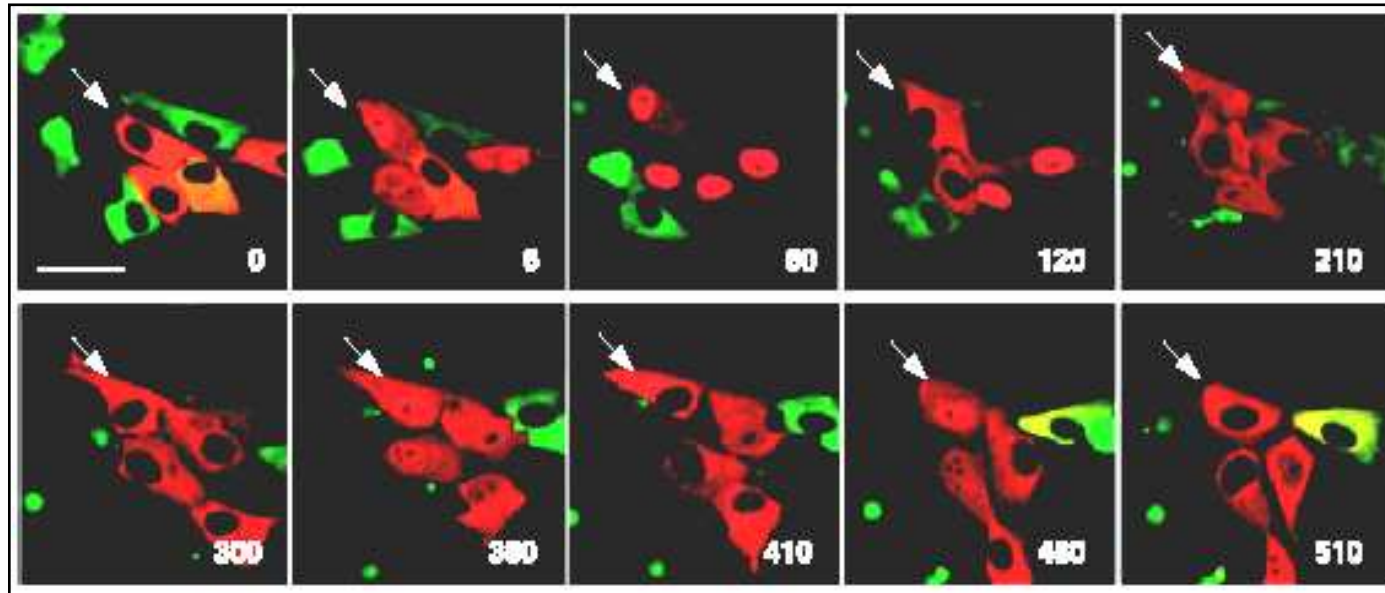# How the number of alleles influences gene expression

Beata Hat
Pawel Paszek
Marek Kimmel
Kazimierz Piechor
Tomasz Lipniacki
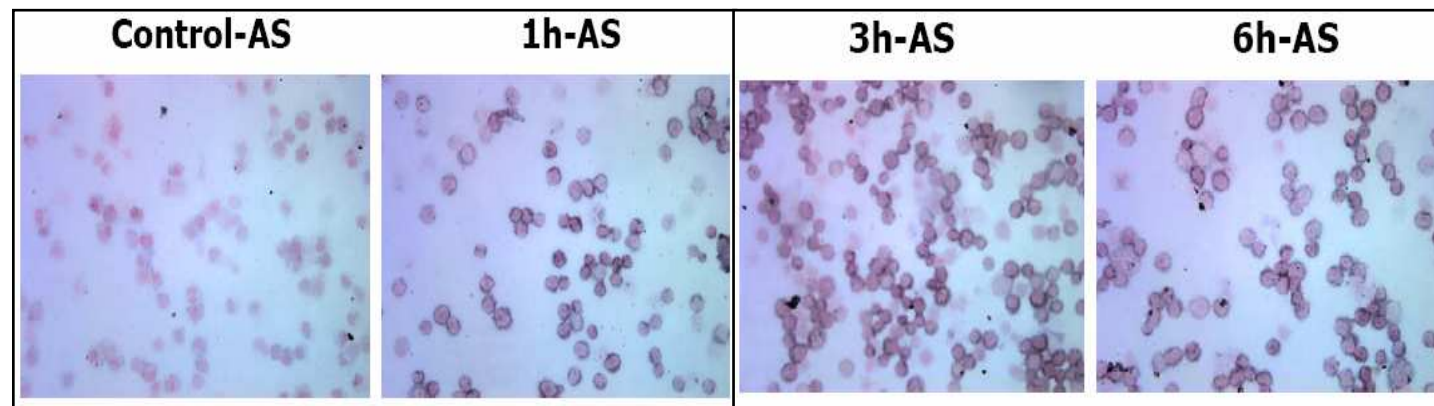
Single cell experiments

Cytoplasm - nucleus oscillations in NF-kB (red) and IkBa (green) system after TNF treatment
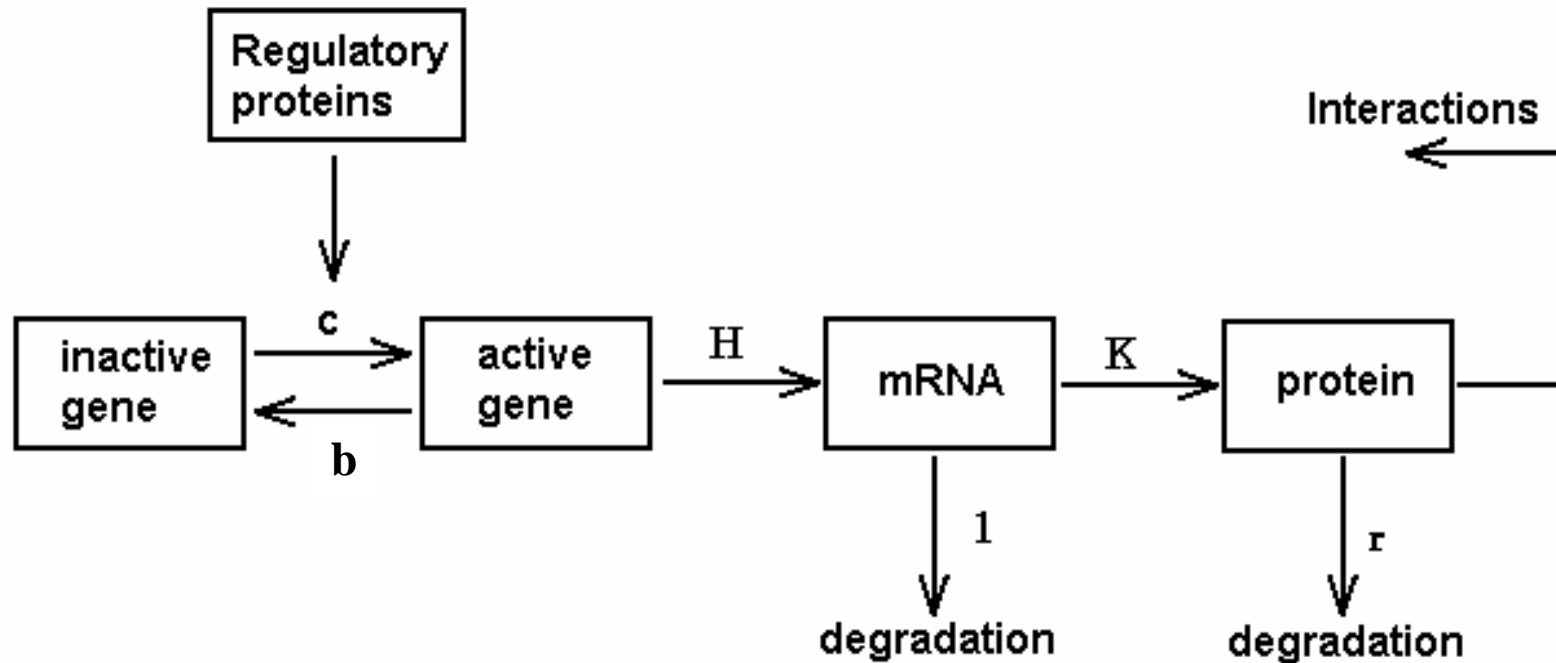

Immortal cancer cell line (SK-N-AS cells), M.R.H. White group

IL8 mRNA level after TNF treatment

Immortal cancer cell line (HeLa cells), A. Brasier group
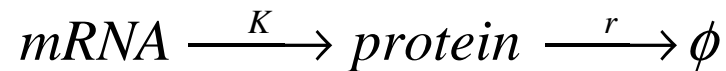
The main steps in gene expression

The number of molecules involved:

$$1 \leq DNA \leq mRNA \leq protein \leq 10^6$$

The aim of this work

- How increase in number of gene alleles due to the cancer development or genome duplication in a cell cycle influences its regulation

- How the loss of one allele or its transcriptional inactivity can result in haploinsufficiency disease for autoregulated genes

- To deduce the behavior of „normal" cells from experiments on the transfected cells

A single haploidal gene without feedback regulation

$$\mathbf{I} \xrightarrow{\ c = c_0\ } \mathbf{A}, \qquad \mathbf{I} \xleftarrow{\ b = b_0\ } \mathbf{A}$$

$$A \xrightarrow{\ GH\ } mRNA \xrightarrow{\ 1\ } \phi$$

$$mRNA \xrightarrow{\ K\ } protein \xrightarrow{\ r\ } \phi$$

$$G(\mathbf{I}) = 0, \qquad G(\mathbf{A}) = 1$$

G - gene state

H - transcription rate

K - translation rate

r - protein degradation rate

Continuous approximation

$$\frac{dx(t)}{dt} = HG(t) - x(t)$$

$$\frac{dy(t)}{dt} = Kx(t) - ry(t)$$

$x(t) := $ # of mRNA molecules

$y(t) := $ # of protein molecules

Probability density functions

$$f = f(x, y, t) = P(\# \, of \, mRNA = x, \; \# \, of \, protein = y, \, G(t) = 0)$$

$$g = g(x, y, t) = P(\# \, of \, mRNA = x, \; \# \, of \, protein = y, \, G(t) = 1)$$

The continuity equations for $f(x,y,t)$ and $g(x,y,t)$

$$\frac{\partial f}{\partial t} + div[f(\frac{dx}{dt}, \frac{dy}{dt})|_{G=0}] = byg - cf$$

$$\frac{\partial g}{\partial t} + div[f(\frac{dx}{dt}, \frac{dy}{dt})|_{G=1}] = -byg + cf$$

$$\frac{\partial f}{\partial t} - \frac{\partial}{\partial x}(xf) + r\frac{\partial}{\partial y}((Kx - ry)f) = byg - cf$$

$$\frac{\partial g}{\partial t} + \frac{\partial}{\partial x}((H - x)g) + r\frac{\partial}{\partial y}((Kx - ry)g) = -byg + cf$$

Haploidal gene with feedback

- The protein degradation time is much larger than mRNA one (the protein is synthetized directly from the gene and regulates its own expression)

$y(t)$ - protein level

$$\mathbf{I} \xrightarrow{\ c(y(t))\ } \mathbf{A}, \qquad \mathbf{I} \xleftarrow{\ b(y(t))\ } \mathbf{A}$$

$$\frac{dy(t)}{dt} = G - y(t)$$

$$\frac{dx(t)}{dt} = HG(t) - x(t)$$

$$\frac{dy(t)}{dt} = Kx(t) - ry(t)$$

We have the following continuity equations for $f(y,t)$ and $g(y,t)$

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial y}(-y \cdot f) = b(y)g - c(y)f$$

$$\frac{\partial g}{\partial t} + \frac{\partial}{\partial y}((1-y) \cdot g) = -b(y)g + c(y)f$$

$f(y,t) := \Pr(prot.\,level = y,\ G = 0)$

$g(y,t) := \Pr(prot.\,level = y,\ G = 1)$

For $\dfrac{\partial f}{\partial t} = \dfrac{\partial g}{\partial t} = 0$ we obtain $f(y) = \exp\left[\int_0^y \left(\frac{b(z)}{1-z} + \frac{c(z)-1}{z}\right)dz\right]$

$$g(y) = \frac{y f(y)}{1-y}$$

Diploidal gene with feedback

$$\frac{d}{dy}\left(-y \cdot f_0\right) = b \cdot f_1 - 2c \cdot f_0$$

$$\frac{d}{dy}\left((1-y) \cdot f_1\right) = 2c \cdot f_0 - (b+c) \cdot f_1 + 2b \cdot f_2$$

$$\frac{d}{dy}\left((2-y) \cdot f_2\right) = c \cdot f_1 - 2b \cdot f_2$$

$f_0(y) := \Pr(prot.level = y, G = 0)$

$f_1(y) := \Pr(prot.level = y, G = 1)$

$f_2(y) := \Pr(prot.level = y, G = 2)$

$$f_0(y) = \int_{s_1}^{s_2} f(z)f(y-z)dz$$

In the case without feedback

$(c(y) = c_0$ and $b(y) = b_0)$

the functions $f_0(y),\ f_1(y),\ f_2(y)$

are given by the convolution formulas

$$f_1(y) = 2\int_{s_1}^{s_2} g(z)f(y-z)dz$$

$$f_2(y) = \int_{s_1}^{s_2} g(z)g(y-z)dz$$

where $s_1 = 0$ and $s_2 = y$ for $0 \le y \le 1$

$s_1 = y - 1$ and $s_2 = 1$ for $1 < y \le 2$

Positive Feedback

We consider the external induction of self-activating gene

$$\mathbf{I} \xrightarrow{\ c(y(t))\ } \mathbf{A}, \qquad \mathbf{I} \xleftarrow{\ b(y(t))\ } \mathbf{A},$$

$$c(y) = c_0 + c_2 \cdot y^2, \qquad b(y) = b_0$$
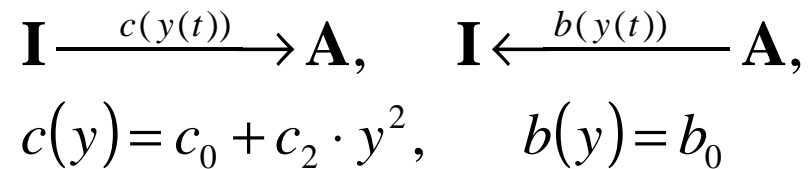
There are three patterns of N - allelic ($N = 1, 2, 4$) gene activation corresponding to different pairs of c_2 and b_0 in (c_2, b_0) - plane

{A} - mode in which gene remains Active
    (i.e. $E_0(y) > \frac{1}{2}$) for all c_0

haplo

{B} - mode in which gene activates for some c_0
    and distribution $\rho(y)$ is transiently Bimodal

diplo

{U} - mode in which gene activates for some c_0
    and its distribution $\rho(y)$ remains Unimodal
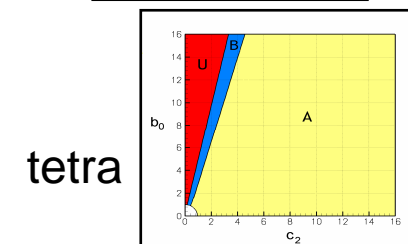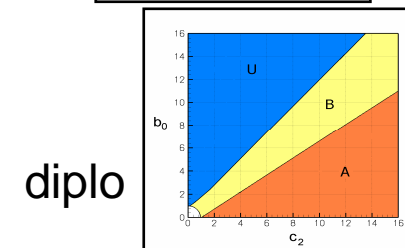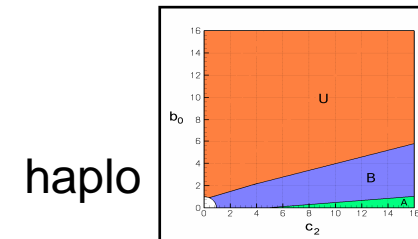
tetra

Protein distributions

3 modes of diploidal gene activation

{A} - mode

[C], diploidal, $c_2=10$, $b_0=3$

c0 = 0.01
c0 = 1
c0 = 5

{B} - mode

[D], diploidal, $c_2=10$, $b_0=10$

c0 = 0.1
c0 = 0.8
c0 = 1.2
c0 = 8

{U} - mode

[E], diploidal, $c_2=3$, $b_0=10$

c0 = 0.5
c0 = 5
c0 = 20
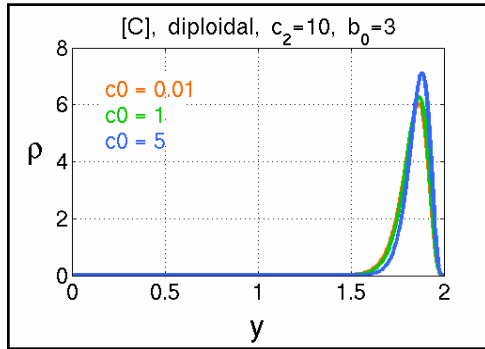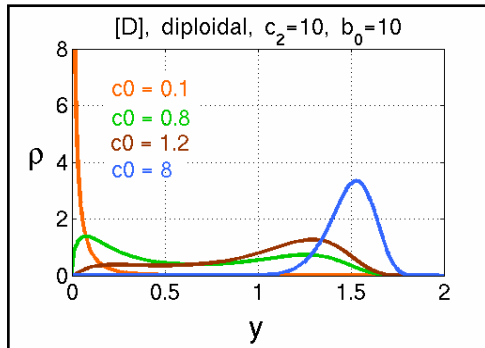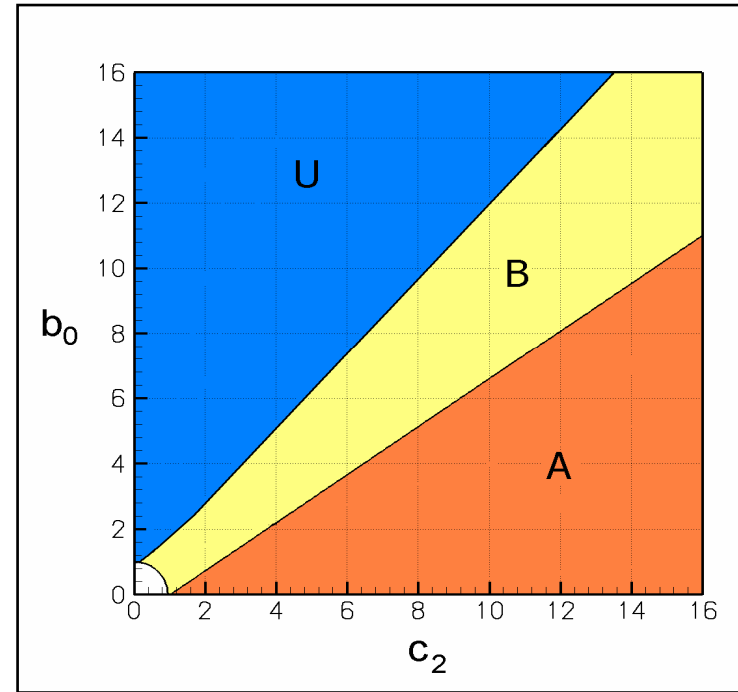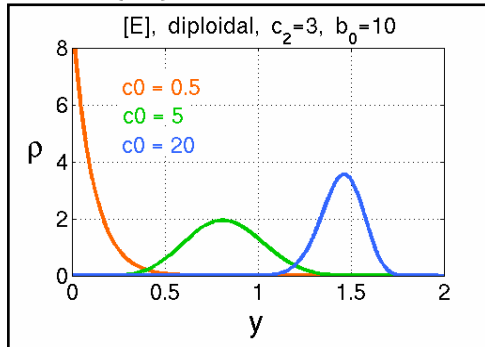
Mean and variance (per gene copy)

[A], $E_0[y]$

c2=10, b0=3
c2=10, b0=10
c2=3, b0=10

[B], $Var_0[y]$
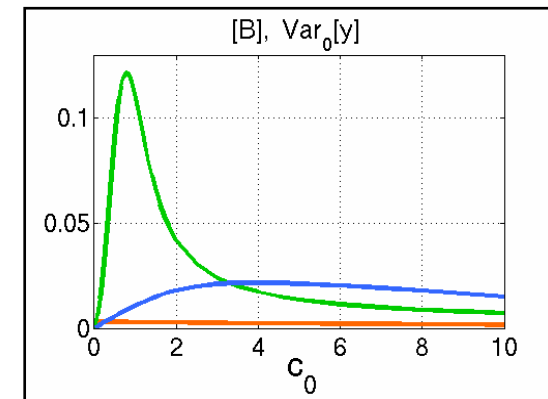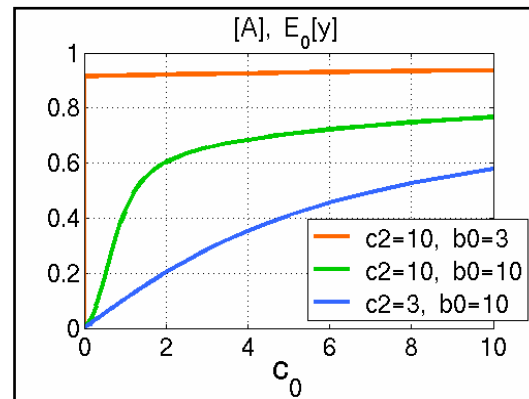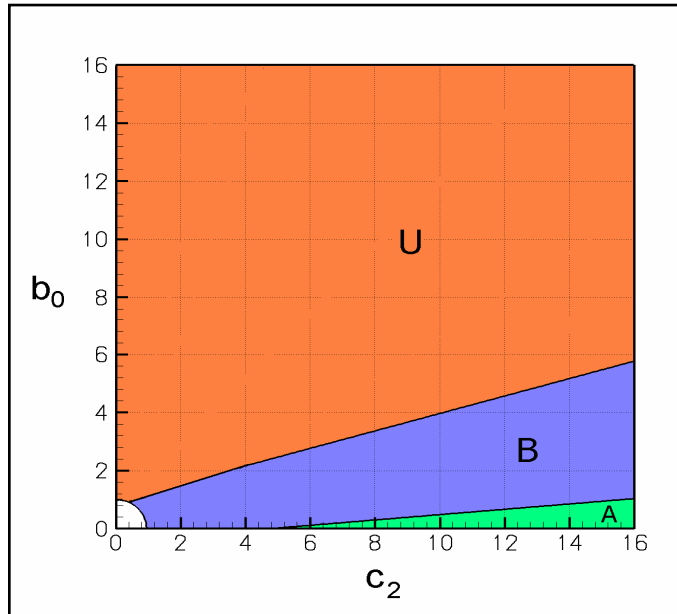
haploidal gene

3 regions in (c_2, b_0) - plane corresponding to 3 modes of activation

diploidal
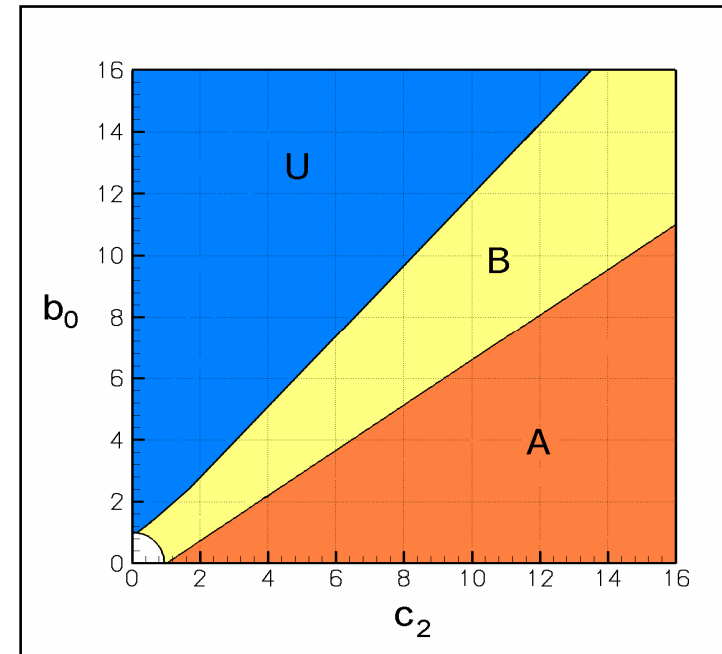
tetraploidal
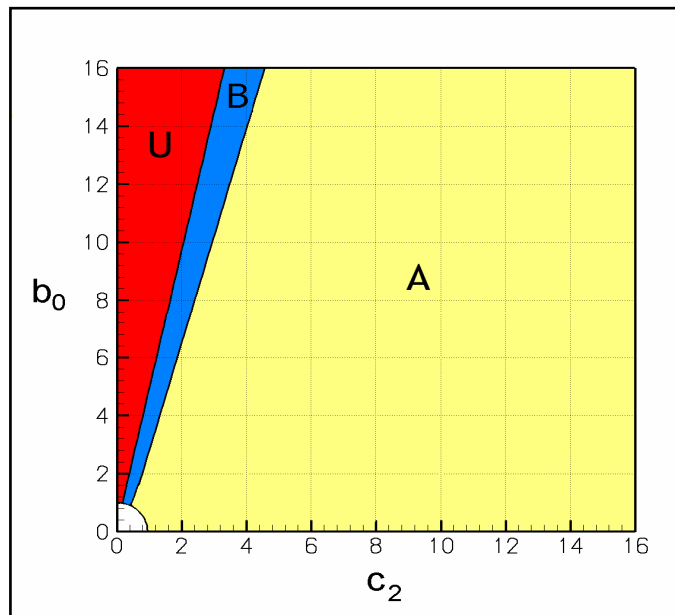
9 possible patterns of 1-,2- and 4-copy-gene-system activation
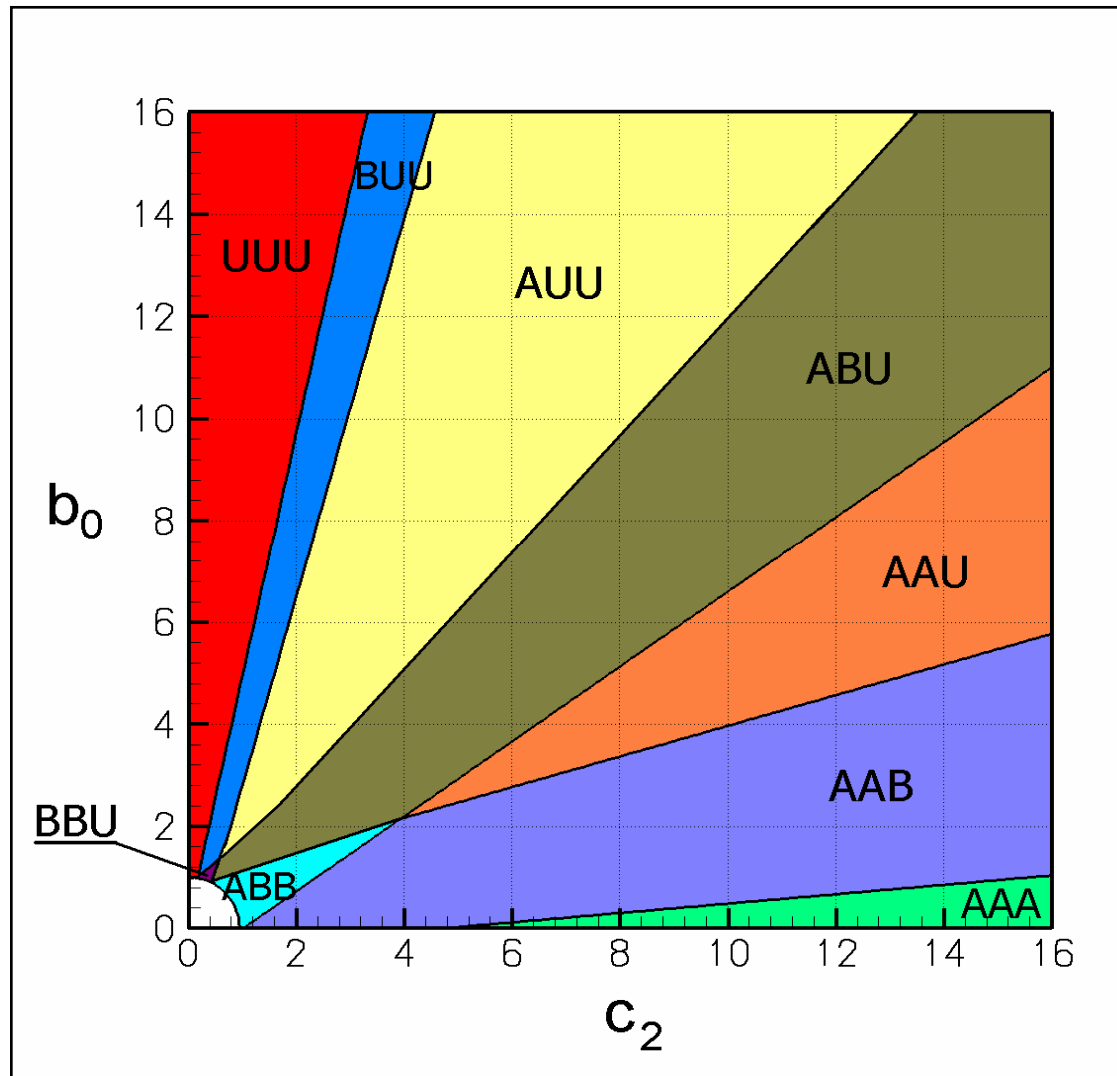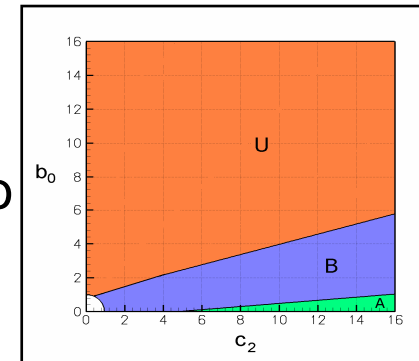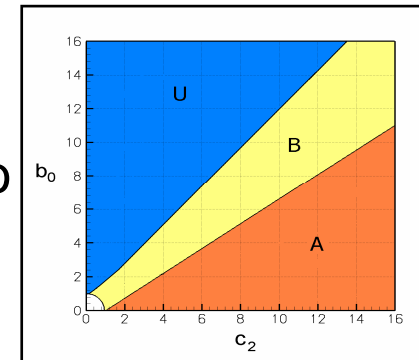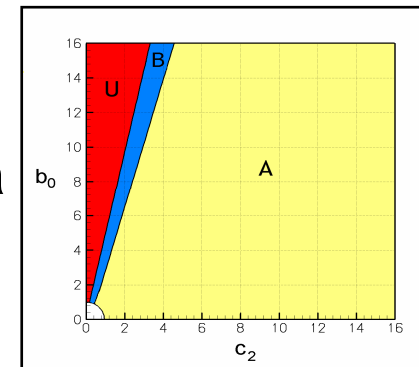
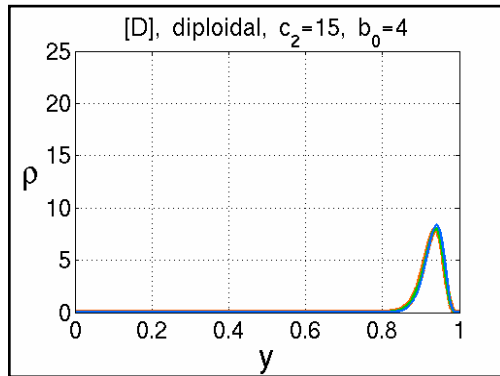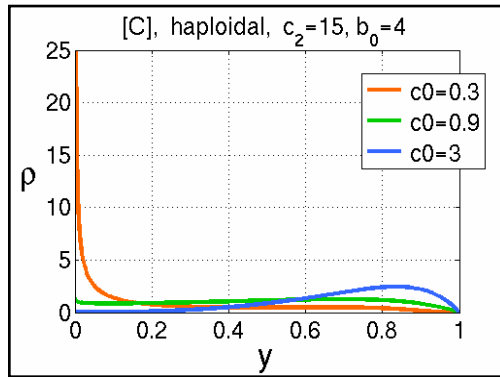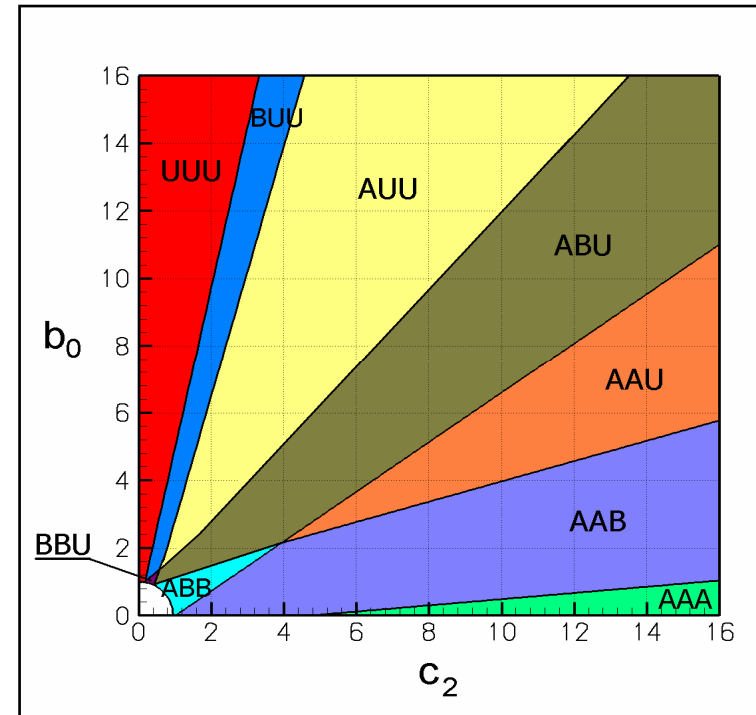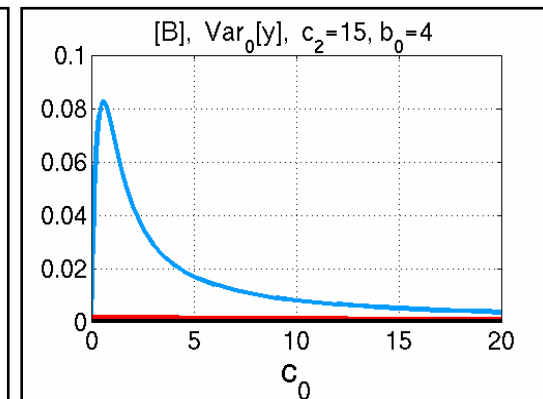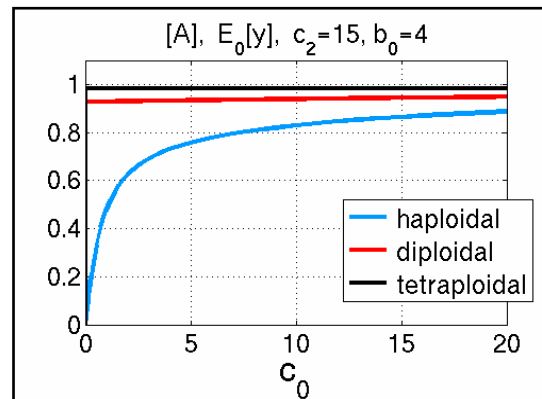{ABU} sector denotes that tetraploidal gene is in the {A}-mode,
diploidal - in the {B}-mode, haploidal - in the {U}-mode

Protein distributions
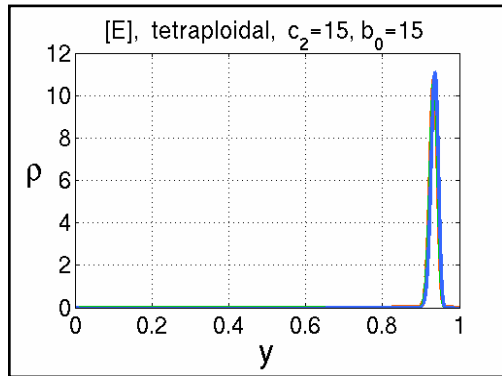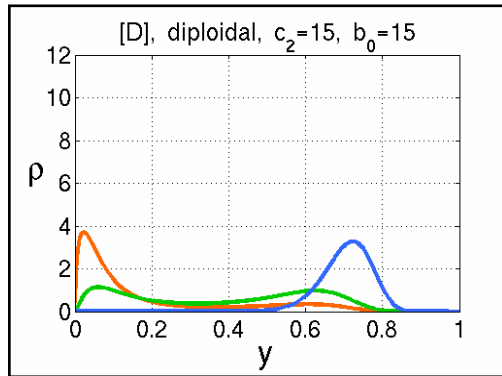
Mode {AAB}, $c_2 = 15$, $b_0 = 4$

[C], haploidal, $c_2=15$, $b_0=4$

[D], diploidal, $c_2=15$, $b_0=4$
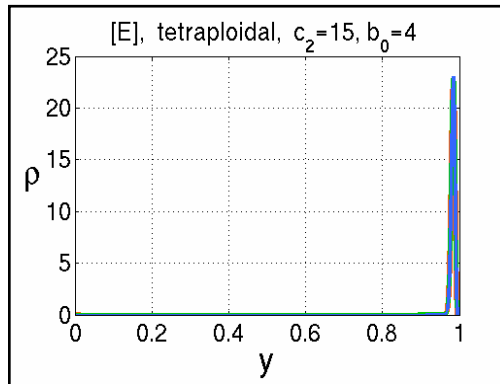
[E], tetraploidal, $c_2=15$, $b_0=4$

Mean and variance of the protein (per gene copy)

[A], $E_0[y]$, $c_2=15$, $b_0=4$

[B], $Var_0[y]$, $c_2=15$, $b_0=4$

Protein distributions

Mode {ABU}, $c_2 = 15$, $b_0 = 15$

[C], haploidal, $c_2 = 15$, $b_0 = 15$

c0=0.8
c0=1.2
c0=7

[D], diploidal, $c_2 = 15$, $b_0 = 15$
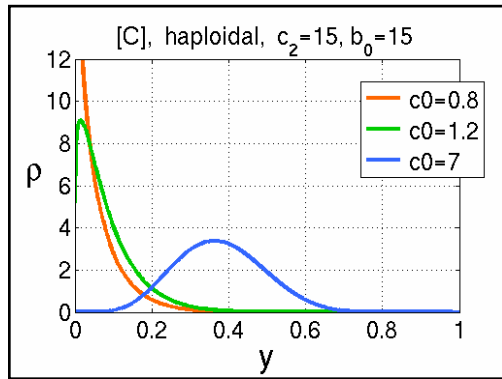
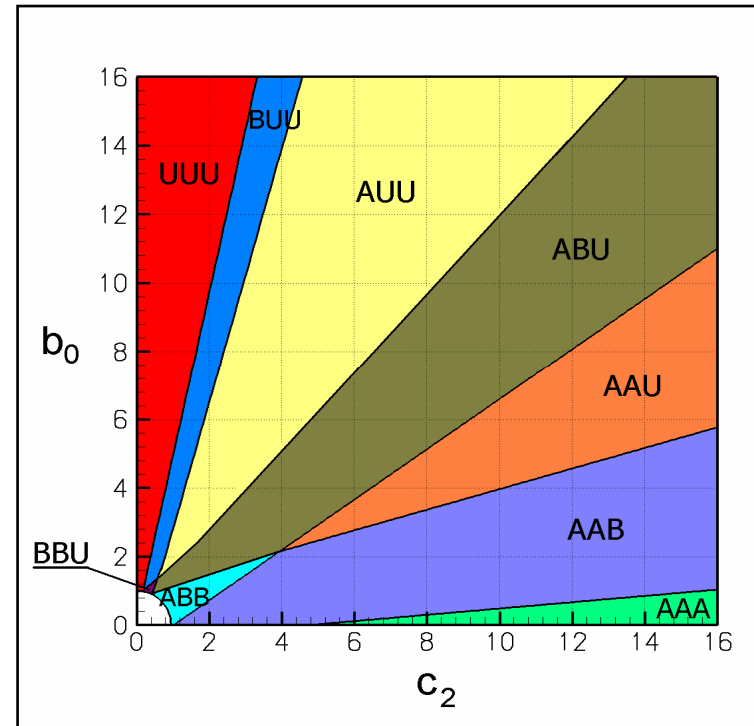[E], tetraploidal, $c_2 = 15$, $b_0 = 15$

Mean and variance of the protein (per gene copy)

[A], $E_0[y]$, $c_2 = 15$, $b_0 = 15$

haploidal
diploidal
tetraploidal

[B], $Var_0[y]$, $c_2 = 15$, $b_0 = 15$
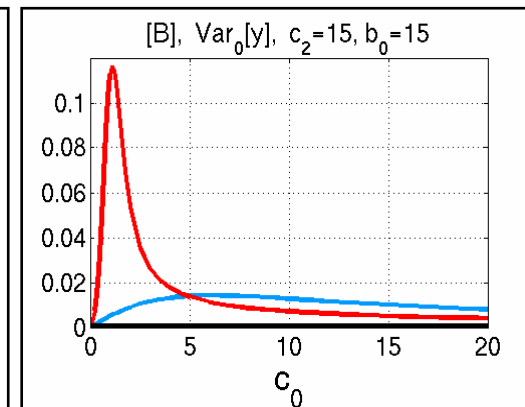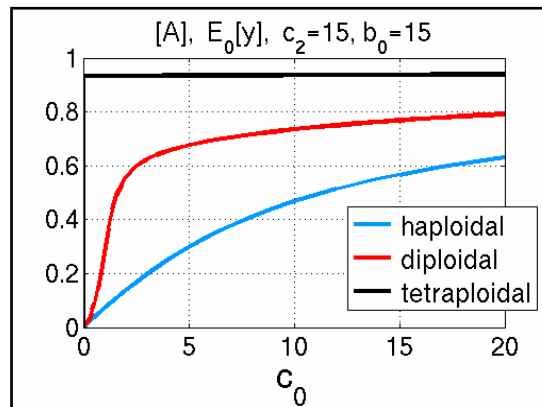
Take Home Conclusions

Considering the simultaneous activation of a haploid,
diploid and tetraploid gene there exist nine modes of
gene activation

- allele loss may stop the persistent gene activity and
  lead to disease if the constant level of gene product is
  required

- gene duplication may result in a persistent activity and
  lead to disease when haploid or diploid gene is „designed"
  to act as a switch