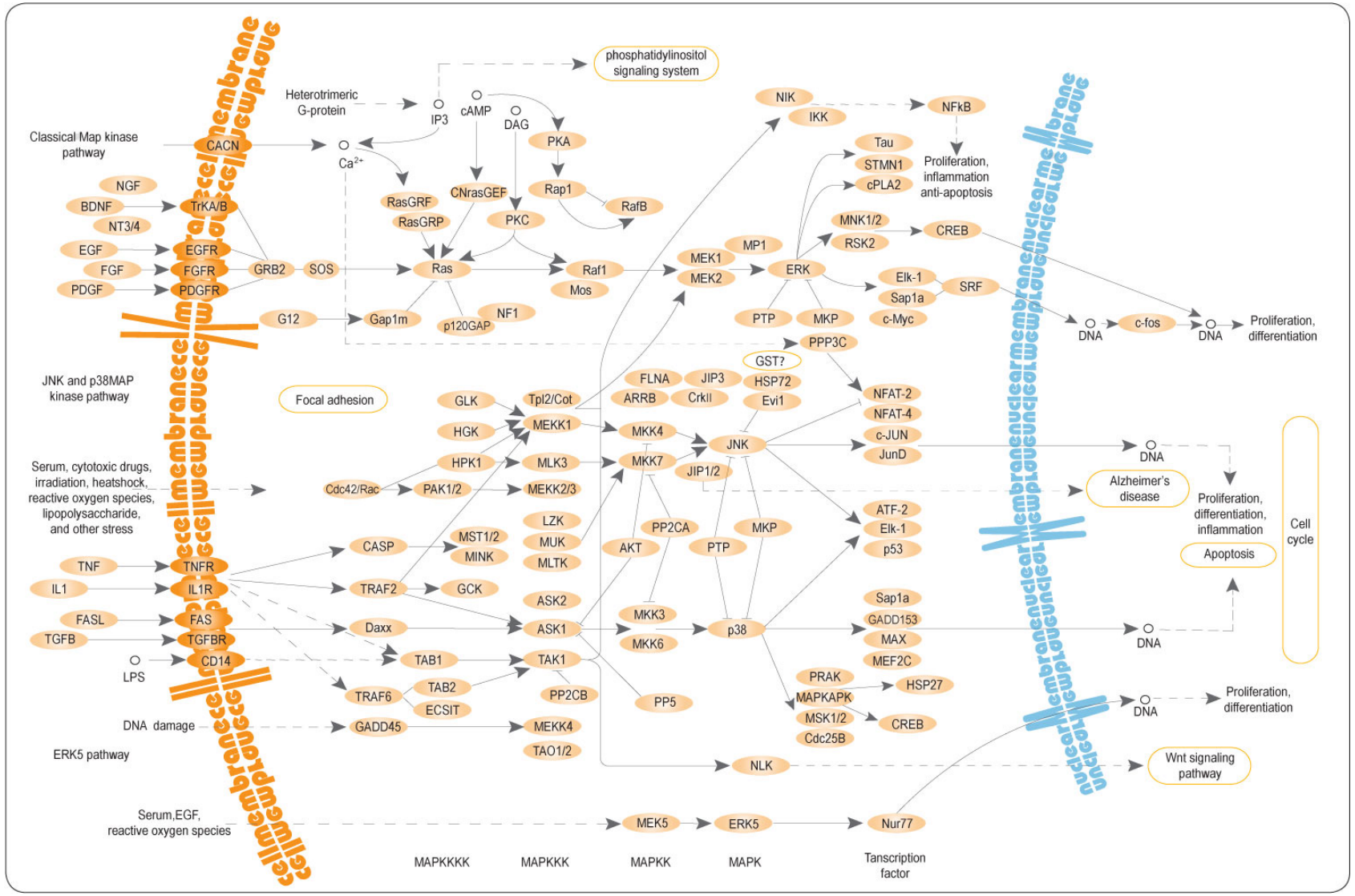
Figure 2: Maltose Present



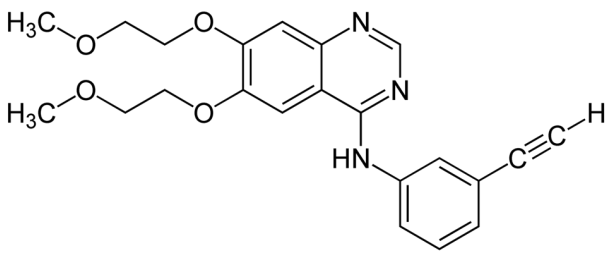


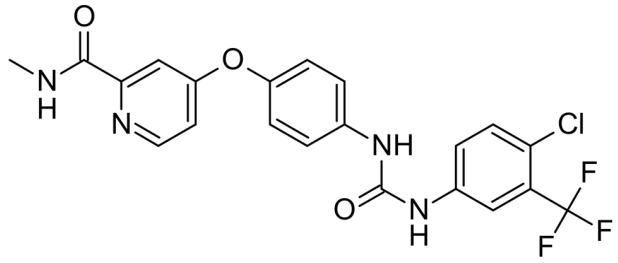


MAP Kinase pathway

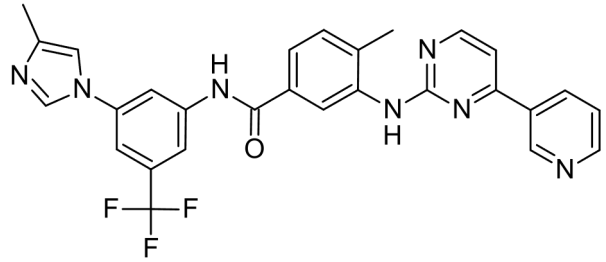


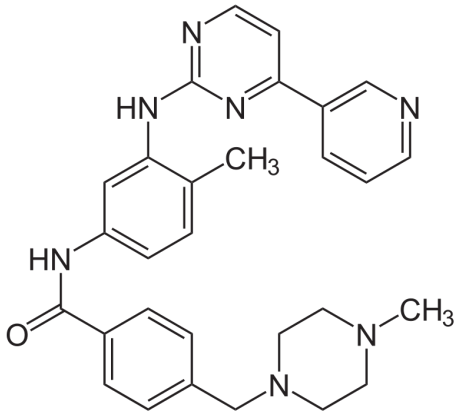
Inhibitors of MAP kinases



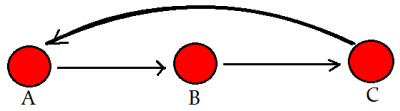


Target family	Specific target	Inhibitor	Vendor (catalog#)
Receptor tyrosine kinase	EGFR	Gefitinib (Iressa®)	American Custom Chemicals (ACC) Corporation (184475-35-2)
	VEGFR	Erlotinib (Tarceva®)	ACC (183321-74-6)
			Axon Medchem (1128)
		Lapatinib (Tykerb®)	SELLECK (S1028)
			ACC (231277-92-2)
	PDGFR	Sunitinib (Sutent®)	ACC (557795-19-4)
			Axon Medchem (1398)
		Sorafenib (Nexavar®)	ACC (284461-73-8)
			Axon Medchem (1397)
Non-receptor and receptor tyrosine kinases	Bcr–Abl,	Nilotinib (Tasigna®)	SELLECK (S1033)
			ACC (64157-10-0)
	Bcr–Abl, c-Src	Dasatinib (Sprycel®)	ACC (302962-49-8)
			Axon Medchem (1392)
	Bcr–Abl, c-SCT, c-Kit, PDGFR	Imatinib (Gleevec®)	ACC (220127-57-1)
			Axon Medchem (1394)
G-proteins	Ras	Tipifarnib (Zarnestra™)	Onicon Pharmachemie (192185-7201)
MAPKKK	Raf	Sorafenib (Nexavar®)	ACC(284461-73-8)
			Axon Medchem (1397)
MAPKK	MEK1/2	U0126	EMD Biosciences (662005)
			SELLECK (S1102)
		PD184352	Axon Medchem (1368)
			SELLECK (S1020)
		AZD6244	SELLECK (S1008)
	MEK5	BIX02188, BIX02189	Boehringer Ingelheim (*not commercially available*)
MAPK	p38	SB203580	EMD Biosciences (55389)
			Axon Medchem (1364)
		SB202190	EMD Biosciences (559388)
			Axon Medchem (1363)
		BIRB-796	Axon Medchem (1353)

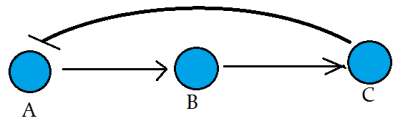




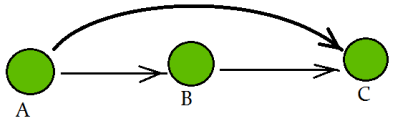
Gene regulatory networks



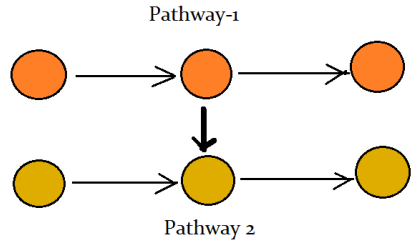
Positive Feedback



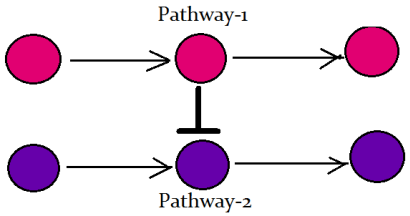
Negative feedback
(Feedback Inhibition)



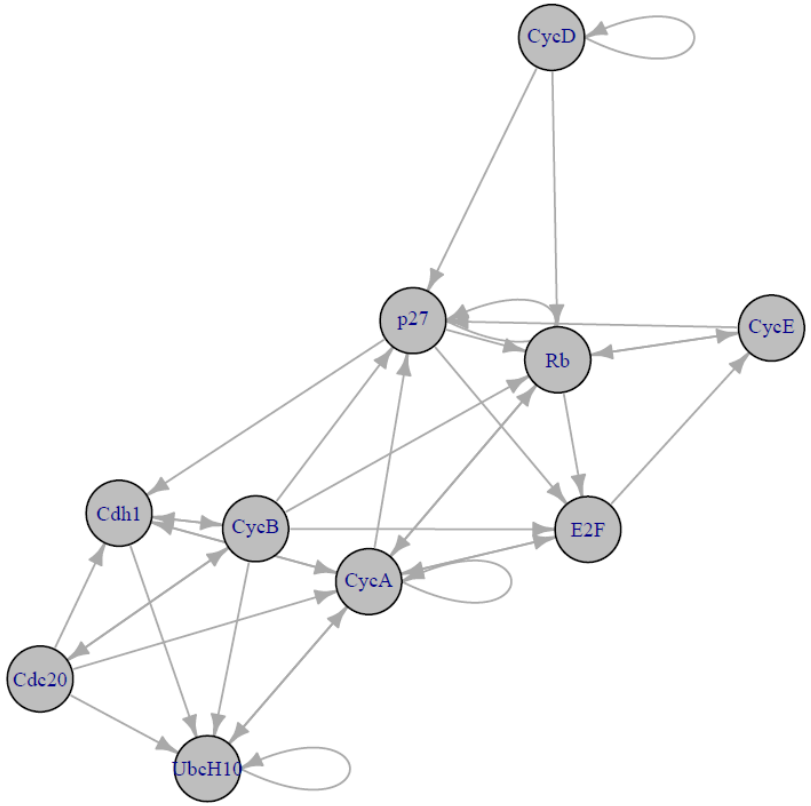
Feedforward relay



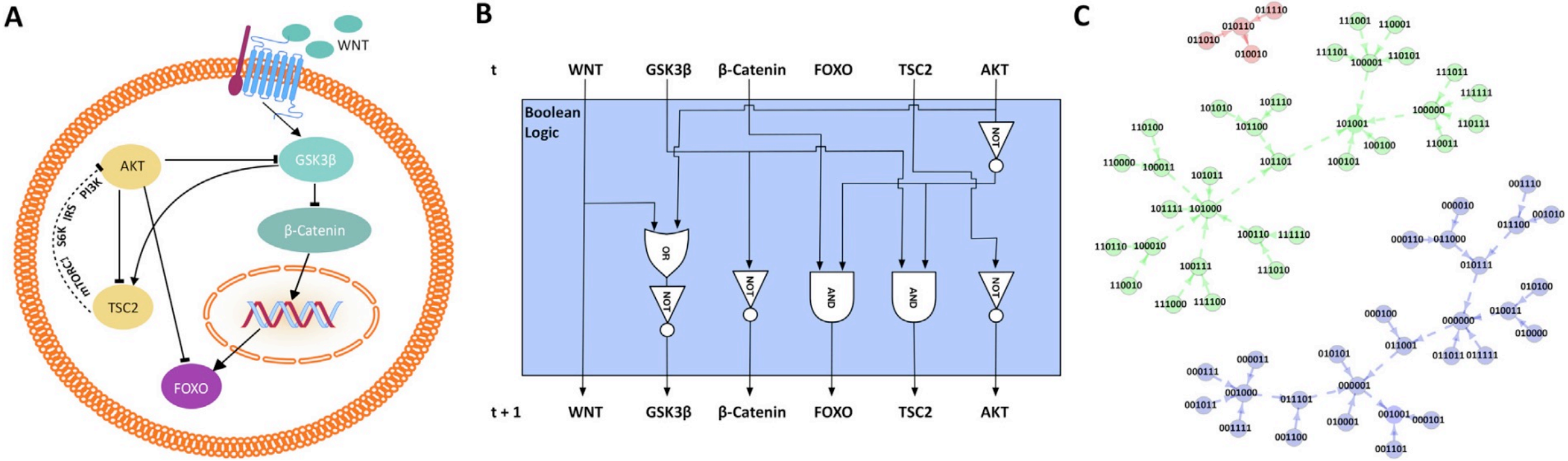
Stimulatory
Crosstalk



Inhibitory
Crosstalk



Boolean Network of the mammalian cell cycle as proposed by Faure et al. [4].
Generated with boolNet-Package [5].



Boolean Networks

discretised variables / components / genes

{ high/low , on/off , 1/0, true/false }

- similar to experts knowledge

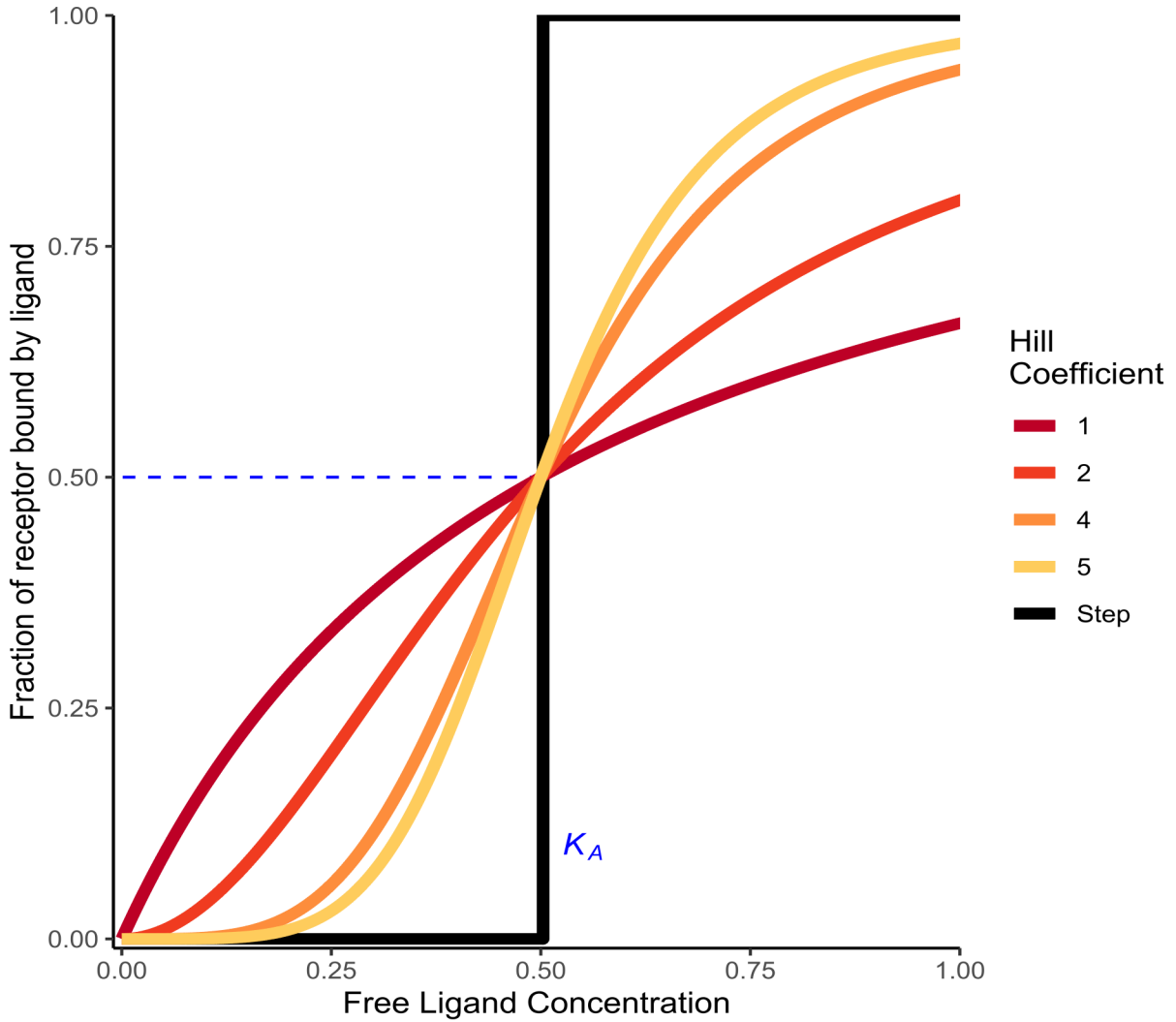
Hill function vs. step function

$$hill([L], K_A, n) = \frac{[L]^n}{[L]^n + {K_A}^n}$$

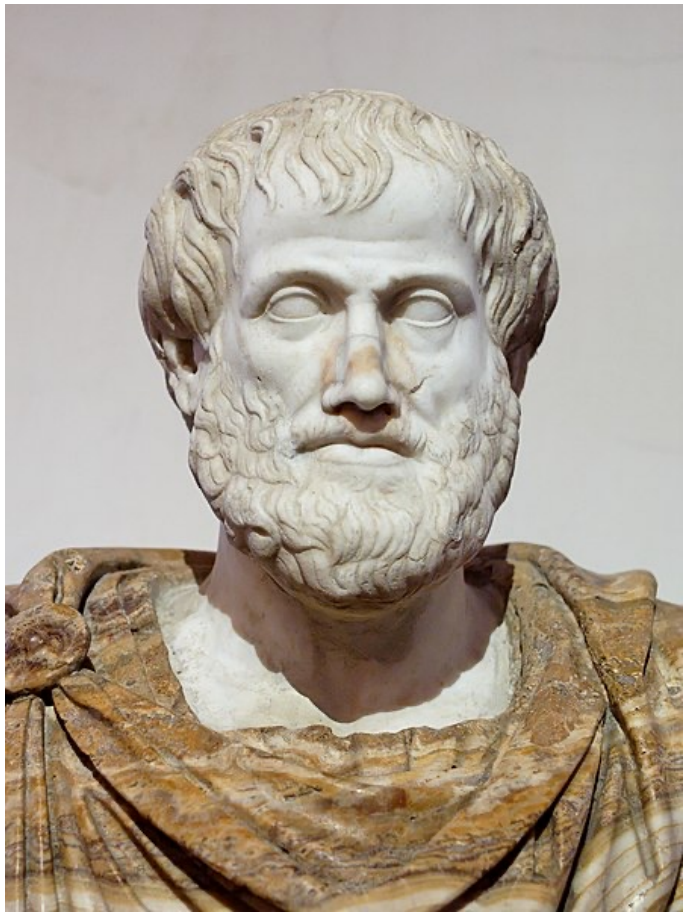
$$step([L], K_A) = \begin{cases} 0 & \text{if } [L] < K_A \\ 1 & \text{if } [L] >= K_A \end{cases}$$

$[L]$:= total ligand concentration

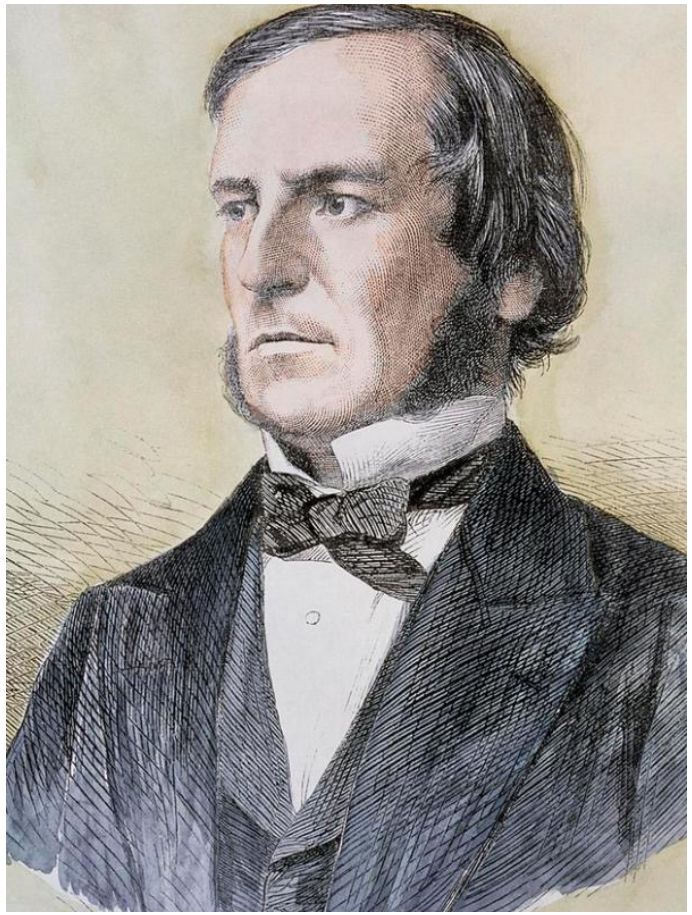
K_A := ligand concentration producing half occupation



Propositional logic



Aristotle (384 – 322 BC)



Georg Bool (1815 –1864)

Propositional logic

x_1	x_2	$\wedge$	$\vee$
0	0	0	0
0	1	0	1
1	0	1	0
1	1	1	1

$$2^{(2^K)} = \begin{cases} 2 & K = 0 \\ 4 & K = 1 \\ 16 & K = 2 \\ 256 & K = 3 \\ \dots & \end{cases}$$

connectivity in gene regulatory networks are usually low K << N ; K ~ 2-5

Boolean network

System $\mathbf{X}$ consists of n genes :

$$\mathbf{X} = \{x_1, x_2, ..., x_n\} \ , \ x_i \in \mathbb{B}$$

with n transistion functions f :

$$\mathbf{F} = \{f_1, f_2, ..., f_n\} \ , \ f_i : \mathbb{B}^{\mathbf{K}} \rightarrow \mathbb{B} \ ,$$

with connectivity $\mathbf{K} \leq n$

updates in descete time steps. State at time t :

$$\vec{\mathbf{x}}(t) = \{x_1(t), x_2(t), ..., x_n(t)\}$$

$$\mathbf{x}_i(t+1) = f_i(\ \vec{\mathbf{x}}(t)\ ) \ , \ f_i \in \mathbf{F}$$

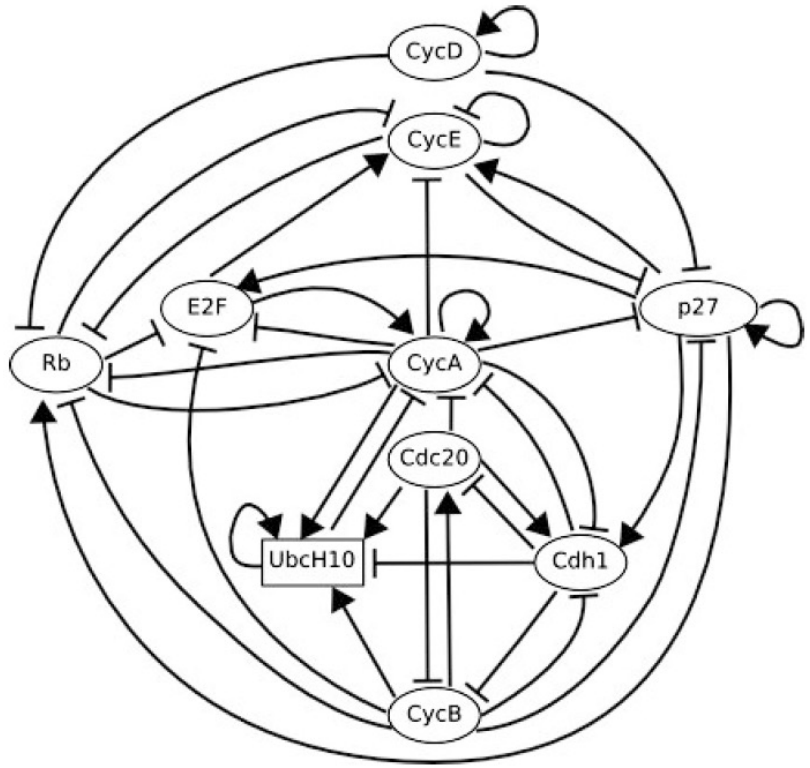


Fig. 1. Logical regulatory graph for the mammalian cell cycle network. Each node represents the activity of a key regulatory element, whereas the edges represent cross-regulations. Blunt arrows stand for inhibitory effects, normal arrows for activations.

State graph

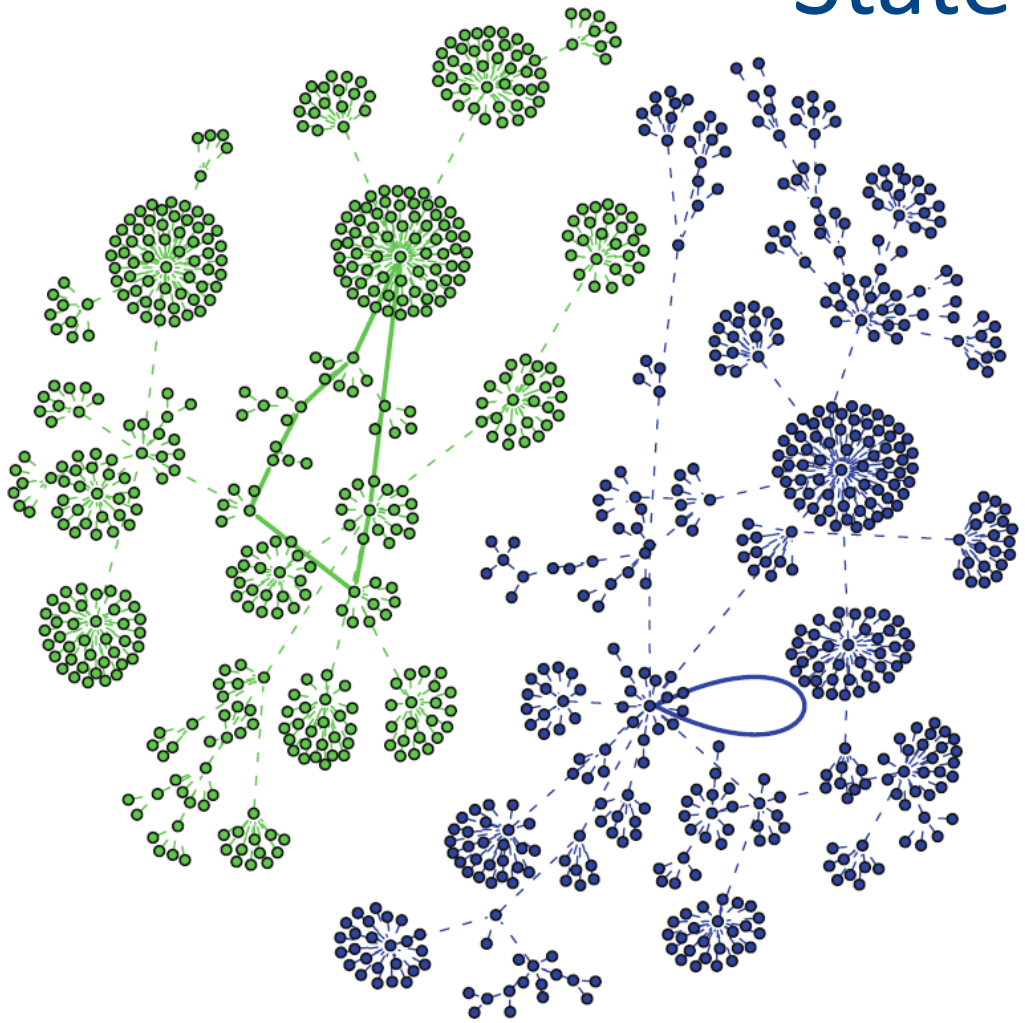
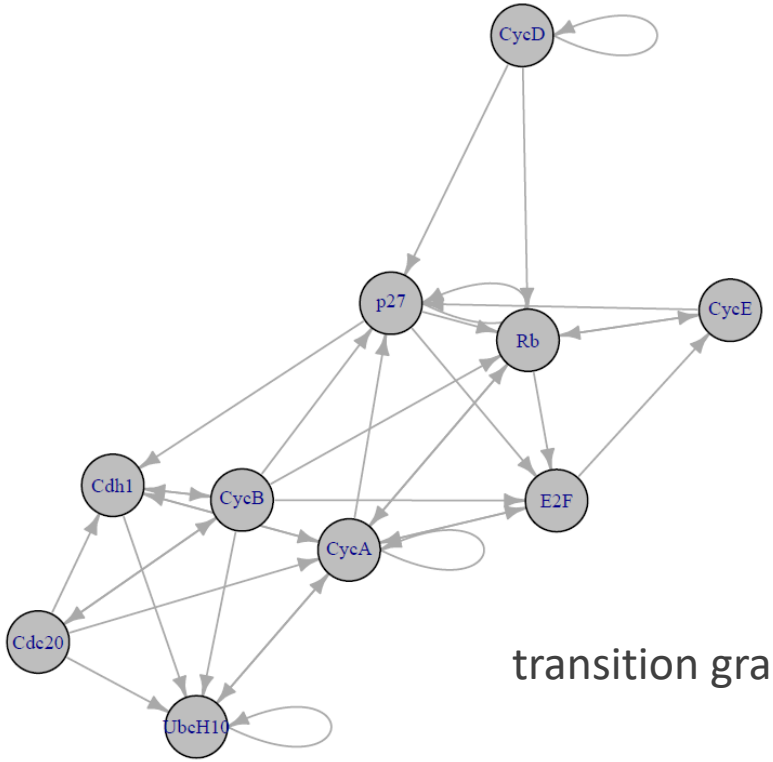
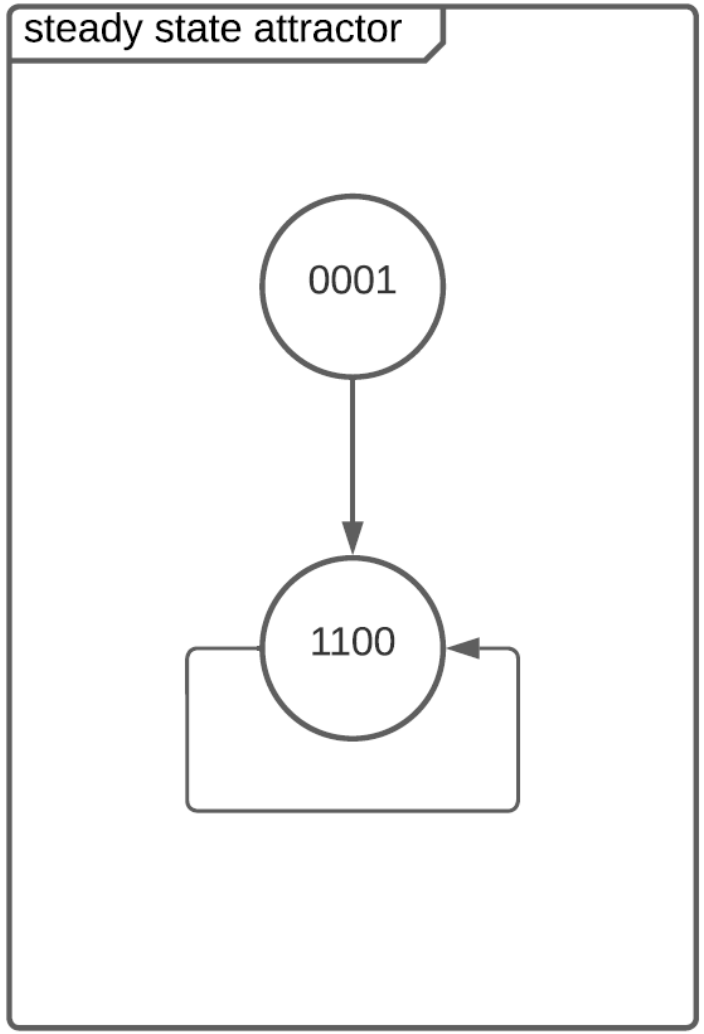


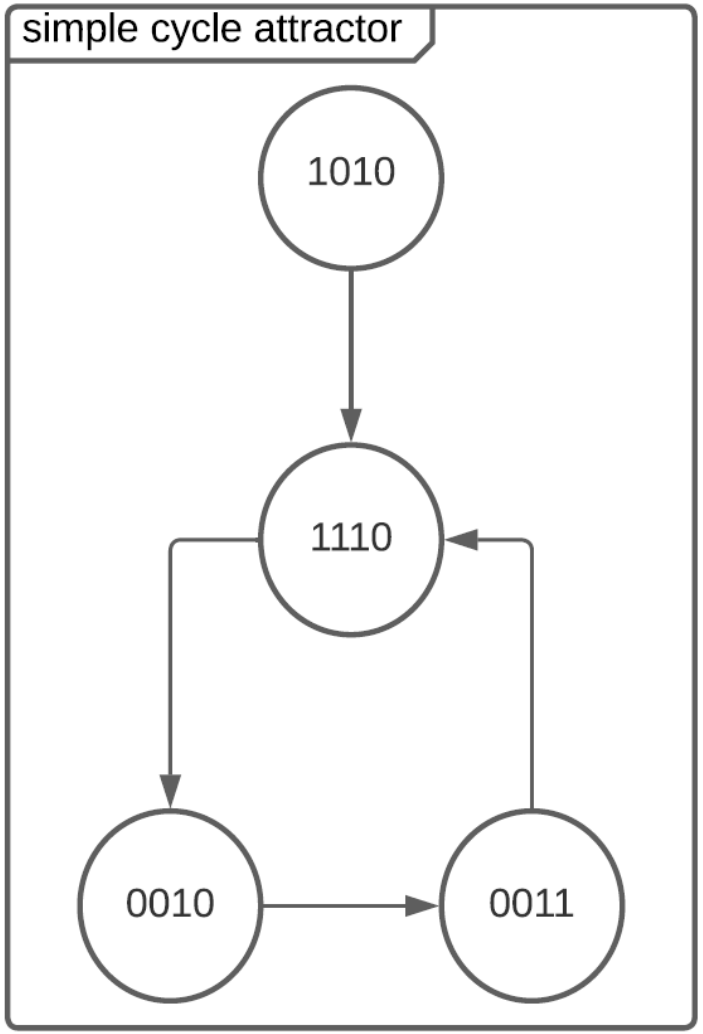
Fig. 3 The state graph of the mammalian cell cycle network. Each *node* represents a state of the network, and each *arrow* is a state transition. The *colors* mark different basins of attraction. Attractors are highlighted using *bold lines* (color figure online)



transition graph

Attractors


steady state attractor
0001
1100


simple cycle attractor
1010
1110
0010
0011

Attractors correspond directly to phenotypes

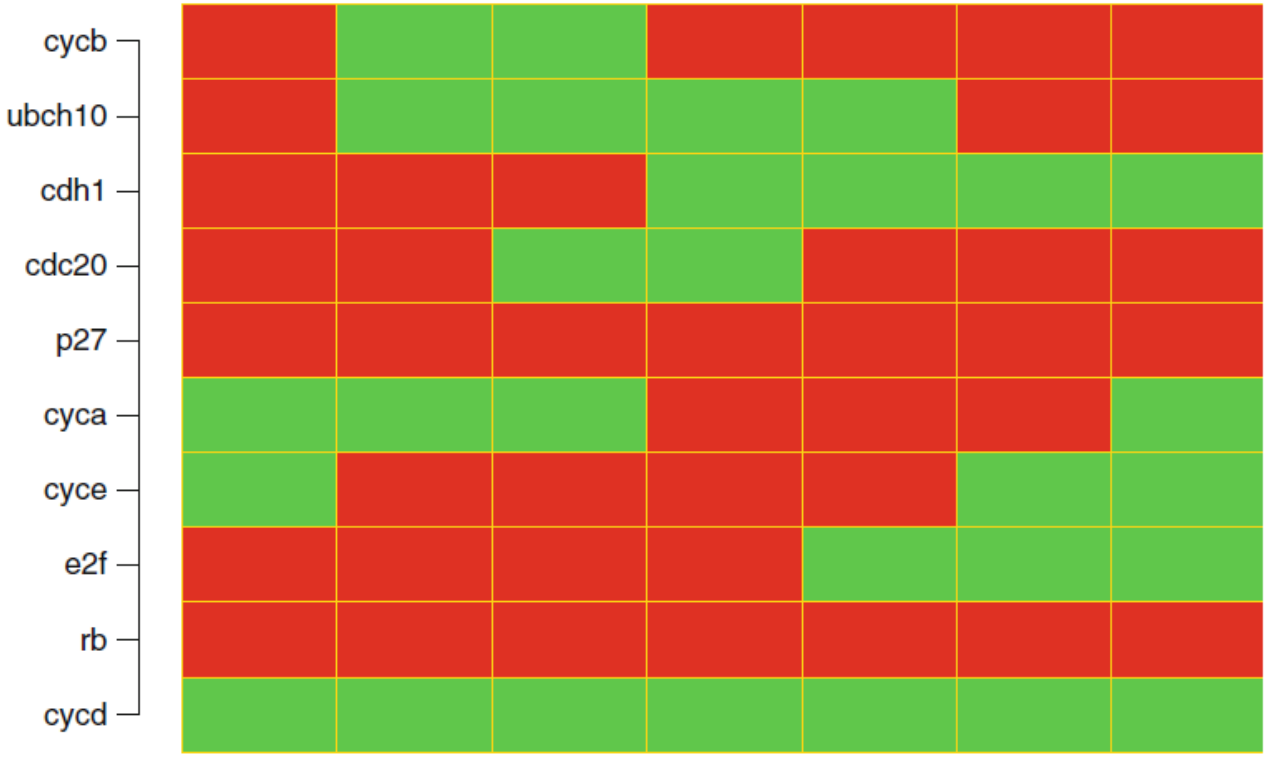


Fig. 4 Visualization of the state changes of the seven-state attractor of the mammalian cell cycle network. The columns of the table represent consecutive states of the attractor. *Green cells* denote an active gene at a certain point of time, whereas *red genes* denote an inactive gene (color figure online)

attractor search

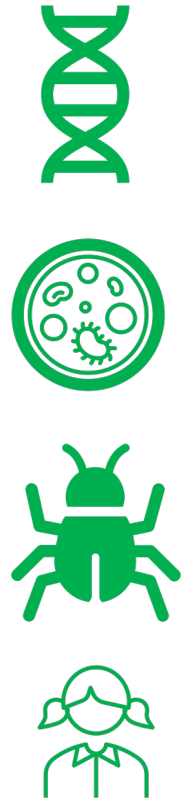
- computationally expensive
- NP complete

Example:
64 genes
=> 2^{64} = 18446744073709551616
exhausting search: ~ 600 years
(calculating: 100,000 states per second)

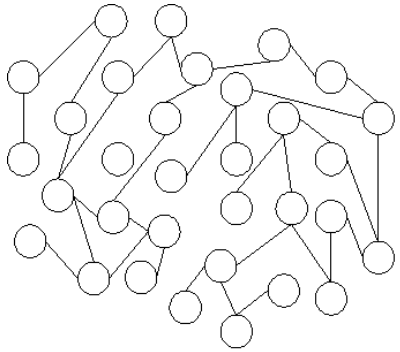
BN raise the Question:
Which properties must a GRN have to enable life?

biological networks should be:

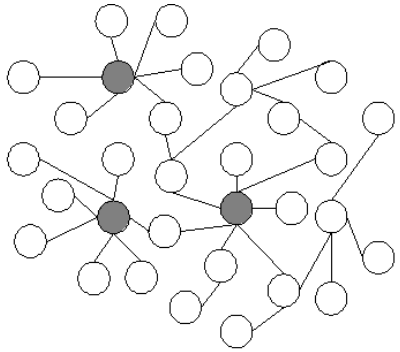
- robust against perturbations
  - e.g. errors, environment, poison, mutations
- flexible enough to allow adaption



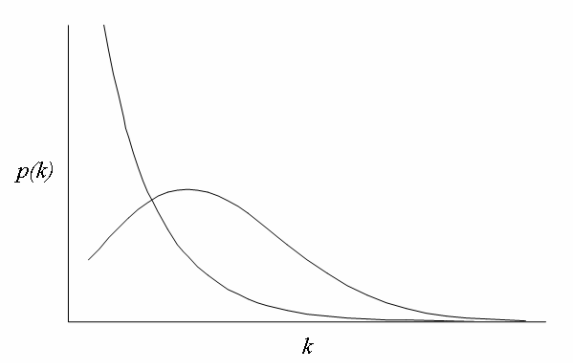
Scale-Free Network



(a) Random network



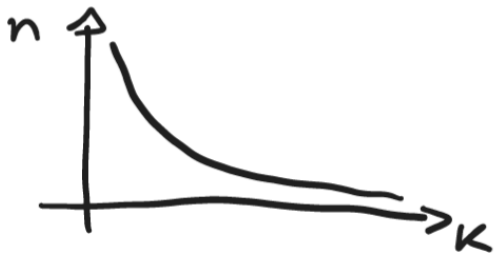
(b) Scale-free network

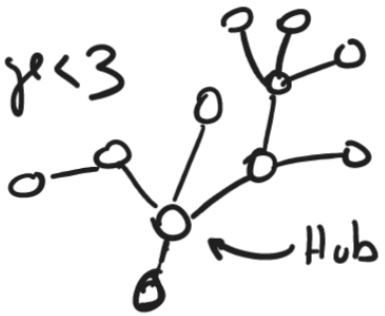


$\frac{N(k)}{n}$; $k = 1, 2, 3, 4 \ldots$

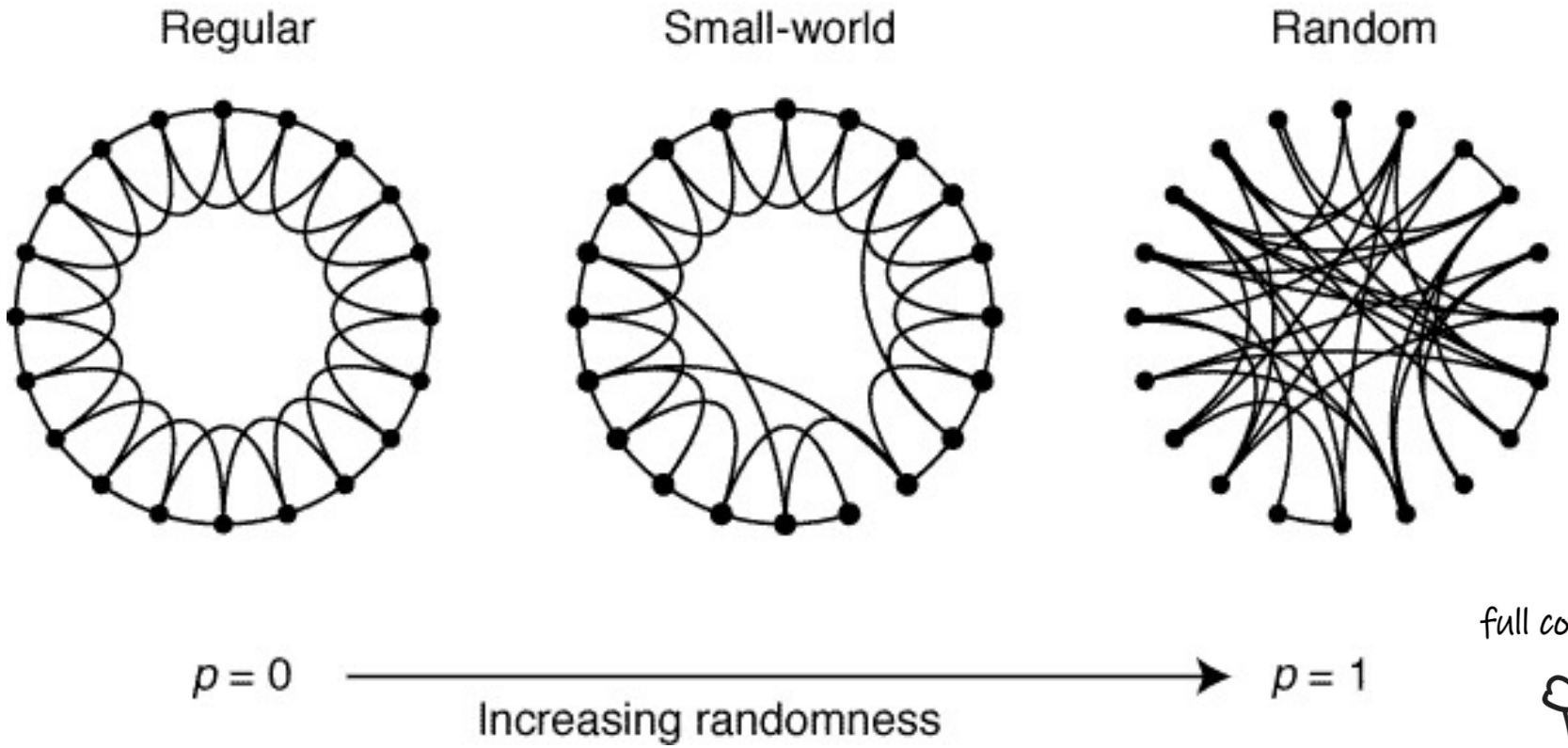
fraction of nodes with degree k

$P(k) \propto k^{-\gamma}$ $2 < \gamma < 3$

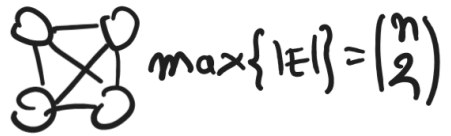




Small-world



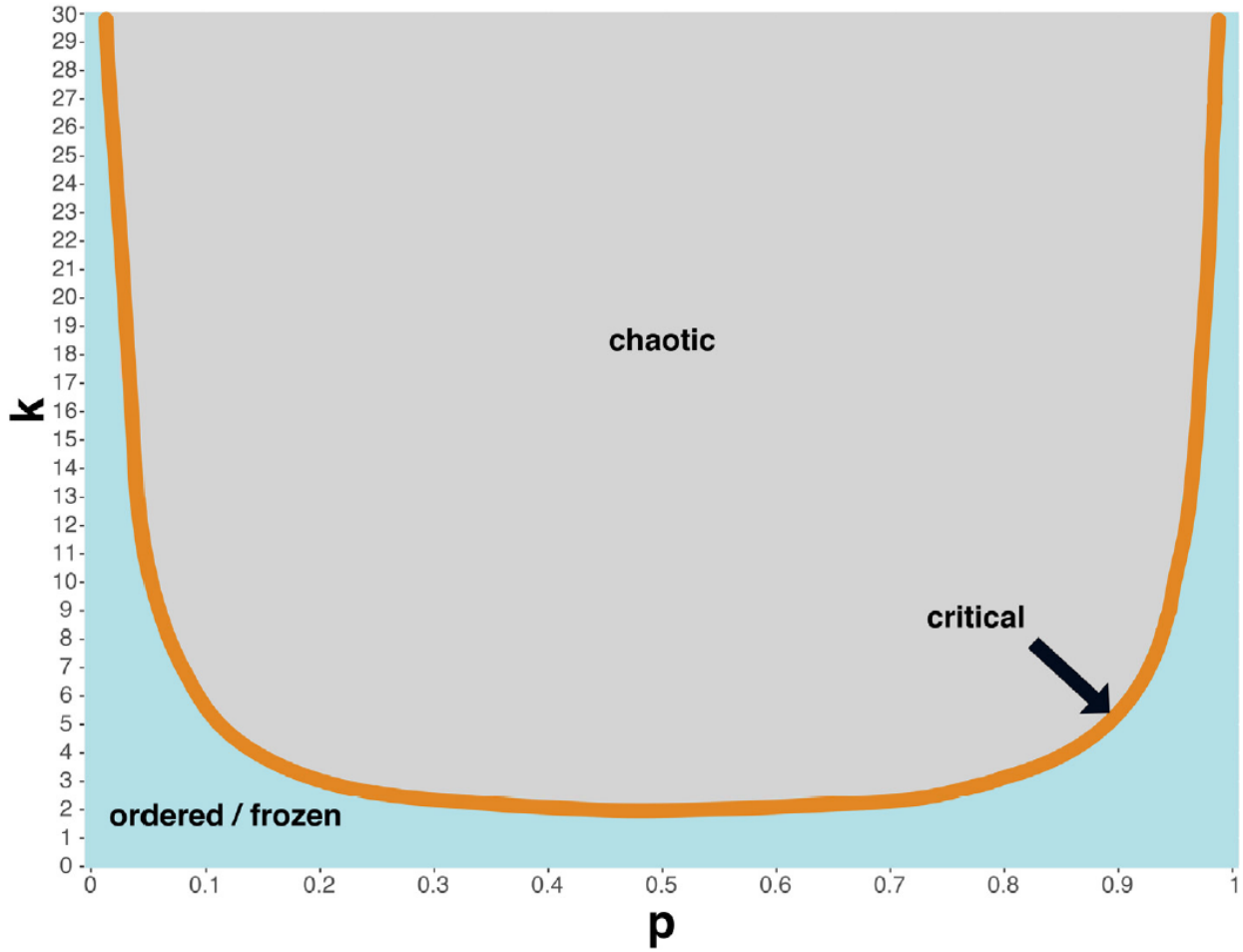
full connectet graph



clustering coefficent

averaged shortest path

“Living at the edge of chaos”

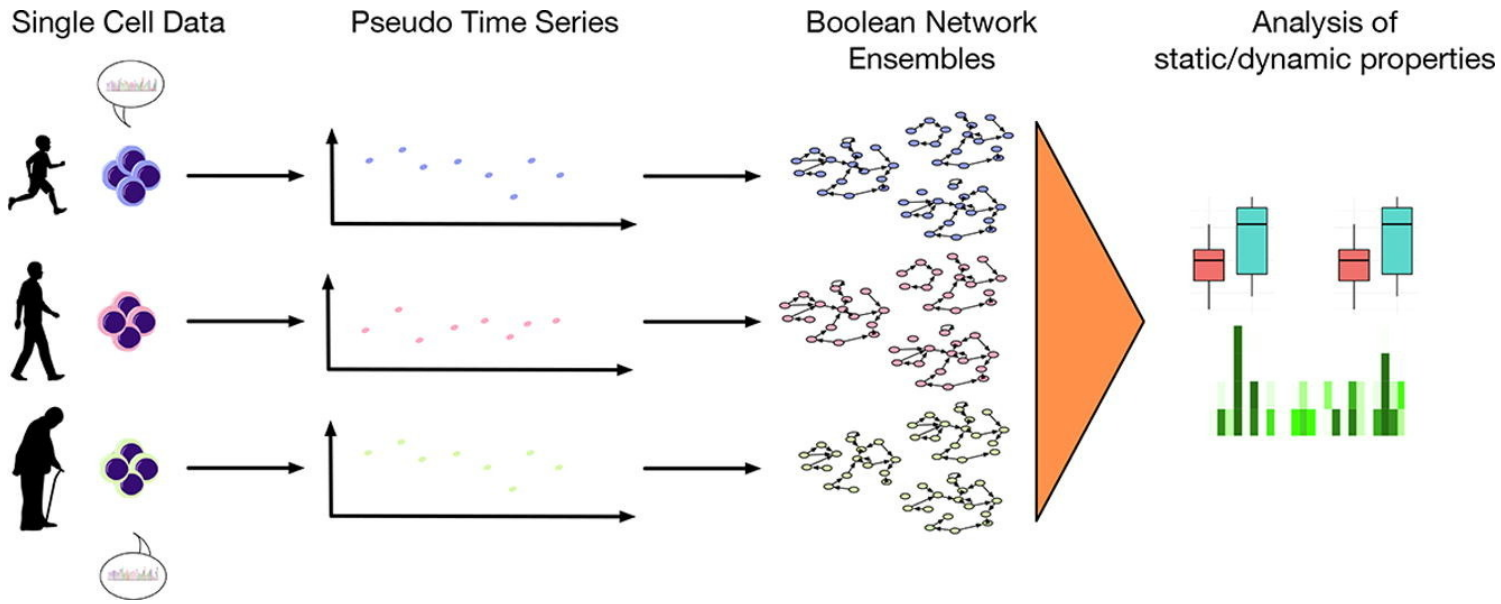


NK Networks
- n genes
- k connection (averaged)
- p (bias percentage of “1”s in truth table)

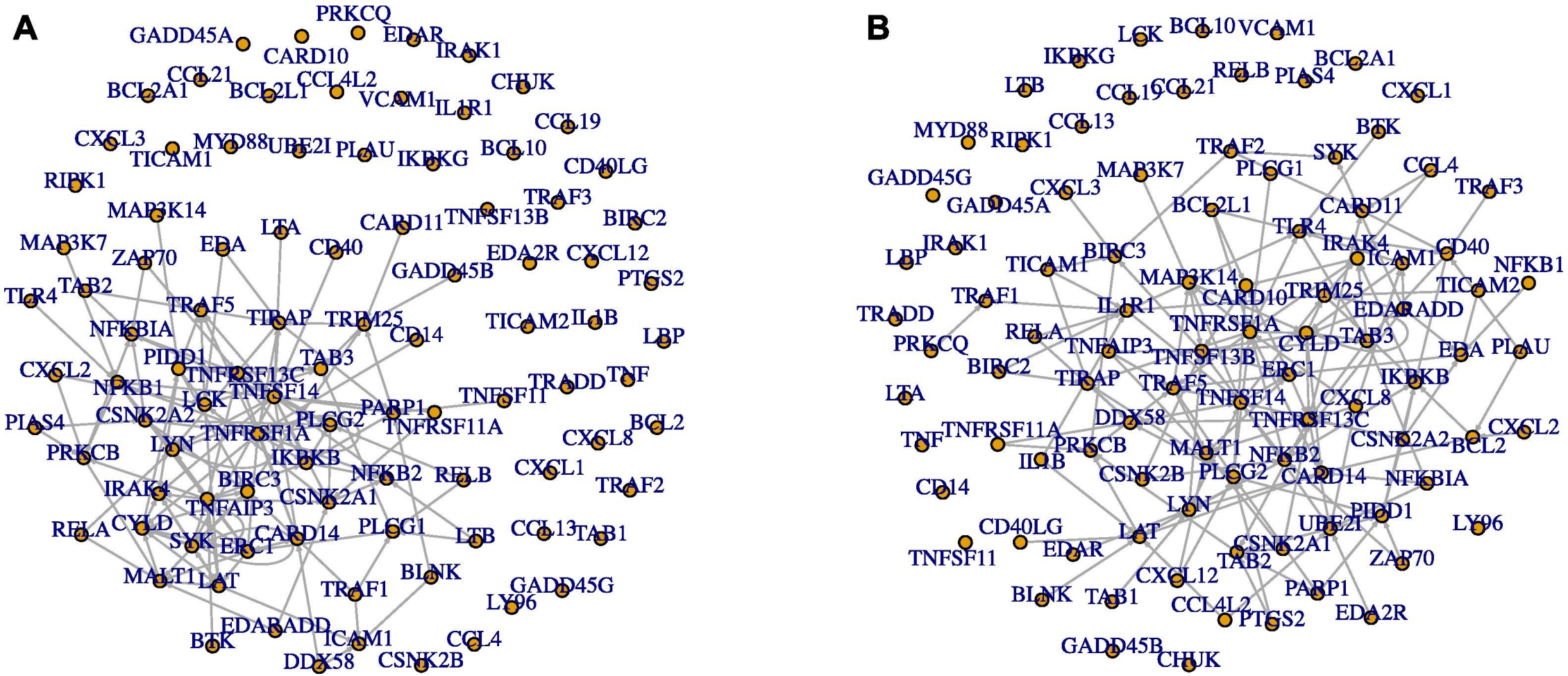
Fig. 5. Regimes of random Boolean networks. The three different regimes of random Boolean networks depending on the two parameters k and p. If $2 \cdot k \cdot p \cdot (1 - p) > 1$ the networks belong to the critical regime, if < 1 to the ordered regime. Networks in between are in the critical regime or at the “edge of chaos”. For $p = 0.5$ networks with $k = 2$ are considered to be at the “edge of chaos”.

Automatically referring from single-cell RNA sequencing data

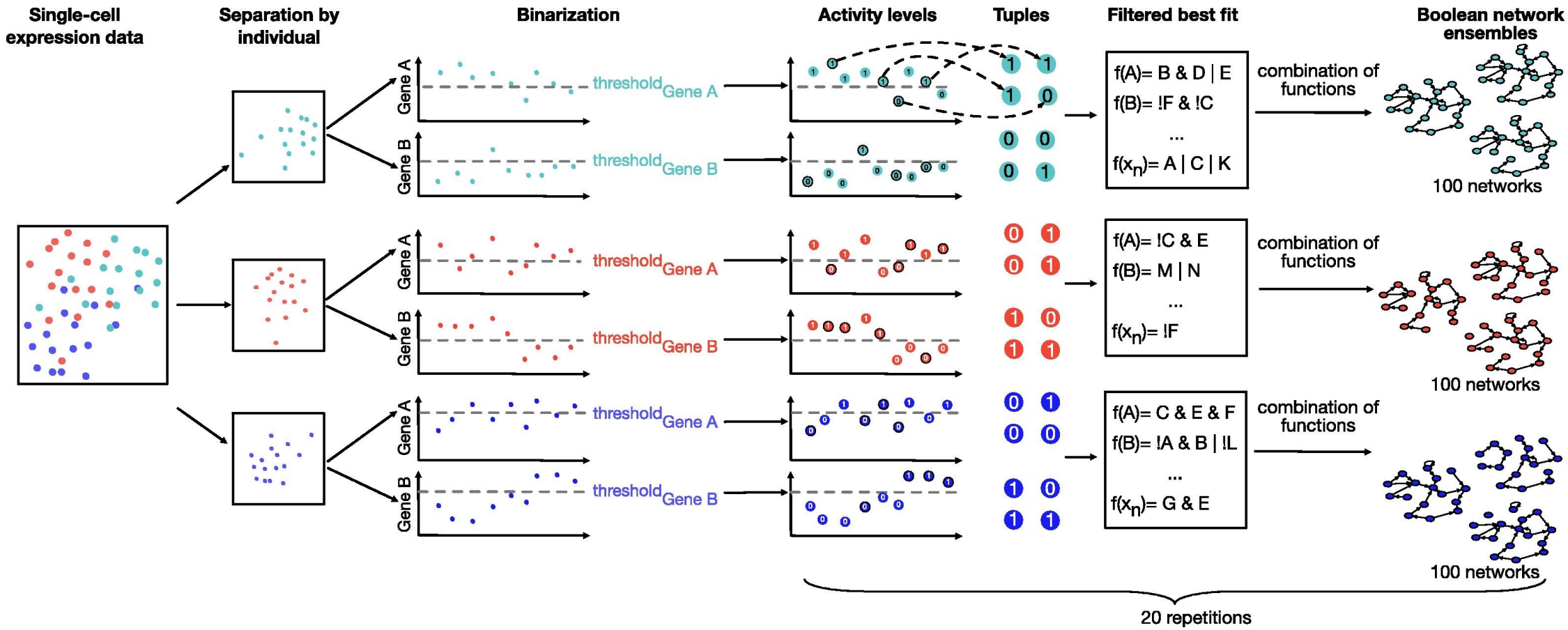
- single cell scRNA-seq data of human hematopoietic stem cells (HSCs)
- aging-associated NF-κB signalling pathway (hsa04064)
- Upon aging, HSCs are increased in number, but their activity is highly heterogeneous and impaired. This impaired function of aged HSCs might influence the immune system (AAIR) and is likely to be associated with leukemia
- ~101 genes

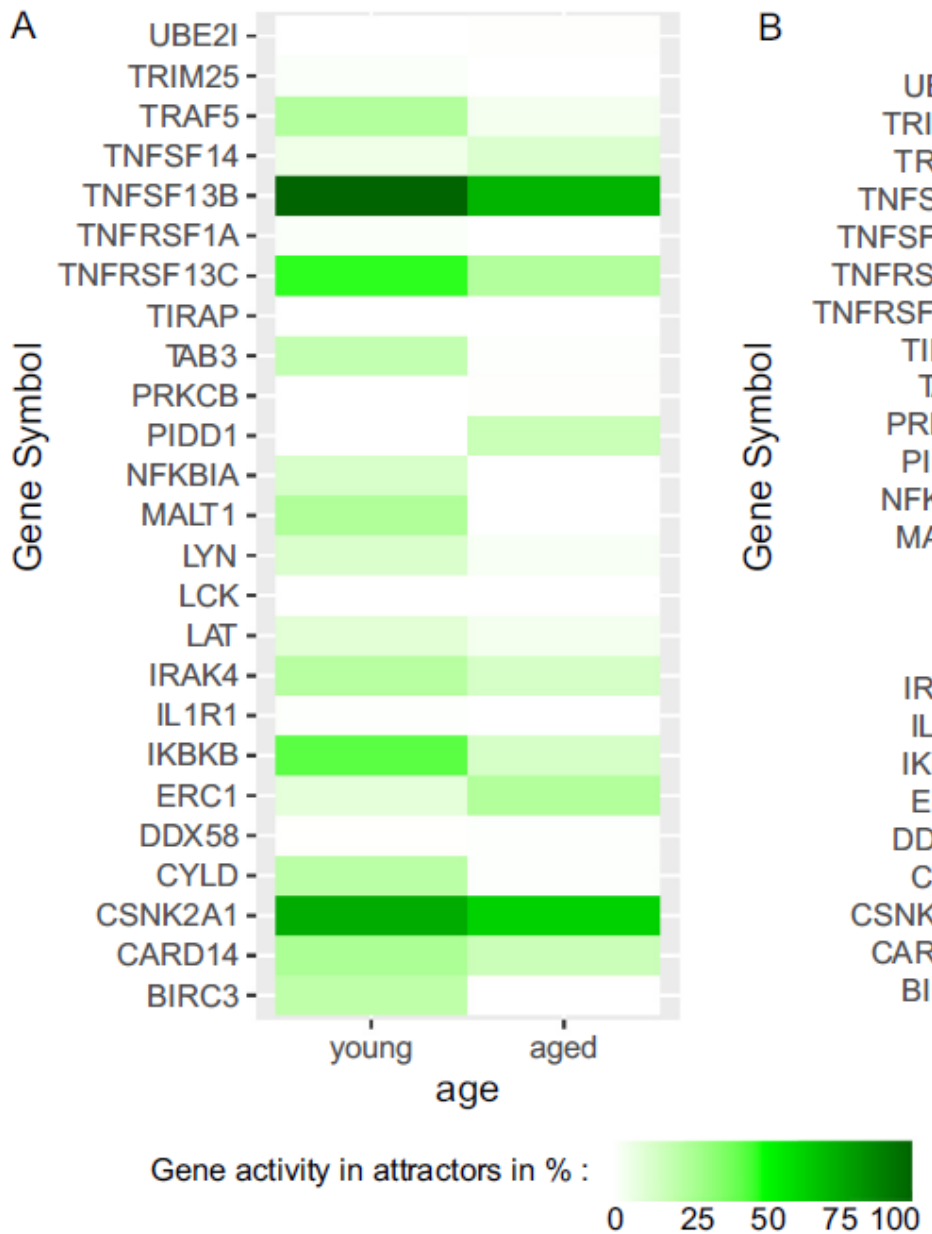


NF-κB signalling pathway (hsa04064)



Automatically referring from single-cell RNA sequencing data





Activity of genes in the aging process.

Activity levels during aging young → aged	Compounds with altering in the Boolean network ensembles	Biological impact of altered activity	References
	BIRC3, IRAK4	Loss of rescue mechanisms for apoptosis causing impaired survival	[68,69]
	CARD14, IKBKB, CYLD	Loss of quiescence	[70-76]
	TRAF5, LAT	Myeloid skewing	[77-80]
	MALT1	Compensatory inhibition of NF-κB activation via induction of quiescence. Loss of NF-κB direct inhibition	[61,62]
	NFKBIA	Propensity to increased inflammatory response via NF-κB activation	[81]
	LYN	Impaired repopulation potential	[82,83]
	ERC1	Cellular polarity alteration causing reduced motility. Presence of ERC1 promotes the turnover of focal adhesions	[84-88]
	TNFRSSF13C	Impaired lymphoid specification	[89,90]
	TNFSF13B	Stable immune response	[63,64]
	CSNK2A1	Resistance to senescence	[65]
	PIDD1	Control of balance between repair and apoptosis after DNA-damage	[66,91]
	TNFSF14	Loss of quiescence due to increased cycling	[92]

Conclusion

Gene regulation

- high complexity
- massive growing data and knowledge

=> computable **gene regulator network** models a necessary

GRN are of highest interest for

- better understanding of pharmacodynamics
- finding new drug targets
- general understanding of life

Boolean networks

- have a long history (about 50 years)
- in the past focus on artificial networks

=> fascinating insights in theoretical biology

become an absolute **hot topic** in recent years

this was made possible by breakthroughs in numerous areas:

- growing interest in GRN as drug targets
- Next Generation Sequencing (NGS)
- growing computing power
- **new sophisticated algorithms**

Literature

[1] Julian D. Schwab, Silke D. Kühlwein, Nensi Ikonomi, Michael Kühl, Hans A. Kestler, Concepts in Boolean network modeling: What do they all mean?, Computational and Structural Biotechnology Journal, Volume 18, 2020, Pages 571-582, ISSN 2001-0370, https://doi.org/10.1016/j.csbj.2020.03.001. (https://www.sciencedirect.com/science/article/pii/S200103701930460X

[2] HOPFENSITZ, Martin, et al. Attractors in Boolean networks: a tutorial. Computational Statistics, 2013, 28. Jg., Nr. 1, S. 19-36. https://link.springer.com/article/10.1007/s00180-012-0324-2

[3] Réka Albert; Scale-free networks in cell biology. *J Cell Sci* 1 November 2005; 118 (21): 4947–4957. doi: https://doi.org/10.1242/jcs.02714

[4] Adrien Fauré et al. “Dynamical analysis of a generic Boolean model for thecontrol of the mammalian cell cycle”. In: *Bioinformatics* 22.14 (July 2006),e124–e131. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btl210. eprint:https://academic.oup.com/bioinformatics/article- pdf/22/14/e124/614634/btl210.pdf. URL:https://doi.org/10.1093/bioinformatics/btl210.

[5] Christoph Müssel, Martin Hopfensitz, and Hans A. Kestler. “BoolNet—an Rpackage for generation, reconstruction and analysis of Boolean networks”. In: Bioinformatics 26.10 (Apr. 2010), pp. 1378–1380. ISSN: 1367-4803. DOI:10.1093/bioinformatics/btq124. eprint: https://academic.oup.com/bioinformatics/article.pdf/26/10/1378/16892352/btq124.pdf. URL:https://doi.org/10.1093/bioinformatics/btq124.

[6] WATTS, Duncan J.; STROGATZ, Steven H. Collective dynamics of ‘small-world’networks. nature, 1998, 393. Jg., Nr. 6684, S. 440-442. https://www.nature.com/articles/30918#citeas

[7] Fauré A, Naldi A, Chaouiya C, Thieffry D. Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. Bioinformatics. 2006 Jul 15;22(14):e124-31. doi: 10.1093/bioinformatics/btl210. PMID: 16873462. https://pubmed.ncbi.nlm.nih.gov/16873462/

[8] Julian D. Schwab, Nensi Ikonomi, Silke D. Werle, Felix M. Weidner, Hartmut Geiger, Hans A. Kestler, Reconstructing Boolean network ensembles from single-cell data for unraveling dynamics in the aging of human hematopoietic stem cells, Computational and Structural Biotechnology Journal, Volume 19,2021,Pages 5321-5332,ISSN 2001-0370,https://doi.org/10.1016/j.csbj.2021.09.012. (https://www.sciencedirect.com/science/article/pii/S2001037021003974)

[9] Lähdesmäki, H., Shmulevich, I. & Yli-Harja, O. On Learning Gene Regulatory Networks Under the Boolean Network Model. Machine Learning 52, 147–167 (2003). https://doi.org/10.1023/A:1023905711304

Literature

[10] Yang, Y., Li, S., Wang, Y. et al. Protein tyrosine kinase inhibitor resistance in malignant tumors: molecular mechanisms and future perspective. Sig Transduct Target Ther 7, 329 (2022). https://doi.org/10.1038/s41392-022-01168-8

[11] Burkhard K, Shapiro P. Use of inhibitors in the study of MAP kinases. Methods Mol Biol. 2010;661:107-22. doi: 10.1007/978-1-60761-795-2_6. PMID: 20811979; PMCID: PMC3727897. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3727897/

main analytics

Questions…