A Collaborative Project Supported by the
NSF Experimental Expeditions Program

Pan Can

SIGNALS & CANCERS

Bud Mishra
(Courant, NYU/CSHL/MSSM/TIFR)

Carnegie Mellon · NASA · Stony Brook University · University of Maryland · Lehman College · NYU New York University · Cornell University · University of Pittsburgh

Morgagni

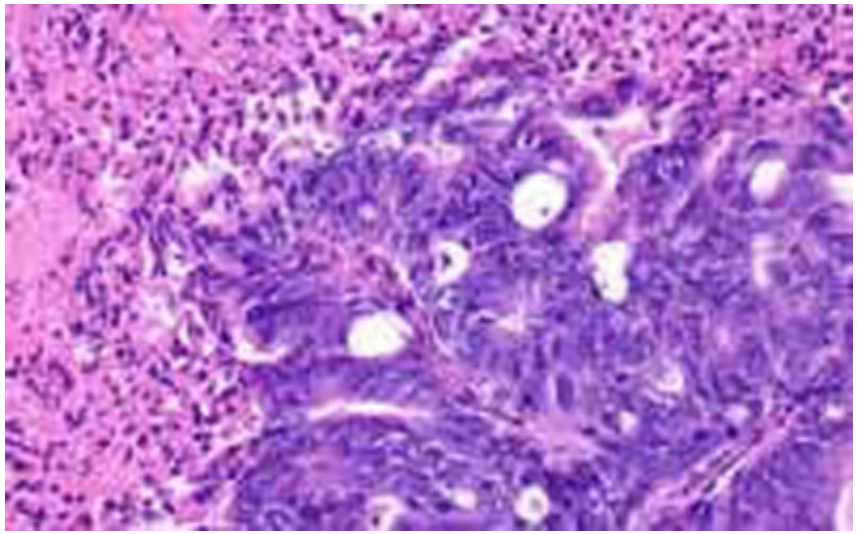
Pan Can

Giovanni Battista Morgagni
1682 -1771

- "…he vomited but little at a time, and seldom and what he did bring up, was watery and for the most part, bitter…
- "Besides this he was troubled with a great thirst and with a kind of frequent swooning, and in particular, with a pain**, just as if he were torn to pieces by dogs…"**
- *This case may have been the first reported pancreatic tumor*

Pancreatic Cancer

- 4th leading cause of cancer death in the US and Europe
- Five-year survival rate is only 4%
- Almost no progress in diagnosis and treatment in the past 40 years



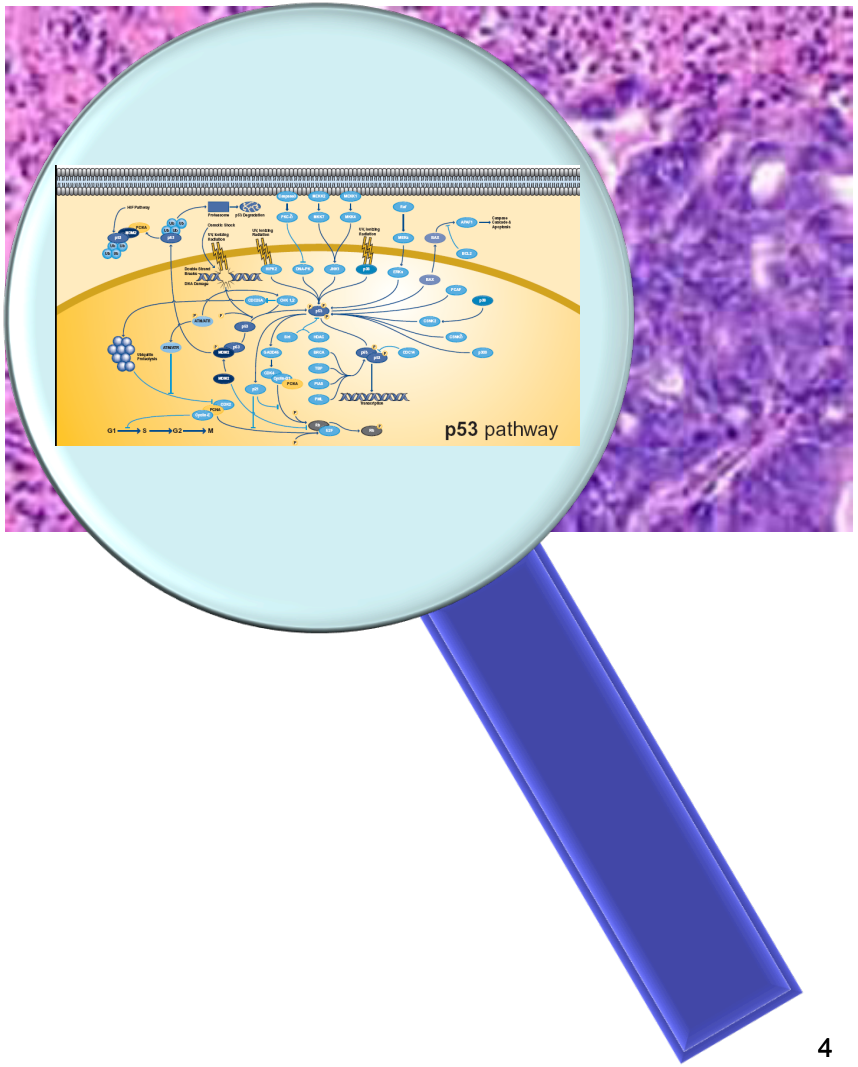
Healthy and diseased pancreas cells

New insights into the dynamics of these deadly diseases are urgently needed!

Why Pancreatic Cancer?

Pan Can

- No animal model, so computational models are needed
- Signaling models from cancer experts at **TGEN** (Translational Genomics)
- We will build new analysis and verification tools
- TGEN collaborators will use tools to better understand cancer dynamics



Model Checking

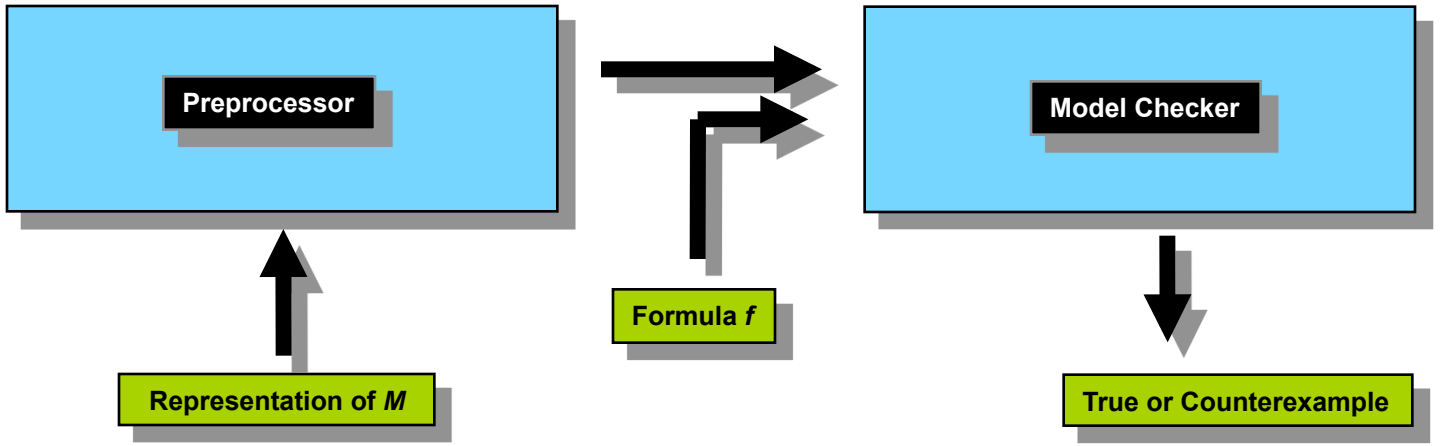
The **Model Checking Problem**:

Let ***M*** be a **state-transition graph**

Let ***f*** be a **formula of temporal logic**

e.g., a U b means "**a** holds true Until **b** becomes true"

Does ***f*** hold along all paths that start at initial state of ***M*** ?



Our Vision

To launch the next generation of research into **revolutionary**, highly **scalable**, and fully **automated** ***Model Checking and Abstract Interpretation*** by extending the reach and application of these techniques into new areas of science and engineering.

Our Vision

Pan Can

Specifically, to undertake a far-reaching and transformative investigation to gain fundamental new insights into the **emergent behaviors** of ***complex embedded and dynamical applications***: namely, in systems biology (**pancreatic cancer** and atrial fibrillation) and in engineering (automotive and aerospace systems).

Primary Challenge: Scalability

Key Scalability Issues:

Spatial Distribution

Stochastic Behavior

Highly Nonlinear Behavior

Mixed (Hybrid) Continuous-Discrete Behavior

Vast Numbers of System State Variables & Components

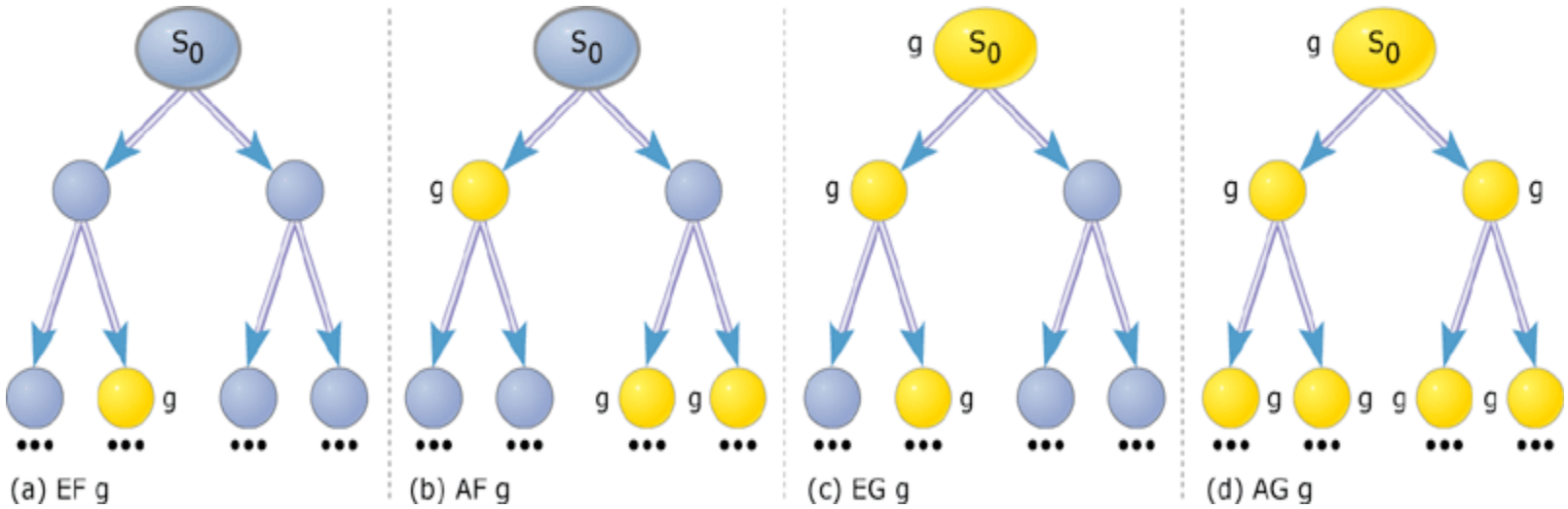
Complex Biological & Embedded Systems can exhibit any combination of these features

Computation Tree Logic (CTL)

- Branching Time temporal logic: interpreted over an ***execution tree*** where branching denotes non-deterministic actions
- Explicitly quantify over two modes – the path and the time
- Each time we talk about a temporal property, we also specify whether it is true on all possible paths or whether it is true on at least one path - ***Path quantifiers***
  - **A** = “for all future paths”
  - **E** = “for some future path”

Some CTL Operators

Pan Can



EF g

AF g

EG g

AG g

CTL Model-Checking

- Straight-forward approach: **Recursive descent on the structure of the query formula**
- Label the states with the terms in the formula:
  - Proceed by marking each point with the set of valid sub-formulas
- **"Global" algorithm:**
  - Iterate on the structure of the property, traversing the whole of the model in each step
  - Use fixed point unfolding to interpret Until:

$$\mathbf{E}(\psi_2 \, \mathbf{U}^+ \, \psi_1) \leftrightarrow \mathbf{E}\,\mathbf{X}(\psi_1 \vee \psi_2 \wedge \mathbf{E}(\psi_2 \, \mathbf{U}^+ \, \psi_1))$$

$$\mathbf{A}(\psi_2 \, \mathbf{U}^+ \, \psi_1) \leftrightarrow \mathbf{A}\,\mathbf{X}(\psi_1 \vee \psi_2 \wedge \mathbf{A}(\psi_2 \, \mathbf{U}^+ \, \psi_1))$$

Engineers Meet Biology

- What is a biological model?
- Who can provide it?
  - ***No Intelligent Designer***

  to tell us the design.
- What is a desired specification?
- What is a biologically important property?
  - ***No Teleological Intent***

  to guide the evolution.

Biological Models of Cancer

Pan Can

- Cancer as a disease of the genome…
- Cancer as a somatic evolutionary process…
- Cancer as a price of symbiosis (mitochondrial)…
- Cancer as a response to multi-cellularity…
- Cancer as a price of repair/regeneration (stem cells)…
- Cancer as a consequence of energy consumption (glucose metabolism)…
- Cancer as a response to external stress…
- Cancer as a response to the micro-environment (hyper- and hypo-methylation)…

Relevant Biological Processes

Pan Can

- Signaling:
  - Kinases…
- Proliferation:
  - Oncogenes and Tumor Suppressor Genes
- Differentiation:
  - Stem Cells…
- Maintenance and Immortality:
  - Autophagy, Necrosis and Apoptosis

War on Cancer

Pan Can

- "… as we know, there are **known knowns**; there are things we know we know.

  "We also know there are **known unknowns**; that is to say we know there are some things we do not know.

  "But there are also **unknown unknowns** – the ones we don't know we don't know."
  - *Ex-US Secretary of Defense, Mr. Donald Rumsfeld, Quoted completely out of context*.

Known Known Biology

- Theory: “World Where There Are Names for Everything.”

"Addicted to Death"

Pan Can

- Cancer is a progressive switch from apoptotic (scheduled) to necrotic (unscheduled) tumor cell death.
- The immunobiology of many intracellular factors are involved:
  - the products of **purine metabolism** (*uric acid, ATP, and adenosine*);
  - the nuclear protein HMGB1; the S100 family members; the heat shock proteins;
- Cancer is the consequence of disordered tumor cell death rather than cell growth
  - Loss of homeostasis
  - A condition called "addicted to death."

Purine Metabolism

Pan Can

- **Purine Metabolism**
  - Provides the organism with building blocks for the synthesis of DNA and RNA.
- **The entire pathway is almost closed but also quite complex. It contains**
  - several feedback loops,
  - cross-activations and
  - reversible reactions

Simple Model

Pan Can

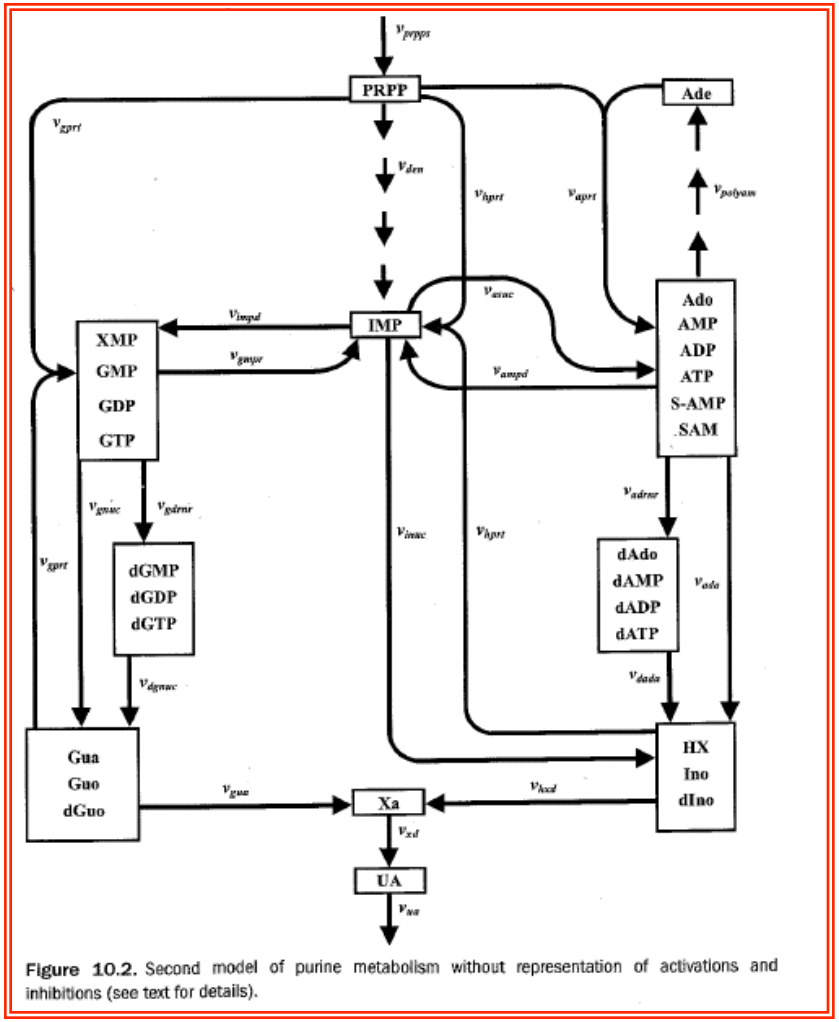
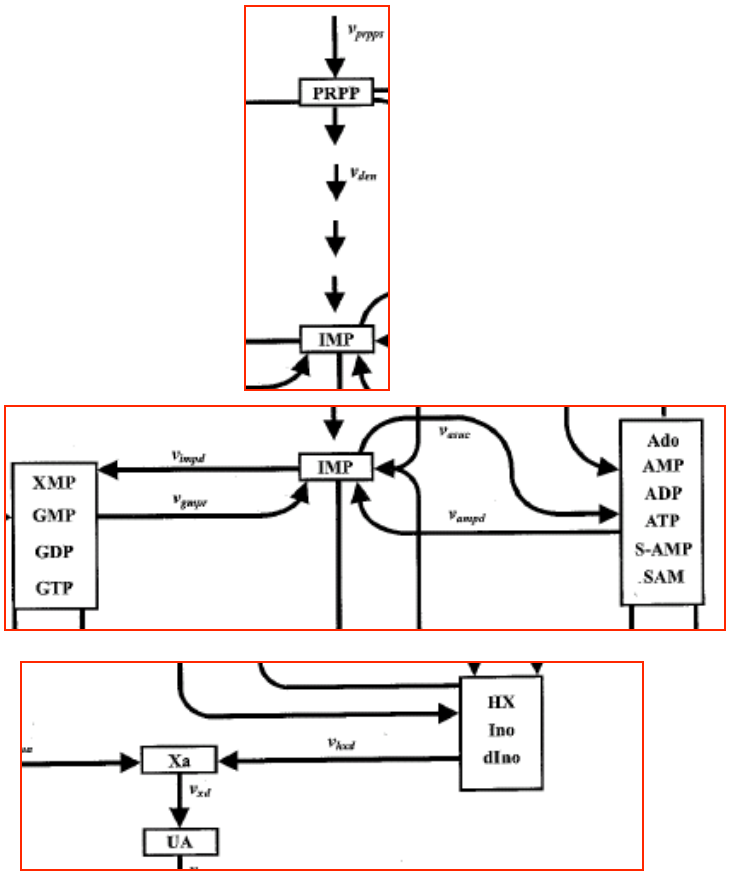


Figure 10.2. Second model of purine metabolism without representation of activations and inhibitions (see text for details).

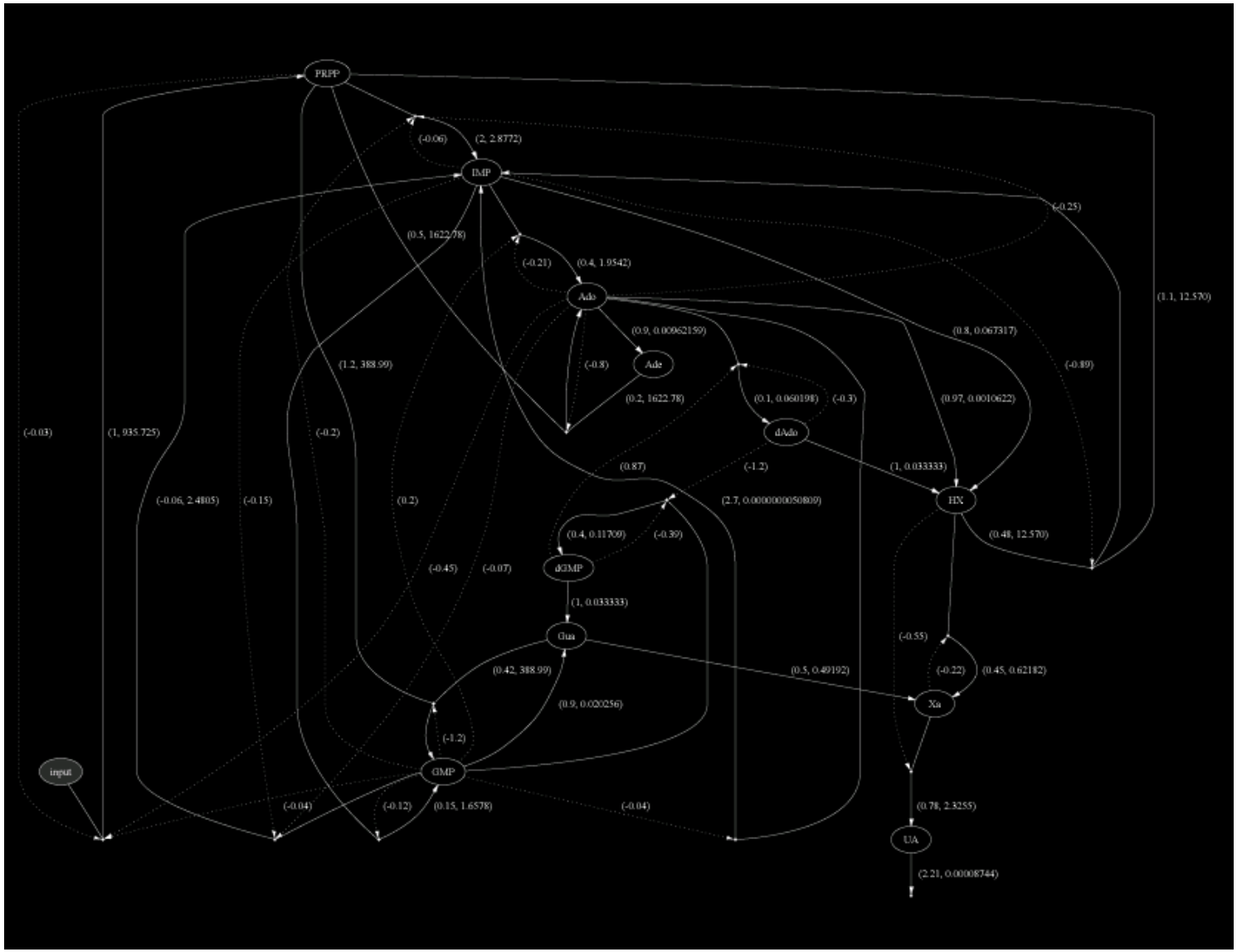
Biochemistry of Purine Metabolism



- The main metabolite in purine biosynthesis is *5-phosphoribosyl-a-1-pyrophosphate* (*PRPP*).
  - A linear cascade of reactions converts PRPP into *inosine monophosphate* (*IMP*).
  - IMP is transformed into AMP and GMP.
  - Guanosine, adenosine and their derivatives are recycled (unless used elsewhere) into *hypoxanthine* (*HX*) and *xanthine* (*XA*).
  - XA is finally oxidized into *uric acid* (*UA*).

Pan Can

Purine Metabolism



Queries

Pan Can

- Variation of the initial concentration of PRPP does not change the steady state. **(PRPP = 10 * PRPP1) implies steady_state()**

TRUE

- Persistent increase in the initial concentration of PRPP does cause unwanted changes in the steady state values of some metabolites.
- If the increase in the level of PRPP is in the order of 70% then the system does reach a steady state, and we expect to see increases in the levels of IMP and of the hypoxanthine pool in a "comparable" order of magnitude.

Always (PRPP = 1.7*PRPP1) implies steady_state()

TRUE

Queries

Pan Can

- Consider the following statement:
- **Eventually**

(Always (PRPP = 1.7 * PRPP1)
implies
steady_state()
and Eventually
Always(IMP < 2* IMP1))
and Eventually (Always
(hx_pool < 10*hx_pool1)))

- where IMP1 and hx_pool1 are the values observed in the unmodified trace.

- The model checker determines that the above statement is false..

- Counter-example: Model checker shows that the increase in IMP is about 6.5 fold while the hypoxanthine pool increase is about 60 fold.
- The model "over-predicts" the increases in products by amounts that are physiologically impossible…

- The model should therefore be amended

False

Final Model

Pan Can

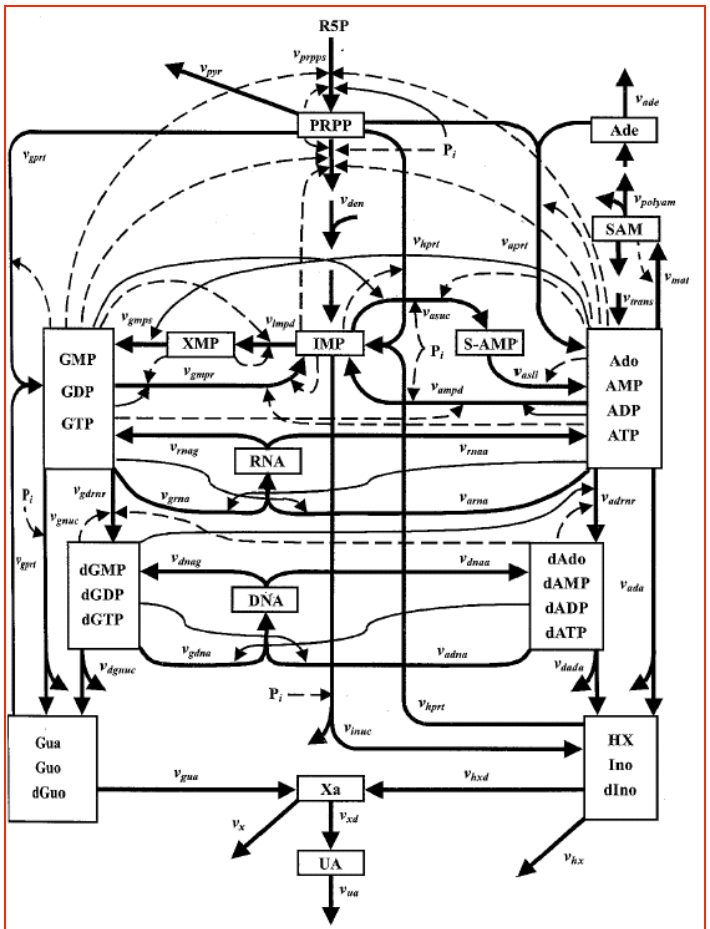
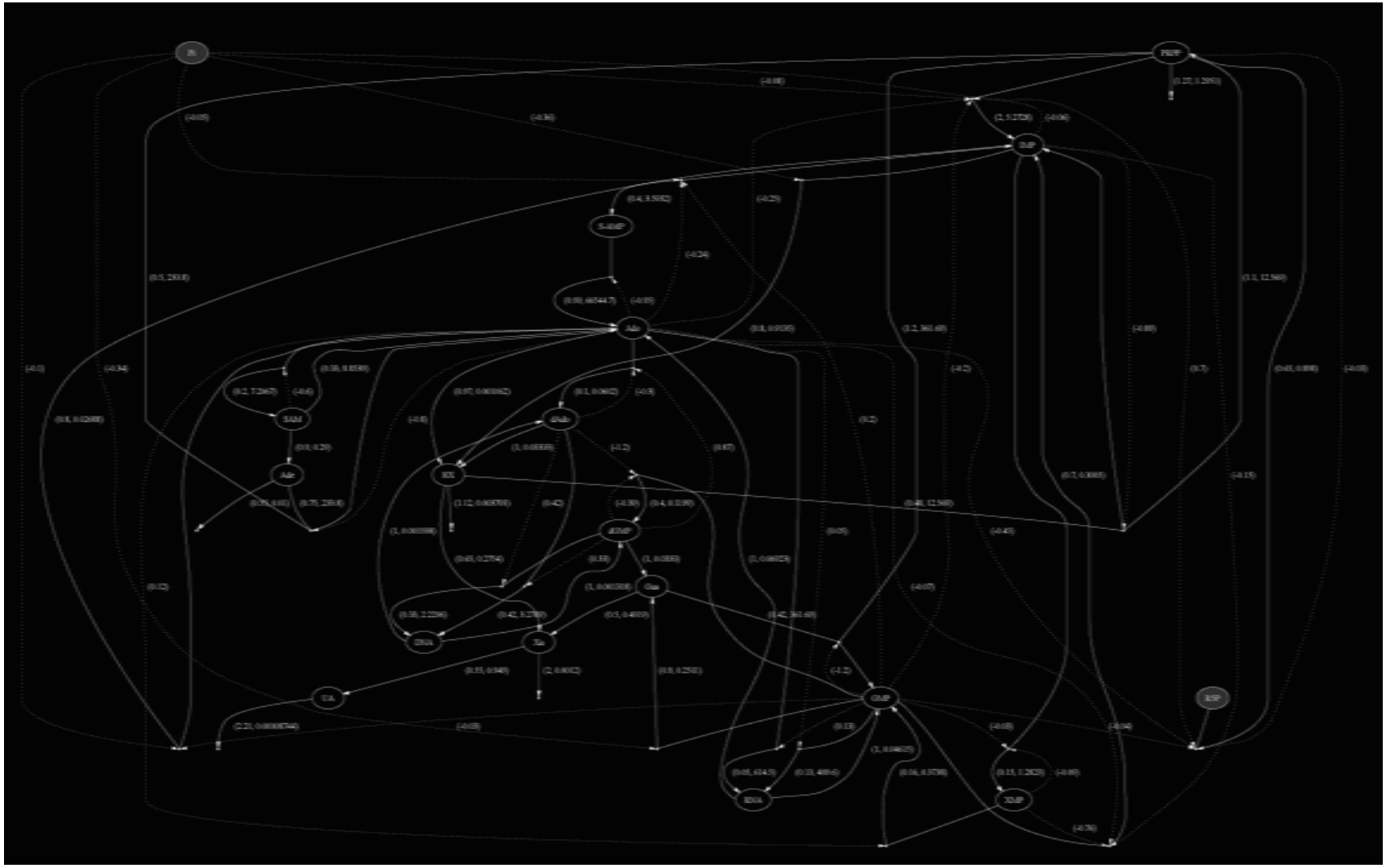


Figure 10.4. Map of the "final" model. Light solid arrows represent activation, while light dashed arrows represent inhibition. Curved heavy arrows entering or leaving the pathway indicate purine ring and ribose moieties that balance the stoichiometry of the system.

Pan Can

Purine Metabolism



XS-Systems:
(AAMC M. et al. 2001-2009)

Pan Can

Canonical Form:

$$\begin{cases} \dot{X}_i = \alpha_i \prod_{j=1}^{n+m} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{ij}} \quad i = 1 \ldots n \\ C_l(X_1(t), \ldots, X_{n+m}(t)) = \sum (\gamma_l \prod_{j=1}^{n+m} X_j^{f_{lj}}) = 0 \end{cases}$$

Characteristics:

- **Predefined Modular Structure**
- **Automated Translation from Graphical to Mathematical Model**
- **Scalability**

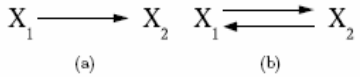


Figure 1: Representation of an unmodified and of a reversible reaction.

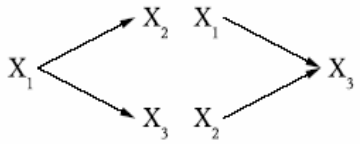


Figure 2: Representation of a divergence and of a convergence branch point (the two processes in each reaction are independent of each other).

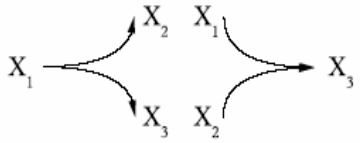


Figure 3: Representation of a single splitting reaction generating two products, X_2 and X_3, in stoichiometric proportions and of a single synthetic reaction involving two source components, X_1 and X_2 always in stoichiometric proportions.

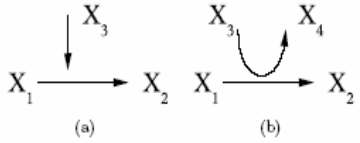
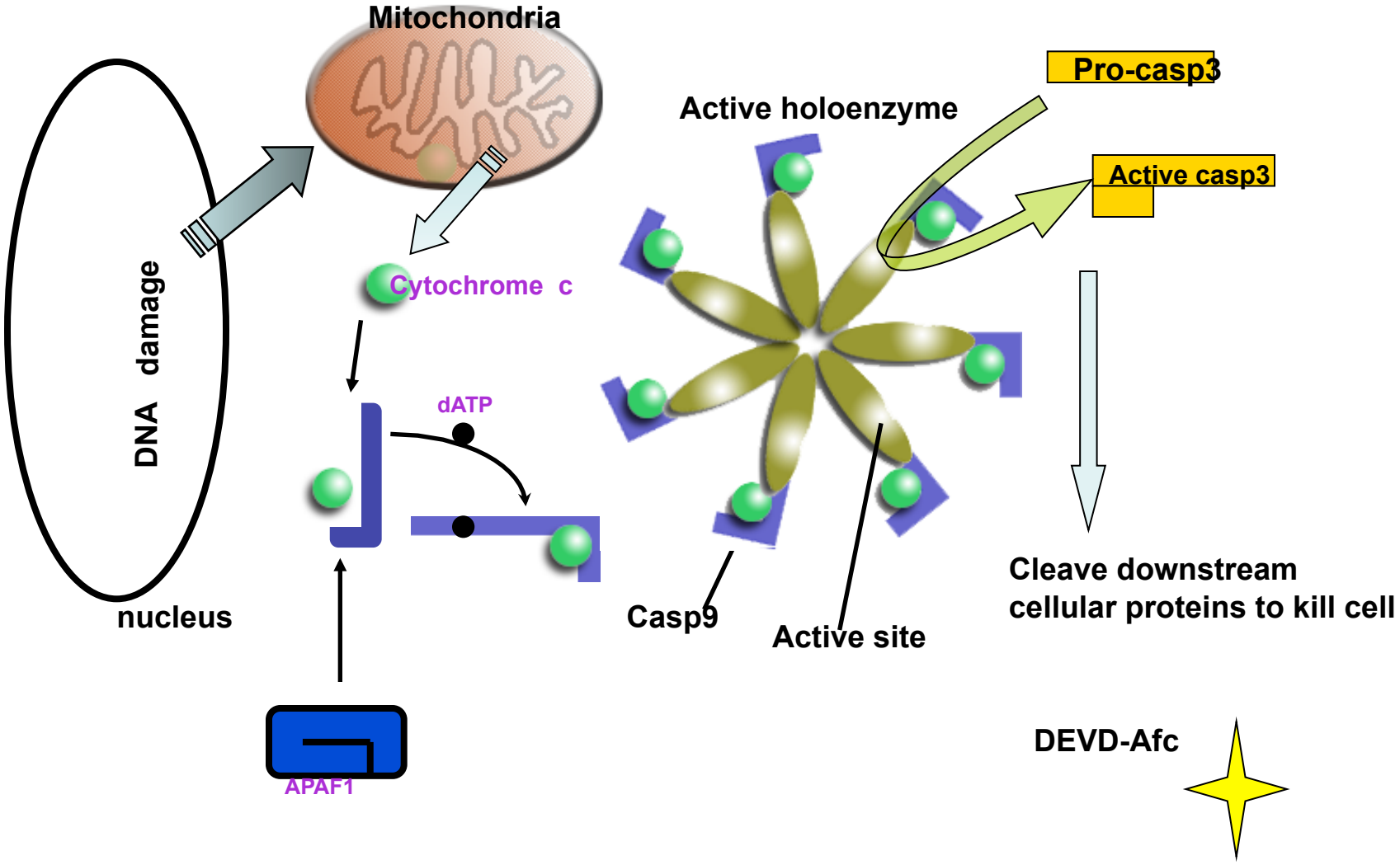


Figure 4: The conversion of X_1 into X_2 is modulated (stimulation or inhibition is represented by the sign of the arrow) by X_3. The reaction between X_1 and X_2 requires coenzyme X_3, which in the process is converted into X_4.

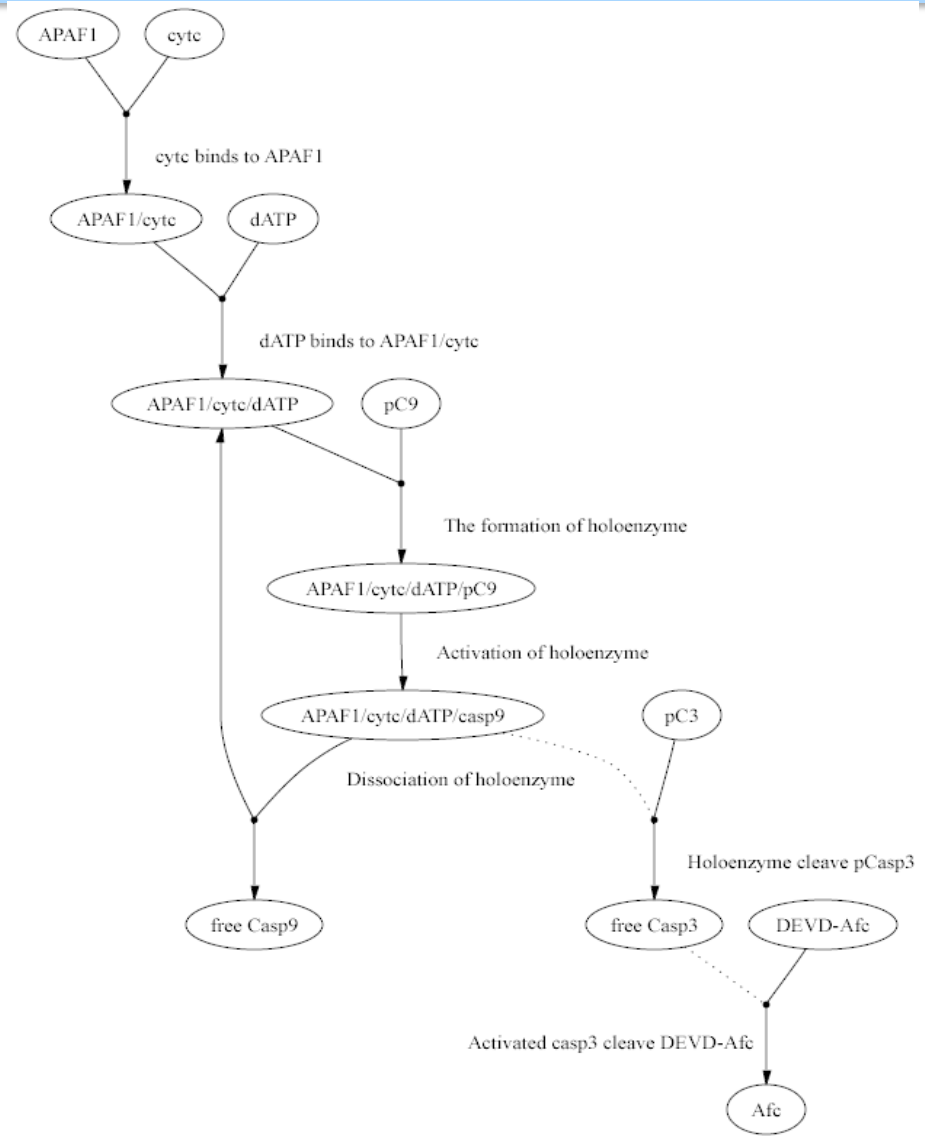
The activation of Casp9 needs APAF1 and cytochrome c

Pan Can

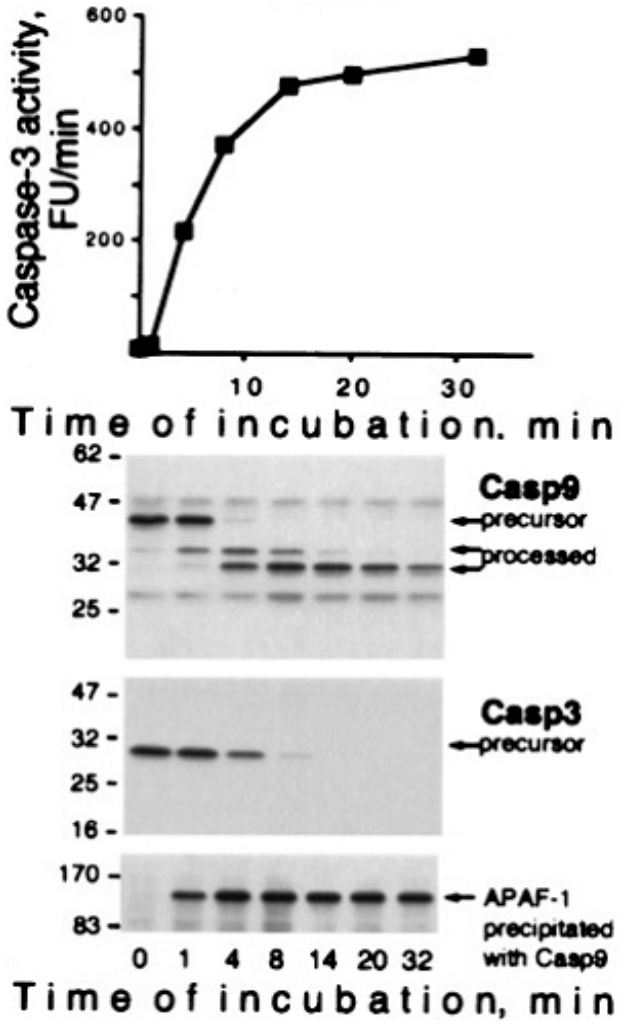


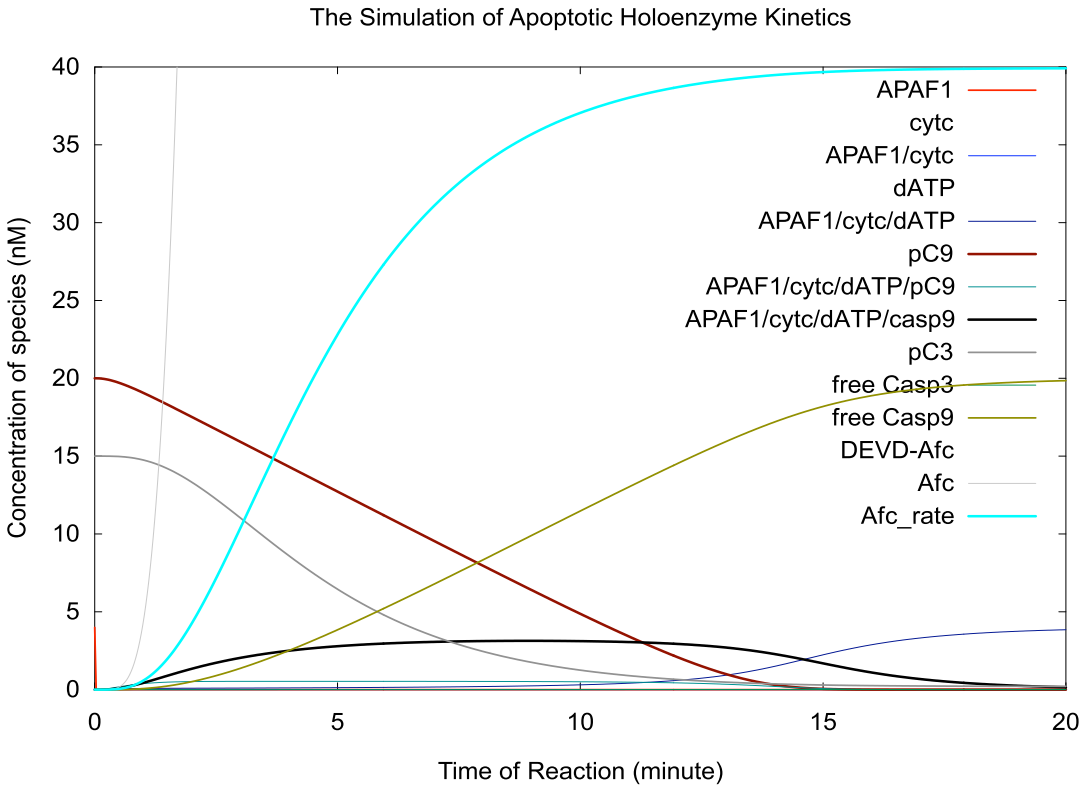
xS-System Model

Pan Can



Simpathica recapitulate the holoenzyme formation process

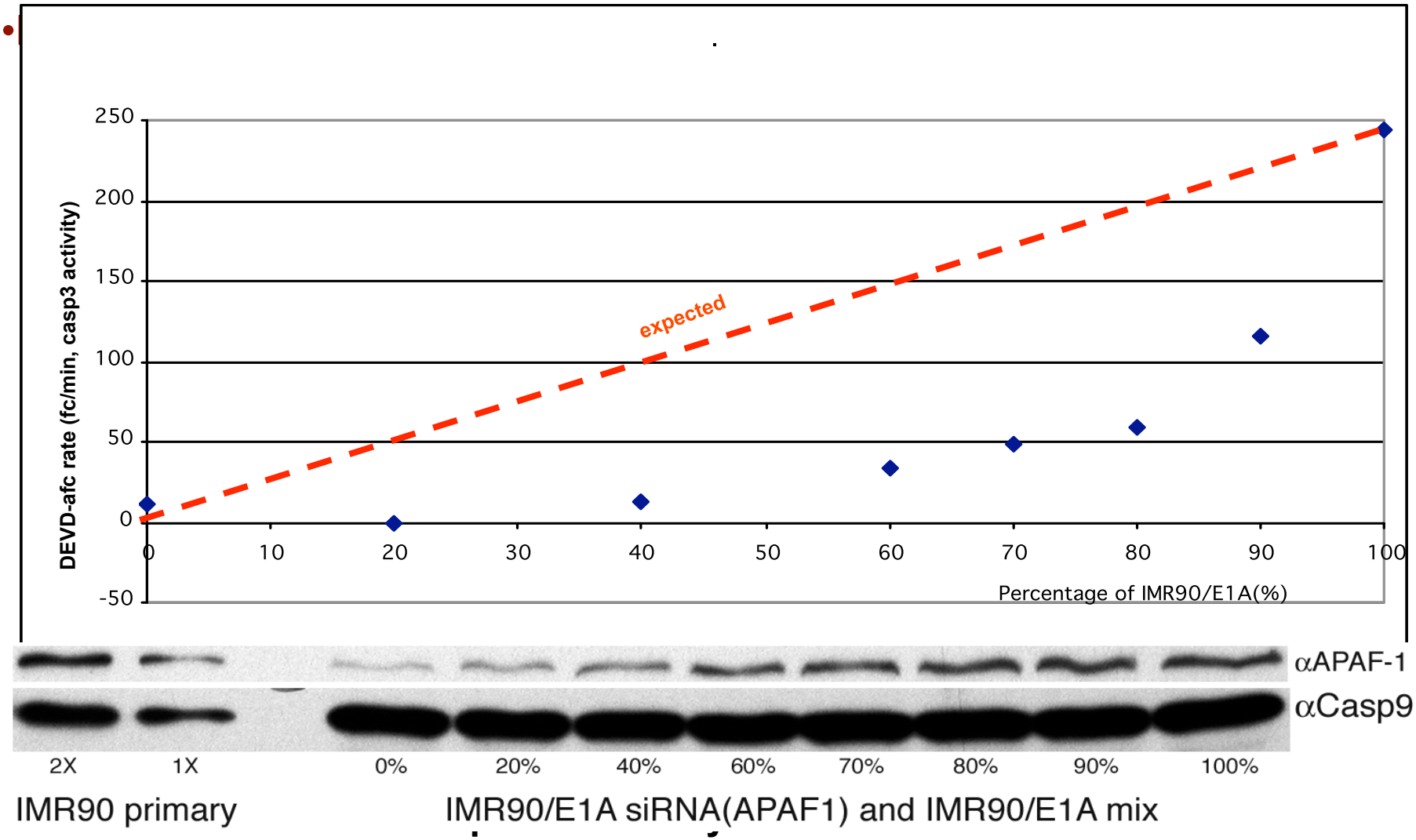




Rodriguez and Lazebnik (1999)

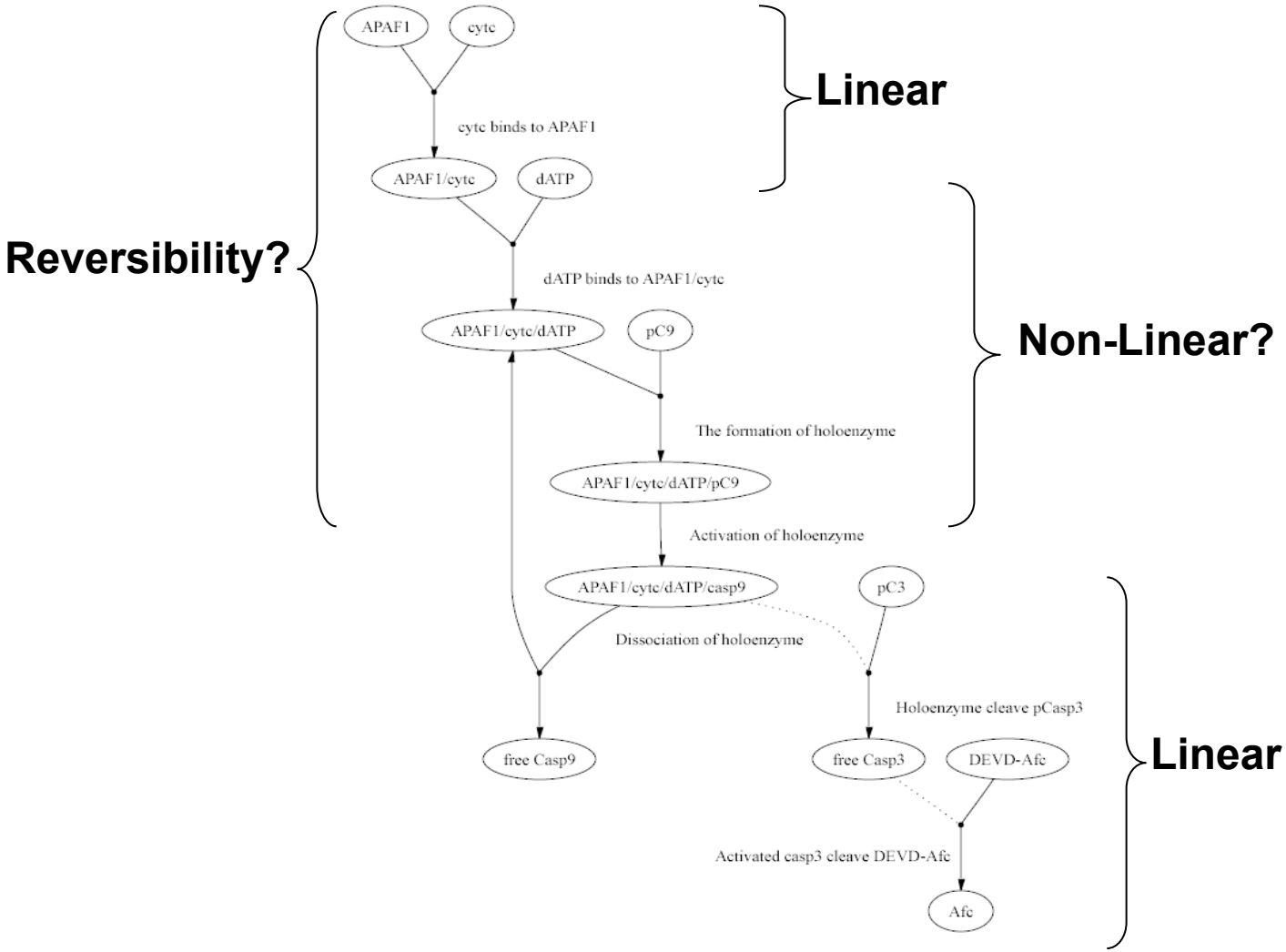
Decreasing [APAF-1] Kill Caspase Activity

Pan Can

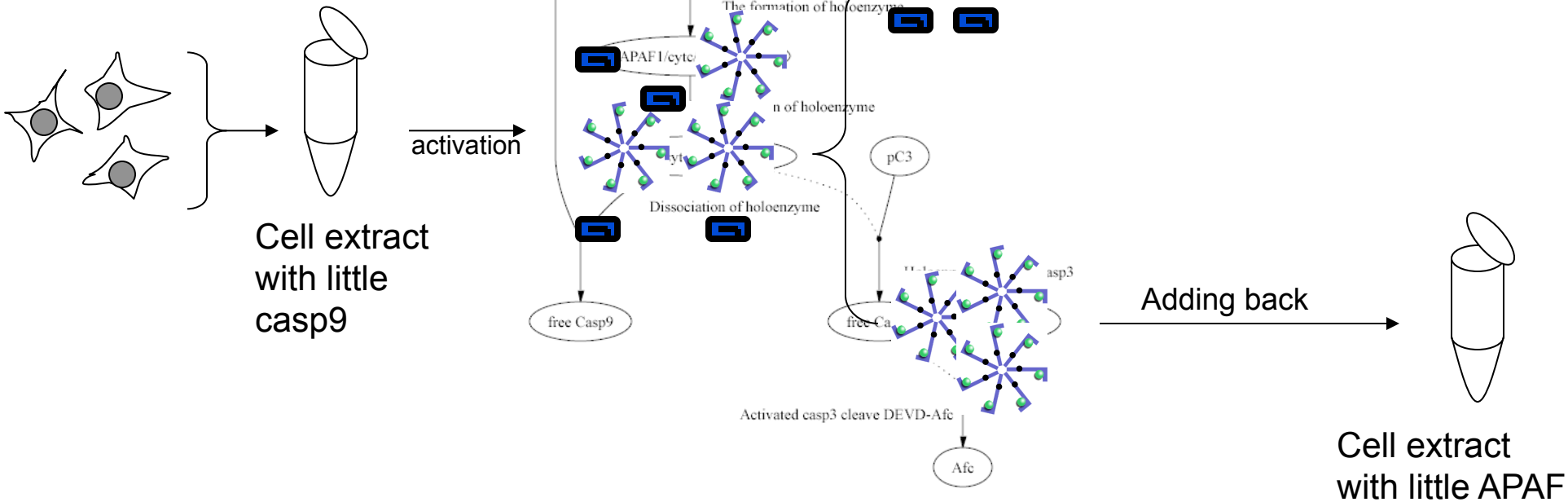


Where to modify the model in Simpathica?

Pan Can



- Recombinant system:
  - cytochrome c, caspase-9, APAF1
- Purification of endogenous APAF1/cytc oligomer



[APAF1/cytc/dATP] ⟷ caspase3 activity

Linear dependence?

XS-Systems:
(AAMC M. et al. 2001-2009)

Pan Can

Canonical Form:

$$\begin{cases} \dot{X}_i = \alpha_i \prod_{j=1}^{n+m} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{ij}} \quad i = 1 \ldots n \\ C_l(X_1(t), \ldots, X_{n+m}(t)) = \sum (\gamma_l \prod_{j=1}^{n+m} X_j^{f_{lj}}) = 0 \end{cases}$$

Characteristics:

- **Predefined Modular Structure**
- **Automated Translation from Graphical to Mathematical Model**
- **Scalability**

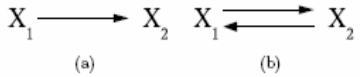


Figure 1: Representation of an unmodified and of a reversible reaction.

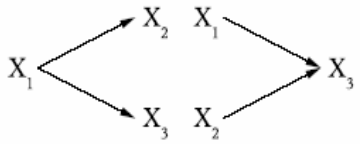


Figure 2: Representation of a divergence and of a convergence branch point (the two processes in each reaction are independent of each other).

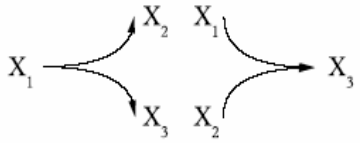


Figure 3: Representation of a single splitting reaction generating two products, X_2 and X_3, in stoichiometric proportions and of a single synthetic reaction involving two source components, X_1 and X_2 always in stoichiometric proportions.

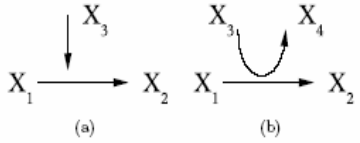


Figure 4: The conversion of X_1 into X_2 is modulated (stimulation or inhibition is represented by the sign of the arrow) by X_3. The reaction between X_1 and X_2 requires coenzyme X_3, which in the process is converted into X_4.

Formal Definition of S-system

Pan Can

Definition 1 (S-system). *An S-system is a quadruple* $S = (DV, IV, DE, C)$ *where:*

- $DV = \{X_1, \ldots, X_n\}$ *is a finite non empty set of* dependent variables *ranging over the domains* $D_1, \ldots, D_n$, *respectively;*
- $IV = \{X_{n+1}, \ldots, X_{n+m}\}$ *is a finite set of* independent variables *ranging over the domains* $D_{n+1}, \ldots, D_{n+m}$, *respectively;*
- DE *is a set of* differential equations, *one for each dependent variable, of the form*

$$\dot{X}_i = \alpha_i \prod_{j=1}^{n+m} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{ij}}$$

  with $\alpha_i, \beta_i \geq 0$ *called* rate constants;
- C *is a set of* algebraic constraints *of the form*

$$C_j(X_1, \ldots, X_{n+m}) = \sum(\gamma_j \prod_{k=1}^{n+m} X_k^{f_{jk}}) = 0$$

  with γ_j *called* rate constraints.

Verifying temporal properties of a reactive system

Step 1. Formally encode the behavior of the system as a semi-algebraic hybrid automaton

Step 2. Formally encode the properties of interest in TCTL

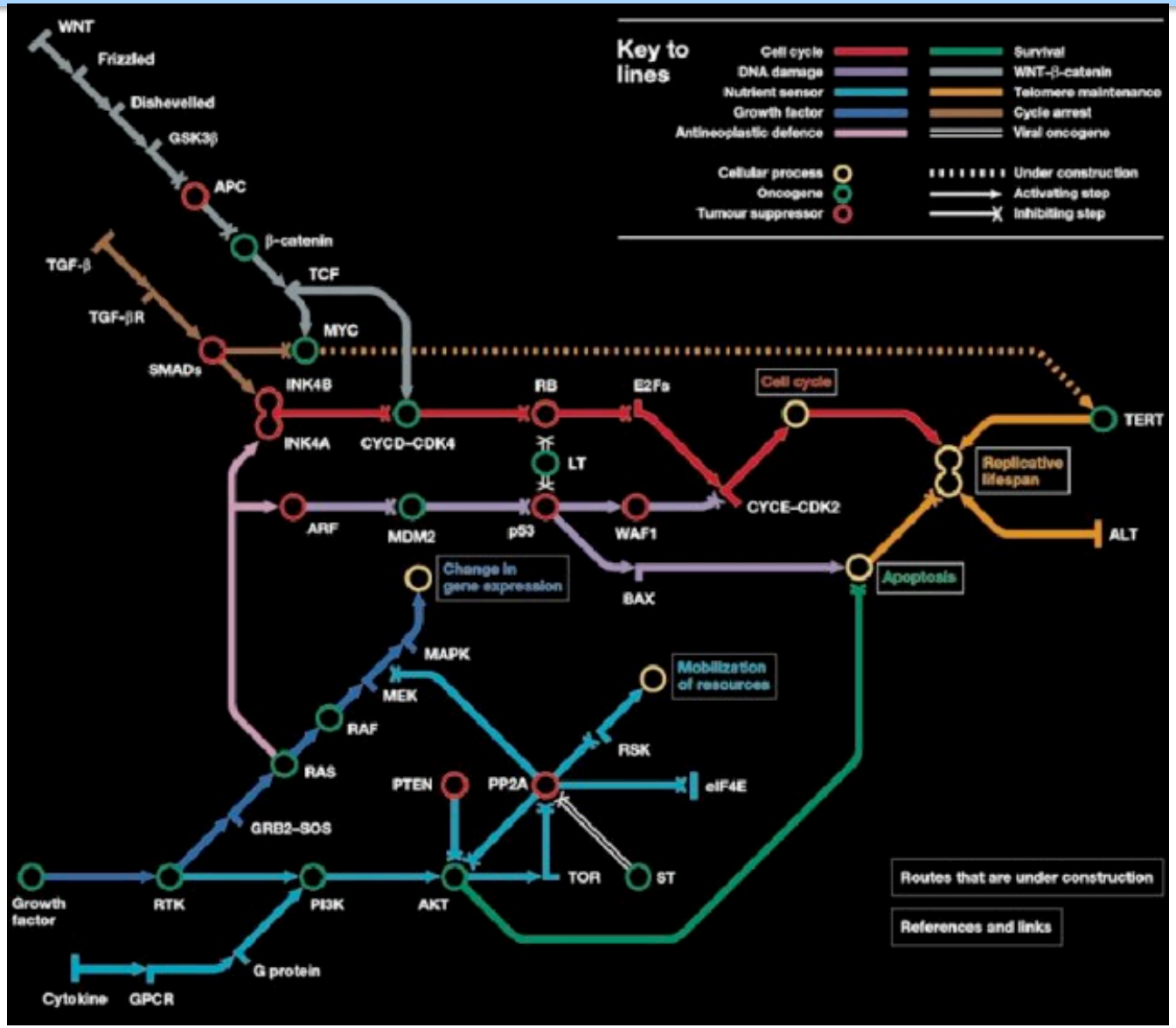
Step 3. Automate the process of checking if the formal model of the system satisfies the formally encoded properties using quantifier elimination

Solution

- Bounded Model Checking
- Constrained Systems
  - Linear Systems
  - O-minimal
  - SACoRe (Semi algebraic Constrained Reset)
  - IDA

Subway Map of Cancer

Pan Can



Is this View of Cancer Necessarily Accurate ?

Pan Can

- "If I said **yes**, that would then suggest that that might be the only place where it might be done which would **not be accurate, necessarily accurate**.
- **"It might also not be inaccurate, but I'm disinclined to mislead anyone."**
  - *Ex-US Secretary of Defense, Mr. Donald Rumsfeld, Once again quoted completely out of context*.

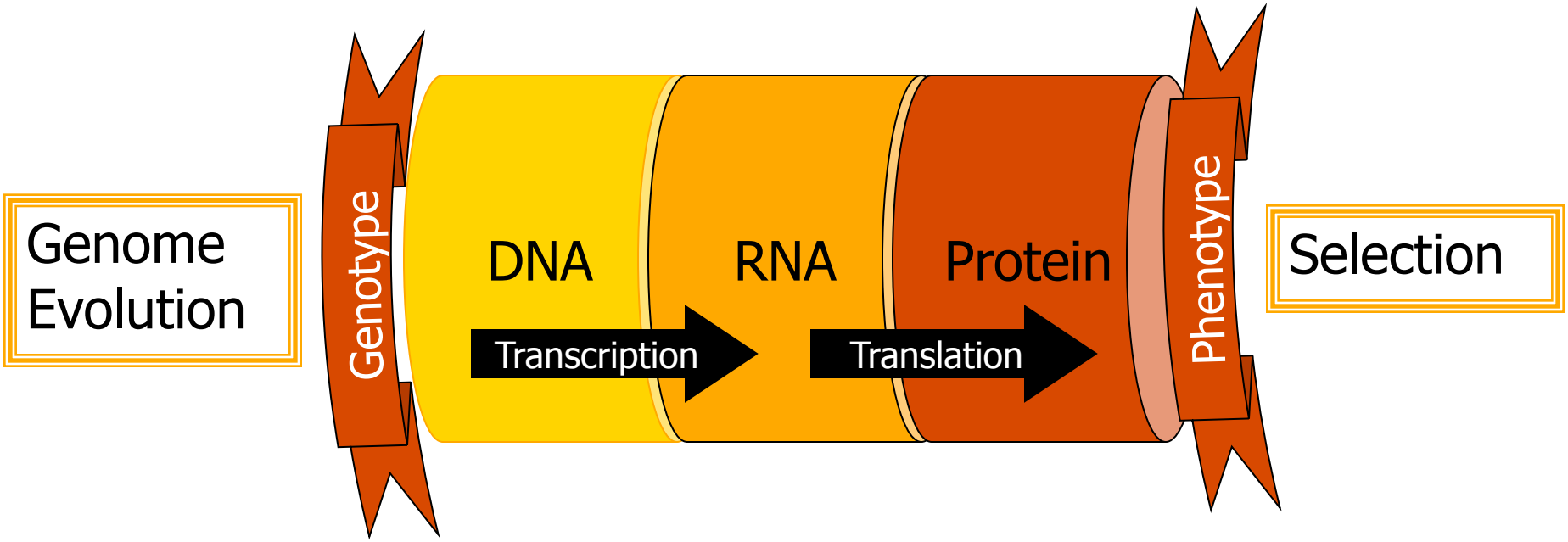
Known Unknown Biology

Pan Can

- Reality: “World Where There Are No Names of Anything.”

The New Synthesis

Pan Can



Cancer Initiation and Progression

Pan Can

Mutations, Translocations, Amplifications, Deletions

Epigenomics (Hyper & Hypo-Methylation)

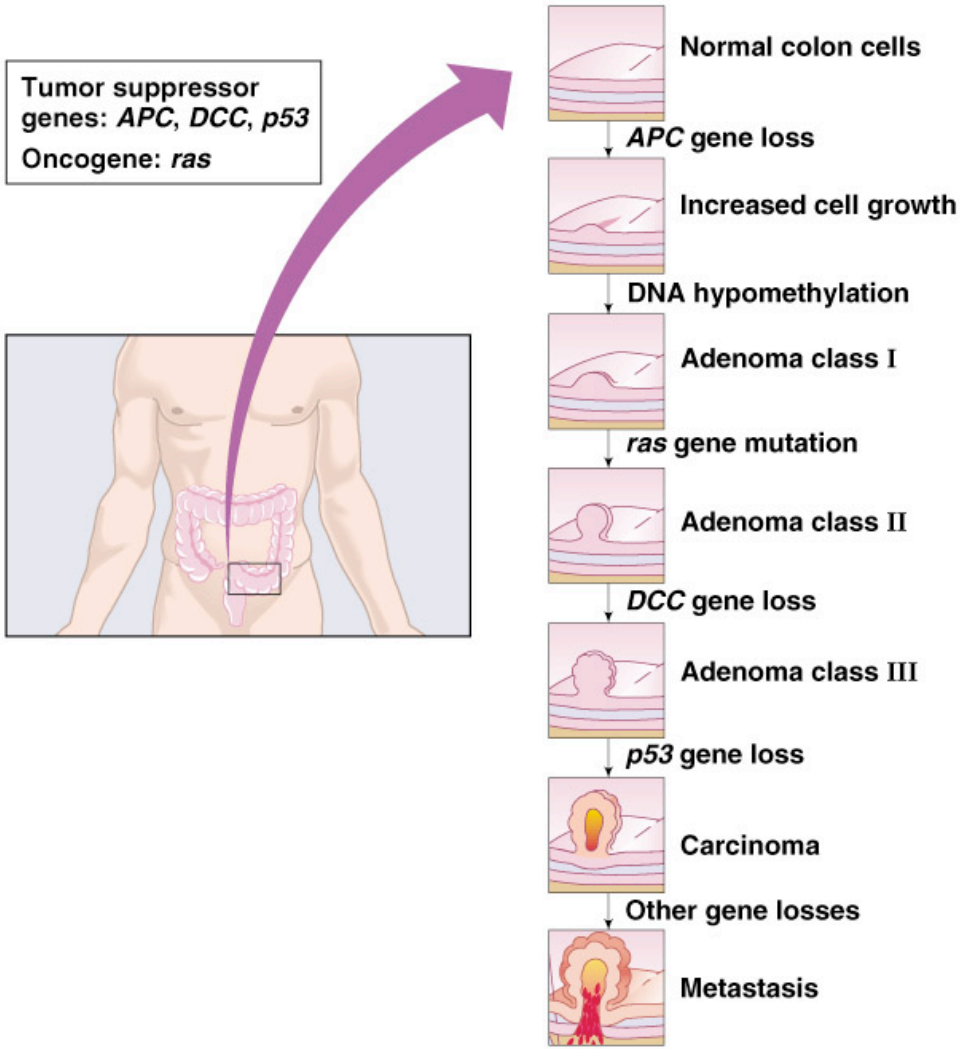
Alternate Splicing

Cancer Initiation and Progression

Proliferation, Motility, Immortality, Metastasis, Signaling, Microenvironment (autophagy)

Amplifications & Deletions

Pan Can



Mutation in a TSG

Epigenomics

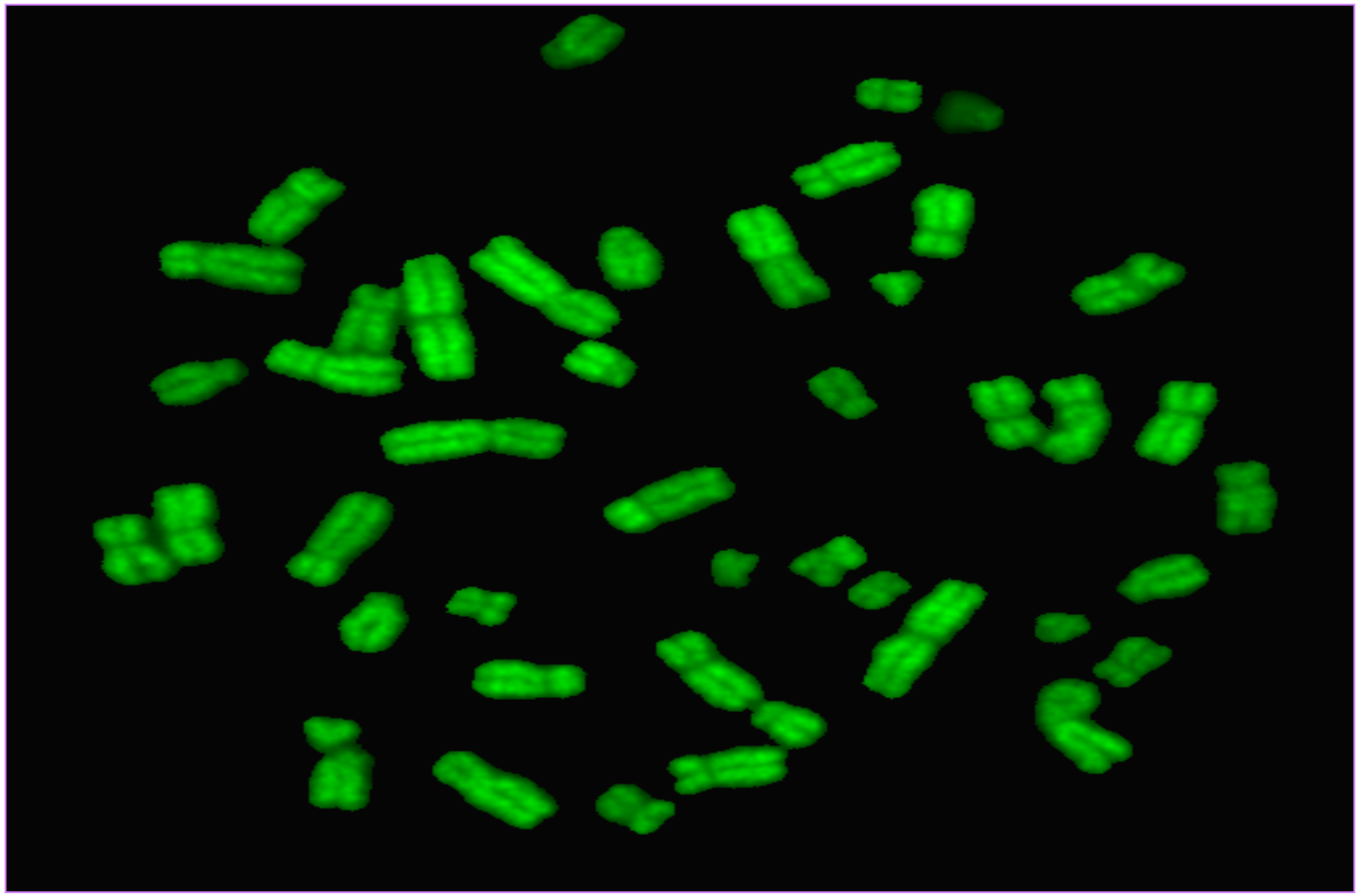
Conversion of a Proto-Oncogene

Deletion of a TSG

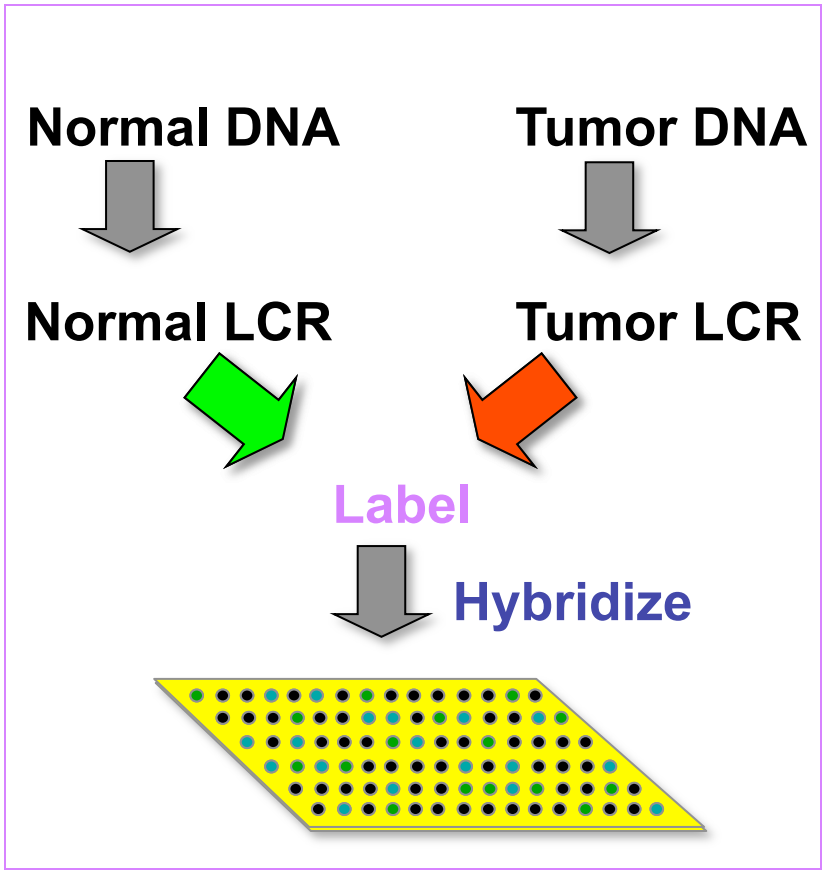
Deletion of a TSG

Karyotyping

Pan Can



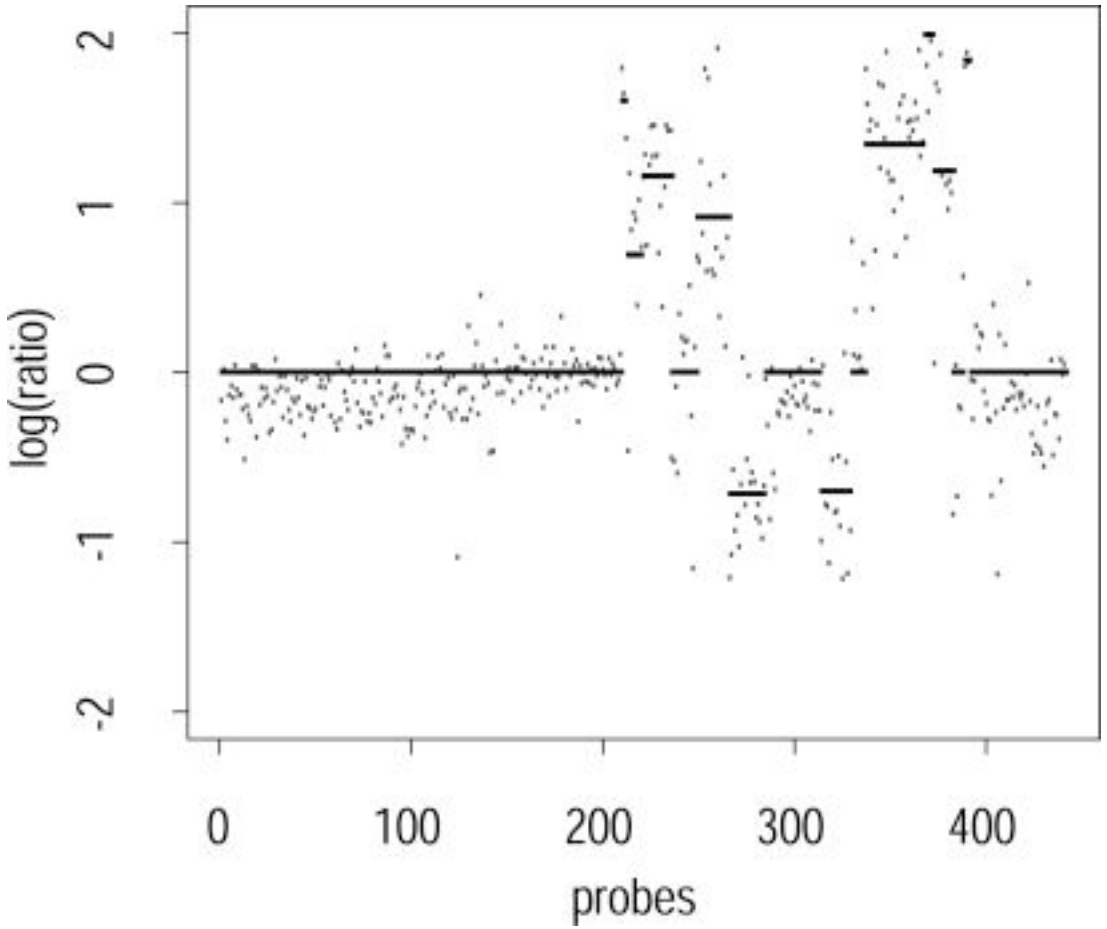
Microarray Analysis of Cancer Genome



- Representations are reproducible samplings of DNA populations in which the resulting DNA has a reduced complexity.
  - Array probes derived from low complexity representations of the normal genome
  - We measure differences in gene copy number between normal and tumor samples ratiometrically

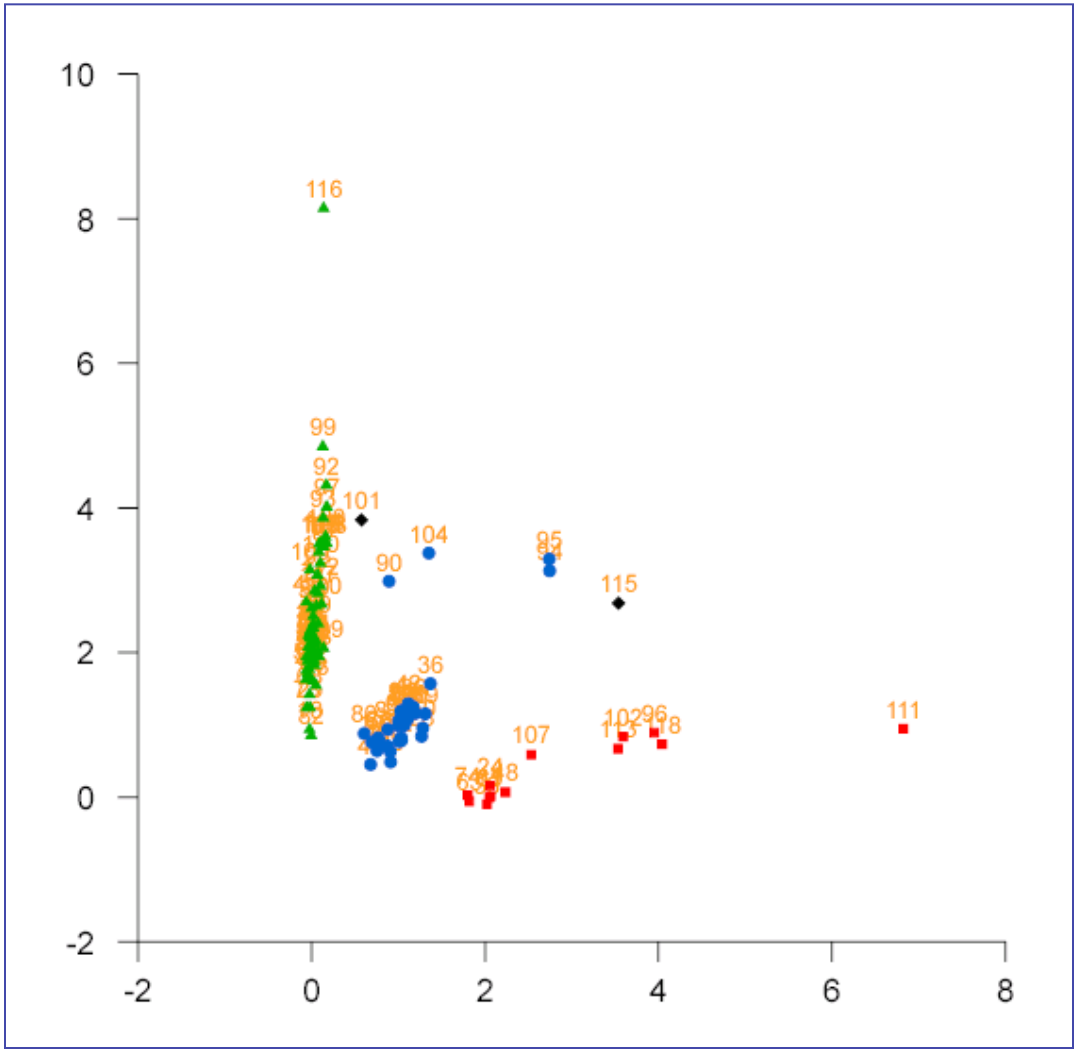
Daruwala et al. (PNAS, 2004)

Pan Can



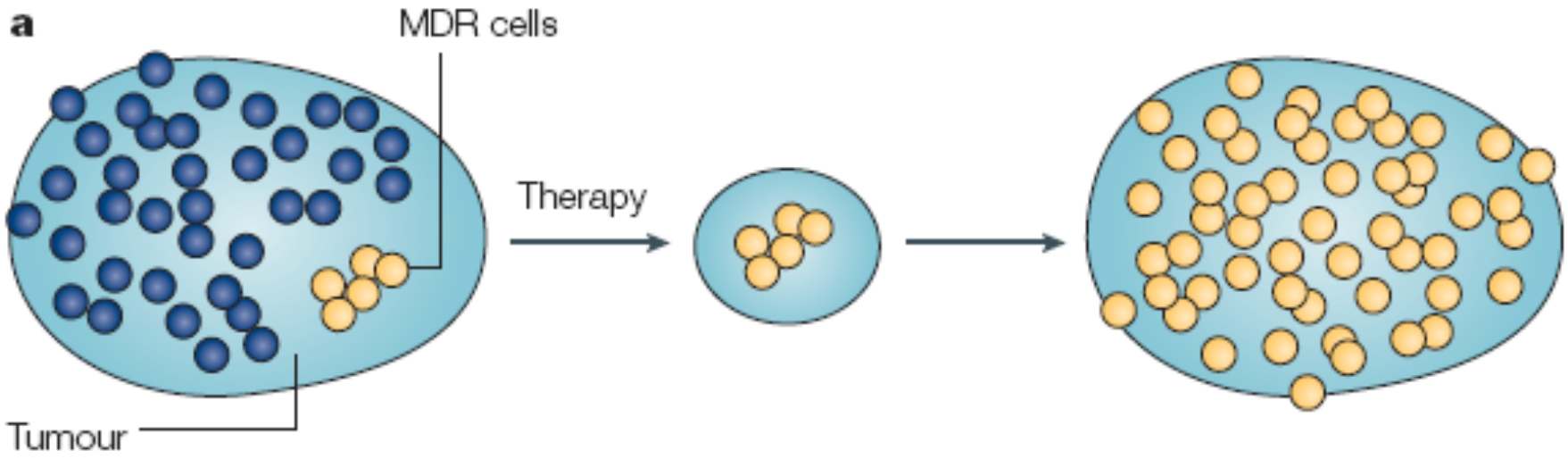
Allelic Frequencies: Cancer & Normal
(Anantharaman et al. unpublished)

Pan Can



Treatment Resistant Cell Subpopulations

Pan Can



Cell Stress: Glycosylation

Pan Can

- Some tumor-specific conditions (e.g., hypoxia, low pH and low level of glucose) commonly cause the glucose-regulated stress response of cancer cells.
- One can induce various stress responses in cancer cells artificially, and study them experimentally.
- For example, Tunicamycin induces (gylycosylation) stress:
  - It blocks the synthesis of all N-linked glycoproteins (N-glycans)
  - And causes cell cycle arrest in G1 phase.

Rapid Mass Change Detectable

Reed et al. (unpublished)

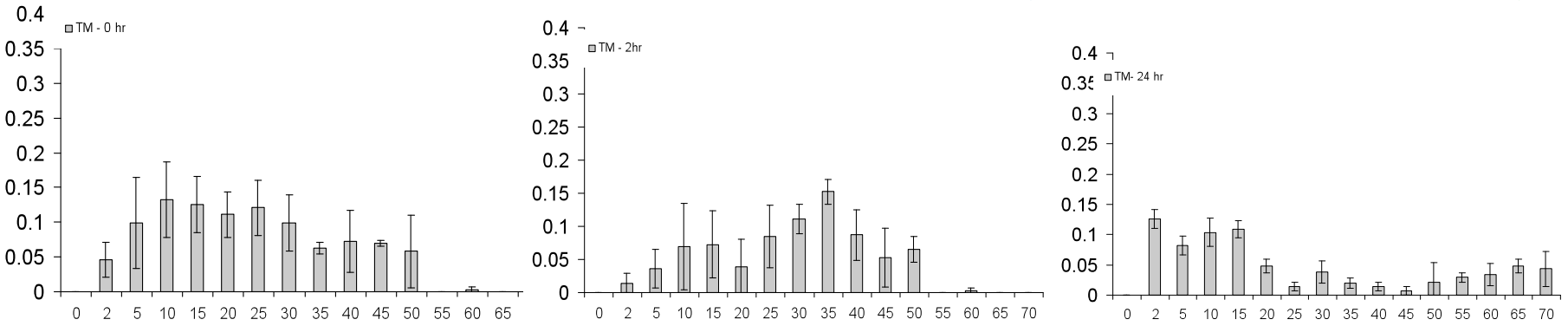
Pan Can

2 μm/mg tunicamycin

T = 0 hr

T = 2 hr

T = 24 hr

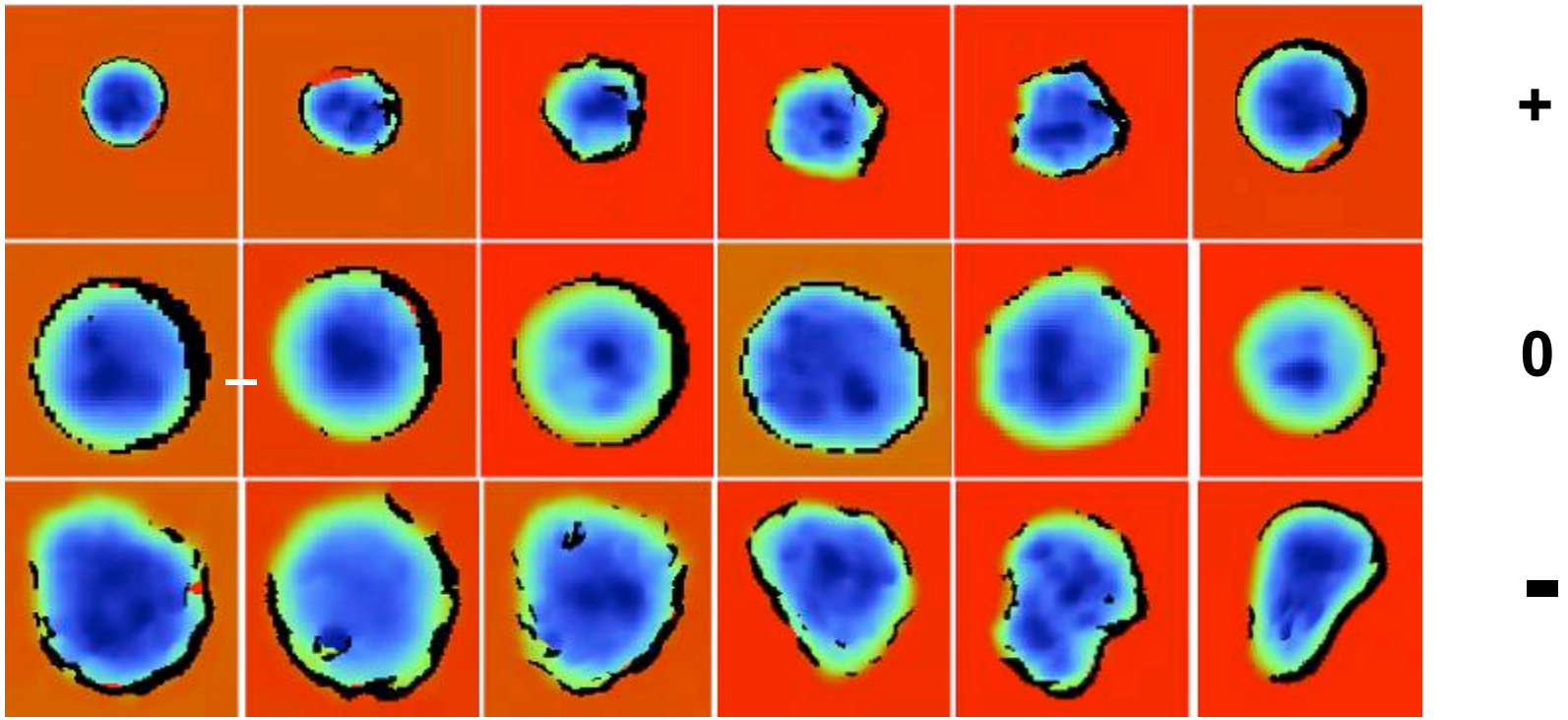


Cell Mass (Optical Density)

H929 myeloma cells

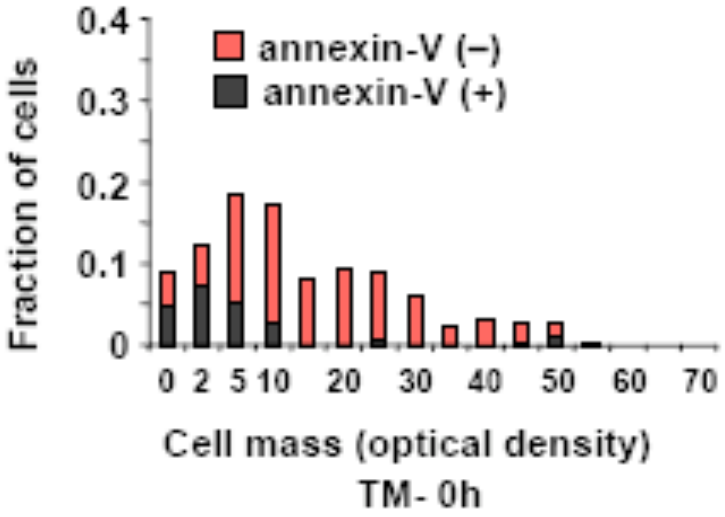
Treated Examples

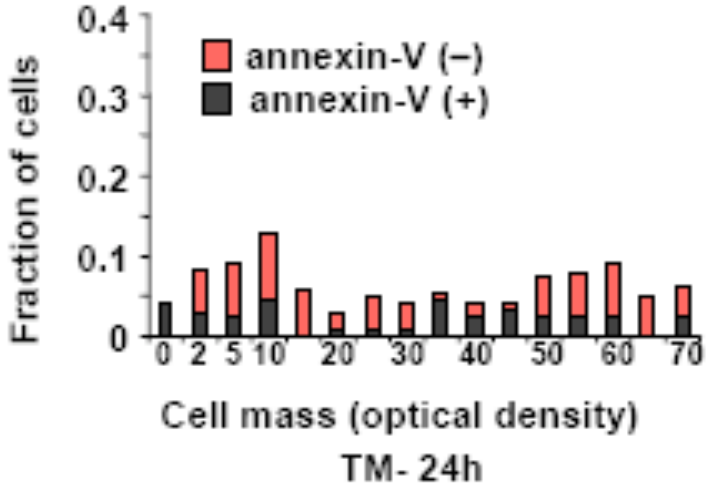
Pan Can



What is going on? Cell Death

Pan Can

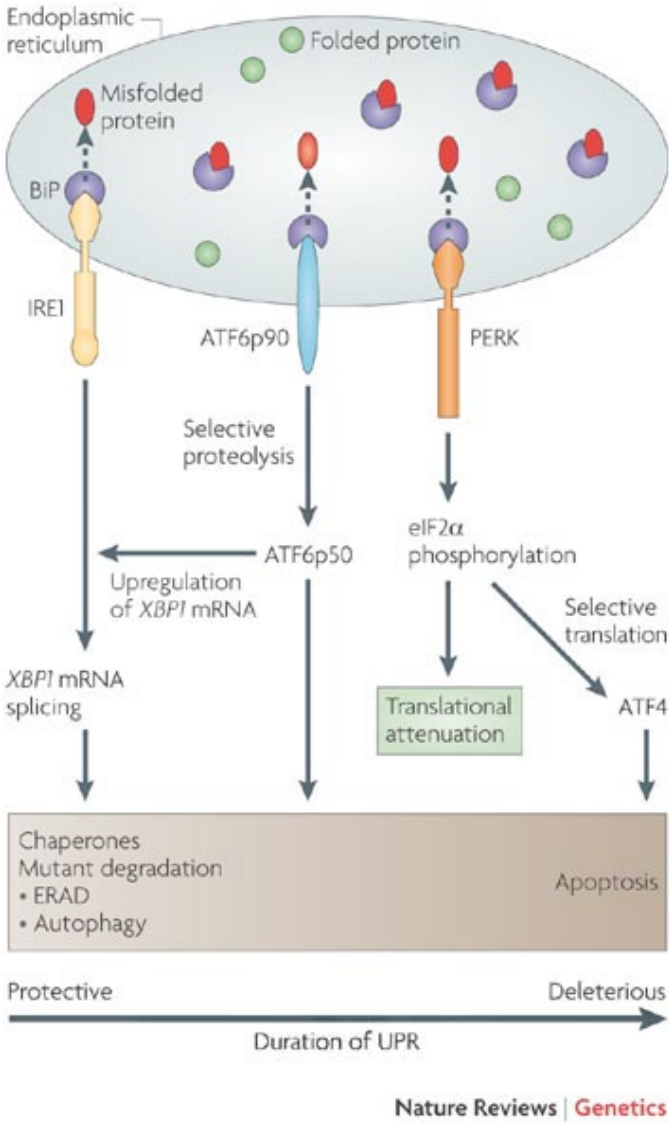




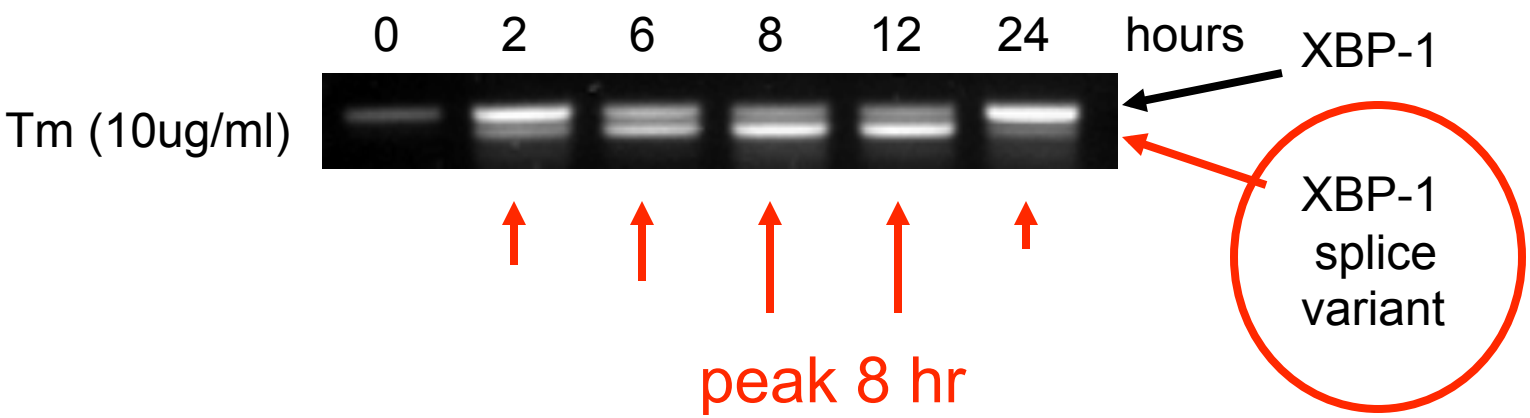
Apoptosis via Interferometry + Fluorescence

Autophagy and Apoptosis

Pan Can



What is going on?



Transcription factor: XBP-1

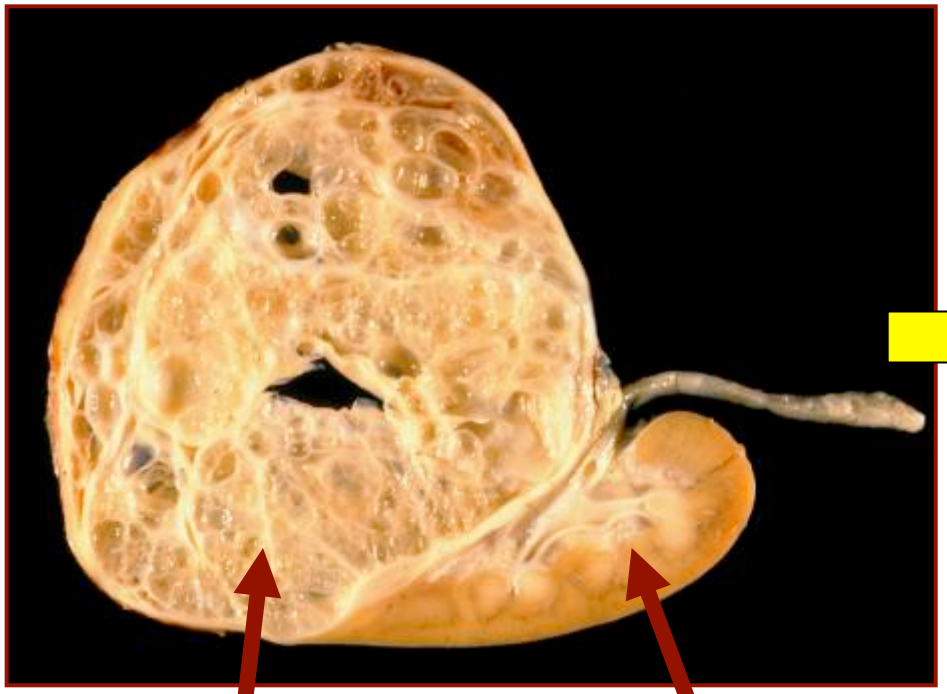
Dunno

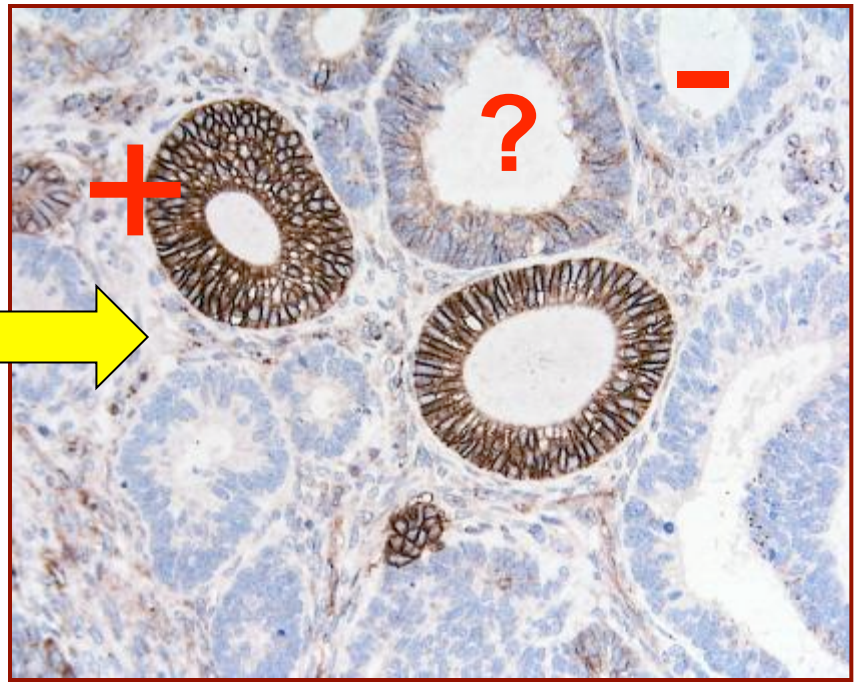
Pan Can

- **“Learn to say ‘I don't know.’**
- **“If used when appropriate, it will be often.”**
  - *Ex-US Secretary of Defense, Mr. Donald Rumsfeld*.

Sampling Problem

Pan Can





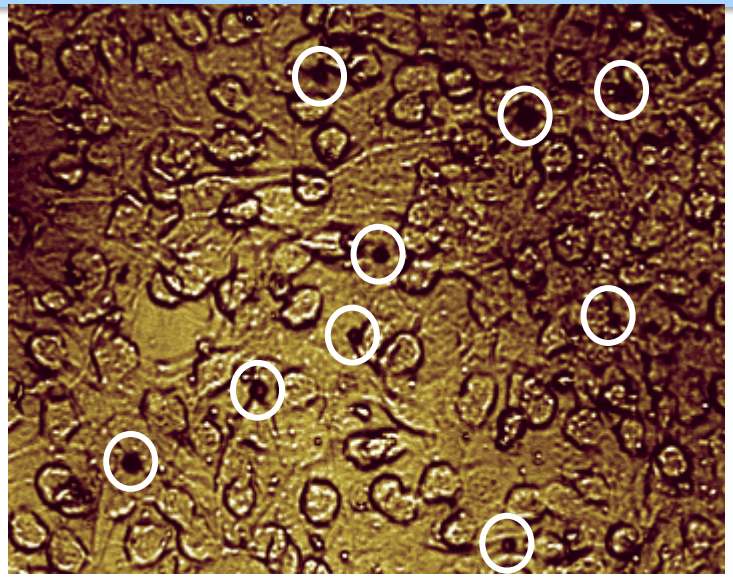
Tumor

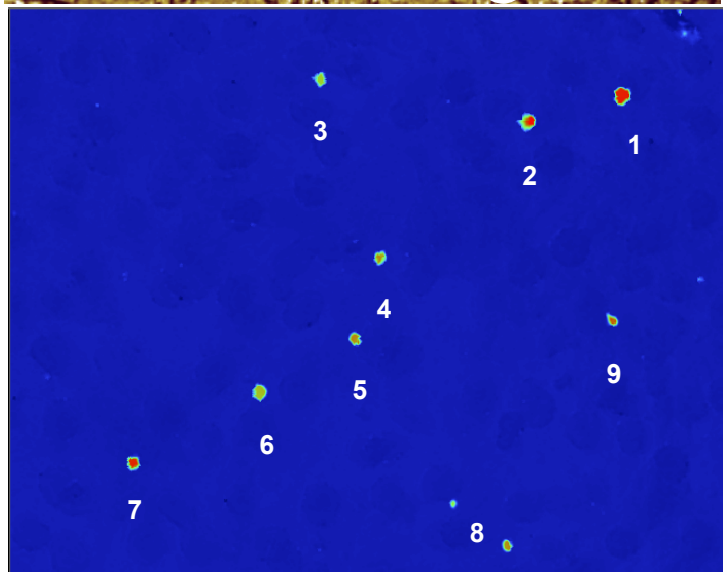
Unaffected kidney

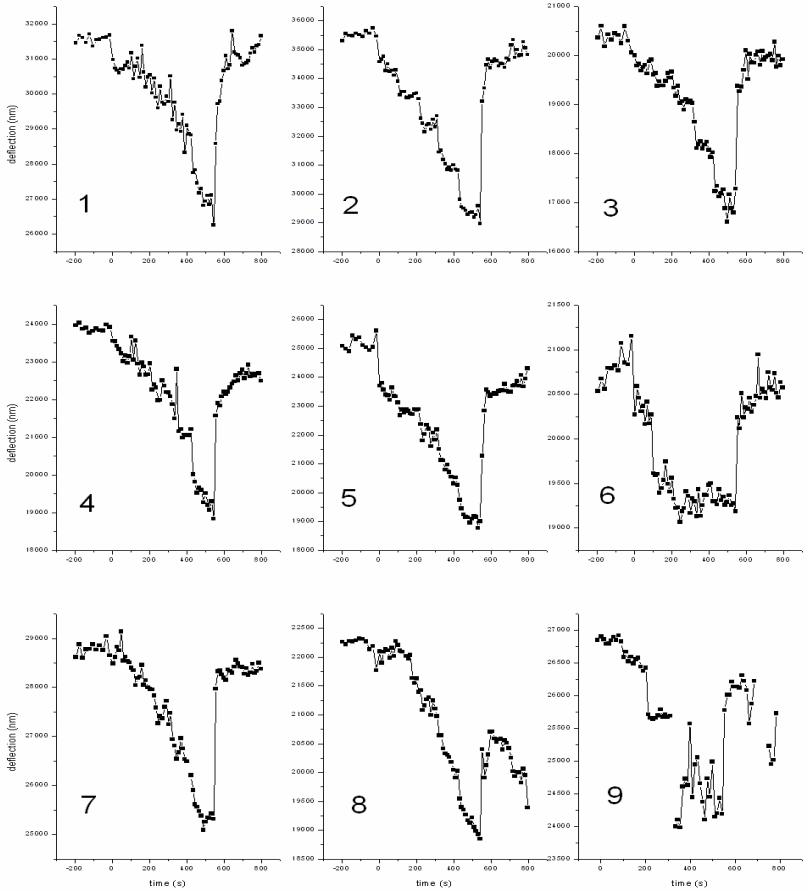
6μM thick tissue section; biomarker + (?)

Viscoelastic Profiling Single Cells

Pan Can



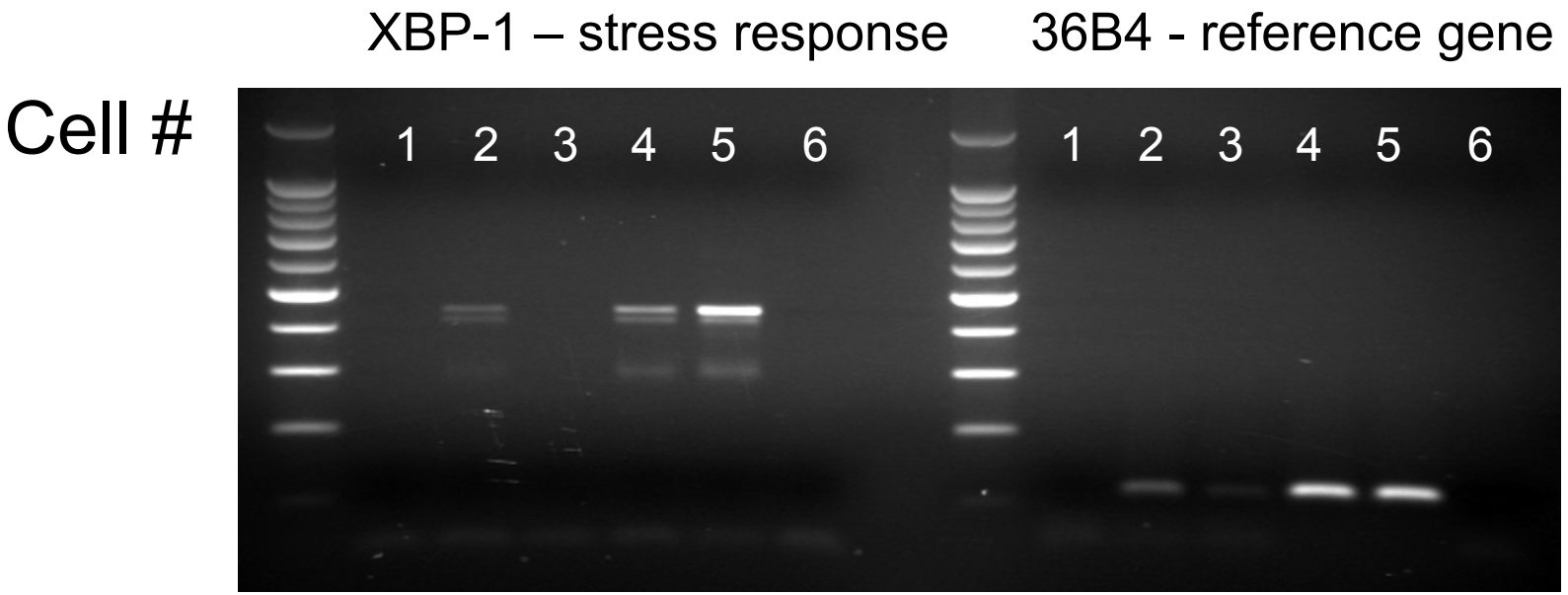




What to Do Next?

Pan Can

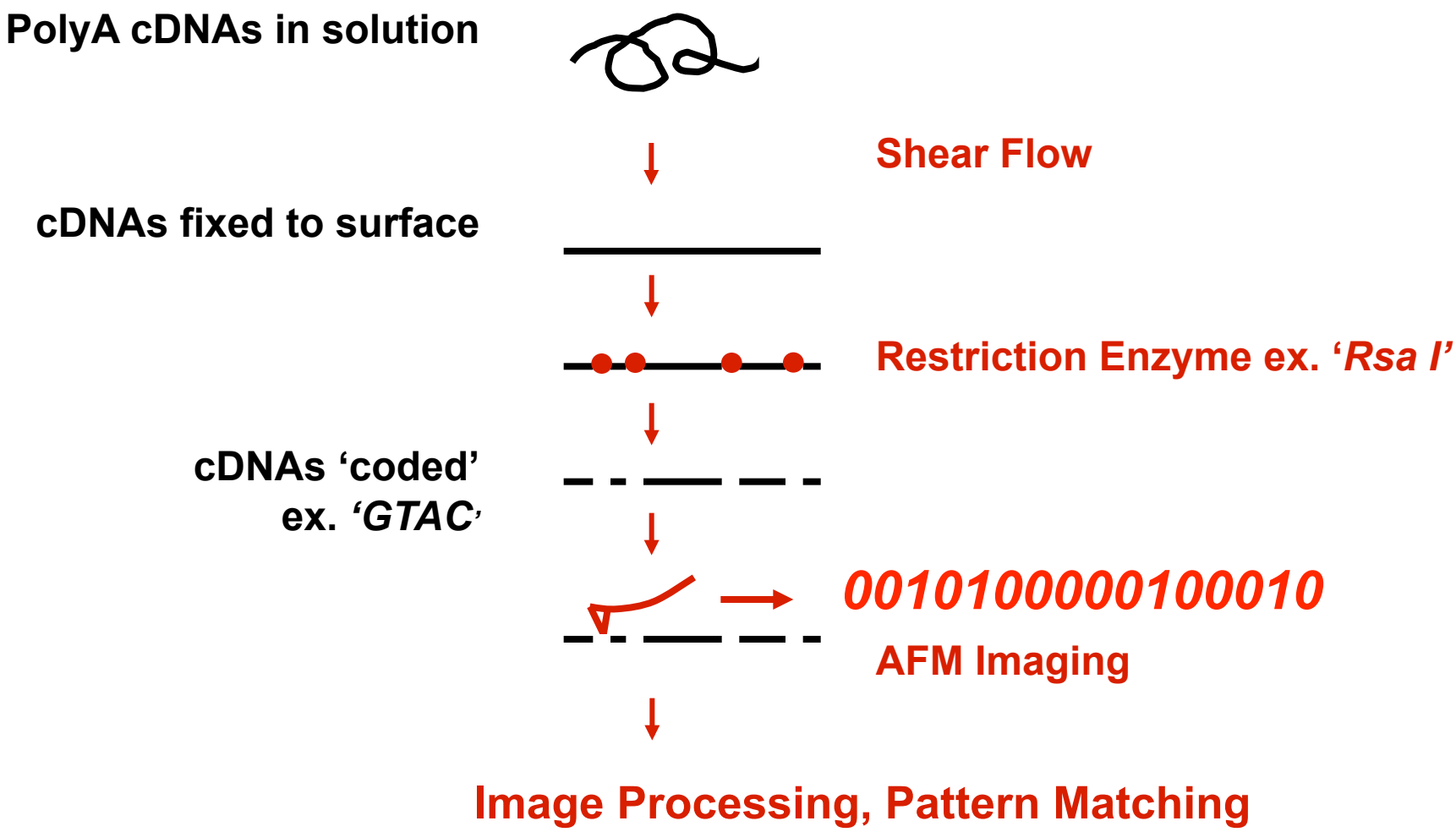
- Single Cell Molecular Profiling via RT-PCR



- Not so easy!

Concept
(M. et al. 2006-2009)

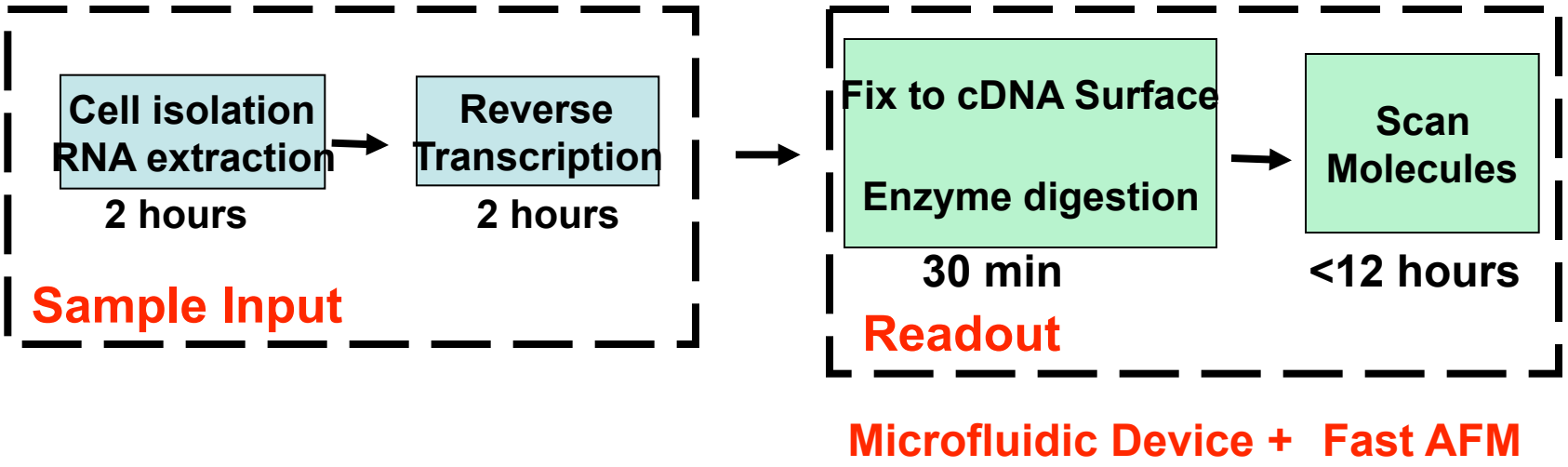
Pan Can



Reed, et al., Nanotechnology, 2007

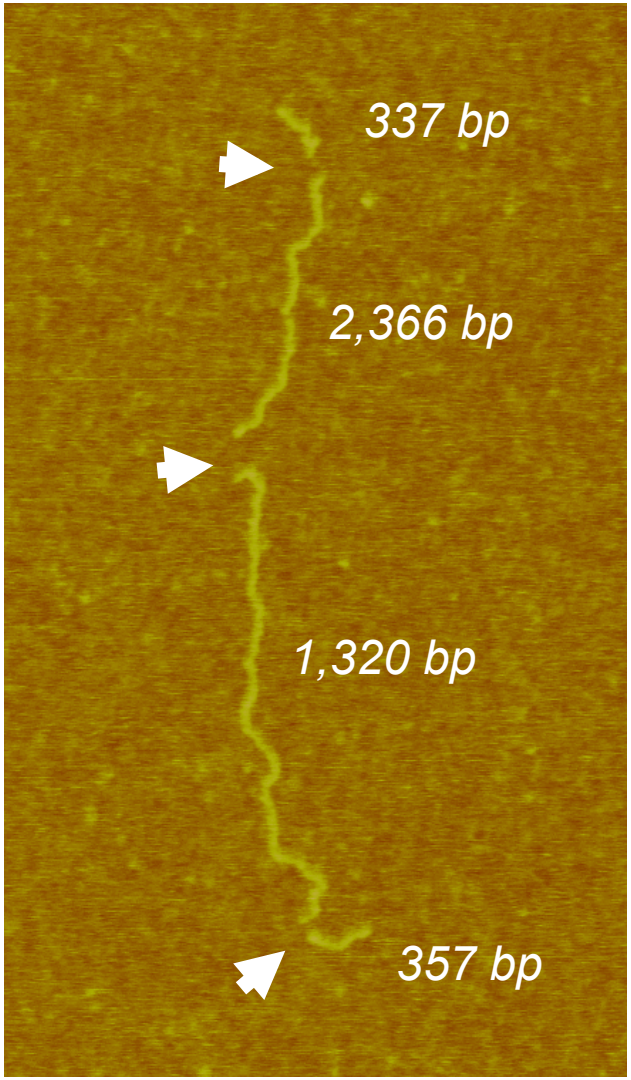
Single Molecule Restriction Map

Pan Can

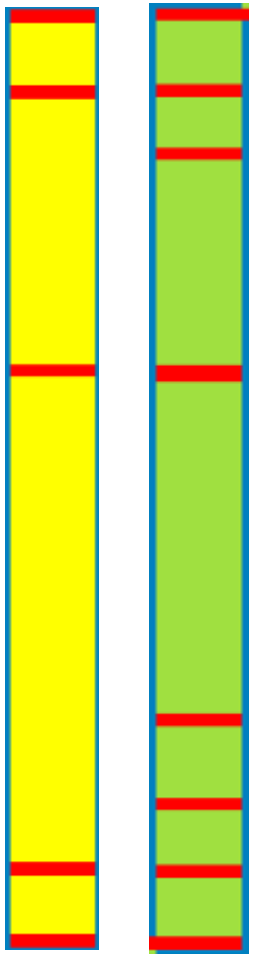


AFM vs Sequence

Pan Can



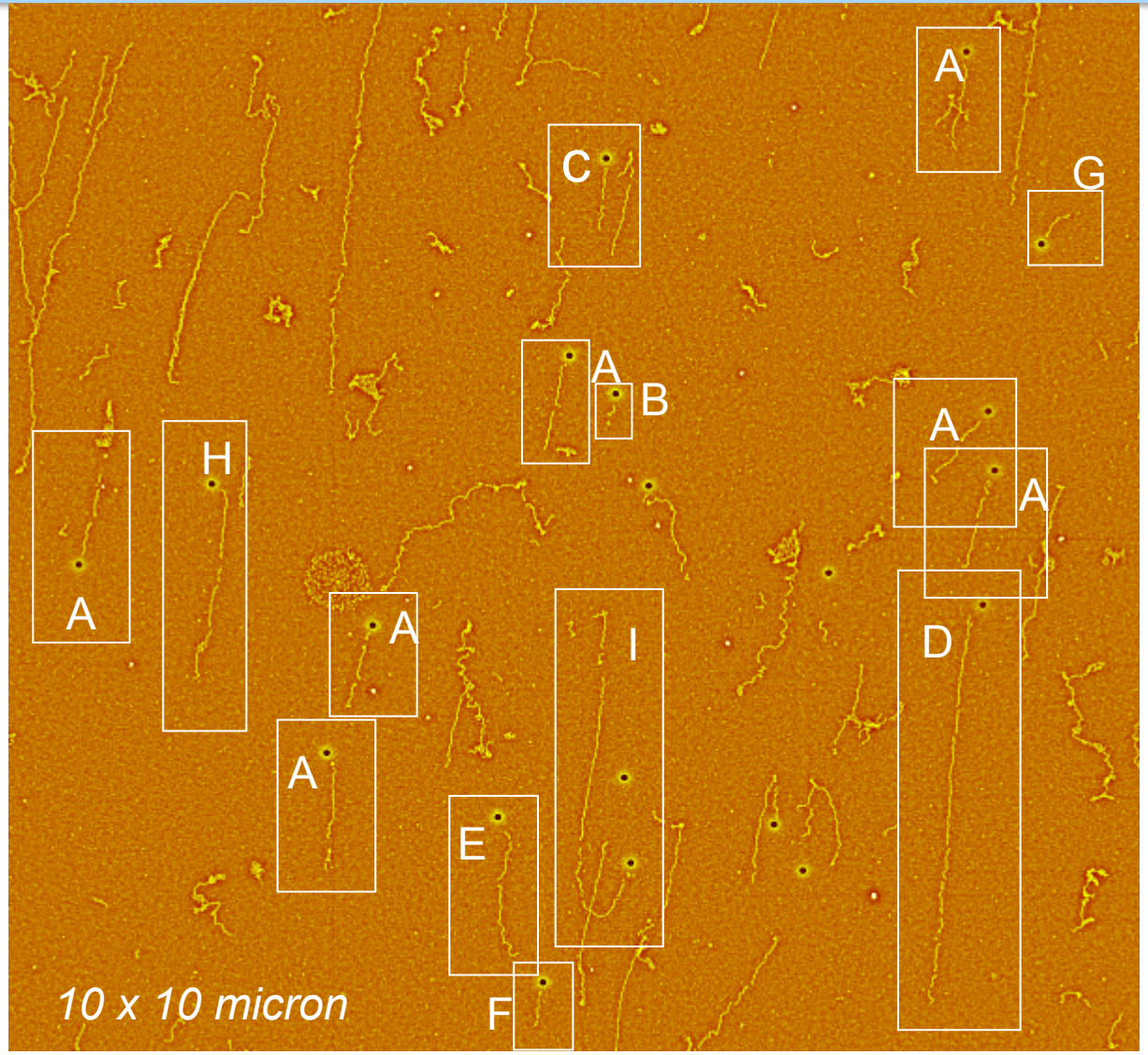
870 x 1,500 nm



AFM **Seq.**

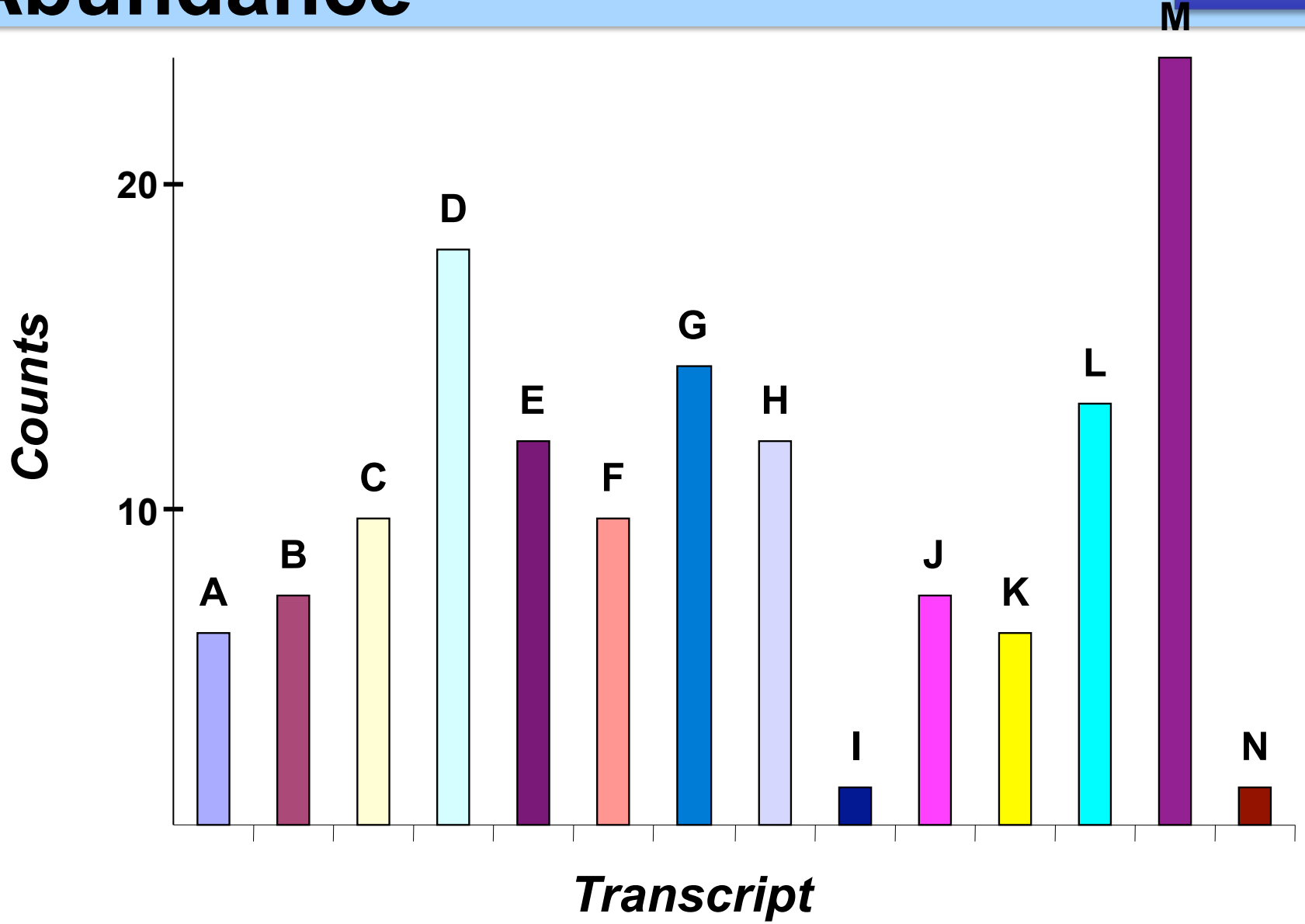
Identify and Count

Pan Can

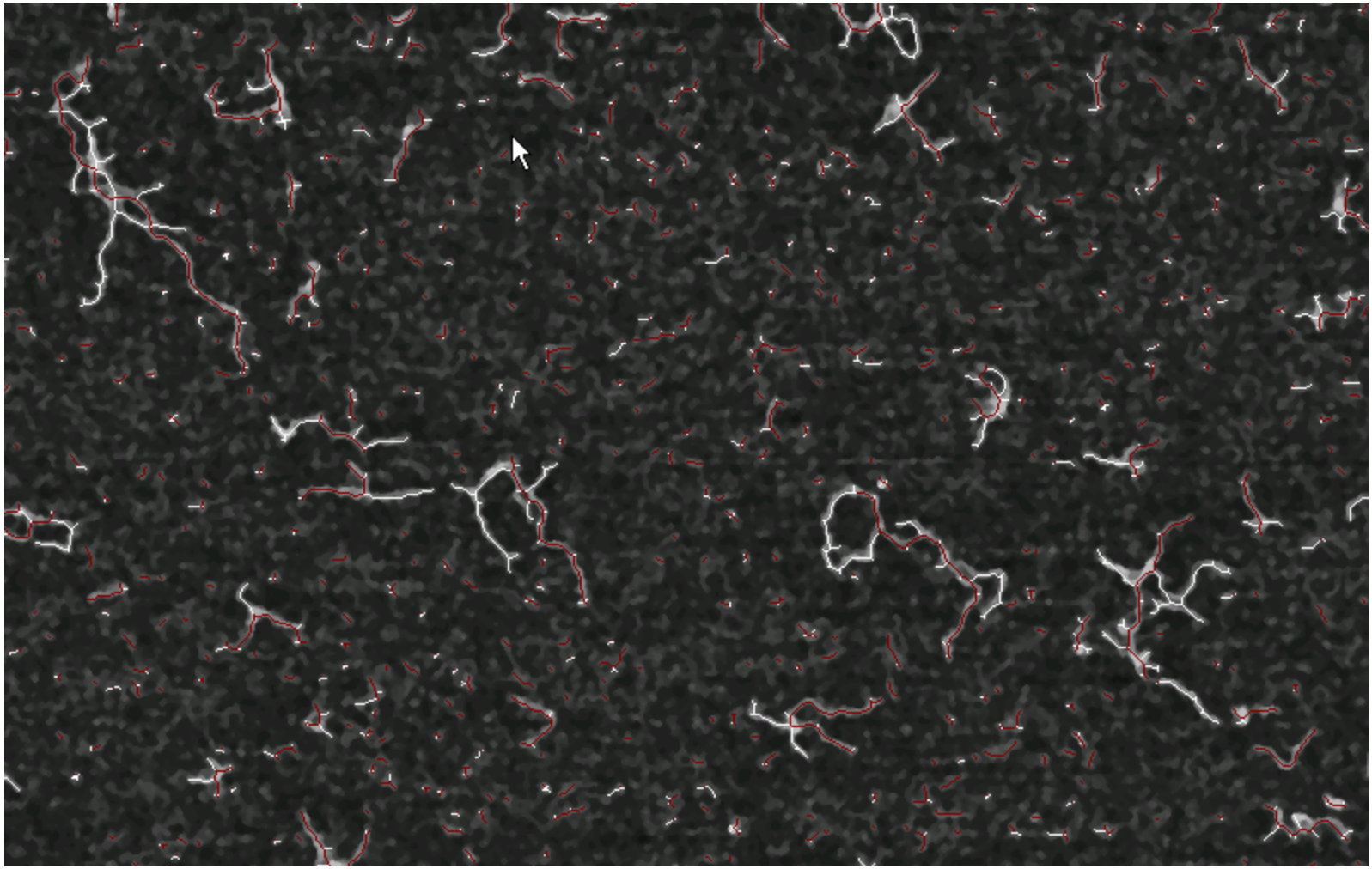


Histogram of Transcript Abundance

Pan Can

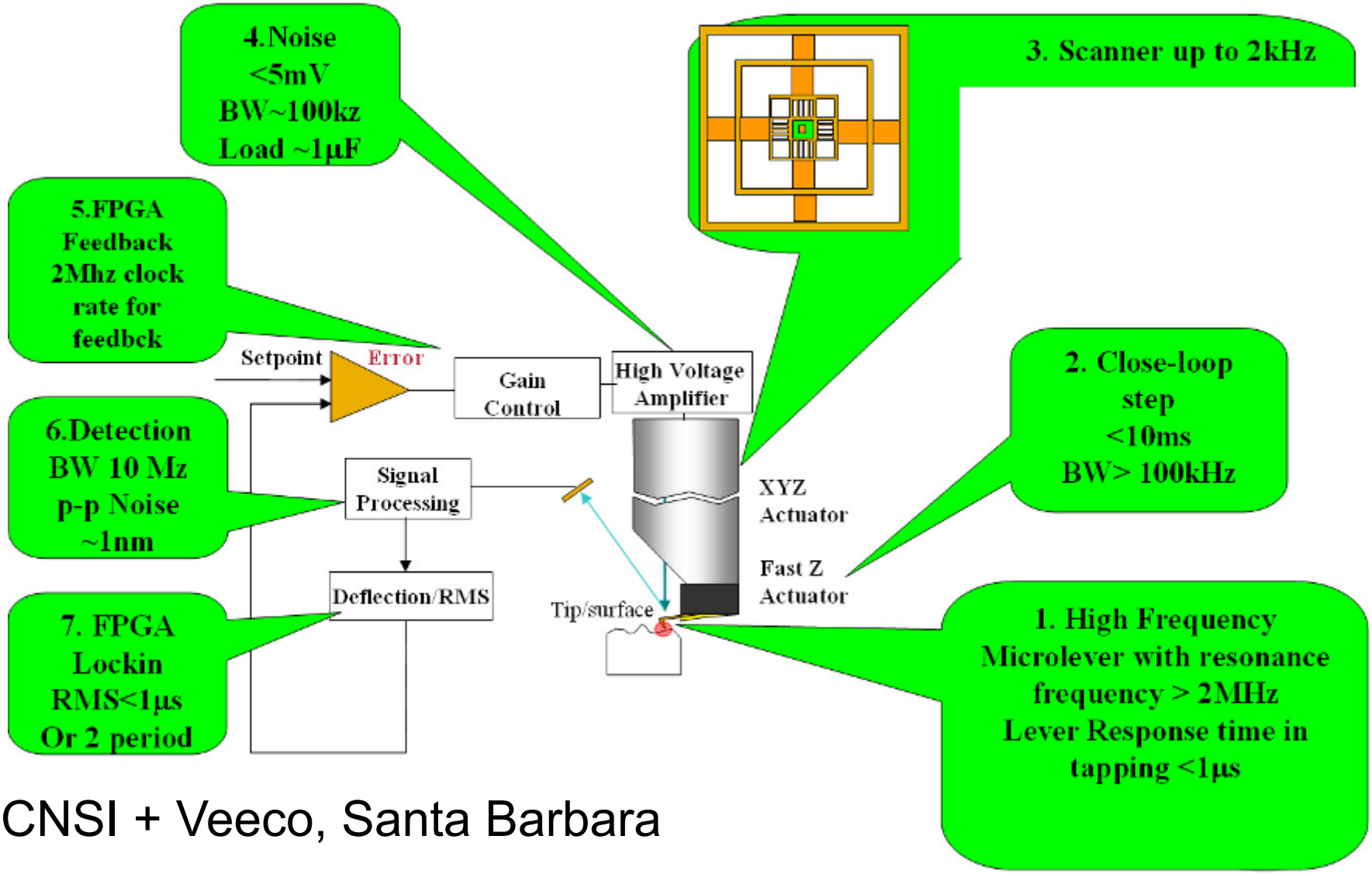


Pan
Can



How to Speed Up an AFM

Pan Can



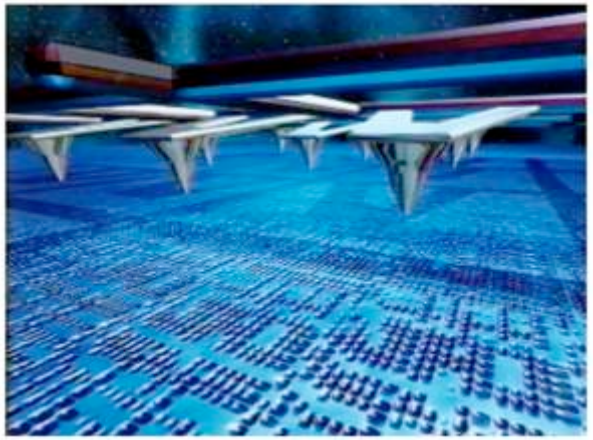
CNSI + Veeco, Santa Barbara

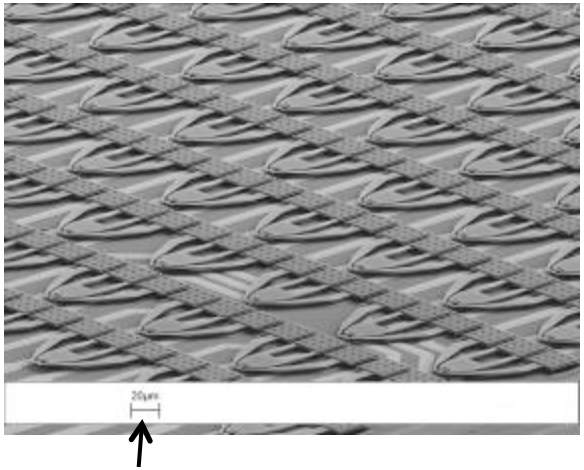
IBM Millipede

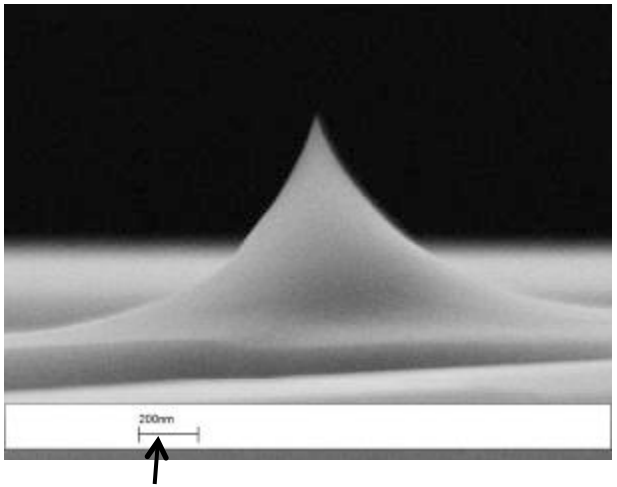
Pan Can

0.05 μsec pixel readout

1,000x faster







Models that are Concepty

Pan Can

- **“I’m not into this detail stuff.**
- **“I'm more concepty.”**
  - *Ex-US Secretary of Defense, Mr. Donald Rumsfeld, Once again quoted completely out of context*.

Hidden Kripke Model

- "Hidden Kripke Model"
  - Reconstruction via ontology based redescription of time-sliced clusters of time-course measurements (arrays)
  - Information Bottle Neck: Parsimony
- Example: Kripke Models
  - Spellman's Yeast Cell Cycle
  - SEB host-pathogen data from WRAIR
  - P. falciparum dataset [Bozdech et al, 1(1):085]
  - Genome Module Map dataset [Segal et al]

Lossy Compression

- Kripke Model as a compressed representation of the true dynamics
- Rate Distortion Theory of Shannon & Kolmogorov (ca. 1948)
  - Trade-off between “rate” (succinctness/compressibility) vs. “distortion”

Information Bottleneck

- Construct the Hidden Kripke Model using the following:
  - the clusters and cluster-edges must optimize the mutual information terms:

minimize:

$$I(D_i;X_i) - \beta_1 I(O|X_i; O|X_{i+1}) - \beta_2 I(O|X_i; O|X_{i-1})$$

- Notice that, conditional on D_i, O is independent of X_i. Blahut-Arimoto reduces to EM-style alternating algorithm
  - First cluster each D_i, identify connections across clusters in neighboring time points
  - Use these connections to derive new constraints on clustering, and re-cluster.

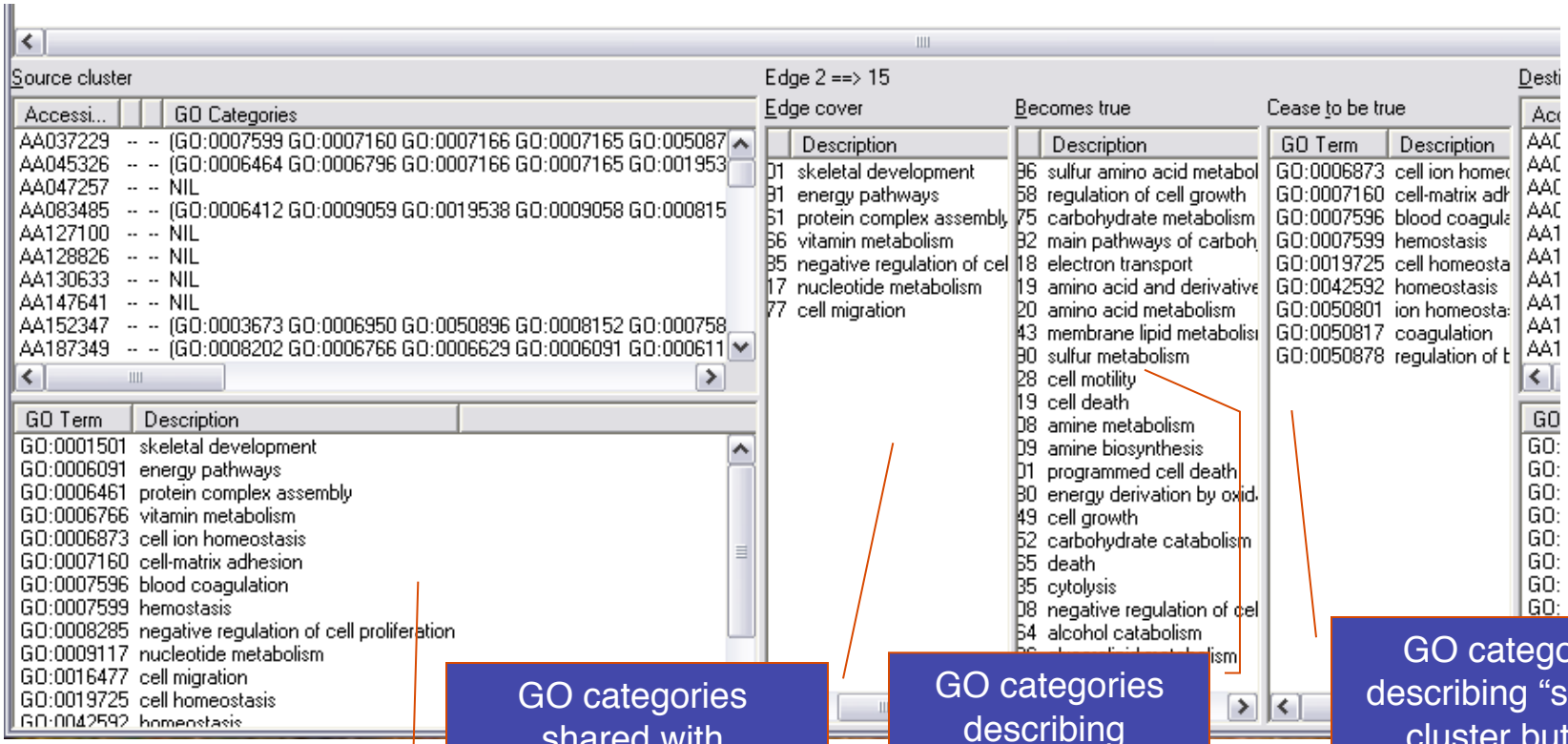
State-Labeling

Pan Can

- Simultaneously test N null hypotheses, one for each gene ontology labeling
  - H_j: no association between the state of the Kripke Model and an associated gene ontology label
  - Because there are many ontology process labels, there is a **large multiplicity issue**
  - **Brad Efron's Empirical Bayes FDR**

GOALIE: GO Algorithmic Logic for Invariant Extraction

Pan Can



GO categories describing genes in "source" cluster

GO categories shared with "destination" cluster

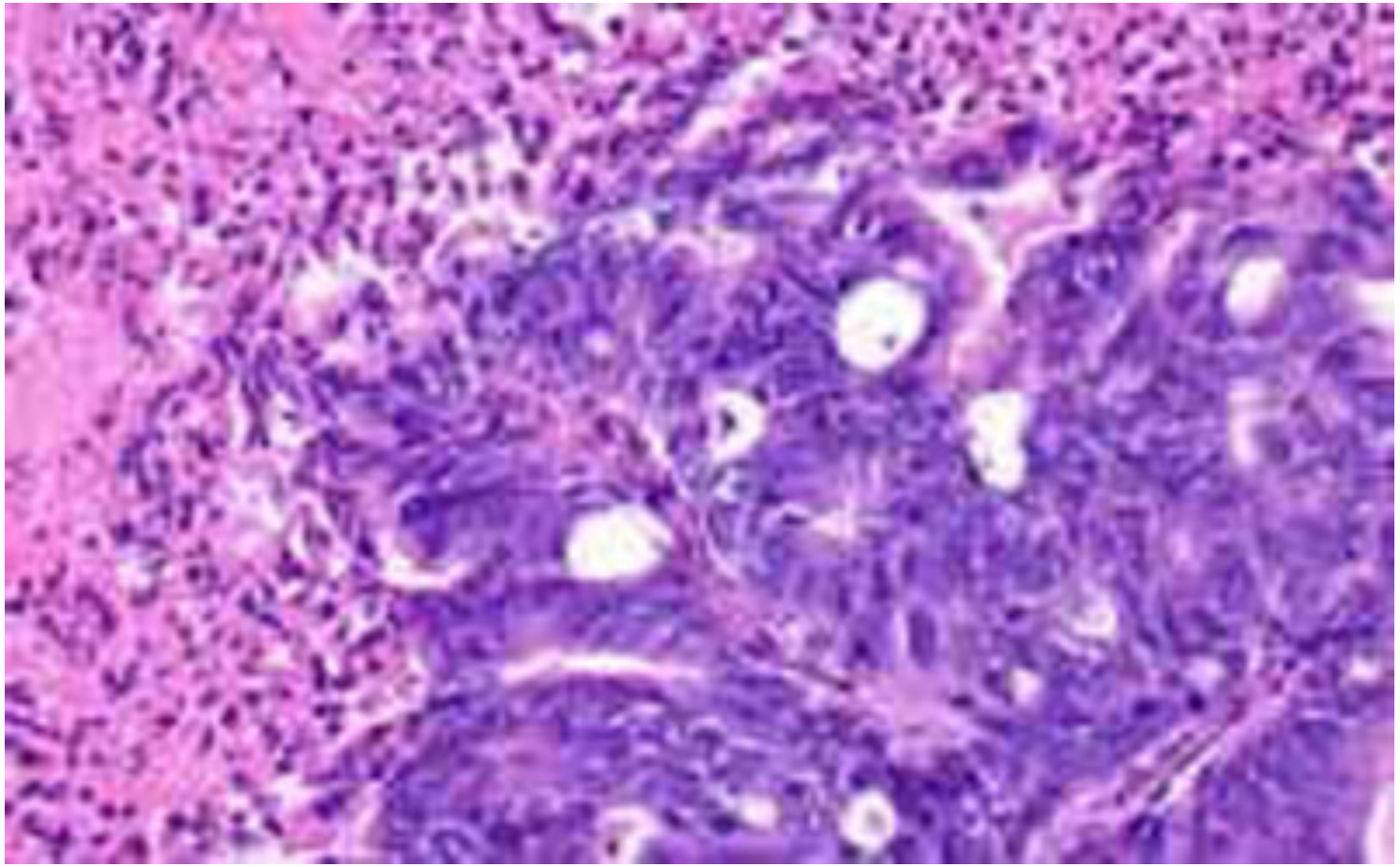
GO categories describing "destination" cluster but not "source"

GO categories describing "source" cluster but not "destination"

Unknown Unknown Biology

Pathologist's View

Pan Can



Healthy and diseased pancreas cells

A Challenge

Pan Can

- **"At present, description of a recently diagnosed tumor in terms of its underlying genetic lesions remains a distant prospect. Nonetheless, we look ahead 10 or 20 years to the time when the diagnosis of all somatically acquired lesions present in a tumor cell genome will become a routine procedure."**
  - Douglas Hanahan and Robert Weinberg
    - *Cell*, Vol. **100**, 57-70, 7 Jan 2000


Blast from the Past

Pan Can

- **"I would not say that the future is necessarily less predictable than the past. I think the past was not predictable when it started."**
  - *Ex-US Secretary of Defense, Mr Donald Rumsfeld.*

Foci

- Measurements
  - Single Cell Single Molecule Experiments
- Modeling & Model Checking
  - Phenomenological & Mechanistic Models
- Mining
  - Hypotheses
- Manipulation
  - Diagnostics and Therapeutics

Translational Systems Biology

- "A Sense of Life: Computational & Experimental Investigations with Models of Biochemical & Evolutionary Processes," (with R. Daruwala, Y. Zhou, N. Ugel, A. Policriti, M. Antoniotti, S. Paxia, M. Rejali, A. Rudra, V. Cherepinsky, N. Silver, W. Casey, C. Piazza, M. Simeoni, P. Barbano, M. Spivak, J-W. Feng, O. Gill, M. Venkatesh, F. Cheng, B. Sun, I. Ioniata, T.S. Anantharaman, E.J.A. Hubbard, A. Pnueli, D. Harel, V. Chandru, R. Hariharan, M. Wigler, F. Park, S.-C.. Lin, Y. Lazebnik, F. Winkler, C. Cantor, A. Carbone, and M. Gromov), *OMICS - A Journal of Integrative Biology*, (Special Issue on BioCOMP, Ed.: S. Kumar), 7(3): 253-268, 2003.
- "From Bytes to Bedside: Computational Biology for Biomedical Translational Research," (with J.P. Mathew, A. Chinnaiyan, G. Bader, S. Pyarajan, B. Taylor, M. Antoniotti, C. Sander and S.J. Burakoff)*, PLoS Computational Biology*, 3(2): 1-12, 2007.
- "Metamorphosis: The Coming Transformation of Translational Systems Biology," (with S. Kleinberg), *ACM Queue* 2009.

Models of Apoptosis

- "Mathematical Modeling of the formation of Apoptosome in Intrinsic Pathway of Apoptosis," (with S. Ryu et al.)*, Systems and Synthetic Biology Journal*, 2009.
- "The Apoptotic Machinery As A Biological Complex System: Analysis Of Its Omics And Evolution, Identification Of Candidate Genes For Fourteen Major Types Of Cancer And Experimental Validation in CML And Neuroblastoma," (with C. Di Pietro et al.), *BMC Medical Genomics*, 2009.

Model Checking in Biology

- "xS-systems: eXtended S-systems and Algebraic Differential Automata for Modeling Cellular Behavior," (with M. Antoniotti, A. Policriti and N. Ugel), *High Performance Computing--HiPC 2002*, (Eds. S. Sahni, V.K. Prasanna & U. Shukla), LNCS 2552:431-442, Springer-Verlag, December 2002.
- "Model Building and Model Checking for Biochemical Processes," (with M. Antoniotti, A. Policriti and N. Ugel), *Cell Biochemistry and Biophysics* (CBB), 38(3): 271-286, Humana Press, June, 2003.
- "Taming the Complexity of Biochemical Models through Bisimulation and Collapsing: Theory and Practice," (with M. Antoniotti, C.Piazza, A. Policriti and M. Simeoni), *Theoretical Computer Science*, 325(1): 45-67, 2004.
- "Simpathica: A Computational Systems Biology Tool within the Valis Bioinformatics Environment," (with M. Antoniotti, S. Paxia and N. Ugel), *Computational Systems Biology*, (Ed. E. Eiles and A. Kriete), Elsevier, 2005.
- "A Coherent Framework for Multi-resolution Analysis of Biological Networks with Memory: RAS pathway, Cell Cycle and Immune System," (with P. Barbano, M. Spivak, J. Feng, and M. Antoniotti), *Proc. National Academy of Science U S A*, 102(18): 6245-6250, 2005.

Algorithmic Algebraic Model Checking

- "Algorithmic Algebraic Model Checking I: Challenges from Systems Biology," (with C. Piazza, M. Antoniotti, V. Mysore, A. Policriti, and F. Winkler), *17th International Conference on Computer Aided Verification*, (The University of Edinburgh, Scotland, UK, July 6 - 10 , 2005), CAV 2005:5-19, 2005.
- "Algorithmic Algebraic Model Checking II: Decidability of Semi-Algebraic Model Checking and its Applications to Systems Biology," (with V. Mysore and C. Piazza), *Automated Technology for Verification and Analysis*: (Taipei, Taiwan, October 4 - 7, 2005), ATVA 2005: 217-233, 2005.
- "Algorithmic Algebraic Model Checking III: Approximate Methods," (with V. Mysore), *7th International Workshop on Verification of Infinite-State Systems*, INFINITY 05, San Francisco, California, USA, August 27, 2005. Electr. Notes Theor. Comput. Sci., 149(1):61-77, 2006.
- "Algorithmic Algebraic Model Checking IV: Characterization of Metabolic Networks," (with V. Mysore), *Algebraic Biology*, AB '2007, Linz, Austria, July 2007.
- "Intelligently Deciphering Unintelligible Designs: Algorithmic Algebraic Model Checking in Systems Biology," (Invited Paper), *Interface: Journal of the Royal Society,* 2009.

Optical Mapping

Pan Can

- “Mapping the Genome One Molecule at a Time -- Optical Mapping,” (with A.H. Samad et al.), *Nature*, 378:516-517, 1995
- “Genomics via Optical Mapping II: Ordered Restriction Maps,” (with T.S. Anantharaman and D.C. Schwartz), *Journal of Computational Biology*, 4(2):91-118, 1997.
- “Genomics via Optical Mapping III: Contiging Genomic DNA and Variations,” (with T.S. Anantharaman and D.C. Schwartz), *Proceedings 7th Intl. Cnf. on Intelligent Systems for Molecular Biology*: ISMB '99 , 7:18-27, AAAI Press, 1999.
- “Optical Mapping and Its Potential for Large-Scale Sequencing Projects,” (with C. Aston and D.C. Schwartz), *Trends in Biotechnology*, 17:297-302, 1999.
- “Optical Mapping,” *Encyclopedia of the Human Genome*, 4: 448-453, Nature Publishing Group, Macmillan Publishers Limited, London, UK, June, 2003.

Copy Number Fluctuations

Pan Can

- “Detecting Gene Copy Number Fluctuations in Tumor Cells by Microarray Analysis of Genomic Representations,” (with R. Lucito et al.), *Genome Research*, 10(11): 1726-1736, 2000.
- “Comparing Genomes,” Special issue on "Biocomputation:" *Computing in Science and Engineering.* , pp 42-49, January/February 2002.
- “A Versatile Statistical Analysis Algorithm to Detect Genome Copy Number Variation,” (with R.-S. Daruwala, A. Rudra, H. Ostrer, R. Lucito, and M. Wigler), *Proc. National Academy of Science U S A*, 101(46): 16292-7, 2004.
- “Mapping Tumor Suppressor Genes using Multipoint Statistics from Copy-Number Variation Data,” (with I. Ionita and R. Daruwala), *American Journal of Human Genetics*, 79(1):13-22, July 2006.
- “Copy Number Variant Analysis of Human Embryonic Stem Cells,” (with H. Wu et al.), *Stem Cells*, 26(6):1484-9, June 2008.

Single Molecule/Single Cell Nanotechnology

Pan Can

- "Single Molecule Transcription Profiling with AFM," (with J. Reed, B. Pittenger, S. Magonov, J. Troke, M.A. Teitell, and J.K. Gimzewski), *Nanotechnology*, 18, 1-15, 2007.
- "Atomic Force Microscope Observation of Branching in Single Transcript Molecules Derived from Human Cardiac Muscle," (with J. Reed, C. Hsueh and J. Gimzewski), *Nanotechnology*, 19 384021 (8pp), 2008.
- "Image Analysis of Single Molecule Transcription Profiles with AFM," (with A. Sundstrom et al.), Submitted, 2009.

Ontology: GOALIE

Pan Can

- "Discovering Relations among GO-annotated Clusters by Graph Kernel Methods," (with I. Zoppis, D. Merico, M. Antoniotti, G. Mauri), *2007 International Symposium on Bioinformatics Research and Application*s, ISBRA '07, Atlanta, GA, 2007.
- "Systems Biology via Redescription and Ontologies (I): Finding Phase Changes with Applications to Malaria Temporal Data," (with S. Kleinberg and K. Casey), *Systems and Synthetic Biology Journal (SSB)*, 1(4): 197-205, 2008.
- "Systems Biology via Redescription and Ontologies (II): A Tool for Discovery in Complex Systems," (with S. Kleinberg et al.), *Proceedings of the International Conference on Complex Systems*, 2008.
- "Systems Biology via Redescription and Ontologies (III): Protein Classification using Malaria Parasite's Temporal Transcriptomic Profiles," (with A. Mitrofanova et al.), *2008 IEEE International Conference on Bioinformatics and Biomedicine*, BIBM'08, 2008.
- "Simultaneously Segmenting Multiple Gene Expression Time Courses by Analyzing Cluster Dynamics," (with S. Tadepalli, N. Ramakrishnan, L.T. Watson, and R.F. Helm), (Invited Paper) *Journal of Bioinformatics and Computational Biology (JBCB)*, 7(2): 339-356, 2009.
- "The Temporal Logic of Causal Structures," (with S. Kleinberg), Uncertainty in Artificial Intelligence, UAI 2009, Montreal, Quebec, Canada, 2009.

Answer to Cancer

Pan Can

- **“If I know the answer I'll tell you the answer, and if I don't, I'll just respond, cleverly.”**
  - *Ex-US Secretary of Defense, Mr. Donald Rumsfeld.*

The end