Cancer Systems Biology

Stuart A. Kauffman
iCore chair
Skauffman@ucalgary.ca

Institute for Biocomplexity and Informatics, University of Calgary, Canada
Department of Physics & Astronomy of the University of Calgary, Canada

University of Calgary
2500 University Drive NW, T2N 1N4
CANADA
http://www.ibi.ucalgary.ca/

Cancer Stem Cell DifferentiationTherapy

1. Cancer stem cells have been discovered in breast, prostate, skin, blood, brain and other tumors.

2. Cancer stem cells are capable of persistent self renewal and differentiating into other cells of limited proliferation potential in the tumor.

3. Recent evidence suggests that specific subsets of genes are abnormally expressed in cancer stem cells.

4. Aim of cancer stem cell therapy is to kill, stop the proliferation of, *or causedifferentiation of cancer stem cells.*

5. IBI lab is focusing on high throughput and specific siRNA, small molecules, and expression vector library screening.

Genetic Regulatory Networks

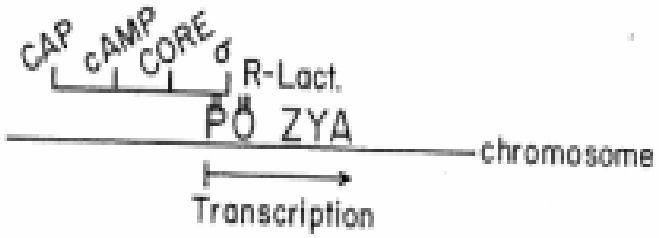
Transcriptome in Yeast regulates 6500 genes

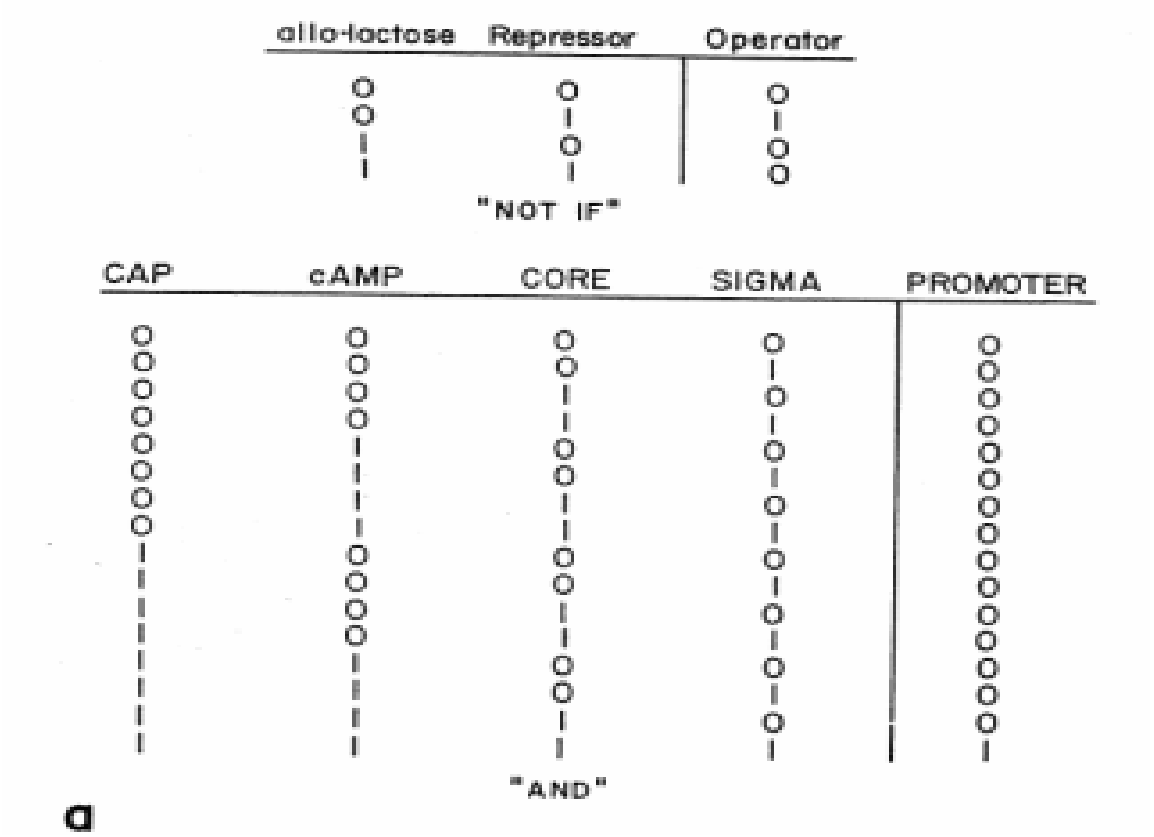
Transcriptome in Humans regulates about 25000 genes

Transcriptome plus Protein signaling network is a parallel processing non linear stochastic dynamical system.

Systems Biology seeks the integrated behavior of this system within and between cells.

Systems Biology also seeks possible general laws.

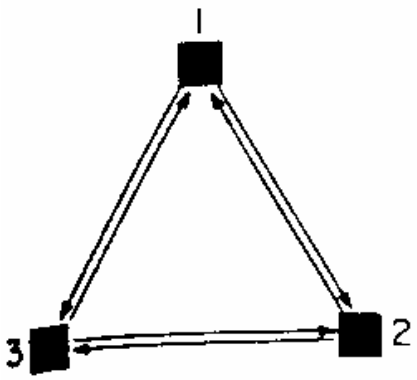




P	Op	ZYA
0	0	0
0	1	0
1	0	1
1	1	0

b

Boolean networks as models of GRN's



2	3	1
0	0	0
0	1	0
1	0	0
1	1	1

"AND"

1	2	3
0	0	0
0	1	1
1	0	1
1	1	1

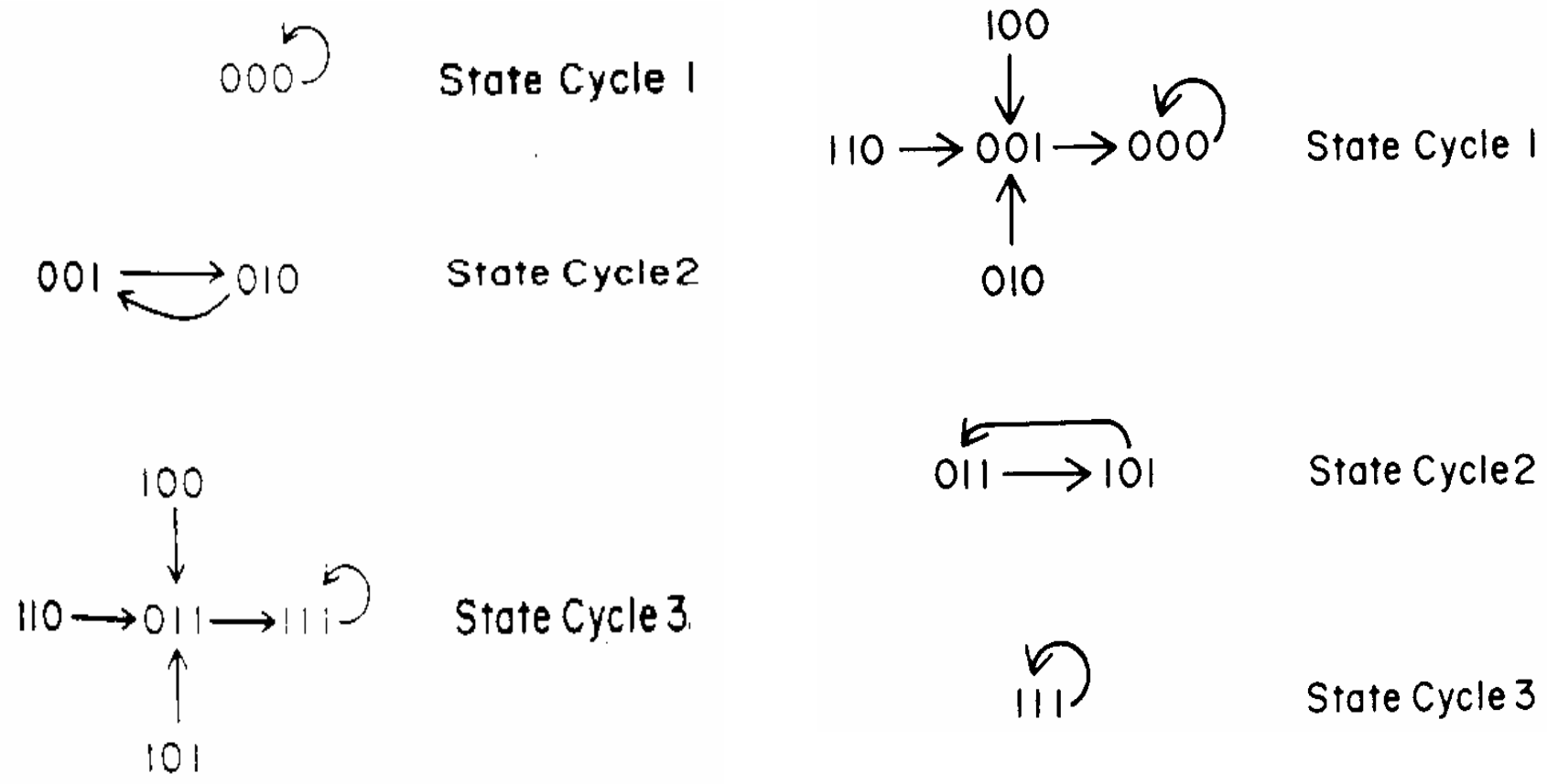
"OR"

1	3	2
0	0	0
0	1	1
1	0	1
1	1	1

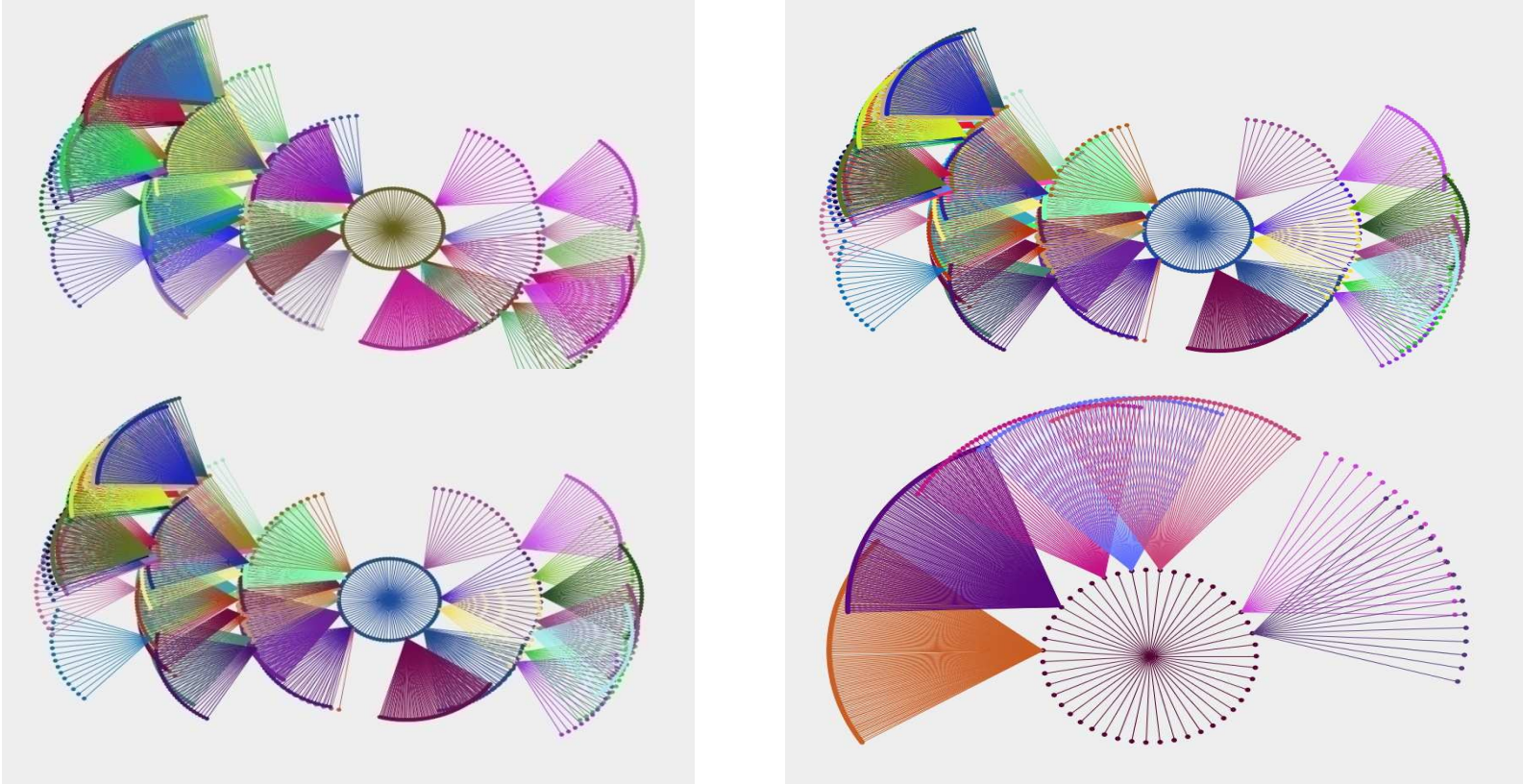
"OR"

T			T + 1		
1	2	3	1	2	3
0	0	0	0	0	0
0	0	1	0	1	0
0	1	0	0	0	1
0	1	1	1	1	1
1	0	0	0	1	1
1	0	1	0	1	1
1	1	0	0	1	1
1	1	1	1	1	1

Boolean networks as models of GRN's



Boolean Networks basins of attraction



All basins of attraction belong to the same network realization with K=2 and N=15.

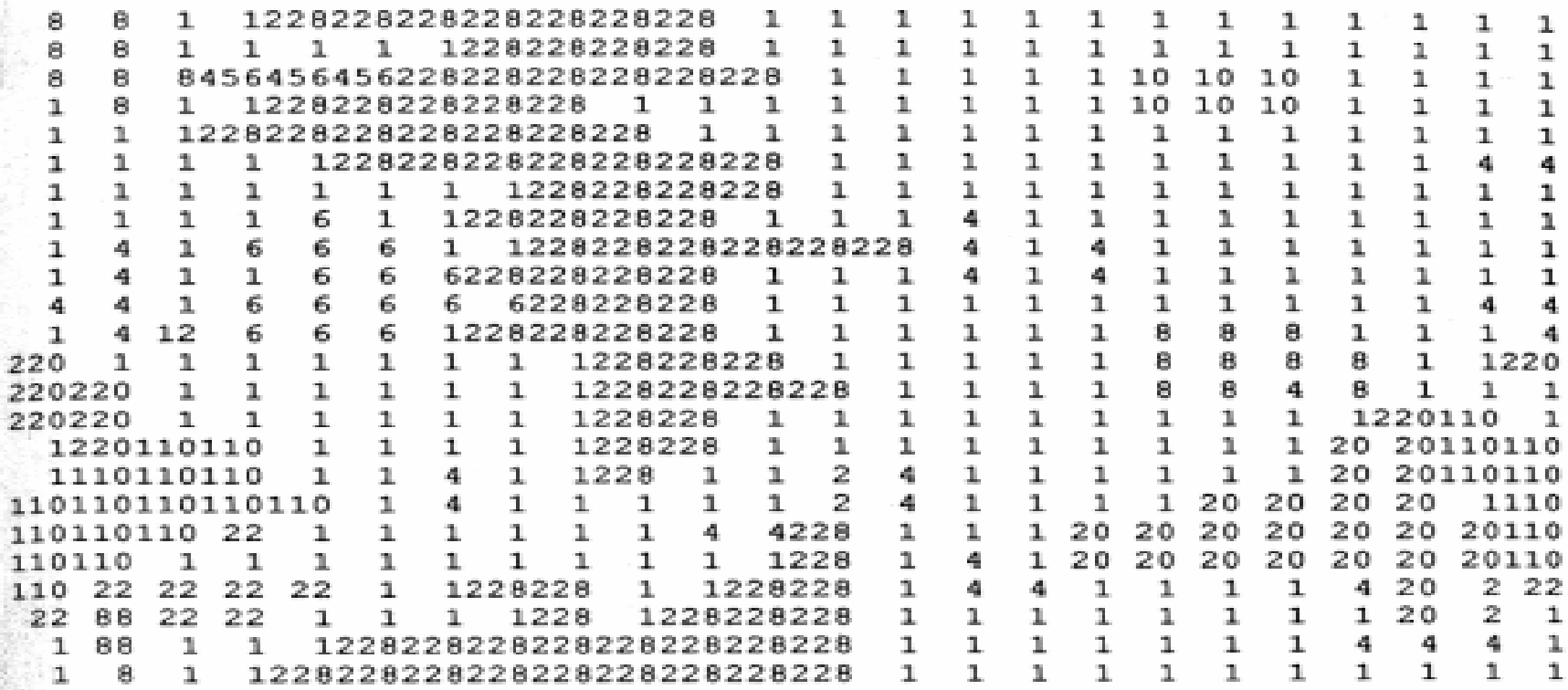
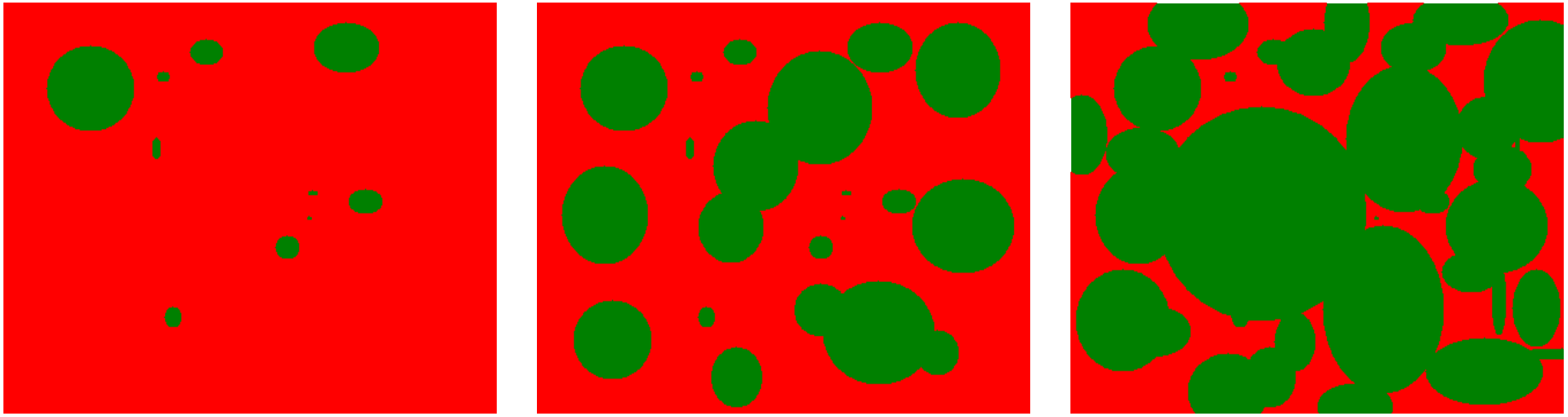


Figure 5.10 Two-dimensional lattice of sites, each a binary state spin which may point up or down. Number at each point in lattice is period of spin on the lattice state cycle. Hence, sites with 1 are frozen active or inactive. Each variable is coupled to its four neighbors and is governed by a Boolean function on those four inputs. When P is increased, the bias in favor of a 1 or a 0 response by any single spin leads, above a critical value P_c, to percolation of a frozen component of spins which spans the lattice and leaves isolated islands of spins free to vary between 0 and 1. (From Weisbuch and Stauffer 1987)

Ordered, Critical and Chaotic Behaviour

Ordered | **Critical** | **Chaotic**



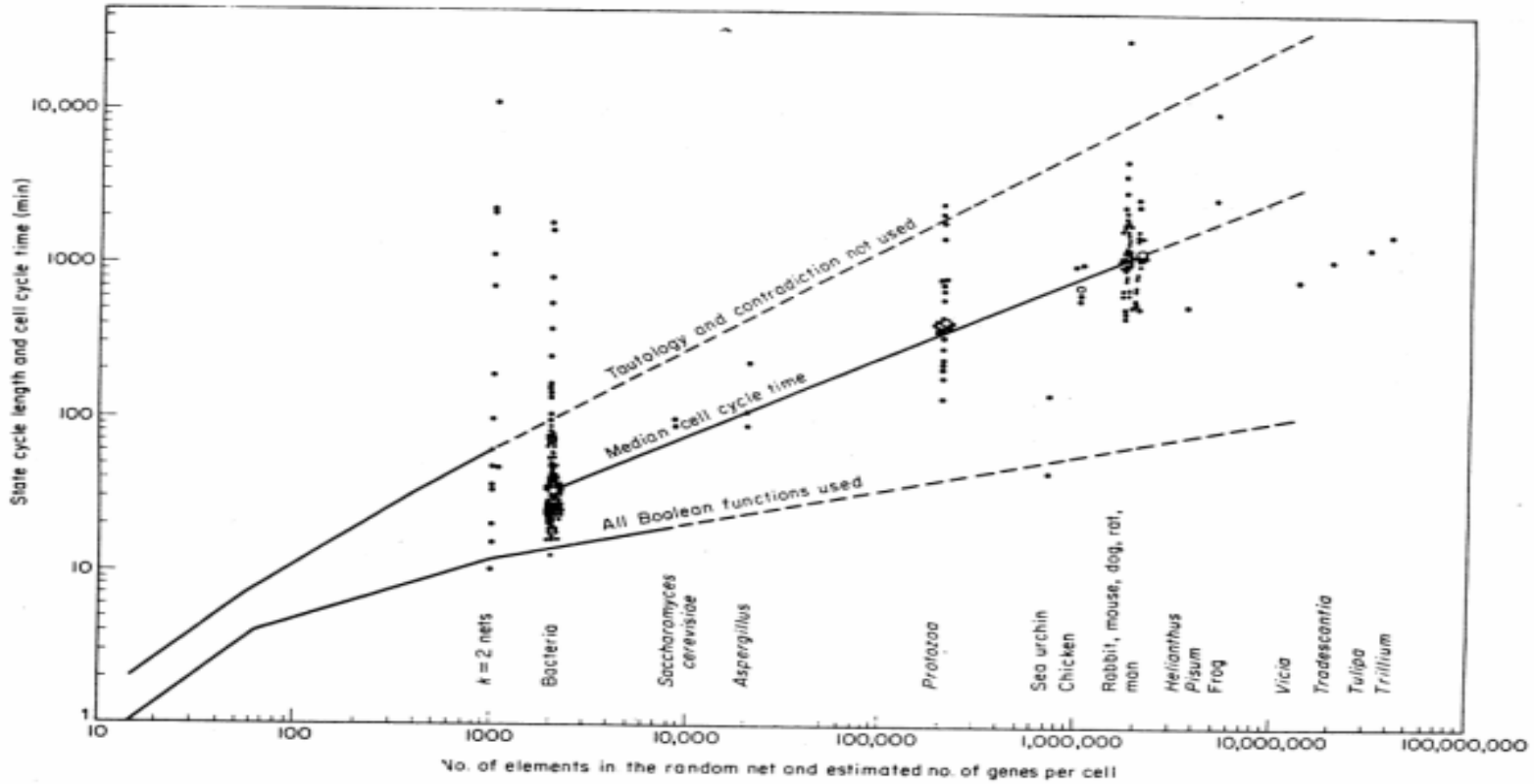


Figure 12.11 Logarithm of cell replication time versus logarithm of the estimated number of genes per cell (assumed proportional to DNA content per cell). Solid line through biological data connects the median replication times. Data from Boolean networks containing 1024 model genes show distribution of state cycle lengths for networks using all Boolean functions of $K = 2$ inputs except "Tautology" and "Contradiction." Median state cycle lengths in Boolean networks with $K = 2$ inputs for different network sizes are shown, using all Boolean functions of two inputs and using all but "Tautology" and "Contradiction." (From Kauffman 1969)

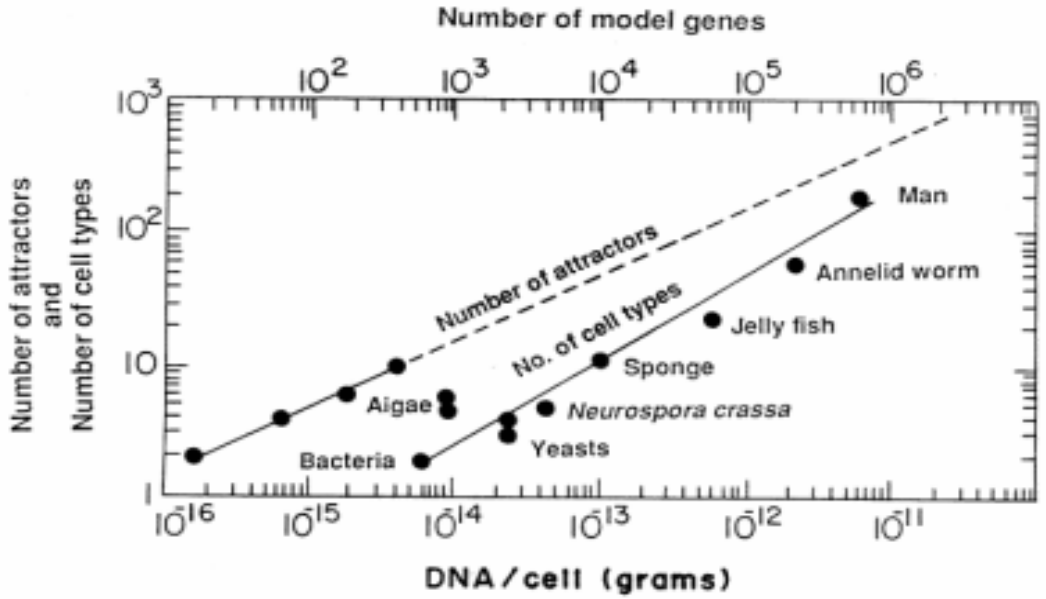
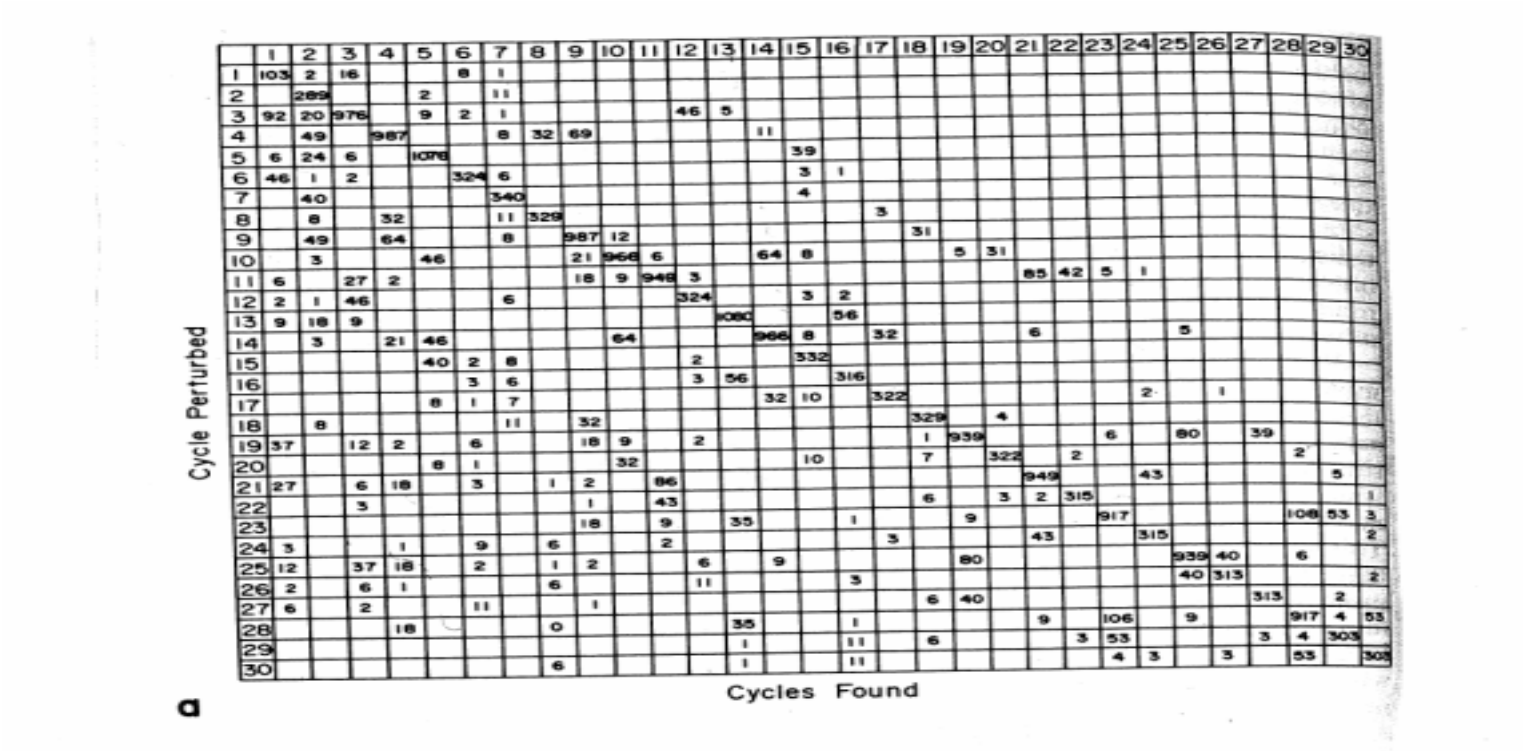
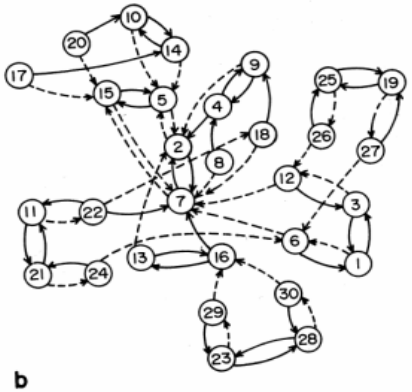


Figure 12.7 Logarithm of the number of cell types in organisms across many phyla plotted against the logarithm of the DNA content per cell. Plot is linear with a slope of 0.5, indicating a power-law relation in which the number of cell types increases as the square root of the amount of DNA per cell. If total number of structural and regulatory genes is assumed proportional to DNA content, then the number of cell types increases as a square-root function of the number of genes. Number of attractors refers to predictions of numbers of model cell types in model genomic regulatory systems having $K = 2$ inputs per gene.


Cycle Perturbed
Cycles Found
a


b

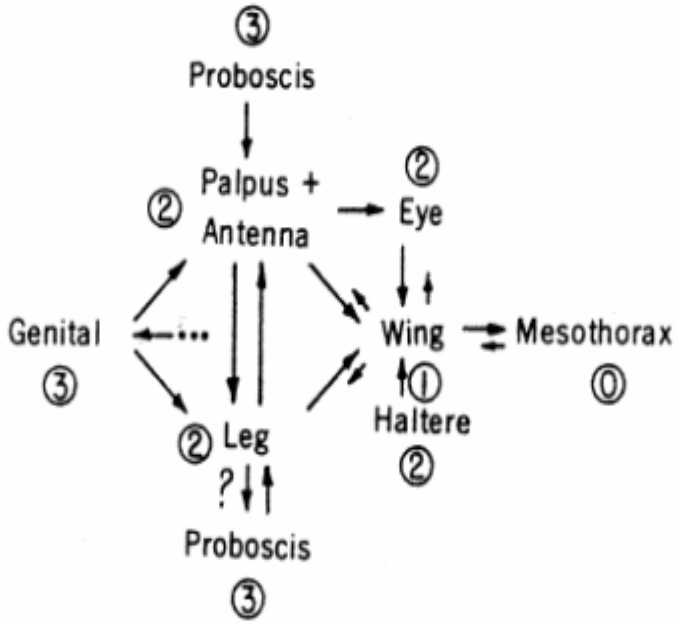


Figure 12.18 Transdetermination between different imaginal-disc-determined states in *D. melanogaster*. Arrows show observed transdetermination steps. Arrow lengths reflect relative probabilities of transitions. Dotted arrow into genital disc indicates that the transdetermination source to genitalia is not certain but is thought to be antenna. Circled numbers indicate minimum number of transdetermination steps separating a disc from the mesothorax determined state. (From Kauffman 1973)

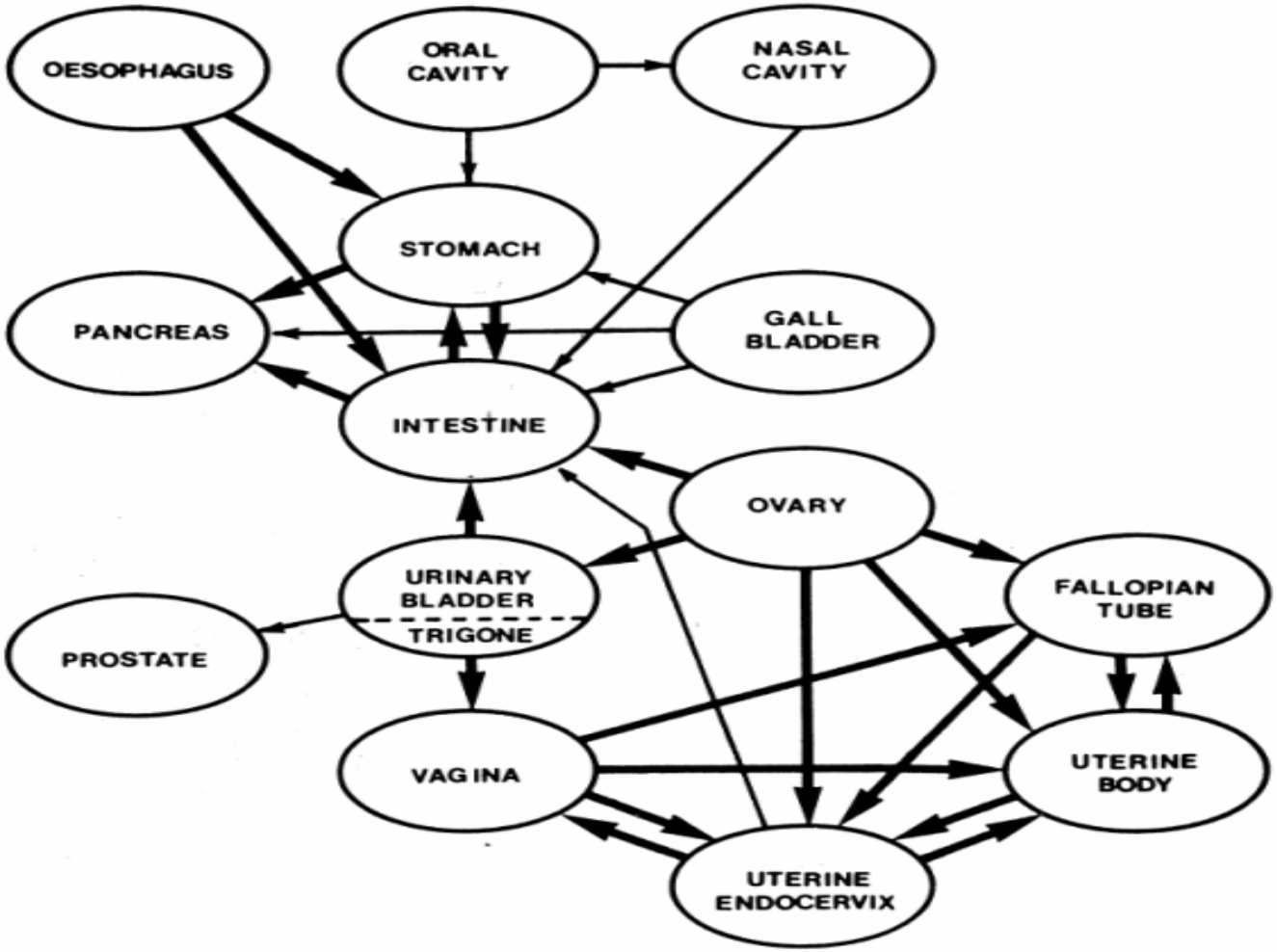
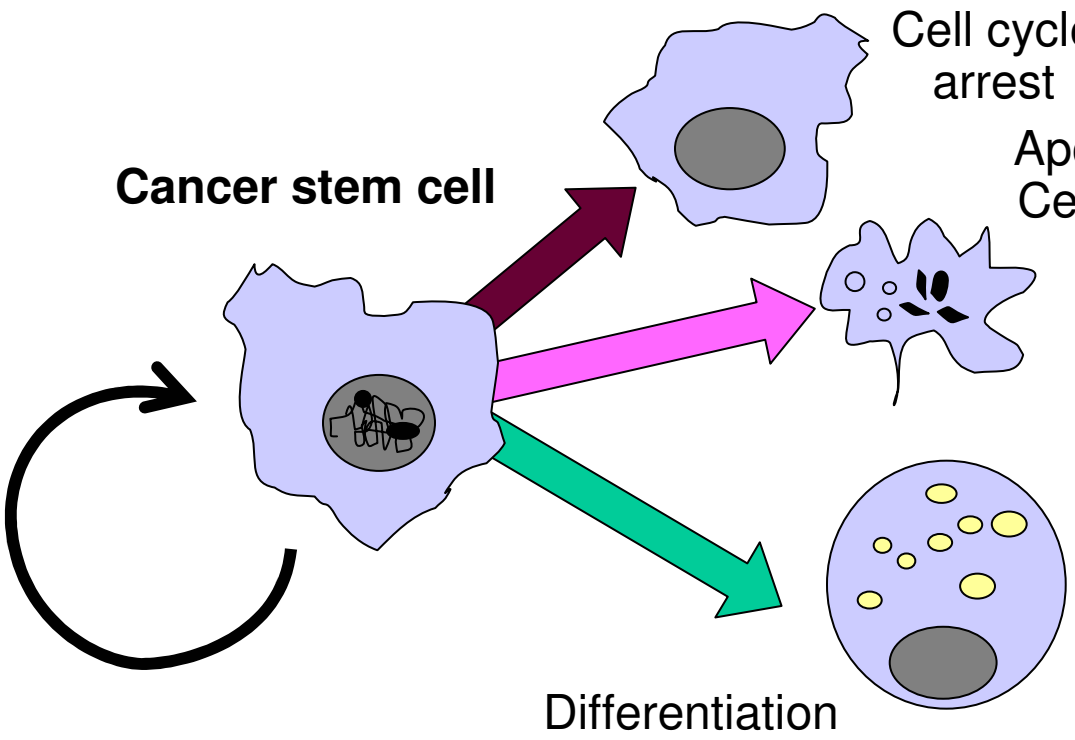


Figure 12.22 Graph of homeotic transformations in humans in the epithelial lining of the digestive, urinary, and female reproductive systems. An arrow from tissue A to tissue B means that patches of B epithelium can be found in the epithelium of A. Thick arrows denote relatively common events, and thin arrows denote very rare ones. Only the epithelial component of each organ is transformed. (From Slack 1985)

Screening for Differentiation therapy of cancer

Differentiation therapy revisited



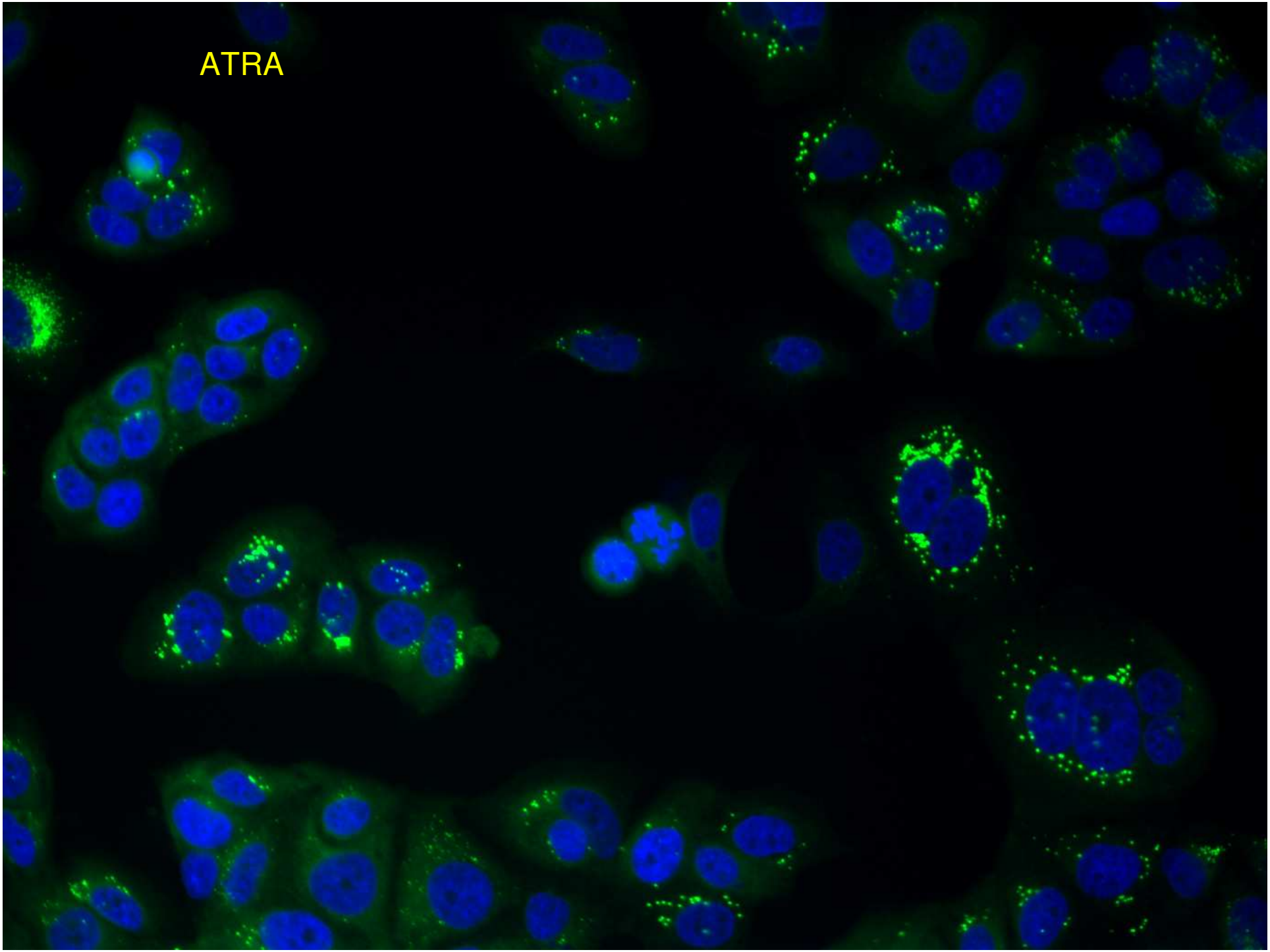
Generic markers

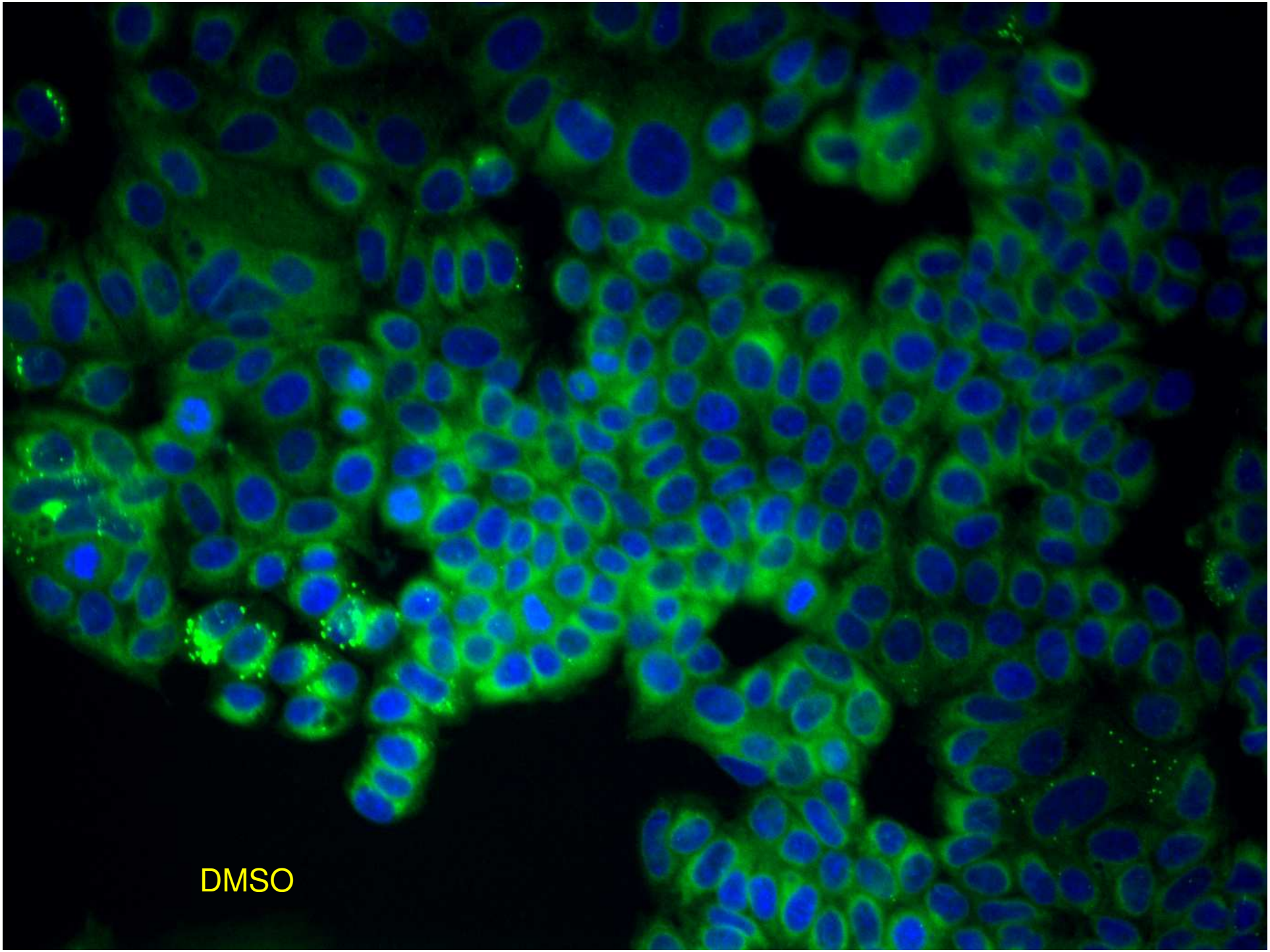
Cell type specific markers
(ofteb unknown)

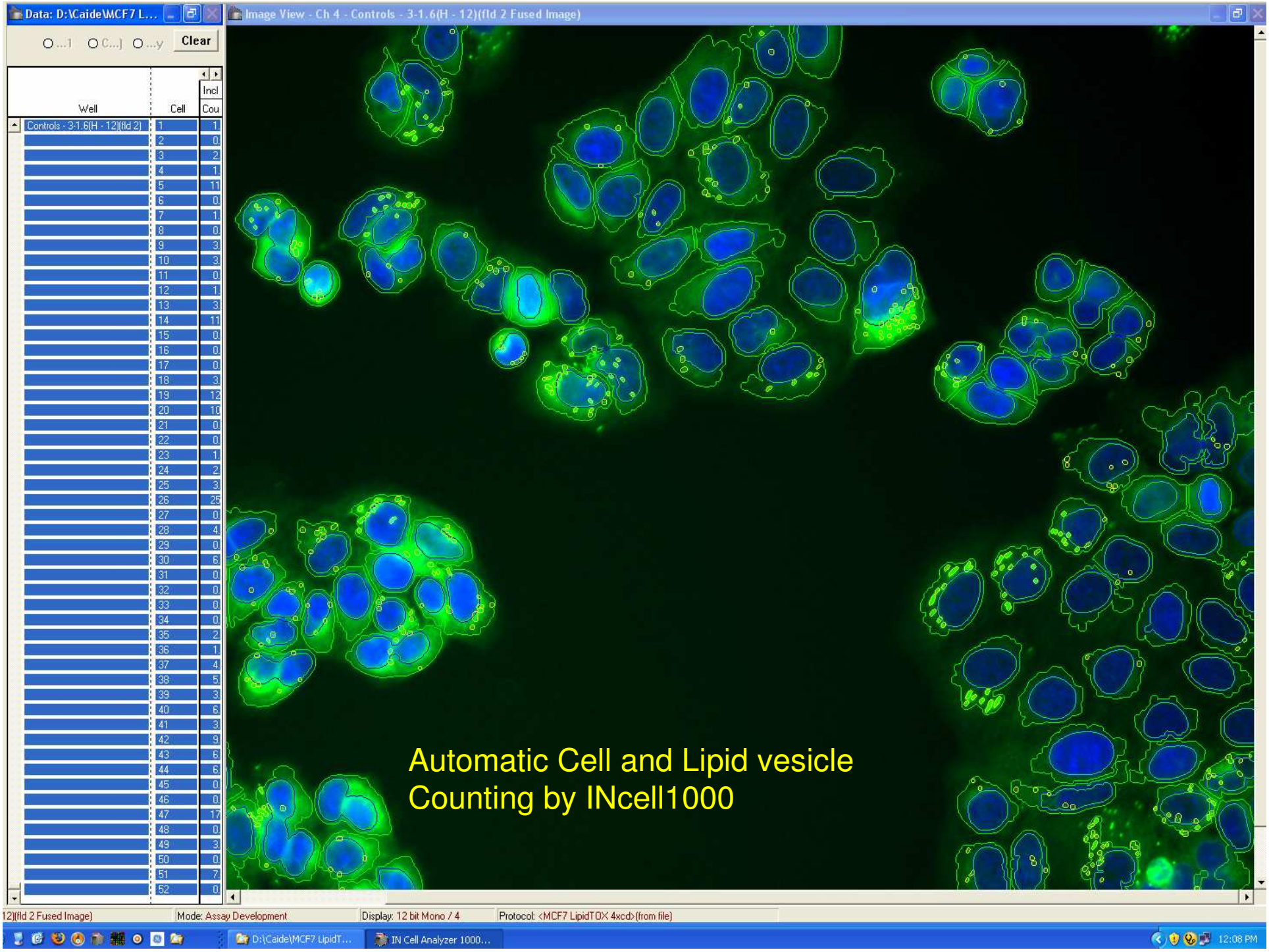
PHENOTYPE SCREENING?

For HTS: must be automated!

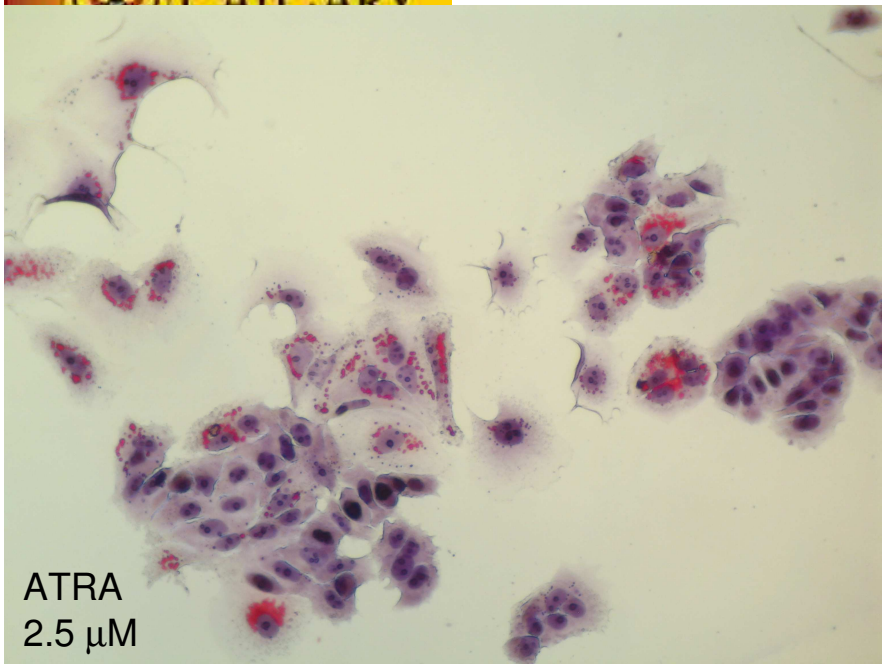
InCell1000 for
High Throughput Image Analysis

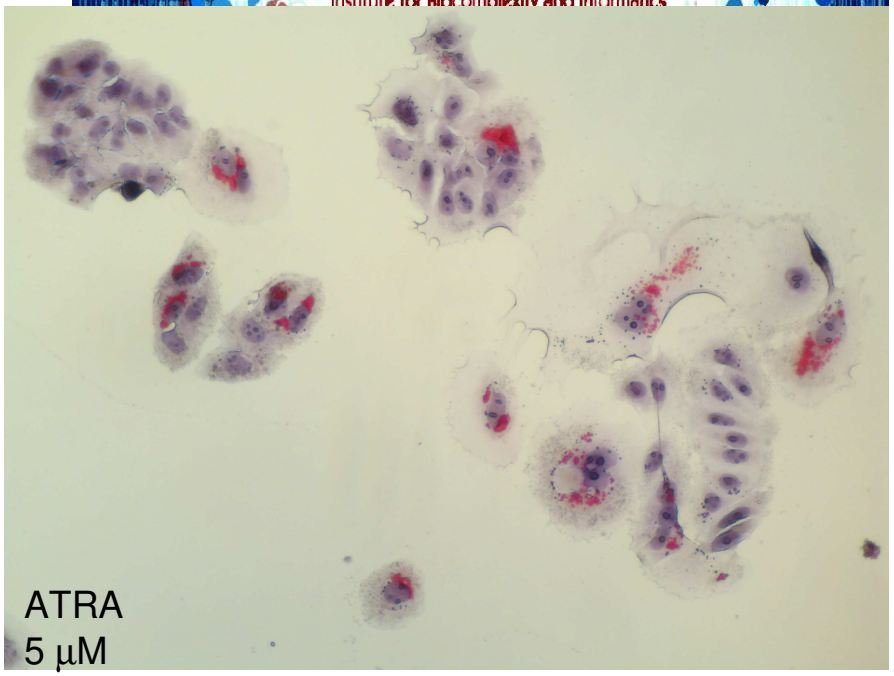

ATRA

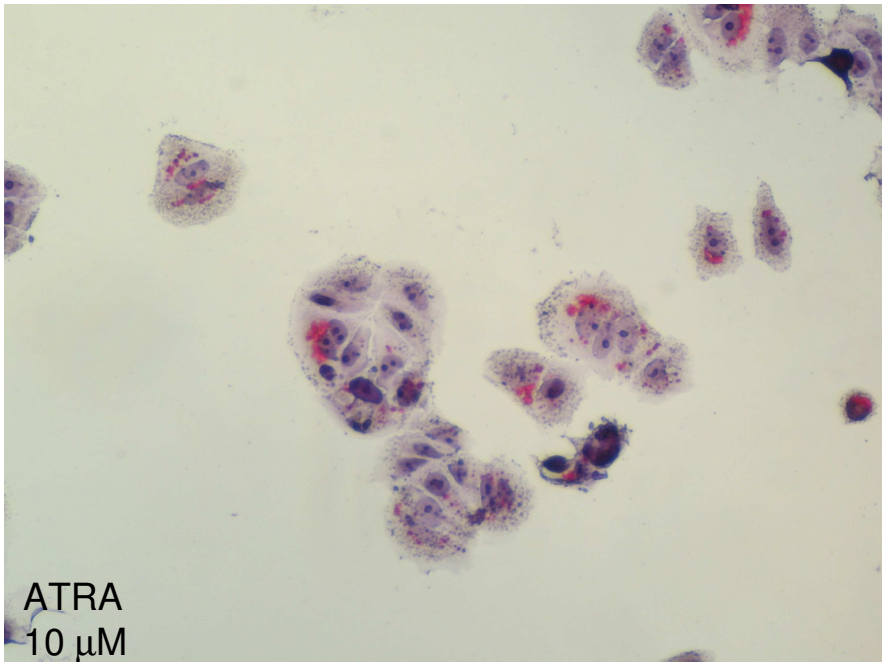

DMSO


Data: D:\Caide\MCF7 L...
Clear
Well
Cell
Incl Cou
Controls - 3-1.6(H - 12)(fld 2)
Image View - Ch 4 - Controls - 3-1.6(H - 12)(fld 2 Fused Image)
Automatic Cell and Lipid vesicle
Counting by INcell1000
12)(fld 2 Fused Image)
Mode: Assay Development
Display: 12 bit Mono / 4
Protocol: <MCF7 LipidTOX 4xcd>(from file)
D:\Caide\MCF7 LipidT...
IN Cell Analyzer 1000...
12:08 PM

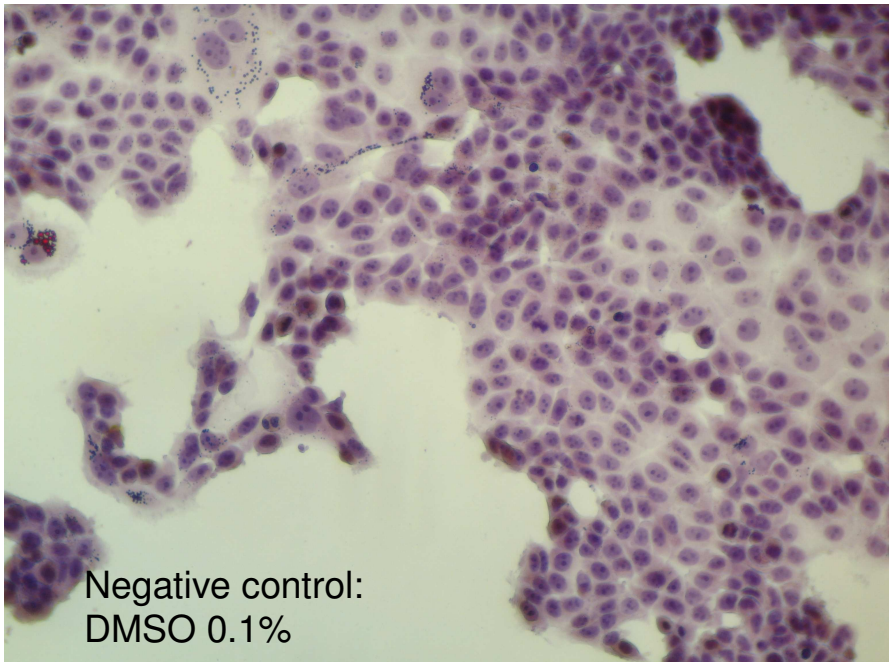
Positive controls





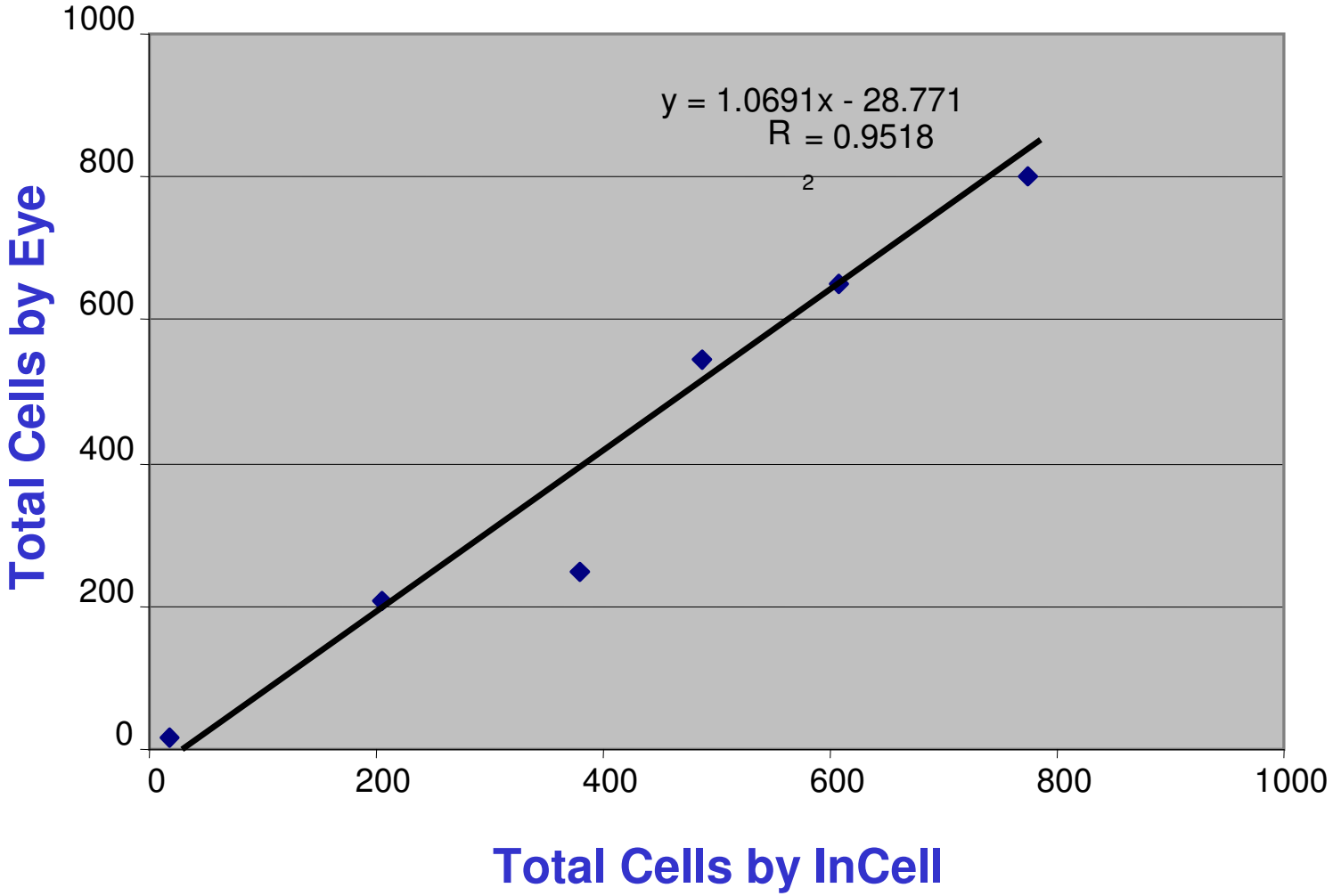


Oil Red O / hematoxylin staining of MCF7 cells:
20x objective, 1.5X digital zoom



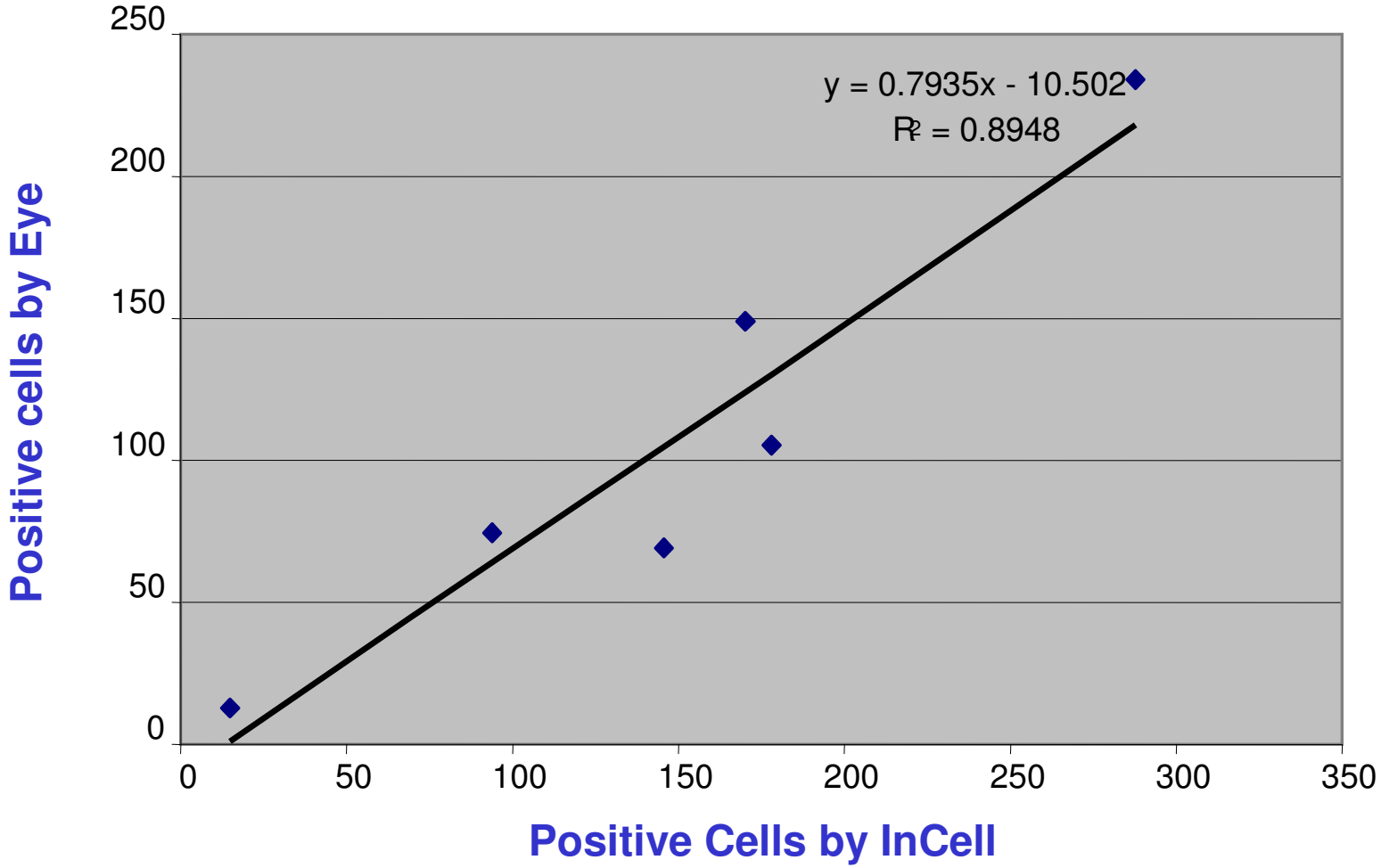
Comparison of Total Cell Number Counting

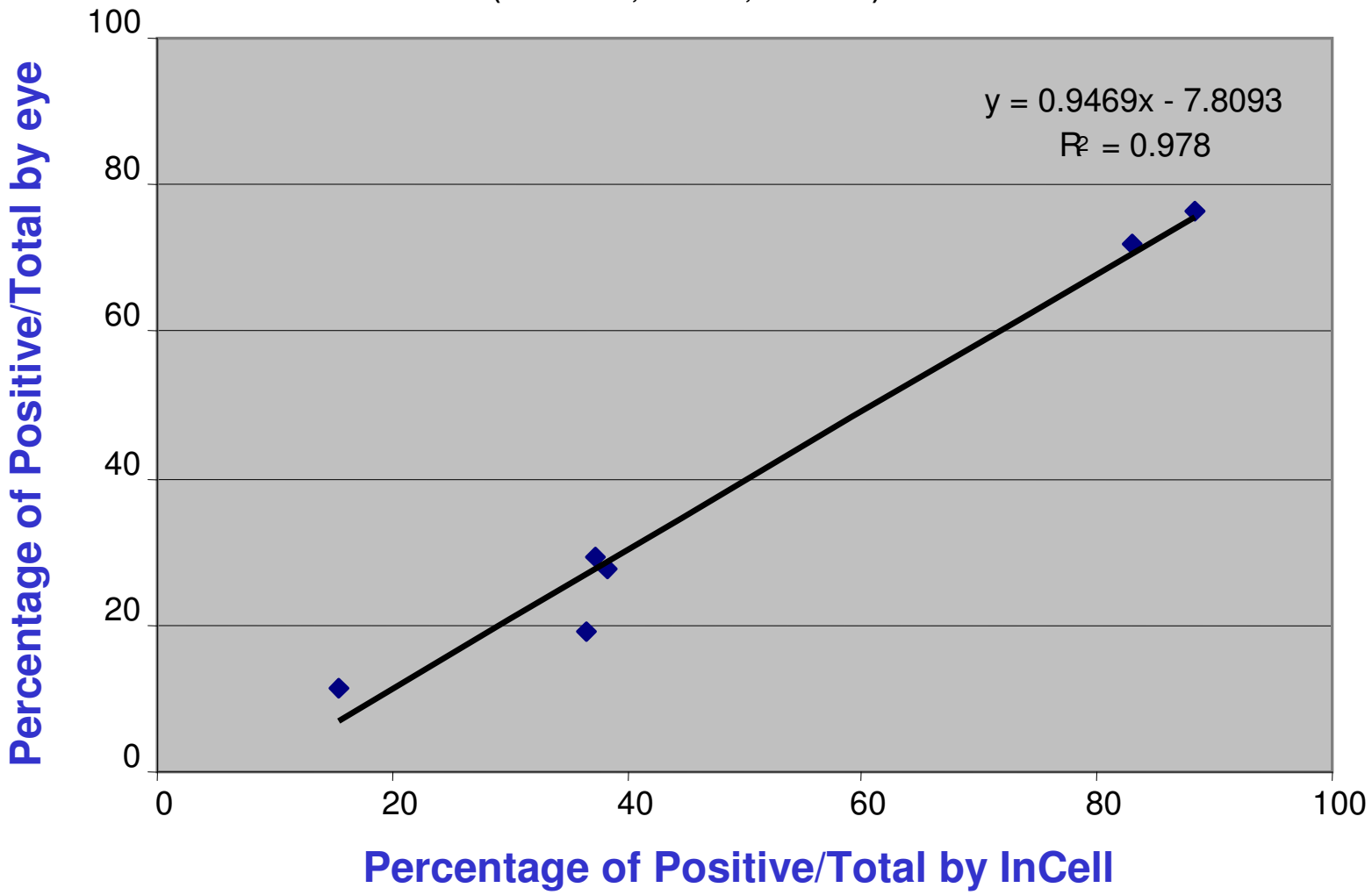
(TL-2 Plate 3, well A1-4, H1 and H8)



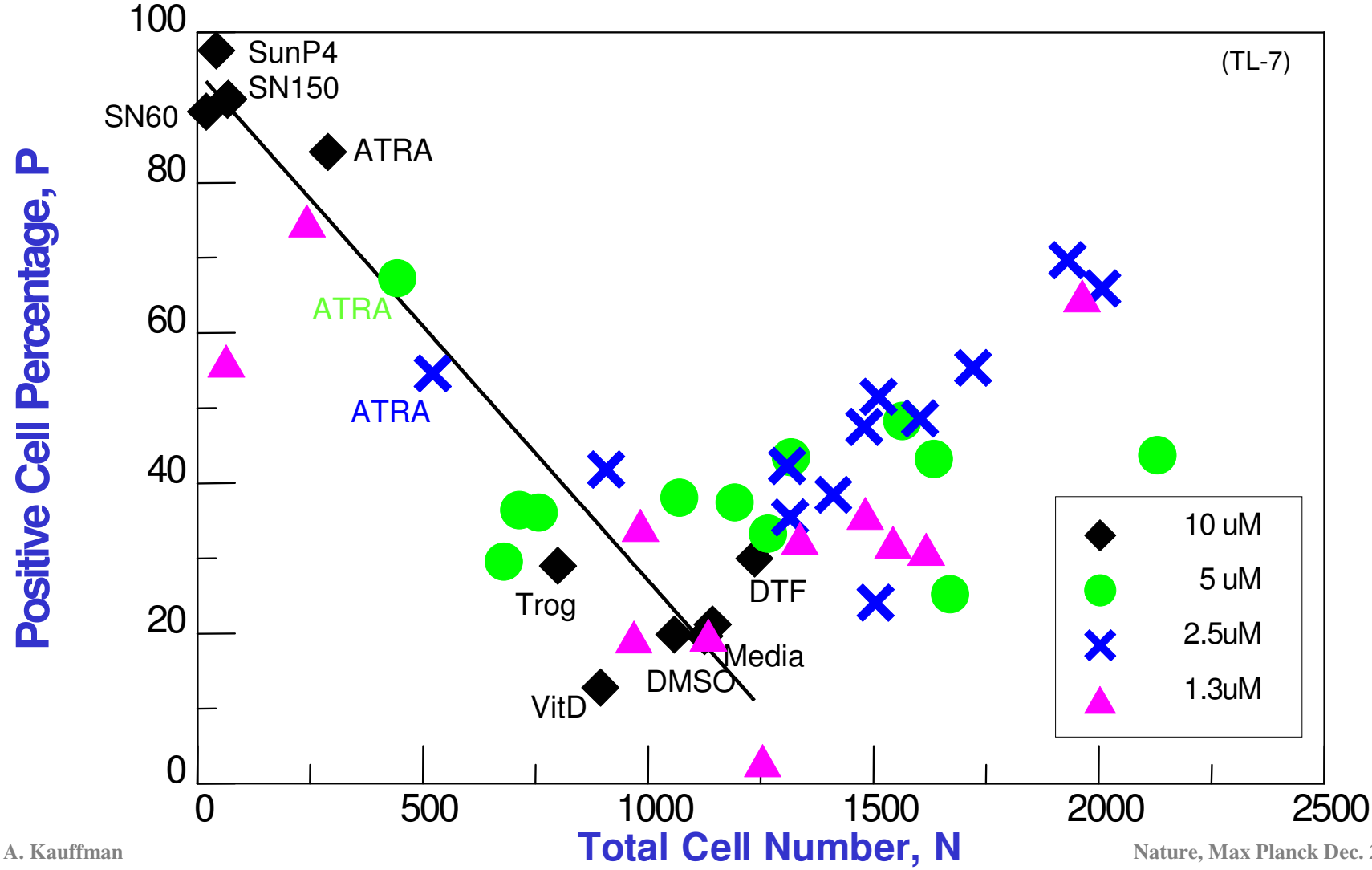
Comparison of Positive Cell Number Counting

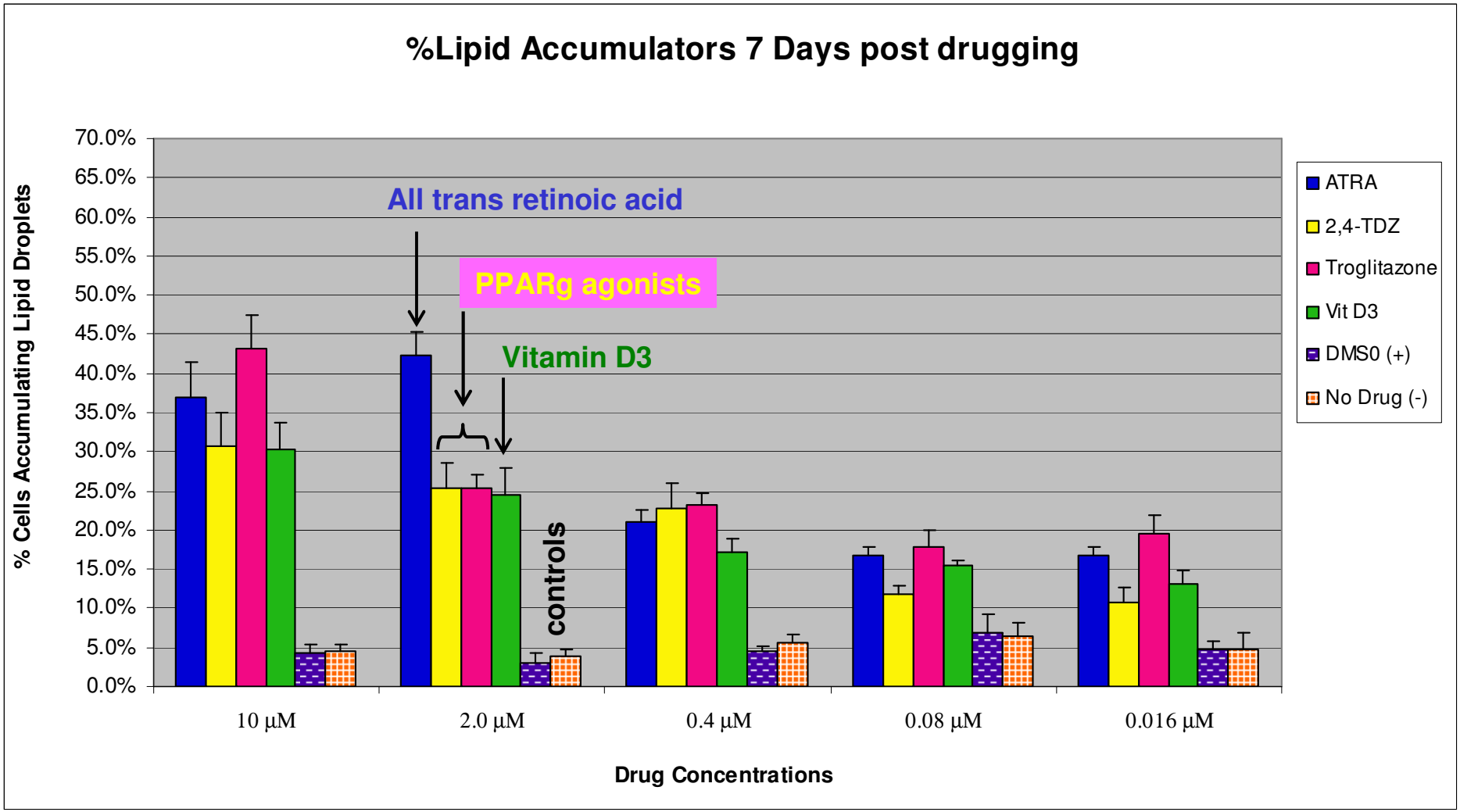
(TL-2 Plate 3, well A1-4, H1 and H8)




Comparison of Positive Cell Percentage
(TL-2 Plate 3, well A1-4, H1 and H8)
Percentage of Positive/Total by eye
Percentage of Positive/Total by InCell
y = 0.9469x - 7.8093
R² = 0.978
0
20
40
60
80
100

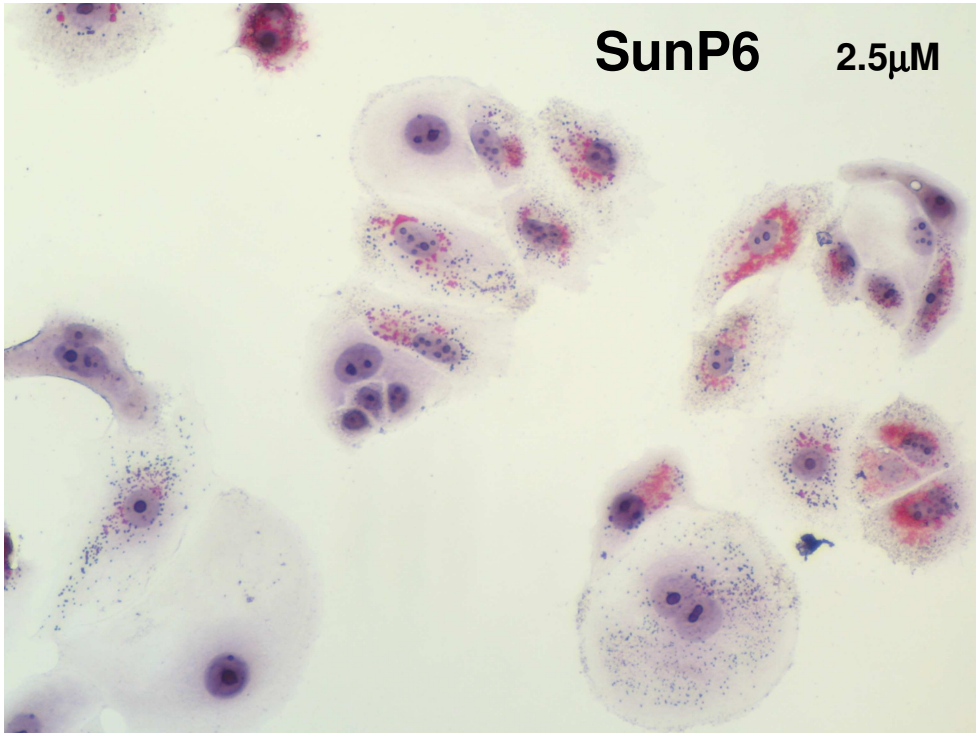
Relationships between Cell Differentiation (% positive) and Cell Proliferation (Total Cell Number)



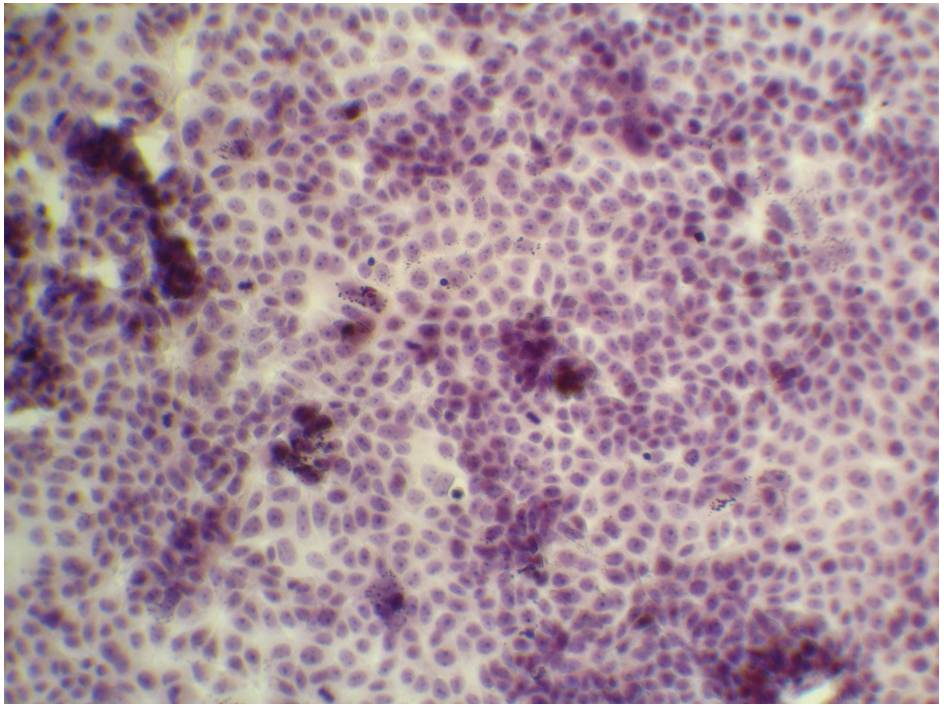


MCF7 cells were seeded in a 96 well plate and treated with various chemicals in a range of concentrations then maintained at cell culture conditions for 7 days. At 7 days post treatment, cells were fixed and stained with a fluorescent green neutral lipid stain and counterstained with Dapi to visualize the nucleus. Five fields in duplicate for each well were analyzed, and a ratio of lipid accumulating cells to total nuclei per field was calculated. Bars represent the average percent of lipid accumulating cells in each test condition. Error bars represent one standard deviation of the mean.

“**Hits**” (from Crooks library)



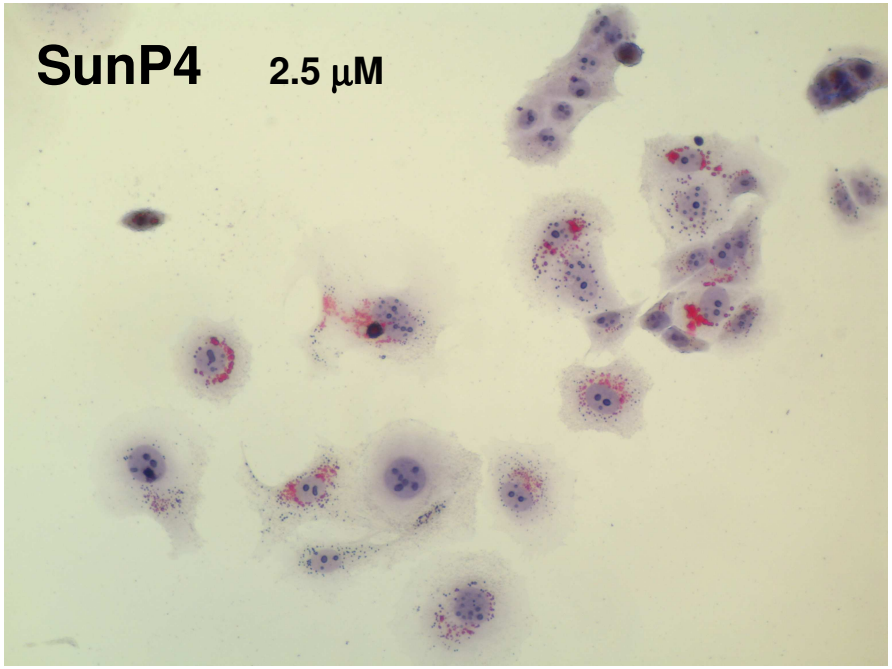
Negative control:
(DMSO 0.1%)

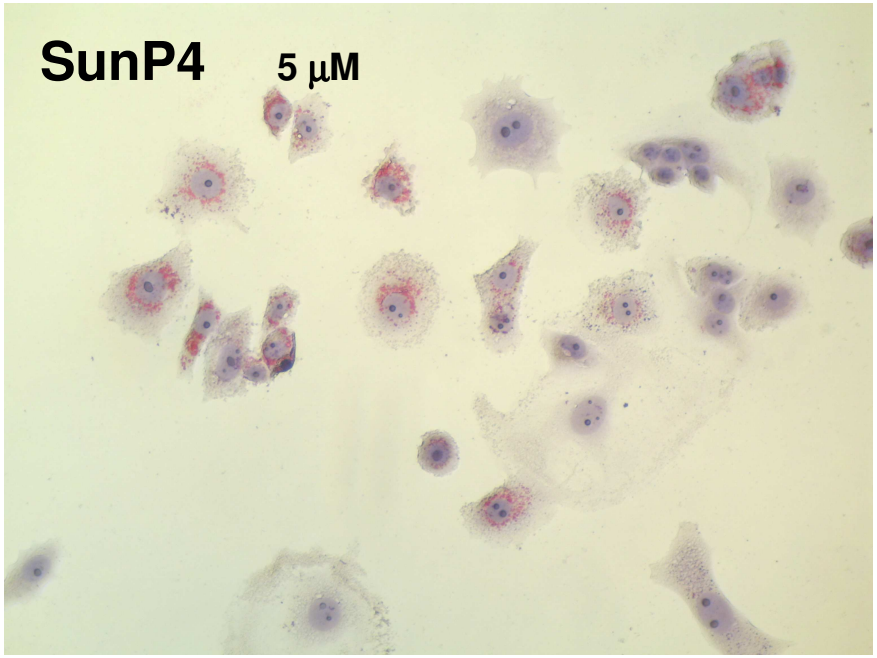


Oil Red O / hematoxylin staining of MCF7 cells:

20x objective, 1.5X digital zoom

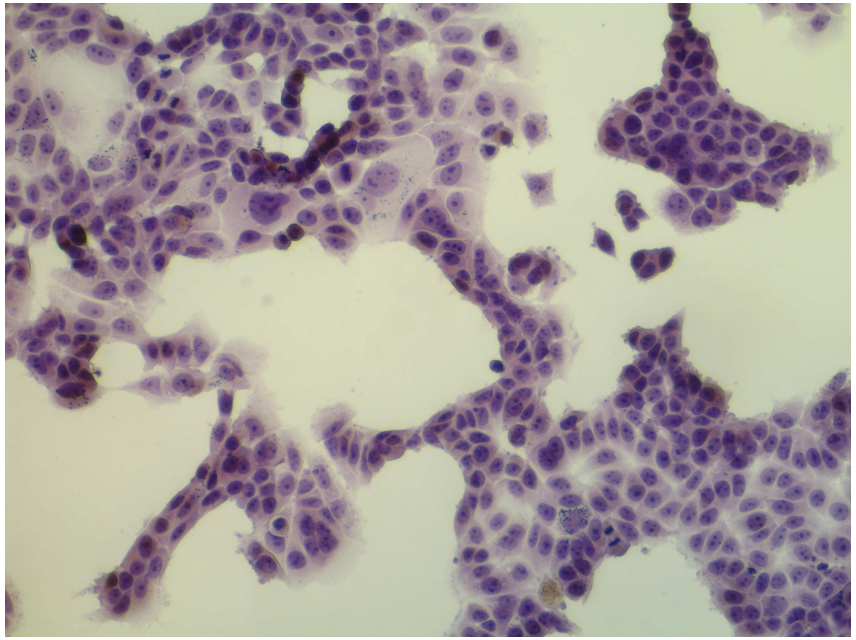
“**Hits**” (from Crooks library)





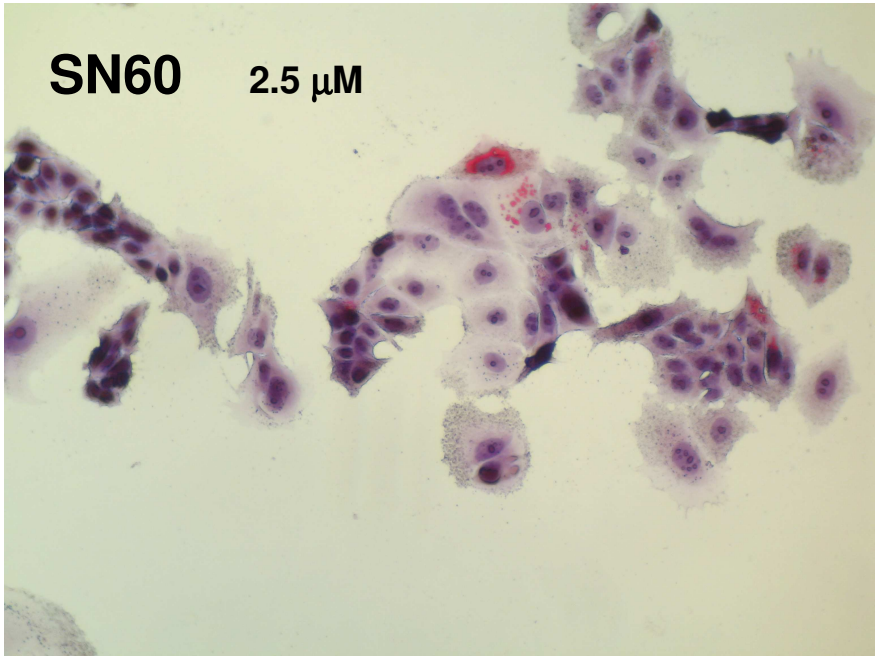
Oil Red O / hematoxylin staining of MCF7 cells:

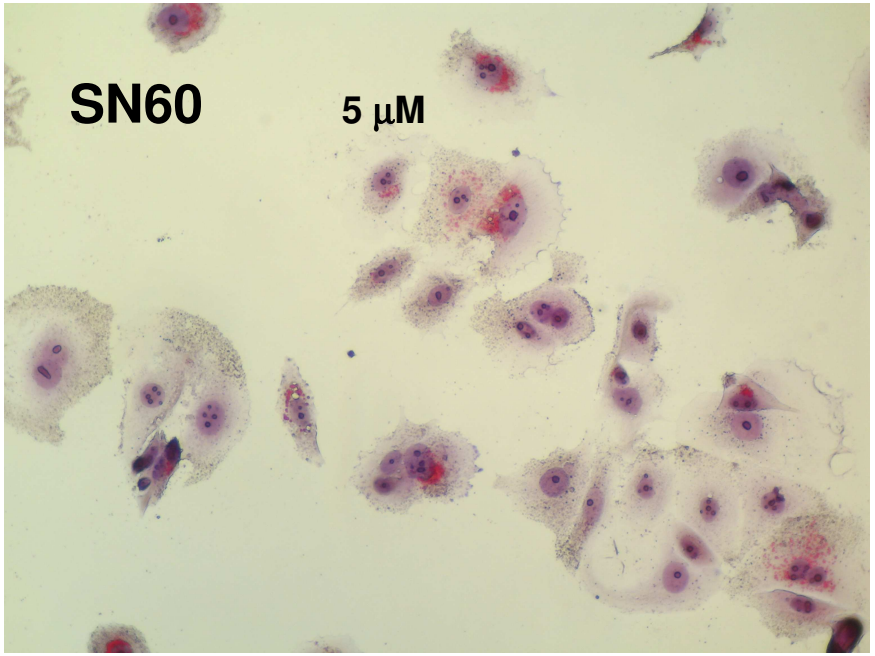
20x objective, 1.5X digital zoom



Negative control:
DMSO 0.1%

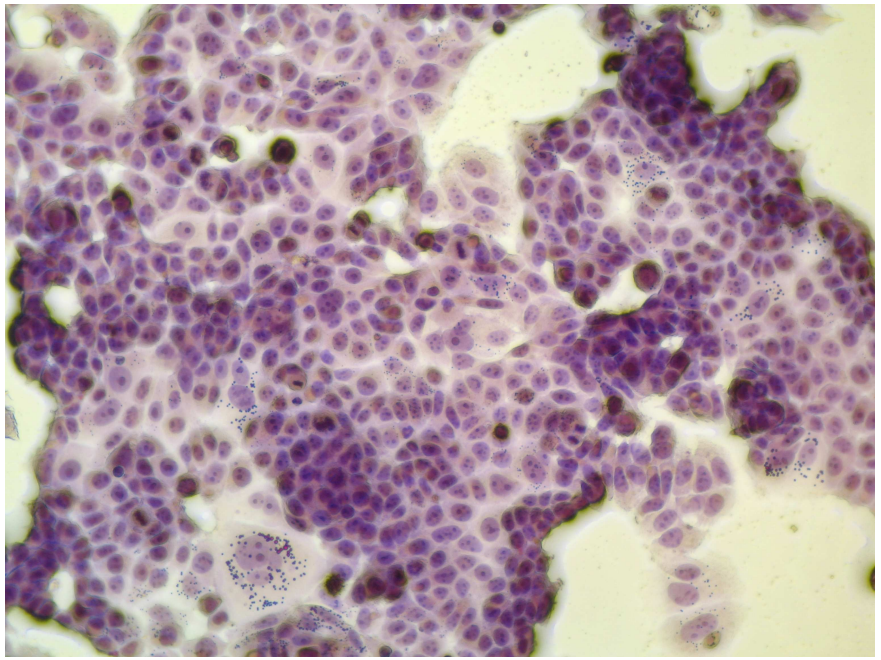
“**Hits**” (from Crooks library)





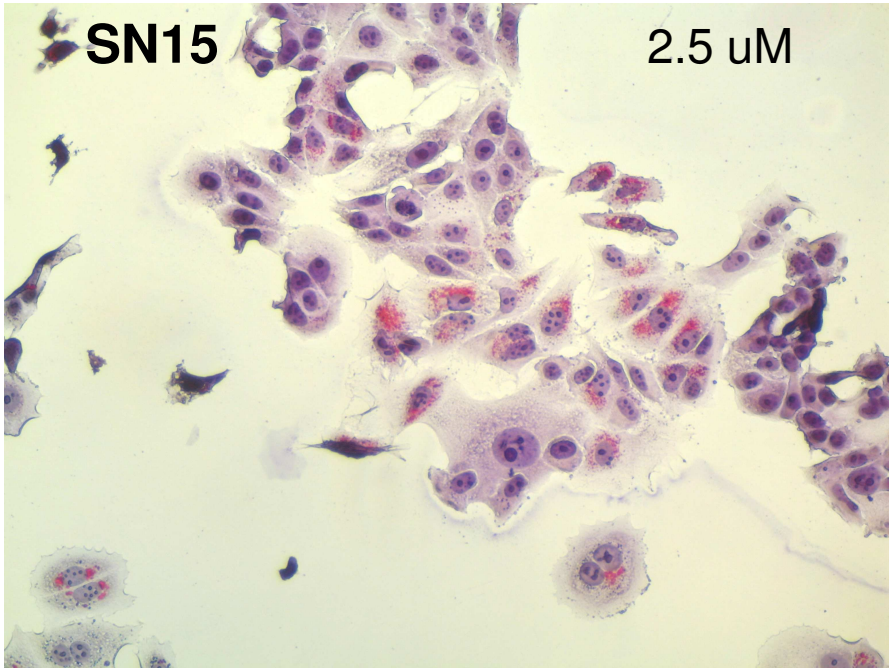
Oil Red O / hematoxylin staining of MCF7 cells:

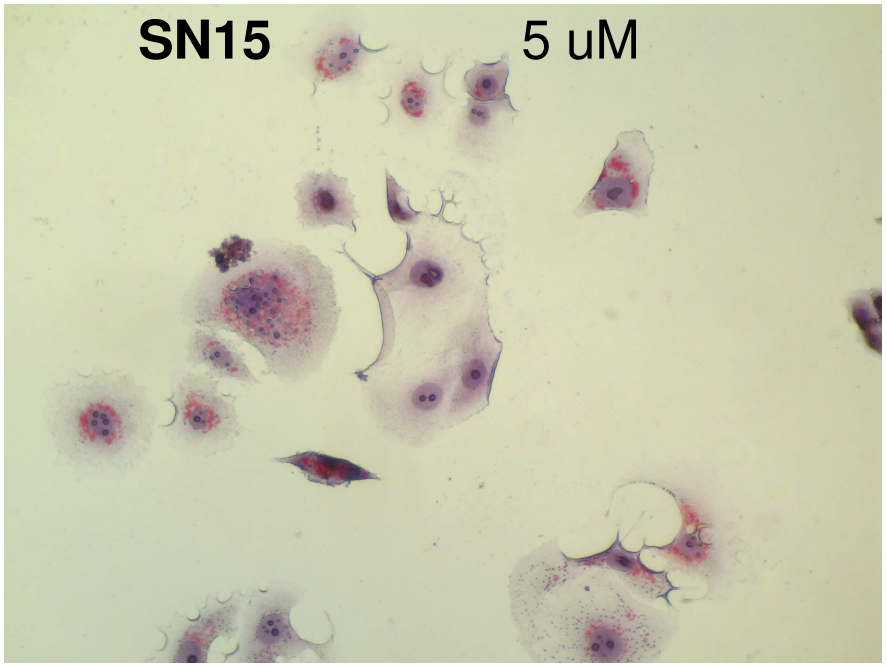
20x objective, 1.5X digital zoom



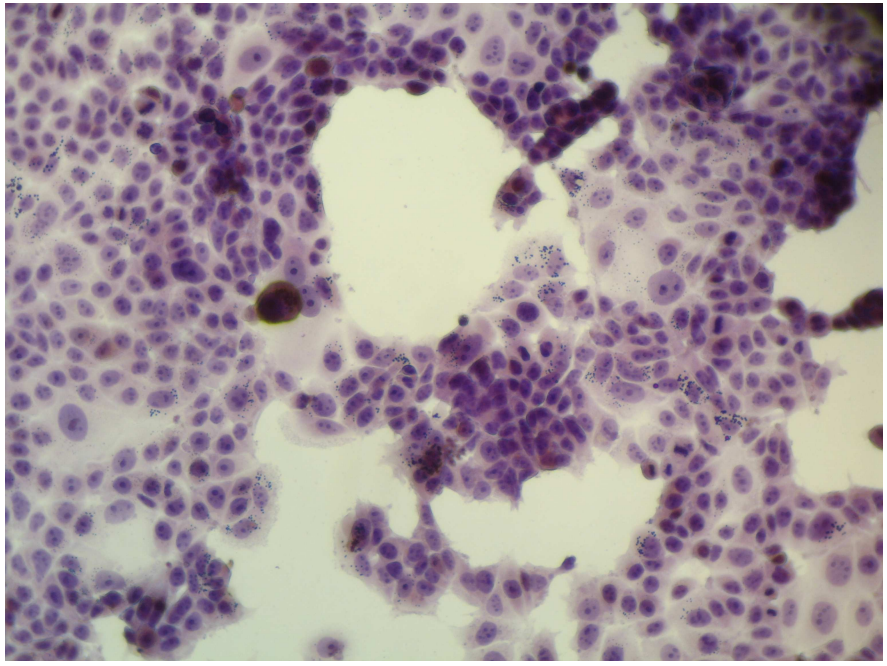
Negative control:
DMSO 0.1%

“**Hits**” (from Crooks library)





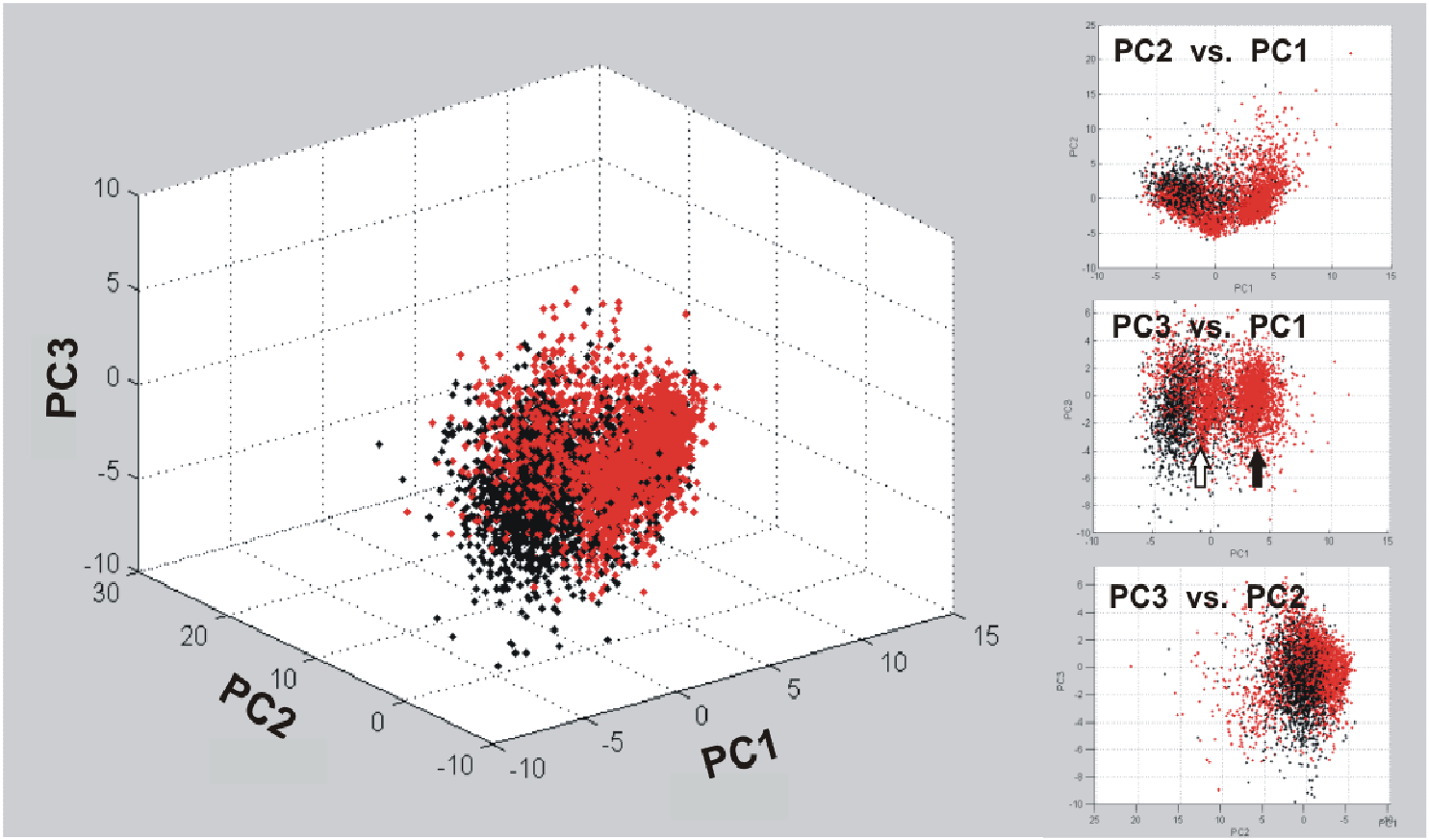
Oil Red O / hematoxylin staining of MCF7 cells:
20x objective, 1.5X digital zoom



Negative control:
DMSO 0.1%

Detection of differentiation by shift in 40-dimensional phenotype space

HL60 ⟶ Differentiated neutrophil



- Approaches to discovering structure and logic of genetic regulatory nets.

1. ChIP-Chip.
2. Inference of transcription factor binding sites.
3. Inference of structure and logic from time series gene expression data.
4. Promoter bashing.
5. Data base integration.

IADGRN

1) Generate network and Boolean functions
2) Generate a random initial state
3) Generate a path of states (affected by noise)
4) Infer the network with pairwise MI and DPI
5) Apply post-inference engine
6) Results Analysis

Memory requirements:

N (number genes), R (number runs), k (connectivity)
Path of States ~ O(N.R)
K functions ~ ($N.2^k$)
Memory usage before inference engine ~ O(N^2+N.R+$N.2^k$)
Adjacency Matrix ~ O(N^2)
Inference engine: O($2.S.N^2$+N^2)

Limits:
50.000 genes,
20 inputs/gene.

Predicting inferability

- Yeast Network, 3459 genes, exponential input distribution: Medusa network.
- Using 600 independent state transitions and mutual information threshold we predicted that we could infer 33% of the regulatory connections and in fact predicted 34% with no false positives.
- Future: Inferring Stochastic Genetic Networks with array noise.
- Long term aim is to use gene expression time series from real cells and be able to estimate inferability of network's structure and logic.
- Inferring personalized structure and logic of cancer stem cell aberrant circuitry for therapy.

JOURNAL OF COMPUTATIONAL BIOLOGY, Volume 15, Number 9, 2008, Pp. 1–14

Heuristic Approach to Sparse Approximation of Gene Regulatory Networks

M. ANDRECUT, S. HUANG, and S.A. KAUFFMAN

ABSTRACT

Determining the structure of the gene regulatory network using the information in genome-wide profiles of mRNA abundance, such as microarray data, poses several challenges. Typically, "static" rather than dynamical profile measurements, such as those taken from steady state tissues in various conditions, are the starting point. This makes the inference of causal relationships between genes difficult. Moreover, the paucity of samples relative to the gene number leads to problems such as overfitting and underconstrained regression analysis. Here we present a novel method for the sparse approximation of gene regulatory networks that addresses these issues. It is formulated as a sparse combinatorial optimization problem which has a globally optimal solution in terms of l_0 norm error. In order to seek an approximate solution of the l_0 optimization problem, we consider a heuristic approach based on iterative greedy algorithms. We apply our method to a set of gene expression profiles comprising of 24,102 genes measured over 79 human tissues. The inferred network is a signed directed graph, hence predicts causal relationships. It exhibits typical characteristics of regulatory networks organism with partially known network topology, such as the average number of inputs per gene as well as the in-degree and out-degree distribution.

Key words: automata, combinatorial optimization, statistical mechanics, stochastic processes.

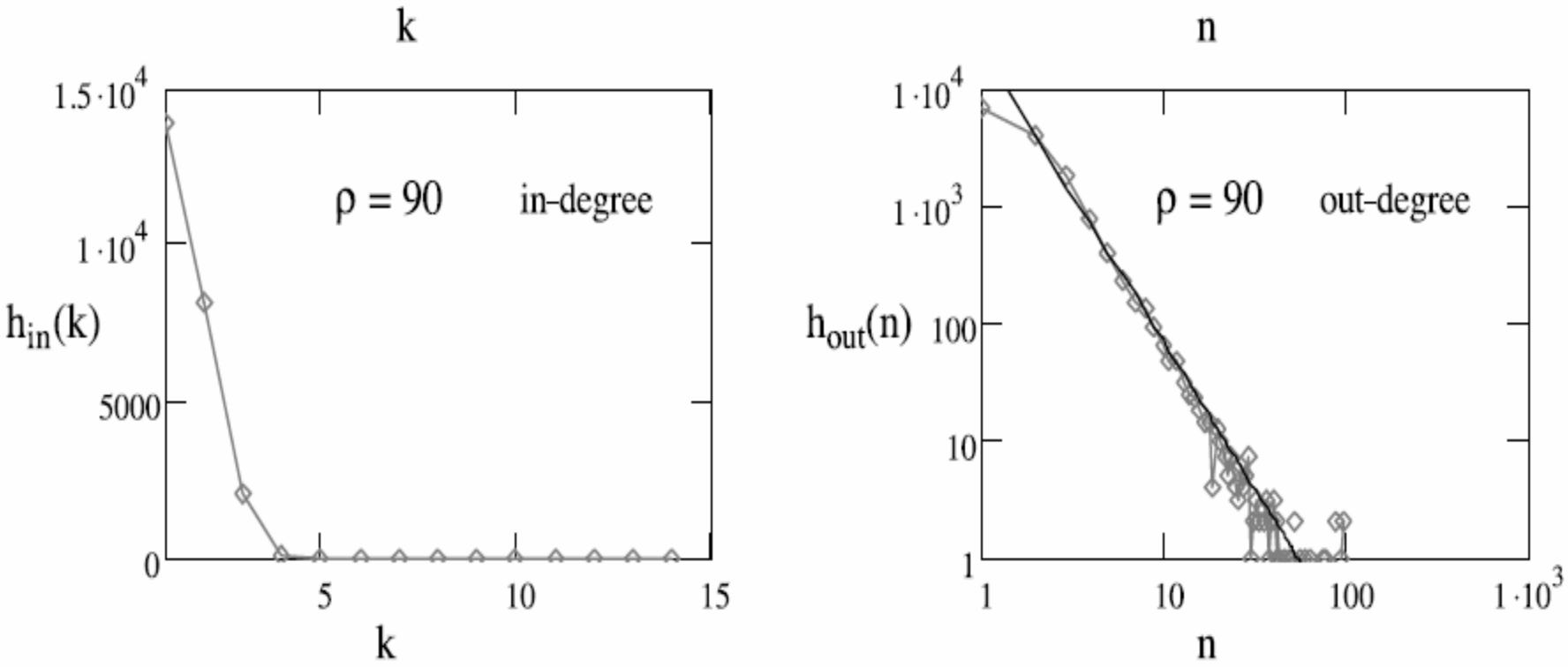


FIG. 4. The in-degree and the out-degree distribution of the reconstructed network: $\rho = 97.5; 95; 90$.

SUMMARY

- 1. Genetic regulatory network is a nonlinear (stochastic) dynamical system.
- 2. Cell types are probably attractors
- Differentiation is i. transition between attractors; ii. Bifurcations to new attractors
- Cancer cells, with or without somatic mutations can be induced to differentiate to non-proliferating cells: Cancer Differentation Therapy may be a major new approach.