CMACS Pancreatic Cancer Challenge

Edmund Clarke James Faeder Haijun Gong Justin Jee
Ilya Korsunsky Bud Mishra Natasa Miskov-Zivanov
Loes Olde Loohuis Andreas Witzel Tongtong Wu Paolo Zuliani

November 2011

Ilya Korsunsky
NYU

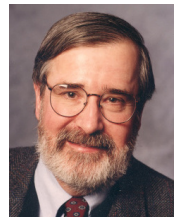
Justin Jee
NYU

Haijun Gong
CMU

James Faeder
U. Pittsburgh

Bud Mishra
NYU

CMACS Pancreatic Cancer Challenge



Ed Clarke
CMU

Natasa
Miskov-Zivanov
U. Pittsburgh

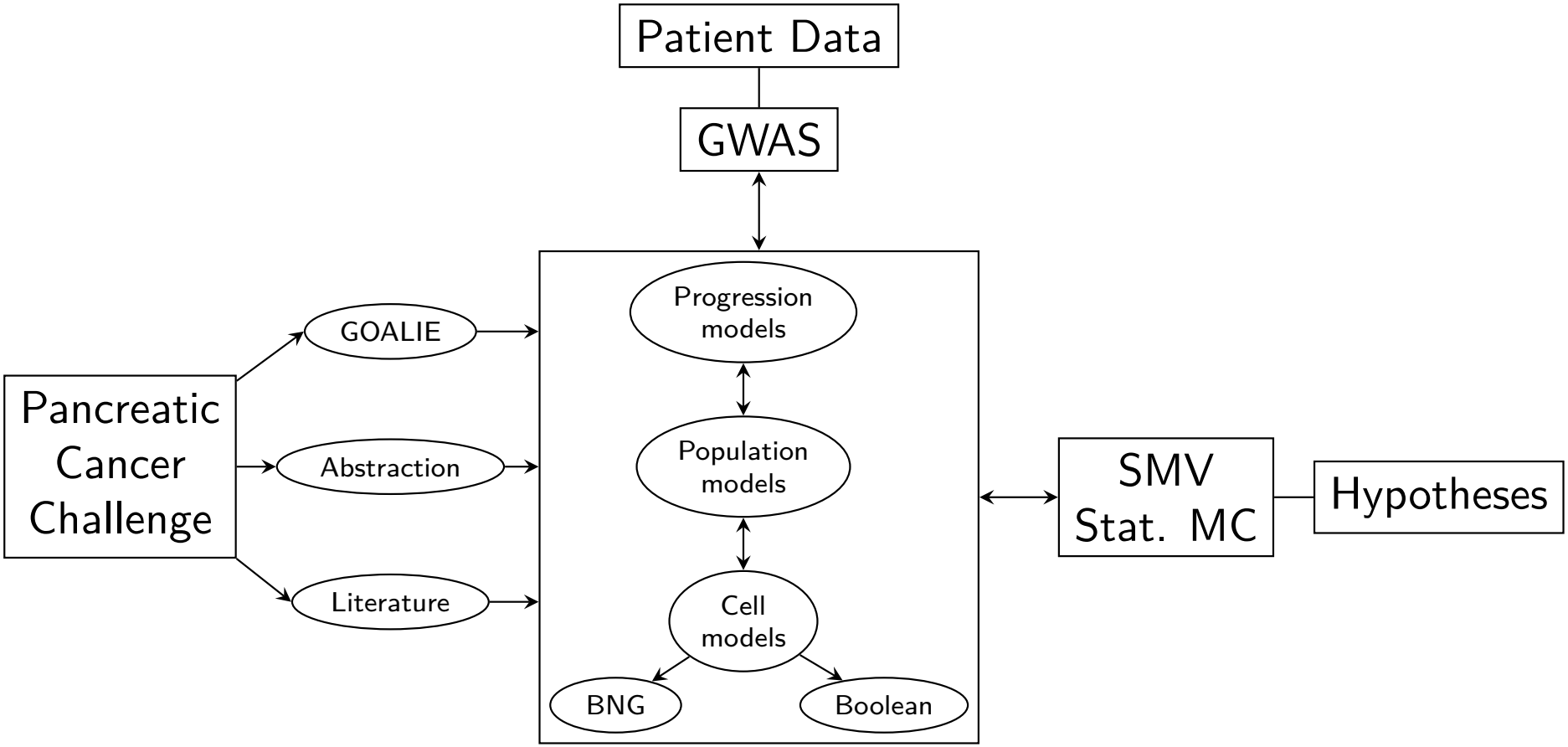
Loes Olde
Loohuis
NYU

Andreas Witzel
NYU

Tongtong Wu
U. Maryland

Paolo Zuliani
CMU

Mission Statement



Build, Reason, Predict, and Manipulate
Models of Pancreatic Cancer
Spanning Molecular, Cellular, Organ, and Population Levels

Outline

Outline

Algorithmic Foundations

- BioNetGen
  -  M. W. Sneddon, **J. R. Faeder**, T. Emonet. *Efficient modeling, simulation and coarse-graining of biological complexity with NFsim*, Nature Methods, Vol. 8, No. 2, 2011.
- Boolean Models
  -  **H. Gong**, **P. Zuliani**, **E. M. Clarke**. *Model Checking of a Diabetes-Cancer Model*, 3rd International Symposium on Computational Models for Life Sciences, 2011.
- Statistical Model Checking
  -  **E. M. Clarke**, **J. R. Faeder**, C. Langmead, L. Harris, S. Jha, A. Legay. *Statistical model checking in biolab: Applications to the automated analysis of t-cell receptor signaling pathway*, Computational Methods in Systems Biology, 2008.

Algorithmic Foundations

- Models from Data
  - Mechanistic
    -  S. Ryu, S. Lin, N. Ugel, M. Antoniotti, **B. Mishra**. *Mathematical modeling of the formation of apoptosome in intrinsic pathway of apoptosis*, Systems and Synthetic Biology Journal, vol. 2, no. 1–2, 2009.
  - Phenomenological
    -  N. Ramakrishnan, S. Tadepalli, L. T. Watson, R. F. Helm, M. Antoniotti, **B. Mishra**. *Reverse Engineering Dynamic Temporal Models of Biological Processes and their Relationships,"* , Proc. National Academy of Science, vol. 107, no. 28, 2010.

Algorithmic Foundations

- Hybrid Model Checking

C. Piazza, M. Antoniotti, V. Mysore, A. Policriti, F. Winkler, **B. Mishra**. *Algorithmic Algebraic Model Checking I: Challenges from Systems Biology*, 17th International Conference on Computer Aided Verification, 2005.

- Supervisory Control

L. Olde Loohuis, **A. Witzel**, **B. Mishra**. *Cancer Hallmark Automata*, manuscript, 2011.

E. Asarin, O. Maler, **A. Pnueli**. *Symbolic controller synthesis for discrete and timed systems*, Hybrid Systems II, 1995.

Outline

Cancer as a Disease of the Genome

- Oncogenes / Tumor Suppressor Genes

Cancer as a Disease of the Genome

- Oncogenes / Tumor Suppressor Genes
- Cancer pathways

Cancer as a Disease of the Genome

- Oncogenes / Tumor Suppressor Genes
- Cancer pathways
- Cancer phenotypes and progression (hallmarks)

Cancer as a Disease of the Genome

- Oncogenes / Tumor Suppressor Genes
- Cancer pathways
- Cancer phenotypes and progression (hallmarks)
- Patient data and personalization

Cancer as a Disease of the Genome

- Oncogenes / Tumor Suppressor Genes
- Cancer pathways
- Cancer phenotypes and progression (hallmarks)
- Patient data and personalization
  - The Cancer Genome Atlas, GOALIE, statistical analysis

Cancer as a Disease of the Genome

- Oncogenes / Tumor Suppressor Genes
- Cancer pathways
- Cancer phenotypes and progression (hallmarks)
- Patient data and personalization
  - The Cancer Genome Atlas, GOALIE, statistical analysis
- Model checking on different levels of abstraction

Cancer as a Disease of the Genome

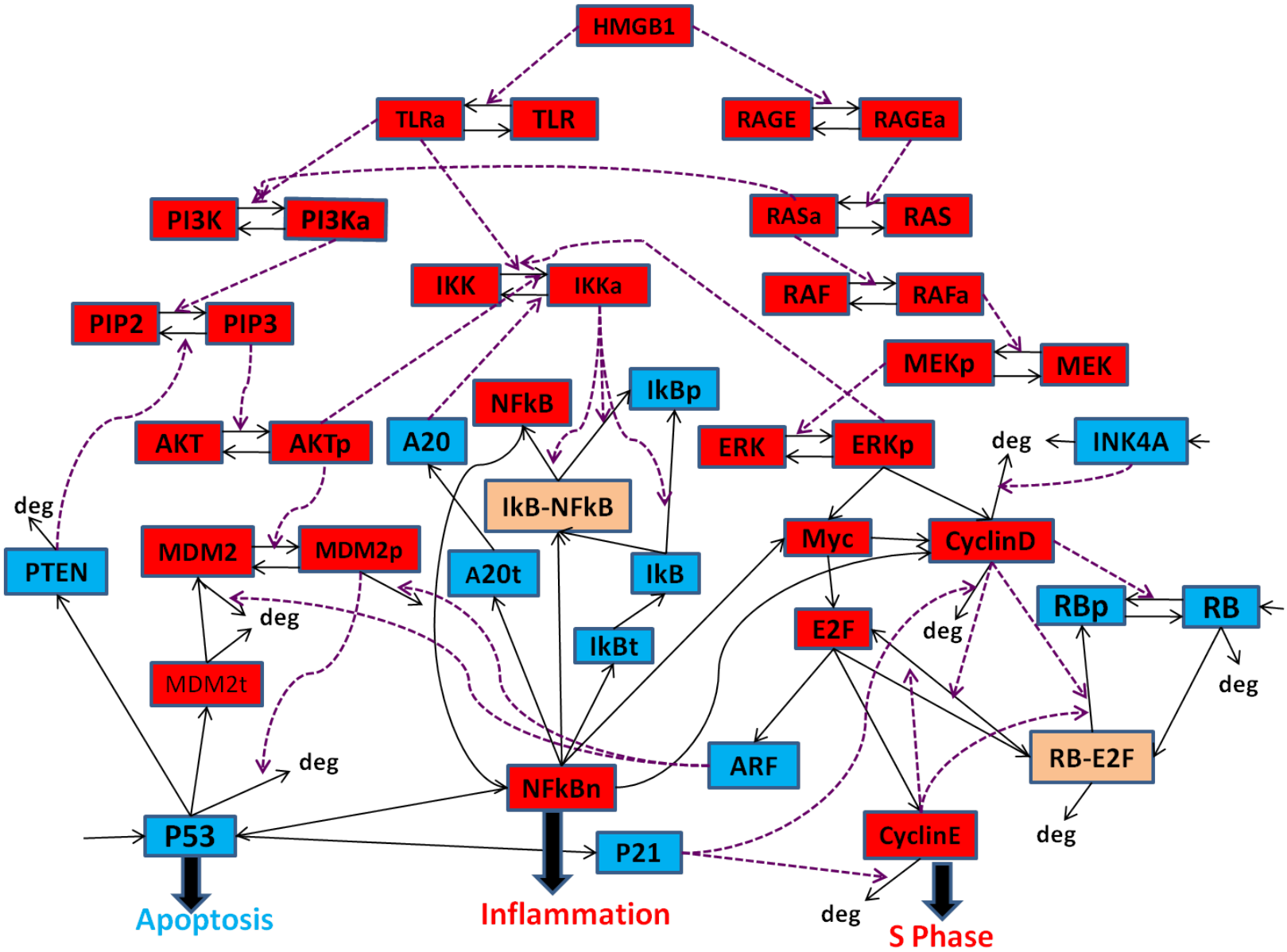
- Oncogenes / Tumor Suppressor Genes
- Cancer pathways
- Cancer phenotypes and progression (hallmarks)
- Patient data and personalization
  - The Cancer Genome Atlas, GOALIE, statistical analysis
- Model checking on different levels of abstraction
- Model-based therapy

Outline

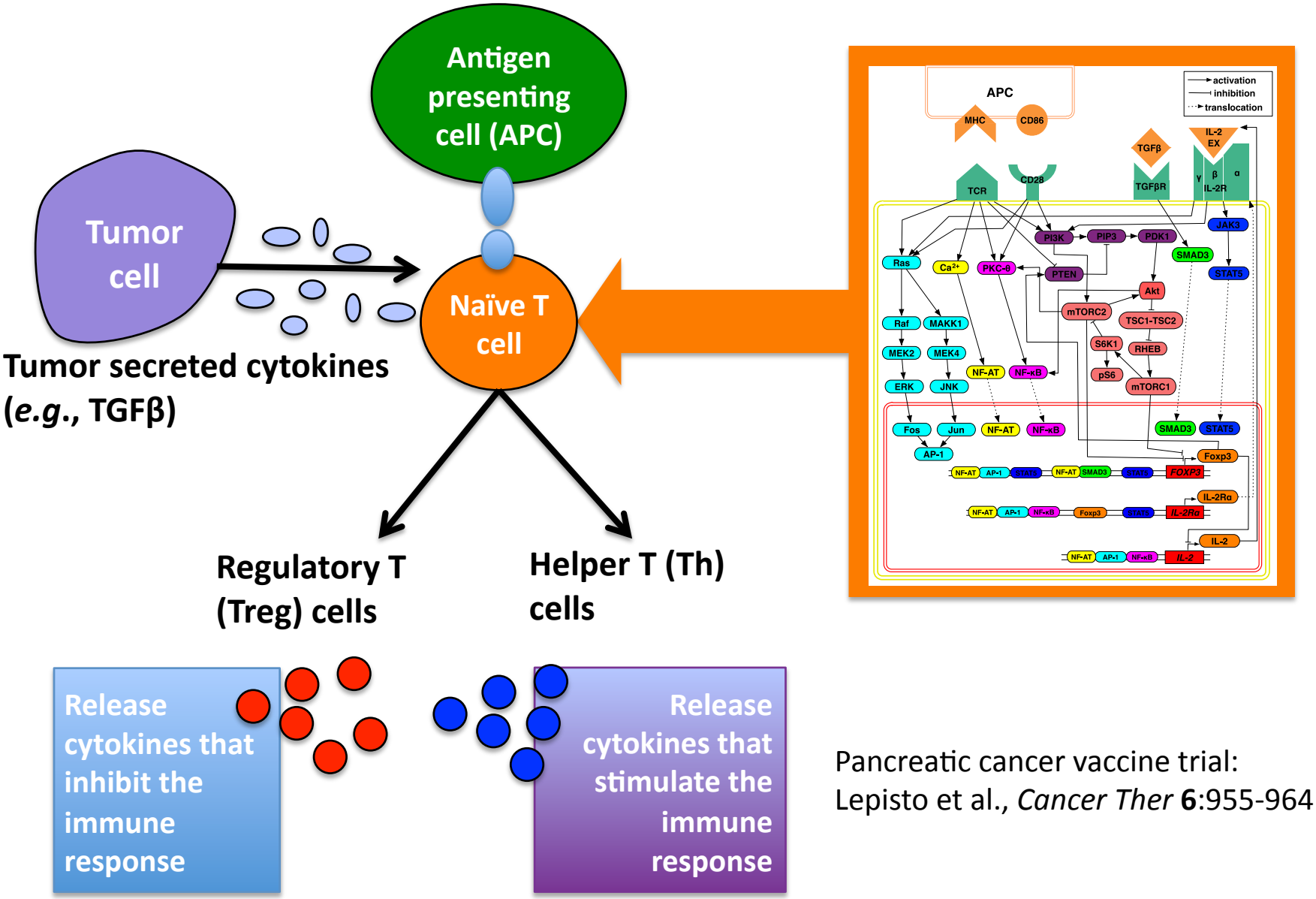
Single-cell Pathways in Cancer

- Cancer can be understood in terms of various cell-autonomous processes: Autophagy, Apoptosis, Mitosis
- There are specific pathways controlling these processes
- We have developed mechanistic models involving these pathways, e.g., ODEs, BioNetGen models, and Boolean models
- Properties of these pathways can be model checked in order to understand them

HMGB1 Model


HMGB1
TLRa
TLR
RAGE
RAGEa
PI3K
PI3Ka
RASa
RAS
IKK
IKKa
RAF
RAFa
PIP2
PIP3
MEKp
MEK
NFkB
IkBp
AKT
AKTp
A20
ERK
ERKp
deg
INK4A
IkB-NFkB
deg
PTEN
MDM2
MDM2p
A20t
IkB
Myc
CyclinD
RBp
RB
E2F
deg
deg
IkBt
MDM2t
deg
deg
ARF
RB-E2F
NFkBn
deg
P53
P21
CyclinE
Apoptosis
Inflammation
deg
S Phase

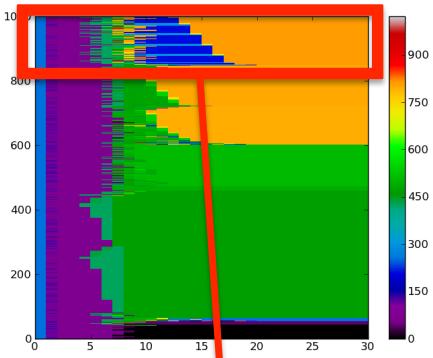
Boolean Model

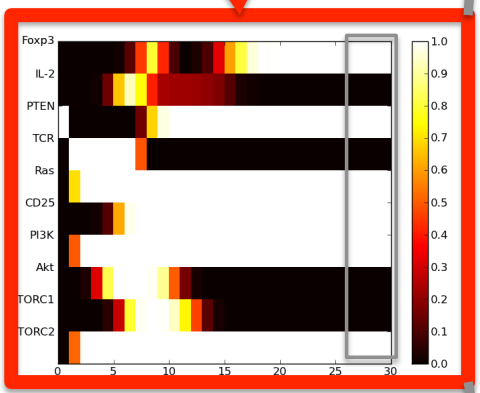


Pancreatic cancer vaccine trial:
Lepisto et al., *Cancer Ther* **6**:955-964 (2008)

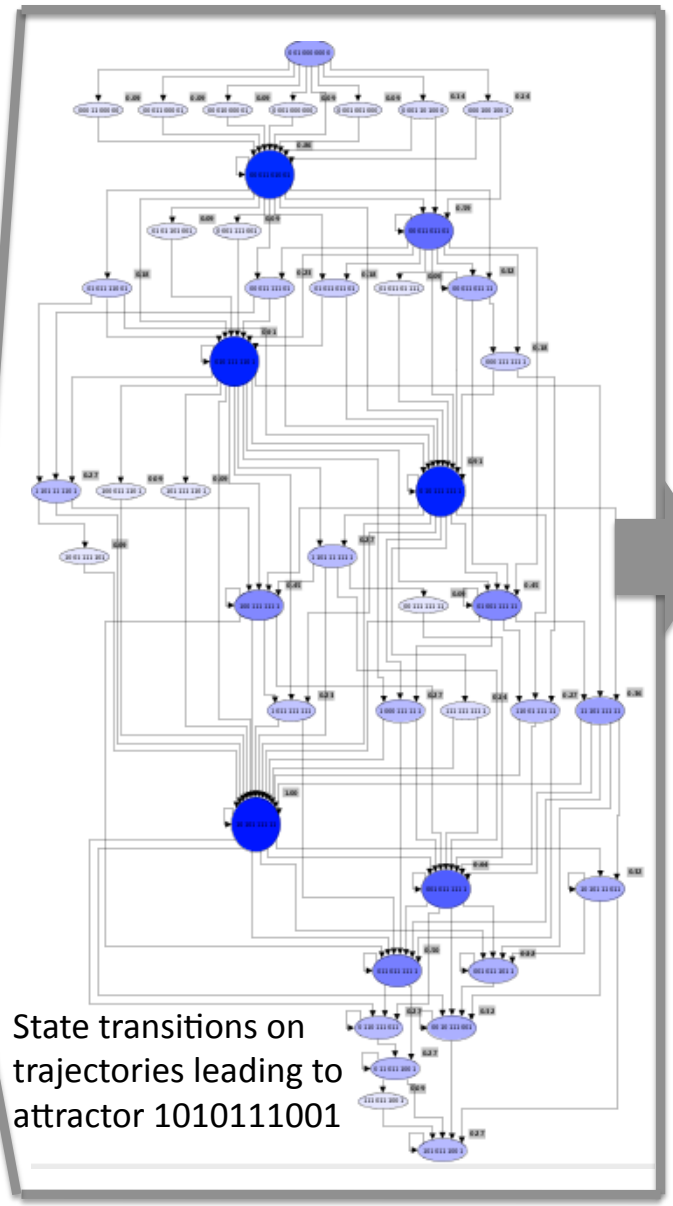
Model Simulation

1000 simulation trajectories



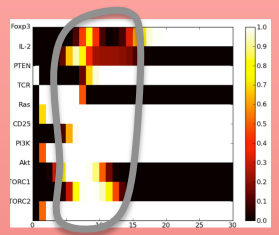


Average trajectories of ten elements for attractor 1010111001



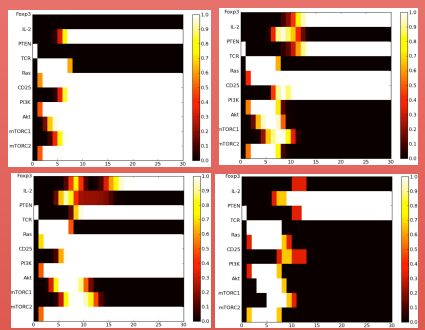
State transitions on trajectories leading to attractor 1010111001

Check properties of the transition diagram (attractor)



- effects of transient changes
- element correlations
- stochasticity in the most connected nodes

Compare and contrast different attractors



Statistical Model Checking

- In English:
  p53 is expressed at low level in normal human cells
- In temporal logic:
  $\text{Prob}_{\geq 0.9}\mathbf{F}^{t}(\mathbf{G}^{900}(\text{p53} < 3.3 \cdot 10^{4}))$
- Verification:

t(min)	# Samples	# Success	Result	Time (s)
400	53	49	True	597.59
500	23	22	True	271.76
600	22	22	True	263.79

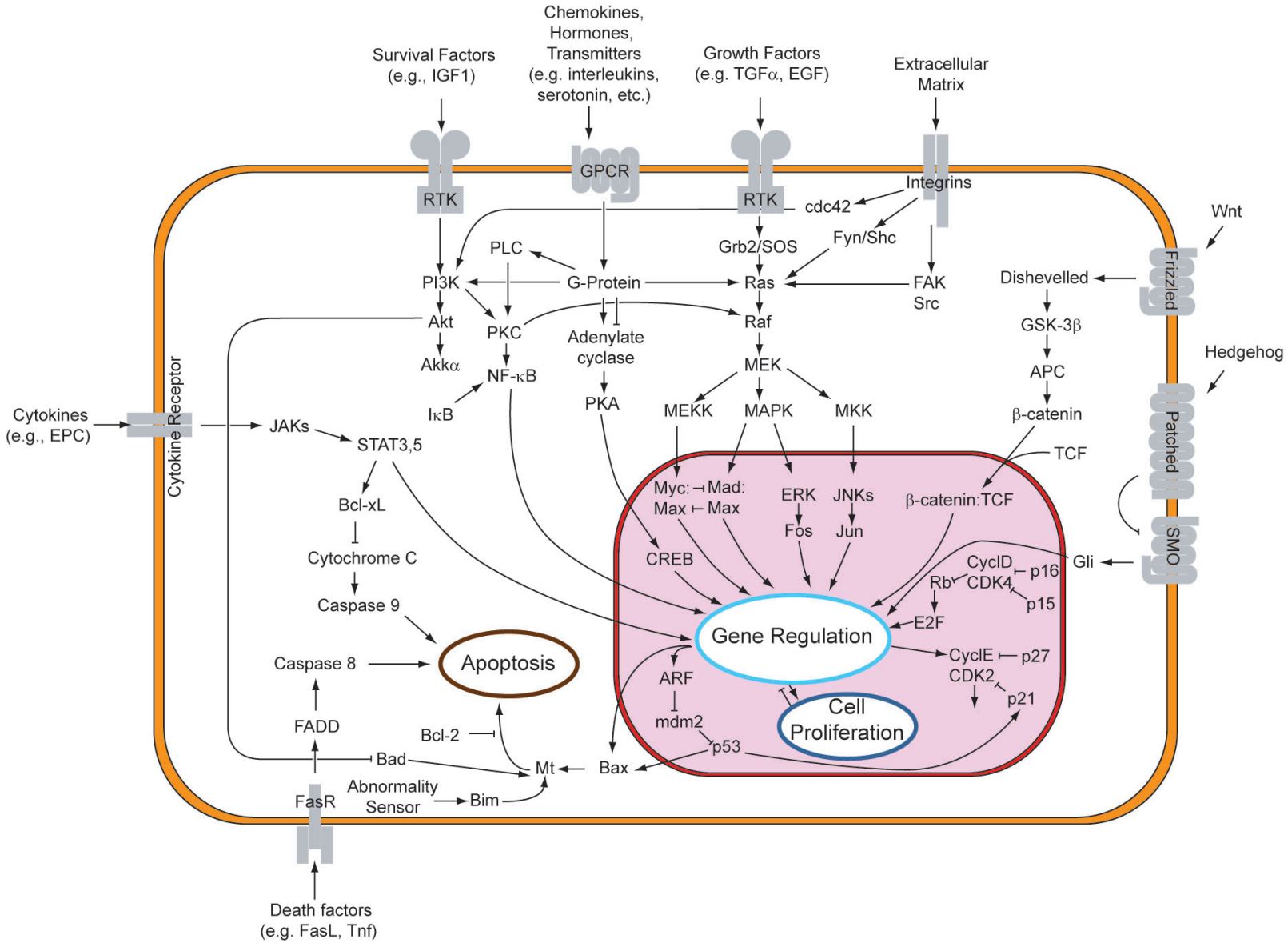
Error probability $= 0.001$

Contribution

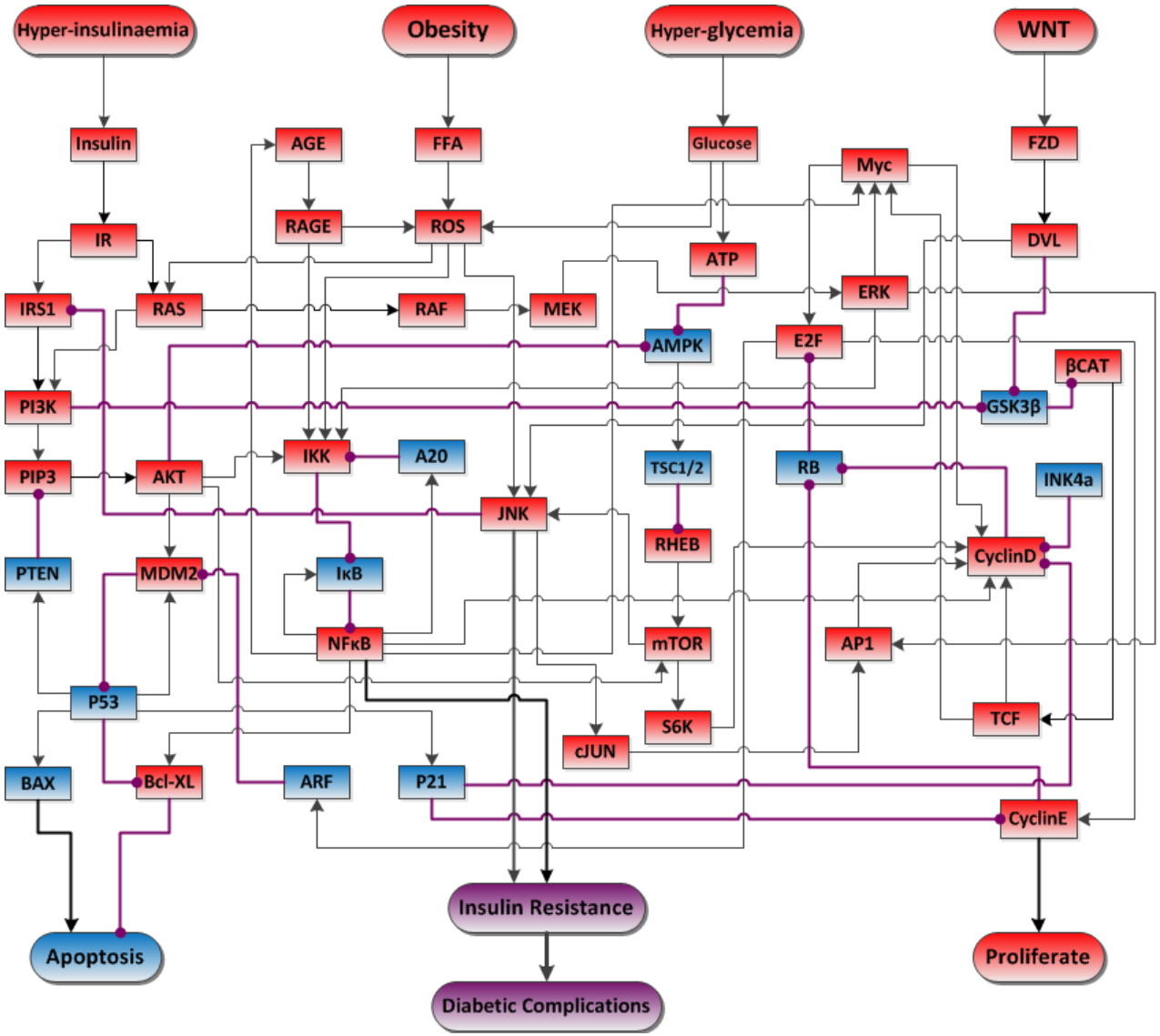
- First computational model for investigating HMGB1 and tumorigenesis; it agrees well with HMGB1 experiments
- Our model suggests a dose-dependent p53, CyclinD/E, NFkB response curve to increasing HMGB1 stimulus
  - this could be tested by future experiments
- The model can provide a guideline for cancer researchers to design new *in vitro* experiments
- Statistical Model Checking automatically validates our model with respect to known experimental results

Outline

Aberrant Inter-cell Signaling



Boolean Network



2^{49} states $\rightarrow$ activation $\multimap$ inhibition

Model Checking

- Do diabetes risk factors influence the risk of cancer or cancer prognosis? We checked the CTL properties:

  (1) **AF**(Proliferate)
  (2) **AF**(Apoptosis)
  (3) **AF**(Resistance)

  (1′) **EF**(Proliferate)
  (2′) **EF**(Apoptosis)
  (3′) **EF**(Resistance)

- Normal Cell:
  - Properties 3 and $2' - 3'$ are true
  - Diabetes risk factors can augment insulin resistance, but cell growth is still regulated by the tumor suppressor proteins
  - Cancer risk might not increase
- Precancerous/Cancerous Cells (INK4a, ARF$= 0$):
  - All but Property 2 are true
  - Diabetes risk factors promote growth in precancerous or cancerous cells and augment insulin resistance

Abstract Signaling Machine (ASM)

ASM simulates few concrete cells in mean field population model

Environment

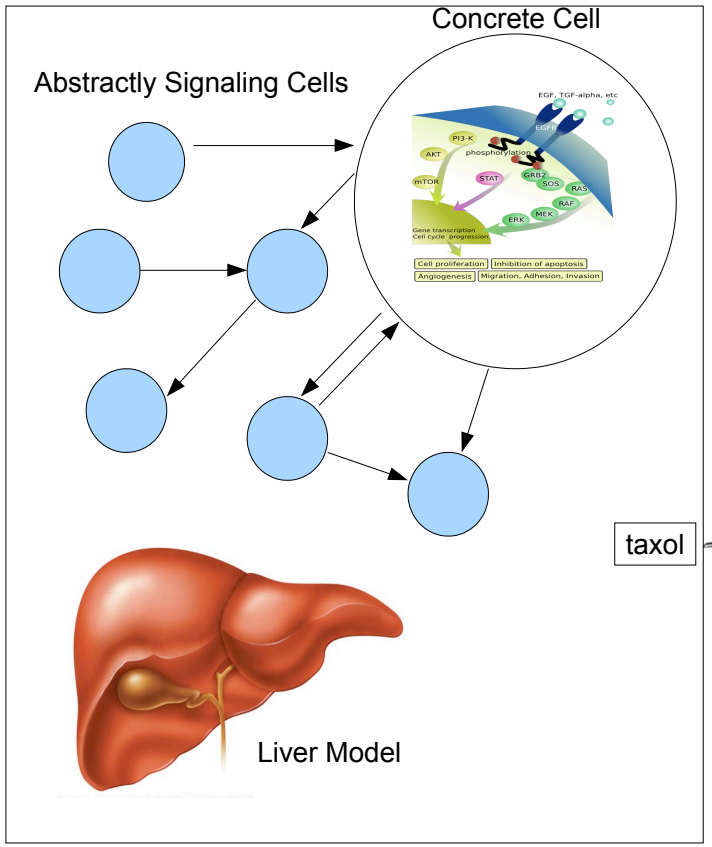
1. Local information $\mathcal{I} =< i_1 \ldots i_n >$
2. Signaling environment $\mathcal{E} =< e_1 \ldots e_m >$

Cells

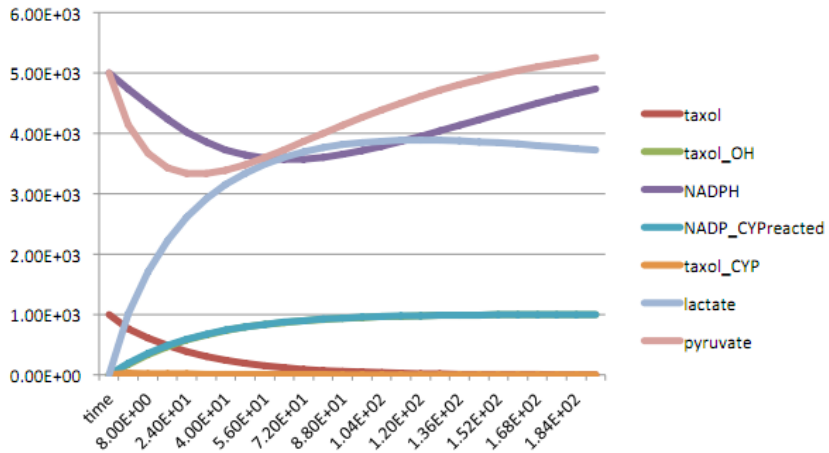
1. Concrete cells: signal transduction pathways, genes etc. state $X =< x_1 \ldots x_r >, x_i \in \mathbb{R}$ and $x_i \geq 0$
2. Abstract cells: abstract internal state $\Sigma \in \mathbb{R}$

	Abstract	Concrete
take action	$\Sigma \to \Sigma + \mathcal{A}$	$X \to X + \mathcal{A}$
send signal	$\Sigma \to \Sigma + \mathcal{S}$	$X \to X + \mathcal{S}$
change state	$\Sigma + \mathcal{I} + \mathcal{E} \to \Sigma$	$X + \mathcal{I} + \mathcal{E} \to X$

Taxol Example

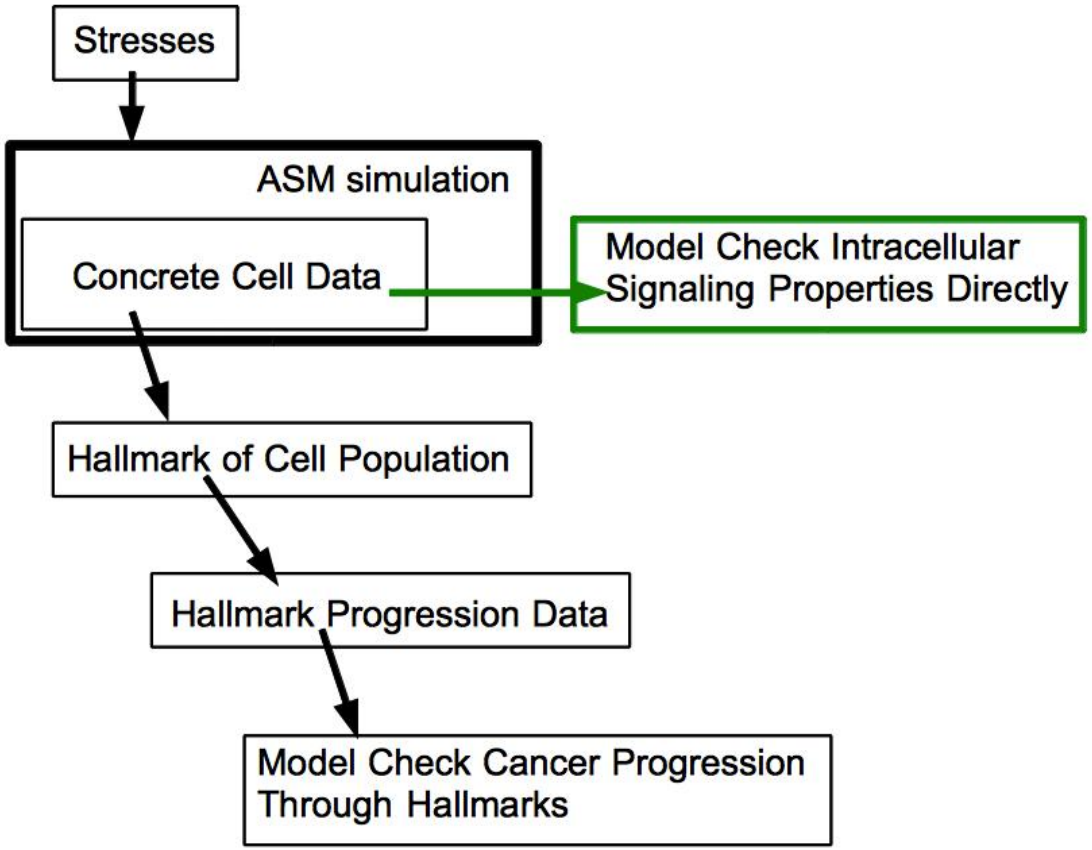


Cancer Cells	Normal Cells	Liver
Dead	Dead	Dead
Dead	Alive	Dead
Dead	Alive	Alive
Alive	Alive	Alive



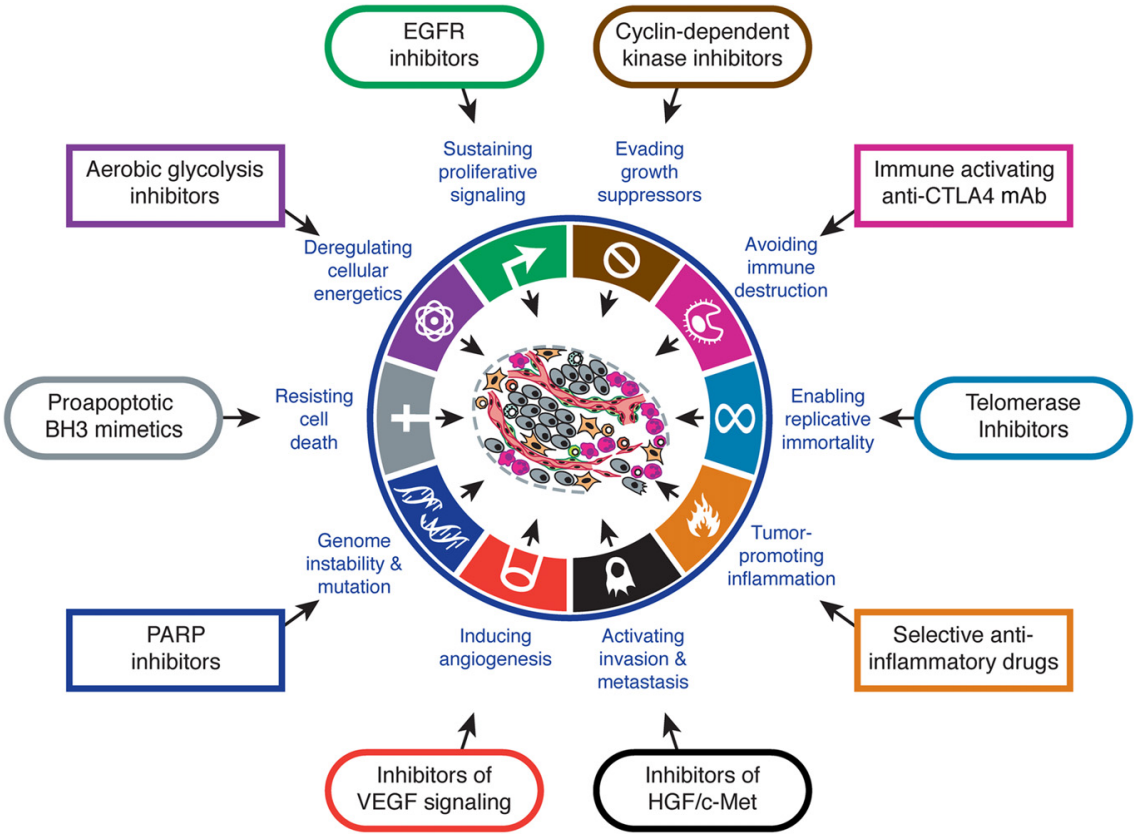
Metabolite concentrations over time

Hallmarks and Model Checking in ASM



Outline

Hallmarks of Cancer



D. Hanahan and R. A. Weinberg. *Hallmarks of Cancer: The Next Generation*, Cell, vol. 144, no. 5, pp. 646-674, 2011.

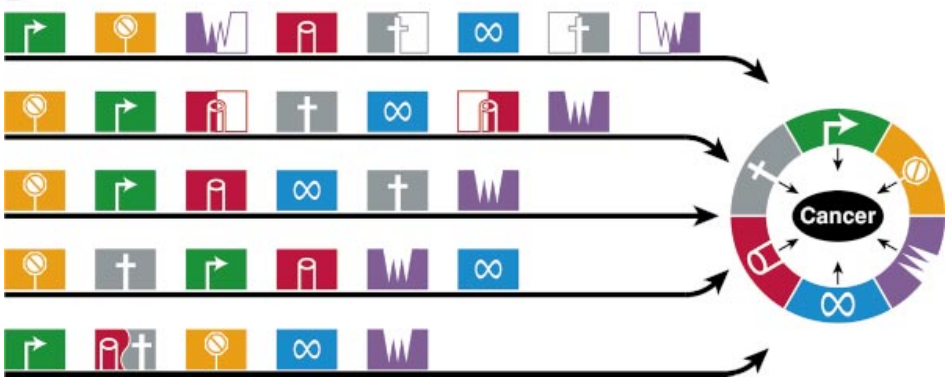
J. Luo, N. L. Solimini, and S. J. Elledge. *Principles of Cancer Therapy: Oncogene and Non-oncogene Addiction*, Cell, vol. 136, no. 5, pp. 823-837, Mar. 2009.

Tumor Progression

A

Component	Acquired Capability	Example of Mechanism
	Self-sufficiency in growth signals	Activate H-Ras oncogene
	Insensitivity to anti-growth signals	Lose retinoblastoma suppressor
	Evading apoptosis	Produce IGF survival factors
	Limitless replicative potential	Turn on telomerase
	Sustained angiogenesis	Produce VEGF inducer
	Tissue invasion & metastasis	Inactivate E-cadherin

B



Growing Lists of Therapies

Agent	Target	Addiction	Hallmarks	Potential mechanisms	References
17AAG (small molecule)	HSP90	NOA		A geldanamycin analog that binds to the ATP-binding pocket of HSP90 and inhibits its catalytic activity	Whitesell and Lindquist, 2005
1MT, MTH-Trp (small molecule)	IDO	NOA		Inhibits tryptophan catabolism in tumor microenvironment to allow T cell proliferation	Muller and Scherle, 2006
5-fluorouracil (small molecule)	DNA	NOA		Inhibits pyrimidine metabolism, incorporation in to DNA and RNA causes cell-cycle arrest	Longley et al., 2003
ABT-737, ABT-263 (small molecule)	BCL-XL, BCL-2	OA		Bind to the BH3 pocket of Bcl-XL and inhibit its antiapoptotic function	Stauffer, 2007
Alvocidib, PD 0332991 (small molecule)	CDKs	OA		Inhibit CDKs and induce cell-cycle arrest	Lee and Sicinski, 2006
AP 12009 (antisense oligo)	TGFβ 2	NOA		Inhibits tumor autocrine and paracrine signaling, reverses immune suppression in the tumor microenvironment	Muller and Scherle, 2006
AZD2281, AG014699 (small molecule)	PARP1	NOA		Inhibit base excision repair in homologous recombination repair-deficient cancer cells	Bryant et al., 2005; Farmer et al., 2005
Bevacizumab (antibody)	VEGF	NOA		Inhibits endothelial cell recruitment and tumor vasculature	Folkman, 2007
BEZ235 (small molecule)	PI3K	OA		Causes cell-cycle arrest in tumor cells and inhibits tumor angiogenesis	Maira et al., 2008
Bortezomib (small molecule)	Proteasome	NOA		Inhibits the catalytic activity of 26S proteasome and induces apoptosis	Roccaro et al., 2006
Celecoxib (small molecule)	COX2	NOA		Reverses immune suppression in the tumor microenvironment, inhibits tumor autocrine and paracrine signaling	Muller and Scherle, 2006
Cisplatin and analogs (small molecule)	DNA	NOA		Induces DNA crosslinks	Siddik, 2003
Erlotinib, Gefitinib (small molecule)	EGFR	OA		Inhibit EGFR tyrosine kinase by competing with ATP binding	Sharma et al., 2007
GRN163L (modified oligo)	hTERT	OA		Mimics telomere sequence and inhibits the hTERT active site	Dikmen et al., 2005; Harley, 2008
GRNVAC1 (cell therapy)	hTERT	OA		Autologous dendritic cells transduced to express an hTERT-LAMP fusion protein to elicit T cell response to hTERT + tumor cells	Harley, 2008; Su et al., 2005

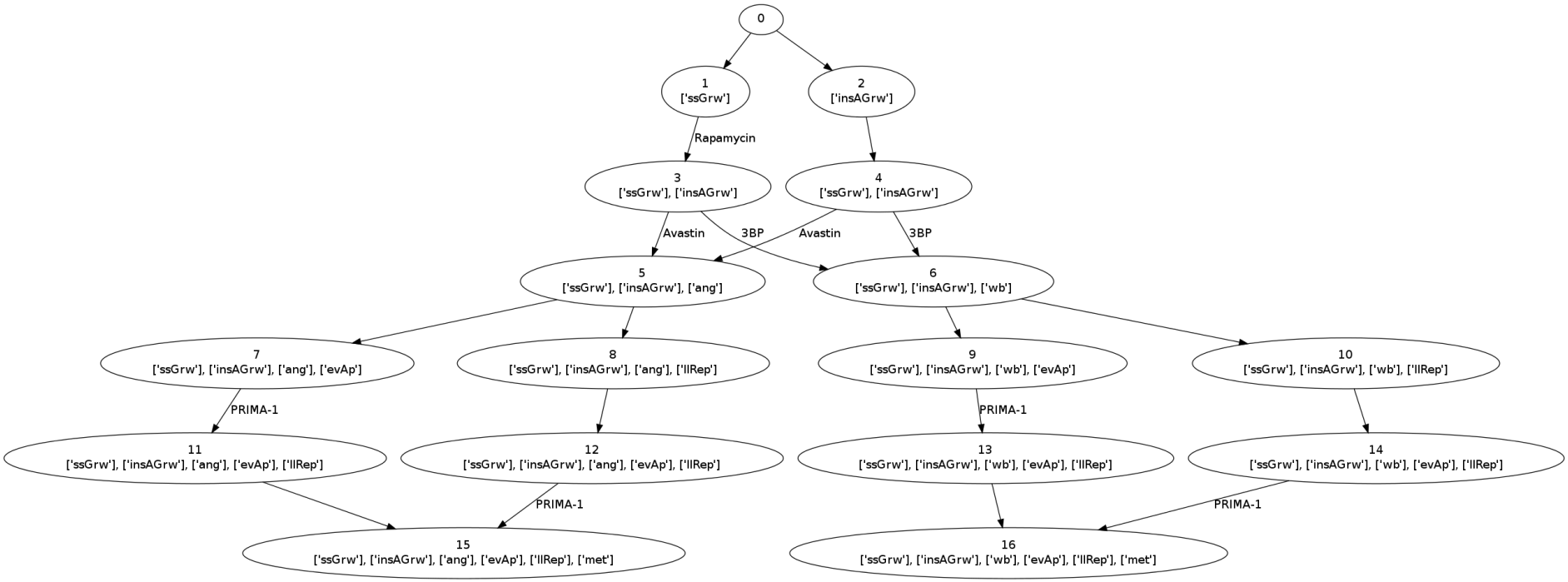
Agent	Target	Addiction	Hallmarks	Potential mechanisms	References
GV1001 (peptide)	hTERT	OA		A short immunogenic peptide from hTERT designed to elicit T cell response against hTERT + tumor cells	Harley, 2008; Nava-Parada and Emens, 2007
Imatinib, Dasatinib (small molecule)	BCR-ABL, c-Kit, Src, PDGFR, other TKs	OA		Tyrosine kinase inhibitor with multiple targets	Quintas-Cardama et al., 2007
Mapatumumab, Lexatumumab (antibody)	TRAIL receptor	NOA		Bind and activate TRAIL receptors to induce apoptosis	Carlo-Stella et al., 2007
Methotrexate (small molecule)	DHFR	NOA		Inhibits thymidine biosynthesis and induces replicative stress	McGuire, 2003
Nutlin-3 (small molecule)	HDM2	OA		Binds to HDM2 and inhibits the binding and ubiquitination of p53	Vassilev, 2007
Oblimersen (antisense oligo)	BCL-2	OA		Inhibits the expression of BCL-2 by blocking translation of its mRNA	Moreira et al., 2006
Paclitaxel, Vinblastine (small molecule)	Mitotic spindle	NOA		Interfere with dynamics and stability of mitotic spindles, activate mitotic checkpoints, and induce chromosome mis-segregation	Weaver and Cleveland, 2005
PF-00477736 (small molecule)	Chk1	NOA		Prevents activation of the DNA damage response, leading to persistent DNA damage and replication stress	Ashwell and Zabludoff, 2008
PRIMA-1, MIRA-1 (small molecule)	Mutant p53	TSGH		Reactivate the function of mutant p53	Selivanova and Wiman, 2007
Rapamycin, RAD001, Temsirolimus (small molecule)	mTOR	NOA		Inhibit protein synthesis	Guertin and Sabatini, 2007
Retinoic acid (small molecule)	RAR, RXR	OA		Induces cellular differentiation	Spira and Carducci, 2003
SAHBs (stapled peptide)	BCL-XL, BCL-2	OA		Stapled BH3 domains that bind to BCL-2 family members and promote apoptosis	Verdine and Walensky, 2007
Sorafenib, Sunitinib (small molecule)	Multiple kinases (VEGFR, RAF, c-Kit, PDGFR)	NOA		Inhibit endothelial cell recruitment and tumor vasculature	Folkman, 2007
Topotecan, Irinotecan (small molecule)	Topo-isomerase I	NOA		Induce DNA breaks	Pommier, 2006
Trastuzumab (antibody)	ERBB2	OA		Inhibits ERBB2 activation and induces immune destruction of cancer cells	Hynes and Lane, 2005

J. Luo, N. L. Solimini, and S. J. Elledge. *Principles of Cancer Therapy: Oncogene and Non-oncogene Addiction*, Cell, vol. 136, no. 5, pp. 823-837, Mar. 2009.

Cancer Hallmark Automata (CHA)

- Formalism to represent the "hallmark view" of cancer
- Represent progression models as Kripke structure / finite automaton
- Personalize model to specific cancer type and stage of patient
- Includes specifications of:
  - disease progression through hallmarks
  - timings of transitions
  - tests to observe disease state
  - effects of drugs on the system
  - costs of hallmarks and drugs (pain, monetary, . . .)

Example CHA



E.g., **AG**¬met will yield therapies that give

- Rapamycin, or Avastin and 3BP, if the patient comes at early stage
- Avastin at stage 3 and 4 and PRIMA-1 at stage 9 and 14 if 3BP has high toxicity
- 3BP at stage 3 and 4 and PRIMA-1 at stage 7 and 12 if the patient's genome indicates adverse reaction to Avastin
- PRIMA-1 if the disease status is advanced but unknown

Timed CHA

A timed CHA consists of

- a set of states, corresponding to hallmarks
- a set of directed edges between states, labeled with clock constraints
- an invariant for each clock and state (time limit)
- a factor for each tuple of drug, clock and state (slow-down or speed-up)

Including Partial Observability

Timed state: pair of state and clock values
Belief set: set of timed states considered possible
Runs: possible sequences of timed states and corresponding belief sets

A therapy maps finite runs to therapeutic actions, namely

- giving a certain drug or a cocktail, or
- performing a test to refine the current belief set

Therapies are assumed to be uniform:
Runs that agree on the belief set sequence map to the same action.

Therapies can be translated into conditional plans.

Epistemic-Temporal Goals

$$K\,\mathbf{AG}_{\leq 20}\neg\mathrm{met}$$

*"It is known that metastasis (*met*) will not be reached within* 20 *years"*

$$\mathbf{AG}(\mathrm{ang} \rightarrow ((\neg\mathrm{met} \wedge \mathbf{AX}\neg\mathrm{met})\ \mathbf{U}\ K\,\mathrm{ang}))$$

"Whenever the tumor acquires angiogenesis, this will be known (strictly) before the tumor reaches metastasis"

Outline

Lasso Penalized Cox Regression for PanCan Survival

- Most of existing studies focusing on the identification of the genetic mutations and not considering the important clinical factor – survival time
- Selection of relevant genes to pancreatic cancer survival from the genome
- Lasso (Least Absolute Shrinkage and Selection Operator) penalized partial likelihood function of the Cox model
- Acceleration of regression coefficient estimation by coordinate descent
- Capacity of handling underdetermined problems where the number of genes far exceeds the number of cases
- Tuning constant chosen by cross-validation (data driven)
- A handful of important genes retained in the final model with nonzero coefficients

T. T. Wu and K. Lange. *Coordinate descent algorithms for lasso penalized regression*, The Annals of Applied Statistics, vol. 2, no. 1, 2008.

Pancreatic Cancer Data Analysis

- Goal: To identify a gene signature of pancreatic cancer survival
- Microarray data: 34 patients with primary PDAC tumors from Johns Hopkins Medical Institutions, 49 from Northwestern Memorial Hospital, and 19 from NorthShore University Health System

  J. K. Stratford, D. J. Bentrem, J. M. Anderson et al. *A Six-Gene Signature Predicts Survival of Patients with Localized Pancreatic Ductal Adenocarcinoma*, PLoS Med., vol. 7, e1000307, 2010.
- 66 out of the 102 PDAC patients died at the end of the study (35% censored)
- 43,376 genes

12-Gene Signature

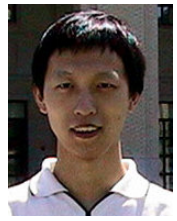
12 genes identified to be directly related to the survival time of the primary PDAC patients, and 8 confirmed to be cancer-related in previous cancer studies:

Genes	Functions
RPS13	Promote cell cycle transition from G1 to S
PCYT1B	Regulates phosphatidylcholine biosynthesis
TREX2	Proapoptotic tumor suppressor, maintain the genomic integrity
ZNF233	Zinc finger protein, deregulated in kidney and pancreatic cancer
ATPAF1	Regulate oxidative phosphorylation pathway
RIMS1	Down-regulated in multidrug resistance gastric carcinoma
SLC43A2	Overexpressed in adenocarcinomas and squamous cell carcinoma
NRAP	Up-regulated in human pancreatic cancer

SLC22A8, C4orf35, C6orf81, and C6orf58

Ilya Korsunsky
NYU

Justin Jee
NYU

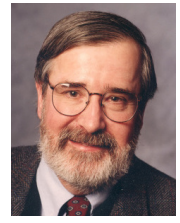


Haijun Gong
CMU

James Faeder
U. Pittsburgh

Bud Mishra
NYU

Thank you!



Ed Clarke
CMU

Natasa
Miskov-Zivanov
U. Pittsburgh

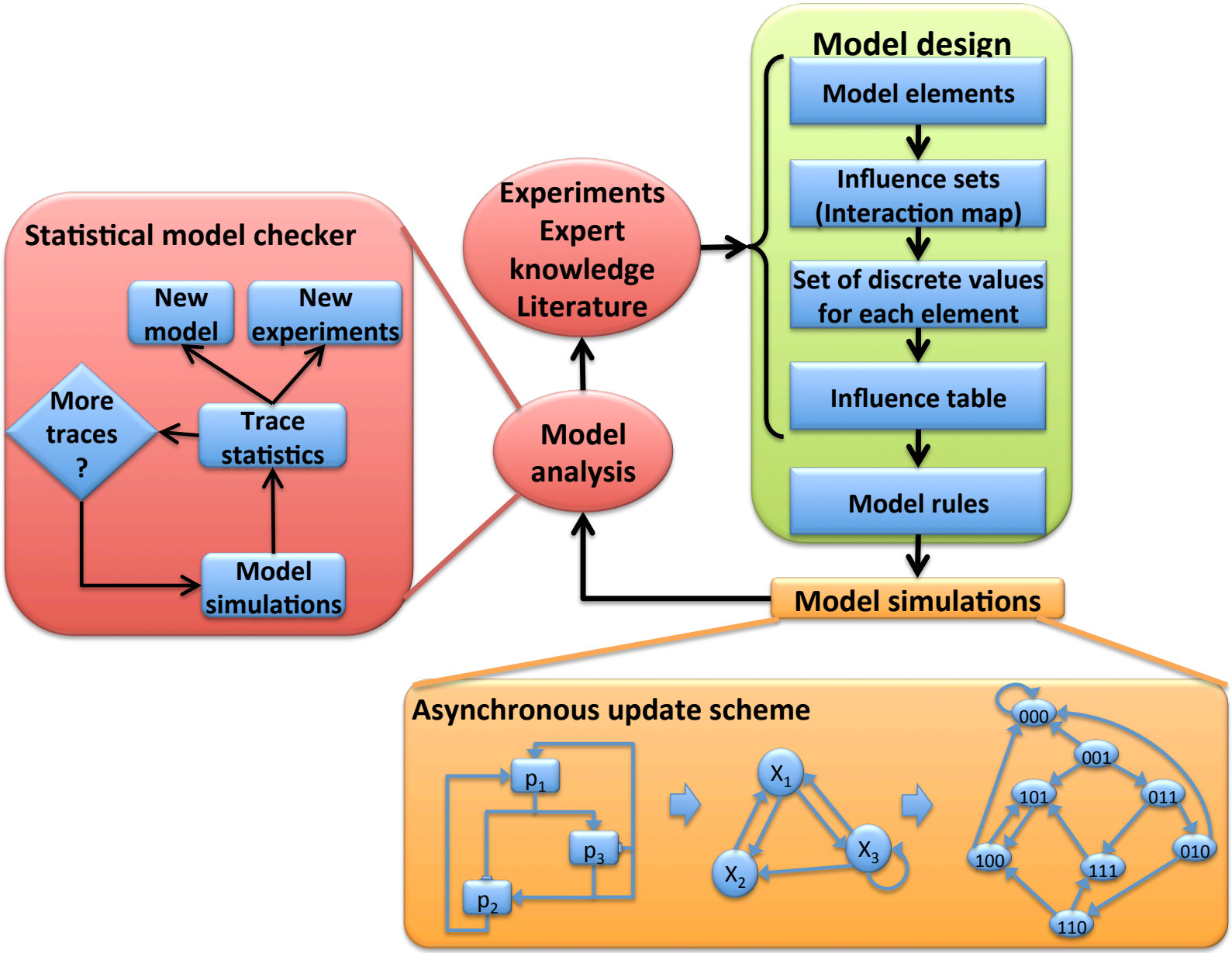
Loes Olde
Loohuis
NYU

Andreas Witzel
NYU

Tongtong Wu
U. Maryland

Paolo Zuliani
CMU

More Details on T-Cell Boolean Model



More Details on T-Cell Boolean Model

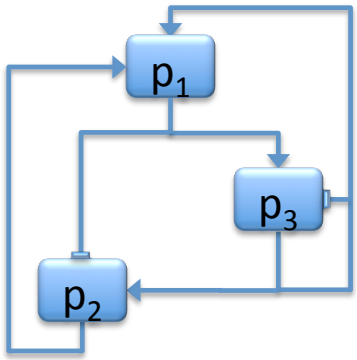
Biological network

Proteins:

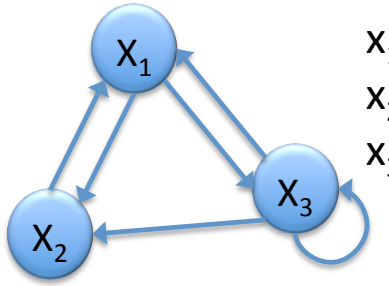
p_1, p_2, p_3

Protein states:

x_1, x_2, x_3



Boolean network

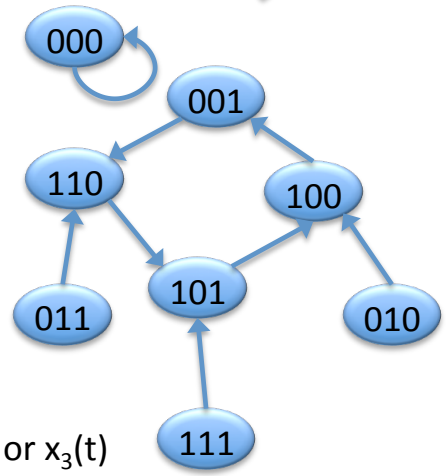


$x_1' = x_2$ or x_3

$x_2' =$ not x_1 and x_3

$x_3' = x_1$ and not x_3

Synchronous updates



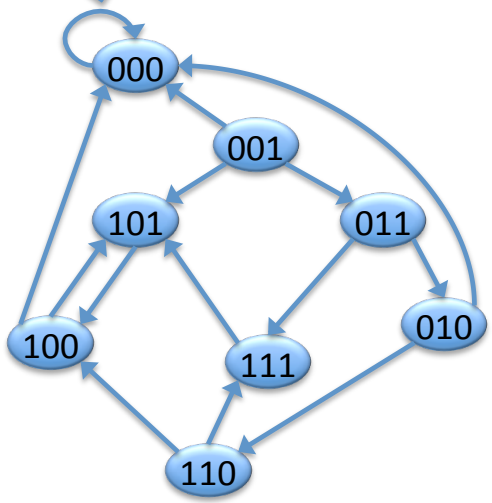
$x_1(t+1) = x_2(t)$ or $x_3(t)$

$x_2(t+1) =$ not $x_1(t)$ and $x_3(t)$

$x_3(t+1) = x_1(t)$ and not $x_3(t)$

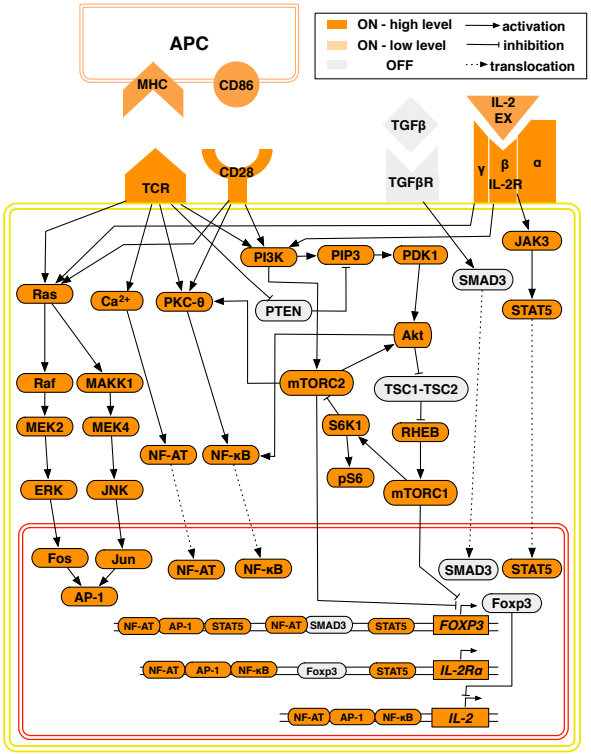
state	$x_1x_2x_3$
s_1	000
s_2	001
s_3	010
s_4	011
s_5	100
s_6	101
s_7	110
s_8	111

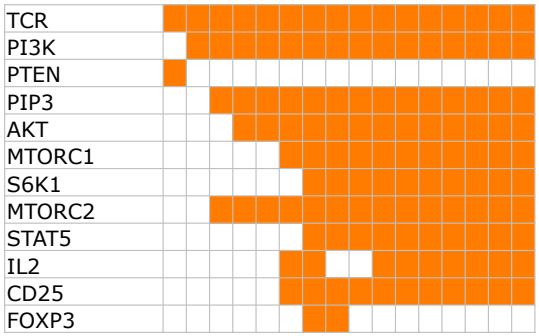
Asynchronous updates

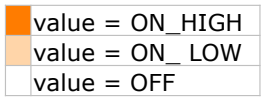


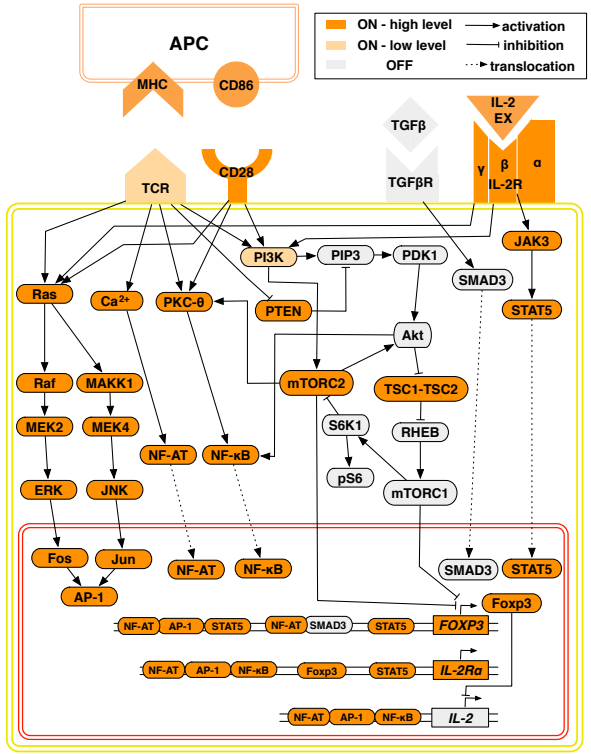
More Details on T-Cell Boolean Model

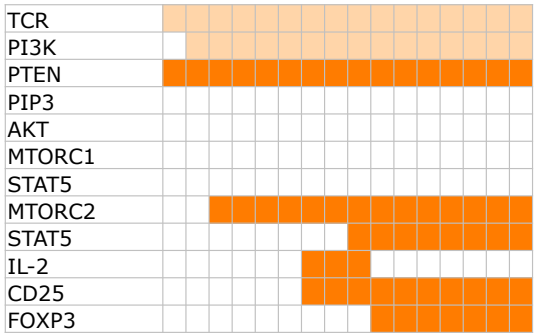
Steady states and trajectories for two different scenarios (high and low antigen dose)











Cox Model for Survival Data

- Observed data: $\{(Y_i, \delta_i, X_i)$, where $Y_i = \min\{T_i, C_i\}$, $\delta_i = I(T_i \leq C_i)$, $X \in R^p$, $i = 1, \ldots, n$
- Cox proportional hazards regression model

$$h(t|X) \quad = \quad h_0(t) \exp\left(\sum_{j=1}^{p} \beta_j X_j\right)$$

D. R. Cox. *Regression models and life-tables*, Journal of the Royal Statistical Society. Series B (Methodological) vol. 34, no. 2, 1972.

- Partial likelihood of the Cox model

$$L_n(\beta) \quad = \quad \prod_{i \in D} \frac{\exp\left(X_i^t \beta\right)}{\sum_{l \in R_i} \exp\left(X_l^t \beta\right)}$$

Lasso Penalized Partial Likelihood

Important genes related to PC survival can be selected via minimizing

$$-\ell_n(\beta) + P_\lambda(\beta)$$

where

- $\ell_n(\beta) = \log\{L_n(\beta)\}/n$ is convex with positive second derivative
- $P_\lambda(\beta)$ is the lasso (Least Absolute Shrinkage and Selection Operator) penalty on β

$$P_\lambda(\beta) = \lambda \sum_{j=1}^{p} |\beta_j|$$

  which is singular at the origin
- Minimizing the above objective function can achieve the desired sparsity hence variable selection

Challenges for High-Dimensional Lasso Penalized Cox Regression

One primary question

What is the most effective method of optimizing the lasso penalized objective function for high-dimensional data?

- High-dimensionality ($p \gg n$)
  - Standard methods of regression
    - Matrix operations
    - Number of arithmetic operations: $O(p^3)$
    - Incapable of handling underdetermined problems with $p \gg n$
- Nondifferentiability of the lasso penalty

Challenges for High-Dimensional Lasso Penalized Cox Regression

One primary question

What is the most effective method of optimizing the lasso penalized objective function for high-dimensional data?

- High-dimensionality ($p \gg n$)
  - Standard methods of regression
    - Matrix operations
    - Number of arithmetic operations: $O(p^3)$
    - Incapable of handling underdetermined problems with $p \gg n$
- Nondifferentiability of the lasso penalty

Solution

Coordinate descent can solve the two problems gracefully (Wu and Lange 2008)