Analysis of cell adhesion during early stages of colon cancer based on an extended multi-valued logic approach†

Daniel V. Guebel,‡[*a] Ulf Schmitz,‡[a] Olaf Wolkenhauer,§[ab] and Julio Vera§[a]

[a]*Department of Systems Biology and Bioinformatics, University of Rostock. 18051 Rostock. Germany.*
[b]*Stellenbosch Institute for Advanced Study (STIAS). Wallenberg Research Centre at Stellenbosch University. Stellenbosch 7600. South Africa.*

‡ DVG and US have equally contributed to this work.
§ OW and JV are equal senior contributors.

E-mail: daniel.guebel@uni-rostock.de

†Electronic Supplementary Information (ESI)

APPENDIX S1. Limitations of classical logic approaches

1.1 ***Motivation.*** We here demonstrate that in case of a generic reaction that is modulated by two or more inhibitors neither the assumption of concerted nor independent inhibition leads to realistic predictions, when using Boolean logic approaches. Because Boolean logic is generalized through the multi-valued logic, we here only present the argumentation for the Boolean case.

1.2 ***Example***. First, we consider a simple case in which a unique substrate (S_1) is transformed into a product (P), while two inhibitors (I_1 and I_2) modulate the reaction. In this situation, inhibition could occur through two mechanisms: (i) by independent action of each inhibitor on the reaction (Table S1-A), or (ii) by a concerted action of both inhibitors (Table S1-B).

Table S1. Truth tables for the formation of a product (P) from a substrate (S), being the process negatively regulated by two inhibitors (I1, I2). **A:** The inhibitors act through independent mechanisms (i.e., I1 OR I2); **B:** The inhibitors act through a concerted mechanism (i.e., I1 AND I2). In both cases, 1 means "present" or "active", while 0 means "absent" or "inactive".

S_1	I_1	I_2	P
1	0	0	1
1	1	0	0
1	0	1	0
1	1	1	0

(Case A)

S_1	I_1	I_2	P
1	0	0	1
1	1	0	1
1	0	1	1
1	1	1	0

(Case B)

As can be seen in Table S1-A, the assumption of independent inhibition frequently leads the full inhibition of the reaction, i.e., in 3 out of 4 cases P is not produced. In contrast to this, the assumption of concerted inhibition leads to a situation where inhibition of P synthesis rarely occurs (3 out of 4 cases). It should be noted, however, that both mechanisms behave similarly when both inhibitors are absent or both inhibitors are present. The trend described in Tables S1-A and S1-B becomes more severe as the number of inhibitors participating in a reaction increases (Figure S1).

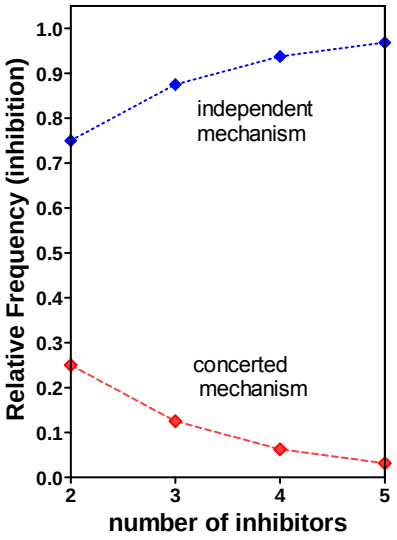


Figure S1: **Outcome of a reaction depending on the number of inhibitors.** We compared the outcome of a reaction, for Boolean logic with concerted and independent inhibition. For a truth table corresponding to a process modulated by several inhibitors, the proportion of cases resulting in effective inhibition was represented as function of the number of participating inhibitors. Blue diamonds represent the results for independent inhibition, whereas red diamonds represent concerted inhibition. Dashed lines are meant only as visual aid to describe the observed trends.

In a reaction influenced by n inhibitors, 2^n possible scenarios of inhibition can occur. If for example, five or more inhibitors are present in a given reaction, the reaction will practically "always" appear as inhibited when inhibitors are assumed to act independently, i.e., the relative frequency of inhibition is $\mathbf{RF}(inhibition) \approx 1$). In contrast, the reaction will practically "never" appear as inhibited in the truth table when inhibitors act through a concerted mechanism ($\mathbf{RF}(inhibition) \approx 0$) (Figure S1). Similar trends can be expected in the case of the multi-valued logic framework.

Truth tables provide an explicit representation of all the space of possible events. Hence, they allow the verification that effective gene repression happens in a unique case out of 2^n possible cases, when concerted inhibition is used. On the other hand, inhibition occurs in (2^n-1) cases for independent inhibition. On the other hand, equations 1 and 2 can be used to determine the probability of inhibition in dependence of the probability of inhibitors to be present. Equation (1) accounts for concerted inhibition, while equation (2) accounts for independent inhibition.

concerted inhibition

$$p(inhibition) = p(I_1)\, p(I_2) \ldots p(I_n) = \prod_{i=1}^{n} p(I_i) \qquad \text{eq.(1)}$$

independent inhibition

$$p(inhibition) = 1 - \prod_{i=1}^{n} (1 - p(I_i)) \qquad \text{eq.(2)}$$

The probability of any event is a positive number and always lesser or equal to 1. Thus, equation (1) indicates, that for concerted inhibition the probability of inhibition decreases as the number of participating inhibitors (n) *increase*. In contrast to that, equation (2) indicates that the probability of inhibition in for independent inhibition increases with the number of participating inhibitors (n). Both conclusions apply for any value $p(I_i)$ smaller than 1 (i.e., $0 \le p(I_i) < 1$). We could verify the trends inferred from these theoretical considerations for one of the scenarios assayed (Boolean analysis assuming concerted inhibition in Table S4). In this case, only 7.7% of the reactions having two inhibitors appear as inhibited. The proportion decreased to 3.8% when the reactions have three or four inhibitors, whereas none of the reactions with five or more inhibitors ever appears inhibited.

1.3 ***Generalization.*** Our initial example can now be generalized to the case of a reaction with *m* substrates and *n* inhibitors (see Table S2). Equations in table S2 (first row) follow the syntax used by CellNetAnalyzer (CNA), where a plus sign (+) represents an "AND" operator (logical concatenation), while an exclamation mark (!) indicates a logical "NOT". In CNA separate reactions leading to the same product are interpreted as "OR" conditions. In the second row of Table S2 the reactions are

represented in common Boolean notation. There, the operators "AND", "OR", and "NOT" are represented by the symbols ($\wedge$), ($\vee$), and (~) respectively.

Importantly, one can see that a unique equation (eq.3) is enough to describe the mechanism of independent inhibition, whereas a system of two equations (eq.4 and eq.5) is required to deal with the concerted inhibition mechanism. While equation (3) is considered in CNA for introducing independent inhibition in signaling networks, we had to combine equations (4) and (5) to come up with a mechanism mimicking concerted inhibition, which has not been explicitly preconceived in CNA.[1,2,3] In logic approaches, natural numbers can be used as coefficients to indicate the *logic level* or *state* of species in a reaction. For example, in the case of the Boolean approach the coefficient w in equation (4) can only have a value of 1 or 0, whereas any positive integer value ($w \geq 0$) can be assigned if multi-valued logic is adopted.

Table S2. The first row shows examples of independent and concerted inhibition using the CNA syntax. In the row below the conditions for the product to be produced are represented in Boolean notation.

Independent Inhibition		Concerted Inhibition	
$S_1 + S_2 + ... + S_m + !I_1 + !I_2 + ... + !I_n = P$	(3)	$I_1 + I_2 + ... + I_n = w\, I_{total}$ $S_1 + S_2 + ... + S_m + w\, !I_{total} = P$	(4) (5)
Condition for product formation: $(\sim I_1) \wedge (\sim I_2) \wedge ... \wedge (\sim I_n) = 0$	(6)	*Condition for product formation*: $(\sim I_1) \vee (\sim I_2) \vee ... \vee (\sim I_n) = 0$	(7)

To prove that the presented equations are in correspondence with the attributed inhibitory mechanisms, we can analyze the conditions required to accomplish them. From the first row in Table S1-A, we concluded that "a reaction modulated by independent inhibitors occurs only if *none* of the potential inhibitors is active". This statement is being formally expressed by equation (6) in Table S2. In contrast, a reaction modulated by concerted inhibition occurs only when the condition given in row 4 in Table S1-B is negated. A formal representation of this condition has been declared in equation (7) in Table S2, which reads "reaction occurs in a concerted inhibition only if *any* of the potential inhibitors are absent or inactive". This conclusion is in full agreement with those inferred from equation (1). Interestingly, observe that a typical biological "OR" concatenation such as the independent inhibition turned into a complementary "AND" concatenation, whereas a typical biological "AND" operation as the concerted inhibition turned into an "OR" operation. This is justified because negations are involved in both cases, being then fully consistent with the De Morgan's laws that govern the logic operations. Moreover, the conclusions arrived for equation (6) and equation (7) are also in full agreement with those previously inferred from equation (2) and equation (1) respectively.

APPENDIX S2: Basis for computing graded inhibition in multi-valued logic

2.1 *Computing basis of the proposed method*

The trends characterized in Figure S1 and equations (1) and (2) show that Boolean logic and also its generalization in form of multi-valued logic, lead to unrealistic results when multiple inhibitors act on one reaction if concerted or independent inhibition is assumed. We here introduce a method named as *graded inhibition* that can deal with the phenomena of multiple inhibitions on the same reaction, something which is not possible with conventional logical approaches. Under the assumption that inhibitors are present or absent according their own physiological regulation, the approach works on the basis of two simple ideas: i) *the more inhibitors present, the higher the inhibitory effect produced*; ii) *inhibitors that are present contribute to the inhibition with equal probability* [(footnote 1)]. These ideas are expressed in logic terms by equations 8a-8b.

$$1\,I_1 + 1\,I_2 + \ldots + 1\,I_n = w_e\,I_{total} \qquad (8a)$$

$$S_1 + S_2 + \ldots + S_m + w\,!I_{total} = (w - w_e)\,P \qquad (8b)$$

The coefficient w is given by the sum of the coefficients from *all* the potential inhibitors able to modulate the analyzed reaction (i.e., $w = \sum_{i=1}^{n} 1 = n$). The coefficient w_e instead is determined by the sum of coefficients of inhibitory species that are active (in equation 8a one inactive inhibitor $I_i=0$ is enough to make w_e lesser than w). Therefore, as can be seen in equation (8b) only in the case when all the potential inhibitors are effectively present ($w_e = w$) the reaction will be fully inhibited ($w\text{-}w_e = 0$). Otherwise, the coefficient ($w\text{-}w_e > 0$) gives the part of non-inhibited substrate that remains to be transformed into product P. Note, that the coefficient w is a fixed property for a given reaction, that represents the *maximum*, *absolute capacity* (or activation level) of the target. The coefficient w_e instead is a variable property of the system that reflects the *current availability* of inhibitors in a given biological scenario. Thus, for a given reaction, the final level of product formed will be a negative function of the coefficient w_e (Figure S2).

In the Figure S2 we have assumed that the amount of product (or the reaction extent) follows a negative, sigmoid dependence on the coefficient w_e. This kind of profile is also frequently used in the field of neural networks and fuzzy logic to deal with phenomena that exhibit saturation and thresholds. Because the multi-valued logic framework cannot process continuous functions, the sigmoid function was quantized by a four-level step-function. Once the coefficients w and w_e are established, their difference ($w\text{-}w_e$) determines univocally the extent of the reaction compatible with the current inhibitory activity, and hence, the level of product that can be formed. Note, that

[1] This is reflected in the fact that in equation 8a all the coefficients of the inhibitors have similar value. This criterion was adopted because the lacking of detailed information about differential effects of inhibitors affecting a given target and due to simplicity reasons. However, within the multi-valued framework is admitted the use of whole values greater than the unit as coefficients. In this case, distinct values for the coefficients of the inhibitors in equation 8a will reflect different probabilities to inhibit the target (e.g, due to different binding or inhibition constants).

our implementation of graded inhibition computes four discrete levels of the product (maximum, high, low, and zero; Figure S2).

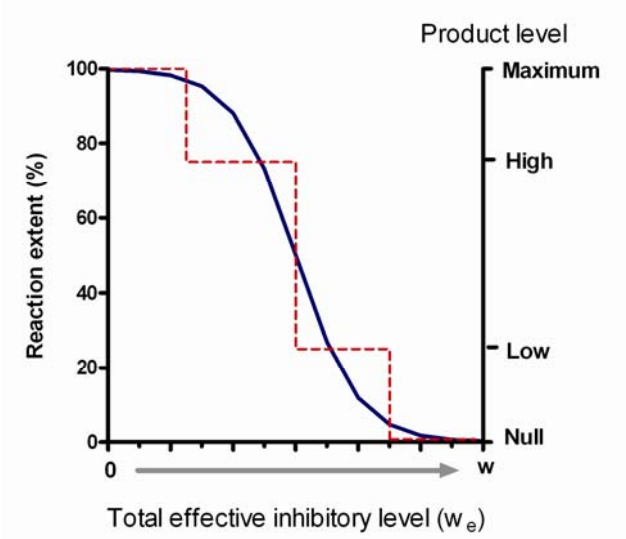


Figure S2: Relationship between the inhibition coefficient determined by the activity of present inhibitors (w_e) and the expected level of synthesized product. In our implementation of graded inhibition the relationship was approximated by a four-level step function (red dashed line), such that four discrete product levels can be computed (Null, Low, High and Maximum).

Equation (8b) maps the variation in the total effective inhibitory level ($0 \le (w - we) \le w$) into levels of product P formed. In Figure S2 we introduce a sigmoid relationship to provide a rational to quantize possible P levels. The process of quantization allowed us to reduce the number of equations to be implemented by our algorithm (see Appendix S3, ESI). After this quantization we only have to deal with four categories of effective inhibitory levels (Null, Low, High, Maximum; see Appendix S3, ESI).

Multi-valued logic supports categorization in more than two levels. We chose four levels of quantization as a trade-off between the computational effort required and the simplicity in the interpretation of the results. We note, that the computational effort grows nearly exponentially for each additional level considered. In addition, the use of four quantization levels contributes to minimize the errors in the assignation of categories (see Appendix S5, ESI), and therefore, to diminish the error propagation in the network (see Table S4).

2.2 *Biological basis of the proposed method*

Due to the additive nature of the coefficient w_e, our method of graded inhibition accounts for the combinatorial, linear effect of the inhibitors rather than for their individual effects. This characteristic can be seen as a consequence of the associative property present in the sum operation. In fact, the same level of inhibition can result from distinct combinations of the inhibitors influencing a given reaction (Figure S2, Appendix S2, ESI). We note, that the number combinations increases with the number of participating inhibitors (Appendix S3, ESI). For simplicity, we neglect differences in the inhibitory capability of individual inhibitors. We rather concentrate on the fraction of active inhibitors from the total number of possible inhibitors to determine the

effective inhibitory level (Equations 8a and 8b). However, the method proposed can handle distinct inhibitory capabilities for individual inhibitors; given that enough biological information is available (see footnote 1, page 5).

Many modeling formalisms are phenomenological. Our approach of *graded inhibition* was conceived as an operational strategy to overcome limitations of current logic formalisms (Appendix S1 and Table S4, ESI). This is in line with other phenomenological modeling formalisms used in biochemical networks, which provide meaningful insights despite their level of abstraction (e.g., Petri nets, fuzzy logic). Interestingly, the widely accepted formalisms of Metabolic Control Theory (MCA)[4] and Biochemical System Theory (BST)[5] allow the analysis of systemic properties of biological networks without any assumptions about the molecular mechanisms governing the network reactions.

No biological evidences for concerted or independent inhibition in transcriptional activity. In the literature there is no evidence about the existence of concerted inhibition for transcription factors and repressors binding to the DNA *in vivo*. Rather than that, the seminal articles discussed in the context of concerted inhibition refer to allosteric enzymatic reactions.[6,7] Similarly, there is no information that supports the *independent* repression by multiple repressors of genes implicated in the regulation of colon cells. Observations of independent effects usually come from studies conducted under non-physiological conditions, such as artificial *in vitro* overexpression of the repressor. *Concerted* and *independent* inhibitions are widely used simplifications, which are extrapolated from enzyme kinetics, rather than a mechanism proven for transcriptional regulation.

The linear additive nature of multiple inhibitions. Under *in vivo* conditions the components of biological networks are frequently sub-saturated and behave with a quasi-linear kinetics with respect to their regulators.[4,5] Cells require having an adjustable kinetic response in order to adapt to the changes in the cellular microenvironment. It has been verified in different biological systems that the control of biological networks is not centralized, but distributed among different players participating in the network regulation. This provides to cell systems an intrinsically robust and redundant structure.[8] It is therefore feasible to consider a *linear*, *additive effect* of the inhibitors with respect to their common target, such as we have assumed in our algorithm of *graded inhibition.*

Synergies in the regulation by multiple inhibitors. Due to the synergistic nature of biochemical networks, the sum of little or moderate contributions from several inhibitors can completely silence the target. Our algorithm can model that kind of synergistic effects given there is enough data to describe them. Unfortunately, there is not enough information in the literature that accounts for these fine effects in case of colon cells (see footnote 1, page 5).

APPENDIX S3: Algorithm to compute multi-valued graded inhibitions

We here describe the rationale of our algorithm that enables the processing of graded inhibition in the multi-valued logic framework. In spite that equations (8a) and (8b) shown in Appendix 2 provide a simple conceptual basis to compute the required algorithm, its implementation in terms of the CNA syntax is not trivial. For this reason, we developed a Python script (named as muval) that works on an equivalent logical structure but is compliant with the CNA syntax. The script can be freely executed under the URL http://www.sbi.uni-rostock.de/muval.

The logic underlying the script is based on the following: a) Instead of a unique virtual entity ($w_e\, I_{\text{total}}$) as in equation (8a), we have to consider explicitly the space of feasible combinations of inhibitors by an equivalent term H_k, where index *k* denotes sum of the coefficients of active inhibitors in the reaction analyzed; b) Instead of a unique reaction accounting for the total inhibitor process ($w\,!I_{\text{total}}$) as in equation (8b), we have to consider explicitly the space of feasible reactions due to the distinct combinations of inhibitors that is accounted by the series of H_k terms; c) In each of the reactions considered, the level of product P formed is given by the coefficient $(w - w_e)$, but referred to the particular combination of inhibitors considered in each reaction.
The practical implementation of our algorithm implies an expansion of the original set of equations in the model, a task that is automatically performed by the Python script. Here, we show how our script generates a set of reactions in CNA notation in order to compute multi-valued logic with *graded inhibitions* (see Table S3).

Table S3. This example shows how the algorithm computes graded-inhibitory responses within multi-valued logic. In this case, a unique reaction (R_1) for the synthesis of product (P) depends on two substrates (S_1, S_2) and three inhibitors (I_1, I_2, and I_3). This requires considering eleven feasible subprocesses.

Cell Net Analyzer Notation		Multi-valued logic with graded inhibitions[1]			
Space of Inhibitors		I_1	I_2	I_3	P
R_{i31}	$I_1 + I_2 + I_3 = H_3$	1	1	1	
R_{i21}	$I_1 + I_2 + !I_3 = H_2$	1	1	0	
R_{i22}	$I_1 + !I_2 + I_3 = H_2$	1	0	1	
R_{i23}	$!I_1 + I_2 + I_3 = H_2$	0	1	1	
R_{i11}	$I_1 + !I_2 + !I_3 = H_1$	1	0	0	
R_{i12}	$!I_1 + I_2 + !I_3 = H_1$	0	1	0	
R_{i13}	$!I_1 + !I_2 + I_3 = H_1$	0	0	1	
Space of Reactions					
R_1	$S_1 + S_2 + !I_1 + !I_2 + !I_3 = 3\ P$	0	0	0	3
R_{1a}	$S_1 + S_2 + H_1 = 2\ P$				2
R_{1b}	$S_1 + S_2 + H_2 = 1\ P$				1
R_{1c}	$S_1 + S_2 + H_3 = 0\ P$				0

[1] For simplicity reasons, the coefficients corresponding to the substrates S_1 and S_2 are omitted

In order to reduce the space of reactions we quantized the level of inhibition into zero, low, high and maximal (Appendix S2). The idea behind this is to come up with four categories for the effective inhibitory level that follow the step-function presented in Figure S2. Our script first computes the values w_e that mark the transitions between two consecutive categories. The first transition (no inhibition →low inhibition) is calculated by the following equation:

$$H_1 = integer\left(\frac{w}{3}\right) \qquad (9)$$

Here, H_1 is a truncated integer that denotes the effective inhibitory level below which no inhibition occurs (Table S3). Above this value low inhibition is induced. The second transition is calculated as follows:

$$H_2 = w - integer\left(\frac{w - H_1}{2}\right) \qquad (10)$$

For example, if a reaction includes 14 potential inhibitors (w=14), it is not needed to consider the complete set of the reactions accounting for all the combinations between 14 inhibitors. In this case, equation (9) allows us to establish that the transition between the category of *no inhibition* and *low inhibition* should be placed at H_1=4. Thus, in the simulations of this reaction, *no inhibition* will occur when the number of active inhibitors is less than four. Because this interval exceeds the mere absence of inhibitors (H_0), it can be considered as a *threshold of inhibition* (see Appendix S5, ESI). Moreover, equation (10) allows us to place second transition at H_2=9. Therefore, when five to nine inhibitors are active, logical simulations will return *low inhibition* for the reaction. In a similar manner ten to thirteen active inhibitors will result in *high inhibition* and *full inhibition* will only occur if all inhibitors are active.

Finally, by defining the input conditions that correspond to the biological scenario to be analyzed, the network can be simulated in CellNetAnalyzer and the corresponding logical steady states computed in terms of multi-valued logical with *graded inhibitions*.

APPENDIX S4: Quality Assessment

4.1 *Comparison of the performance of logical methods*

Our algorithm for multi-valued logic does not use the current concepts of concerted and independent inhibition, but instead uses the concept of graded inhibition (Appendix S2, ESI). Therefore, we want to test if our approach can produce better predictions compared to the others. Towards this aim, we compared the results obtained by our algorithm against those obtained by current Boolean approaches (see Table S4). We chose Scenario 4 (normal, colonic differentiated cells) as basis for the comparison because the expected profile of the chemical species in this scenario is much better known than the ones for the other scenarios. Thus, Scenario 4 is considered as control condition for the following analysis.

Mol. Biosys., 2012, DOI:10.1039/c2mb05277f

Table S4: Results from the simulation of Scenario 4. The proposed method (multi-valued logic with graded inhibition) is compared with the Boolean logic approaches (assuming concerted or independent inhibition). Green cells indicate simulation results that are in accordance to what can be expected, whereas red cells indicate wrong results. The values of the results coming from the multi-valued model were normalized to lie between 0 and 1 by expressing the ratio between the computed level and the maximum level achievable for the indicated species. When the resulting quotient was a fractional number it was underlined to indicate that it is more precise than what the other Boolean approaches produce.

		Boolean		Multi-valued
Function	**Species**	**Concerted inhibition**	**Independent inhibition**	**Graded inhibition**
Mature Adhesion Structures	Tight_Junction(pm)	0	0	1
	Desmosomes(pm)	0	0	1
	CdhE_ß-catenin_Complex$_{(AJ)}$	0	0	1
Adhesion Molecules	CdhE_ß-catenin_IQGAP(pm)	1	0	1
	CdhE_ß-catenin(pm) and CdhE_ ß-catenin$_{(cis)}$	0	0	1
	CdhE_ ß-catenin$_{(trans)}$	0	0	1
	c-Src*$_{(AJ)}$	0	0	1
	ZO-1$_{(AJ)}$	1	1	1
	Nectin-like1/4(pm) and Nectin-like1/4_Afadin$_{(cis)}$	1	1	1
	Nectin-like1/4$_{(trans)}$ and Nectin-like1/4$_{(trans)}$_RAP1*	1	1	1
Actin Structures	Bundled_Actin	0	0	1
	Branched_Polymerized_Actin	0	0	0
Regulatory & Intermediary Species	RAC1*$_{(AJ)}$ and CDC42*$_{(AJ)}$	0	0	1
	RhoA*$_{(AJ)}$	0	0	1
	RAC1*$_{(llamelopodia)}$	0	0	0
	CDC42*$_{(llamelopodia)}$	0	0	0
	RhoA*$_{(llamelopodia)}$	1	0	0.5
	CdhE(c)	1	0	0.83
	CdhEP(Ser683,686,992)	1	0	0.66
	SNAIL	1	0	0.66
	CdhE$_{(endocyted)}$	0	0	0
	CdhE$_{(recycled)}$	0	0	0
	CdhE$_{(catabolized)}$	0	0	0
	HAKAI	1	1	1
	DSG$_{(shedding)}$	0	0	0
	ILK*(pm)	1	1	0.33

	ZEB1/2(n)	1	1	0
	FascinP(c)	1	1	0
	FAK1(c)	0	0	1
	FAK1P(Tyr397)(pm)	0	0	1
	FAK1P(Ser910,843)(pm)	0	0	0.33
	Caveolin(c)	1	1	1
	Caveolin(pm)	1	1	1
	α-actininP(Tyr12)(pm)	0	0	0
	α-actininP(Tyr4,31)(pm)	0	0	0.66
	IQGAP_CaMK(c)	0	0	0
	Vinculin(c)	1	1	1
	PlakoglobinP(Tyr549)(c)	1	1	0
	PlakoglobinP(Tyr643)(c)	0	0	1
ß-catenin (nuclear activity)	ß-catenin_TCF4(n)	0	0	0.25
	Survivin	0	0	0.62

In Table S4 it can be seen that our approach of multi-valued logic with graded inhibitions generated results that match with those physiologically feasible for this scenario, whereas both Boolean approaches considered (concerted or independent inhibition) produced many false predictions. From 45 chemical species listed in Table S4, 17 cases (37.8%) showed full agreement between the three tested methods, while in 19 cases (42.2%) both Boolean approaches failed to produce the expected results, which were obtained by the multi-valued. Only in 1 case (2.2%), one of the two Boolean approaches yielded results in agreement with the multi-valued. The 17 cases with full agreement between the three methods are explained by the occurrence of reactions in the model, which due to their simple structure do not depend on the inhibition criterion used. The 1 case, in which one of the Boolean showed to provide similar results that multi-valued, merely reflects a sub-set of reactions for which the multi-valued formalism was not fully exploited during the building of the network. Importantly, the 19 cases in which neither of the Boolean approaches tested could parallel the performance of the multi-valued reflect the limitations of the Boolean approaches as well as the advantages of the multi-valued (see Appendix S1, ESI).

From a biological perspective, the crucial point in Table S4 is that neither of the two Boolean approaches tested could predict the occurrence of desmosomes, tight-junctions, and the mature complex of CdheE_ß-catenin$_{(AJ)}$ as it has been proven in the literature.

In Textbox S1 we present a discussion of the limitations of the Boolean approaches, which provoke the differences in the outputs generated by the multi valued approach in Table S4.

Textbox S1: Situations leading to erroneous model inferences in Boolean approaches as compared with the multi-valued logic

1. Loss of inhibition thresholds when reactions are given in Boolean representation:

   S + **2** !I = P (multi-valued) S + !I = P (Boolean)

2. Loss of activation thresholds when reactions are given Boolean representation:

   3 S + !I = P (multi-valued) S + !I = P (Boolean)

3. Undue gene repression in case of independent inhibition, when most of the other repressors are absent or inactive:

   $P_{(gen)} + S + !I_1 + !I_2 + \ldots + !I_n = P$; $I_i=1$, $I_j=0$ for $j \neq i \rightarrow P=0$

4. Undue gene expression in case of concerted inhibition, when only one of the inhibitors is inactive:

   $I_1 + I_2 + \ldots + I_n = I_{Total}$ and $P_{(gen)} + S + !I_{Total} = P$;

   $I_i=0$, $I_j=1$ for $j \neq i \rightarrow P=1$

5. Impossibility of Boolean approaches to discern between different biological scenarios (e.g., normal and pathological) when the same species is expressed at different levels:

   Multi-valued $\begin{cases} P_{(gen)} + S_1 = P \text{ and } P = Q \\ P_{(gen)} + S_2 = 3P \text{ and } 3P = R \end{cases}$; $Q \neq R$

   Boolean $\begin{cases} P_{(gen)} + S_1 = P \text{ and } P = Q \\ P_{(gen)} + S_2 = P \text{ and } P = R \end{cases}$; $Q = R$

6. Combinations of the previously enumerated problems propagate across the network.

4.2 *Tracing the causes of discrepancies in Table S4*

Here, we describe how the use of Boolean approaches induces errors like those shown in Table S4. In this section we illustrate the statements given in Textbox S1 with examples from our model simulations.

Most of the observed problems are a direct consequence of the use of Boolean logic approaches to describe given processes. This is for example the case for the discrepancies observed in SNAIL *activation.* In Table S4 it can be seen that SNAIL protein was not produced when using Boolean logic with independent inhibition. We note that according to our model SNAIL production can be repressed by up to six different inhibitors (see reactions R1-R6 in the Table S5, ESI). In the simulated biological scenario three of them were active (Delta-Gli-3, miR-let-7,

p53Tetramerized). Under the assumption of independent inhibition one repressor active is enough to silence the SNAIL gene. Whereas, three out of six active repressors, such as computed, are not enough to repress the SNAIL gene in the case of concerted inhibition. Hence, maximum SNAIL is produced under this criterion.

According to point 6 in Textbox S1 combinations of errors associated to the use of Boolean approaches propagate across the network and provoke phenomena as can be observed in Table S4. For instance, when both Boolean formalisms produce identical results, although opposite assumptions about the mechanisms of multiple inhibition were considered. This is the reason for the discrepancy observed for the case of tight junctions. According to our model, the formation of tight junctions requires the presence of at least four substrates (Occludin, Claudin, ZO-1, Contactin), while being repressed by nine inhibitors (SNAIL, c-Met*, $(\text{alpha-catenin})_2$, alpha-catenin, TNFalpha, ROS, c-Src*(pm), FascinP(c), and Claudin-1) (see reaction R67 in Table S5, ESI). In case of Boolean logic with concerted inhibition, the absence of tight junctions is explained by the lack of the protein Claudin, one of the substrates needed. We note, that the presence of SNAIL leads to the repression of Claudin. This is because partial repression is not allowed in the Boolean approaches (see item 1 in Textbox S1 and reactions R72-R75 in the Table S5, ESI). With the multi-valued treatment, SNAIL was present, but not at maximum possible level. Hence, Claudin was available and lead to the formation of tight junctions.

In the case of Boolean logic with independent inhibition the cause for the absence of tight junctions is different. SNAIL was not produced, and hence Claudin was not repressed. However, the problem emerges from the loss of the induction threshold for ILK* (Point 2 in Textbox S1 and reaction R206 in Table S5, ESI). This error results in the false activation of FascinP(c) (reaction R214 in the Table S5, ESI), which in the case of Boolean logic with independent inhibition is sufficient to block tight junction formation, even if other inhibitors are absent. The abnormal presence of FascinP(c) in the Boolean simulation, among other factors, also masks the formation of bundled actin, which is characteristic for normal differentiated colon cells (see Table S4 and reaction R36b in the Table S5, ESI). In brief, by some combination of the items described in Textbox S1 is also possible to explain the other Boolean discrepancies observed in Table S4 ($\text{CdhE\_ß-catenin\_Complex}_{(\text{AJ})}$, CdhE_ß-catenin(pm), CdhE_ß-catenin(cis), CdhE_ß-catenin(trans), $\text{c-Src*}_{(\text{AJ}),}$ ZEB1/2(n), FAK1(c), FAK1P(Tyr397)(pm), PlakoglobinP(Tyr549)(c), PlakoglobinP(Tyr643)(c), etc).

4.3 *Distinguishability in the extended multi-valued formalism*

Data in Table S4 provides strong evidence supporting that the proposed method of multi-valued graded inhibition correctly predicts the expected outcomes for Scenario 4. We further analyzed whether our method preserves the grouping and distance between the input/output patterns of the model across the six scenarios tested (Figures S3-A and S3-B). This was done by comparing the relationships between all the tested scenarios, first in terms of inputs applied (Table 1) and subsequently for the outputs generated from the simulations (Table 2). The rational for this, is to verify whether the proposed method affects the *distinguishability* of the scenarios considered. To implement this test, each one of the matrices compared was subjected to principal component analysis (PCA).[9,10] In PCA each data matrix is factorized as the product of a matrix of *t*-scores

by the transpose of the matrix of eigenvectors. The *t*-scores are the values of the coordinates for the scenarios as single points in the space delimited by the eigenvectors of the data. Interestingly, the dimensionality of the spaces delimited by the dominating eigenvectors was lower than the dimensionality of the spaces generated by the original variables in each case. In fact, the maximum number of eigenvectors to be considered is dictated by the rank of the matrices containing the data (ρ). In our case, both matrices showed to have the same rank (ρ=6). Having obtained a distribution of points in a sub-space of dimensionality six for two different conditions (inputs and outputs), the first six *t*-scores belonging to each set were used to compute the *distance* among the tested scenarios in each condition. This was done using the standard formula of the Euclidean's distance:

$$\text{distance}_{A,B} = \sqrt{\sum_{i=1}^{m} (x_{i,A} - x_{i,B})^2} \text{ , where } 1 \le i \le m = 6 .$$

Finally, after transforming the initial data matrices into distance matrices, they were used to perform a hierarchical clustering,[11,12] which allows to establish and visualize the relationship between the scenarios (Figure S3-A and S3-B). PCA and cluster analysis were performed using Matlab® v.13. The power, versatility, and validity of the sequential use of PCA and cluster analysis have been proven in other biological investigations.[13,14]

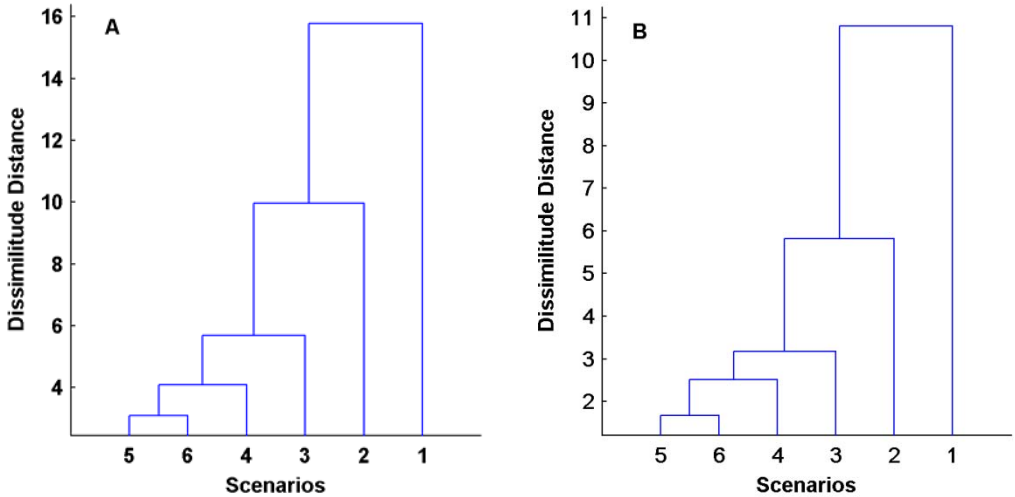


Figure S3: Comparison of the six assayed scenarios after hierarchical clustering (Ward's method, Euclidean distance). **A.** Dendrogram showing the relationship between the scenarios such as can be inferred from the matrix of inputs (Table 1). **B.** Dendrogram showing the relationship between the scenarios such as can be inferred from the matrix of outputs generated by running the model of multi-valued logic with graded inhibitions (Table 2). *Scenario 1*: normal colon stem cells; *Scenario 2*: stressed colon stem cells; *Scenario 3*: colon tumor stem cells; *Scenario 4*: normal differentiated colon cells; *Scenario 5:* stressed differentiated colon cells; *Scenario 6*: differentiated colon tumor cell.

The comparison of dendrograms from Figure S3-A and S3-B confirms that the reciprocal relationship between the six scenarios assayed was maintained after the computation of our algorithm. Moreover, the input matrix (Table 1) and the output matrix (Table 2) proved to have the same rank (ρ=6). Taken together, we conclude that our method does not introduce distortions. Firstly, because the individuality of the tested scenarios was guaranteed; the conservation of the rank in the output matrix

supports the mathematical independence of the scenarios after the computing. Additionally, because the degree of relatedness was also preserved; the dendrograms showed similar hierarchical structure before and after the analysis.

The unique change observed in Figure S3-B was 30% shrinkage in the range of variation in the y-axis. However, this is not a distortion, but a consequence of the predominantly linear processes in the network. This overall linear-behavior was not an obstacle for dealing with the non-linear processes occurring in stem cells. Some cases of those non-linear processes appear in Scenarios 1 and 3 (see Table 2). Examples for those processes are: the change in the ratio of FAK1P(Tyr397) to FAK1P(Ser910,842), the nuclear translocation of ZO-1 instead to its incorporation to adhesive structures as ZO-$1_{(AJ)}$, the induction of Claudin-1 instead of the synthesis of other Claudin isoforms leading to tight-junction formation, the translocation of ß-catenin from the cytoplasm to the nucleus instead to its incorporation to adhesive complexes of CdhE at the plasmatic membrane, and more.

Appendix S5: Thresholds in Boolean and multi-valued logic

A distinctive feature of multi-valued logic is the possibility to define thresholds.[15] We have already shown how the four categories of discretization implemented through equation (9) allow placing a threshold of inhibition (Appendix S3, ESI). It can be argued that Boolean logic can also be adapted to acquire an active state (on) above a given threshold and an inactive state (off) below the threshold. Figure S2 (Appendix S2, ESI) shows that it is difficult to define a reasonable cut-off value when using the Boolean approach. In fact, it can be moved from smaller to larger numbers of active inhibitors, depending on the kind of inhibitory mechanism modeled. We note that, the cut-off position chosen can provoke *classification problems*[16,17,18] like the following ones: scenarios falsely classified as inactive repression (computed as active, when actually being 'partially' repressed; false negatives) and scenarios falsely classified as active repression (computed as fully repressed, when actually being partially or not repressed, false positives).
The classification problem is intrinsic to any quantization procedure, hence it is also present in the multi-valued logic approach proposed. However, false classifications errors are more frequent in Boolean logic because the problem is amplified when fewer categories are considered. Compared to Boolean logic the problem is reduced when four ordinal categories of inhibition are used, as has been proposed in our algorithm. We note, that the practical impact of the classification error is higher for the extreme categories (null and maximum) rather than for central categories (low, high). In terms of our simulation the classification problems that are important are the distinction between *null* and *low level* inhibition and between *high level* and *full inhibition*.

Appendix S6. In Table S5 (see Excel file) is given the detail about the reactions included in our model of cell adhesion in colon. The reactions are written following the CNA syntax. This file also contains a complete list with common and scientific names of all the chemical species and genes participating in the model, and 229 bibliographic references which in addition to the 111 references given in the main text, justify the reactions considered by the model.

Appendix S7

To the aim to illustrate the complexity and integrative nature of our model, in Figure S4 we present an image of the Cytoscape map of our model. This allows the visualization of the main functional regions considered in our network. In addition, to document the network we also provide the supplementary file ADHESION_MAP_CYTOSCAPE.cys. To visualize this file, it must be load on the free software Cytoscape[19] (available from Web site: http://www.cytoscape.org).

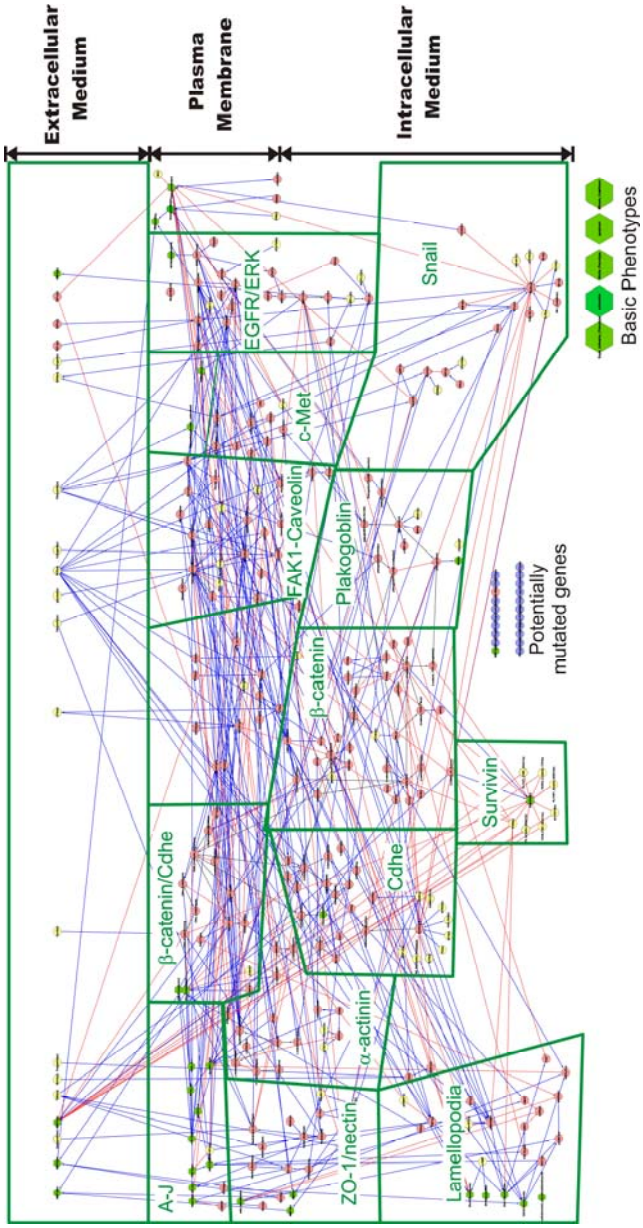
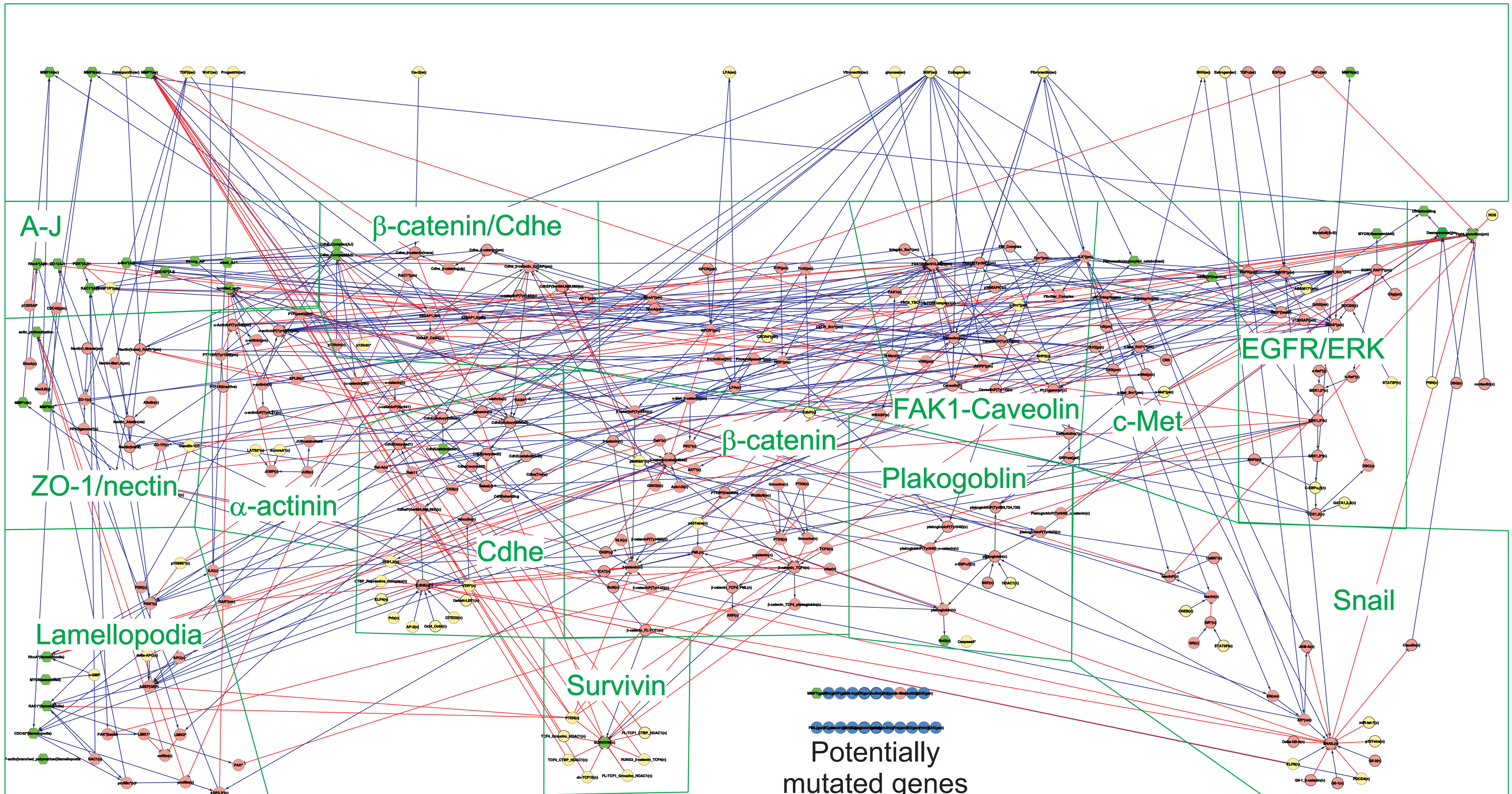


Figure S4: Cytoscape map of the cell adhesion model. The diagram is organized to show the compartments where processes take place (extracellular medium, plasma membrane, intracellular medium). Hierarchical order is given by the inputs to the system (yellow circles), the intermediate layer of the model (red nodes), and read-outs (green hexagons). Blue arrows account for activation processes, while red hammerheaded lines represent inhibition. Green polygons delimit functional areas, while illustrating the high degree of interconnectedness among the different sub-systems in the network. For details see the scalable high resolution image below.



REFERENCES

1 Saez-Rodriguez L, Simeoni J, Lindquist R, Hemenway U, Bommhardt B, Arndt U, Haus U, Weismantel R, Gilles ED, Klamt S, Schraven B. A logical model provides insights into T-cell signaling. PLoS Comp Biol. 2007, 3(8): e163.

2. Samaga R, Saez-Rodriguez J, Alexopoulos L., Sorger P. K., Klamt S. The logic of EGR/ErbB signaling: theoretical properties and analysis of high-throughput data. PLoS Comp Biol. 2009, 5(8):e1000438.

3. Schlatter R, Schmich K, Avalos Vizcarra I, Scheurich P, Sauter T, Borner C, Ederer M, Merfort I, Sawodny O. ON/OFF and Beyond -A Boolean Model of Apoptosis. PLoS Comput Biol. 2009, 5:e1000595.

4. Visser D, Heijnen JJ. The mathematics of metabolic control analysis revisited. Metabolic Engineering 2002, 4:114-123

5. Savageau MA. Biochemical Systems Theory. Operational differences among variants representations and their significance. J Theor Biol. 1991, 15: 509-530.

6. Monod J, Wyman J, Changeux JP. On the nature of allosteric transitions: A plausible model. J. Mol. Biol. 1965, 12:88-118.

7. Koshland DE, Némethy G, Filmer D Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry. 1966, 5:365-8.

8. Moreno-Sanchez R, Saavedra E, Rodriguez-Enríquez S, Galllardo-Perez JC, Quesada H, Westherhoff H.V. Mitochondrion. 2010, 10:626-639.

9. Johnson RA, Wichner DW. Applied multivariate statistical analysis. 4th edition. Englewood Cliffs, NJ: Prentice Hall. 1998, pp.458–512.

10. Wall ME, Rechtsteiner A, Rocha LM. Singular value decomposition and principal component analysis: IN A practical approach to microarray data analysis. D.P. Berrar, W. Dubitzky, M. Granzow, eds.pp. 91-109, Kluwer: Norwell, 2003. Available in http://public.lanl.gov/mewall/kluwer2002.html.

11. D'haeselee P. How does gene expression clustering work? Nat Biot. 2005, 23:1499–1501.

12. Johnson RA, Wichner DW. Applied multivariate statistical analysis. 4th edition. Englewood Cliffs, NJ: Prentice Hall. 1998b, pp. 726–798.

13. Guebel DV, Cánovas M, Torres NV. Analysis of the *Escherichia coli* response to glycerol pulse in a continuous, high cellular density culture by a multivariate approach. Biotechnology Bioengineering. 2008, 102: 910-922.

14. Guebel DV, Cánovas M, Torres NV. Model identification in presence of incomplete information by Generalized Principal Component Analysis: Application to the determination of the common and specific responses of *Escherichia coli* to multiple types of pulses in a continuous, high-cellular density culture**.** Biotechnology Bioengineering. 2009, 104: 785-95**.**

15. de Jong H. Modeling and simulation of genetic regulatory systems: A literature review. J Comput Biol. 2002, 9: 69-105.

16. Smith M.K. Type I and II errors and significance levels. Web site devoted to Common mistakes in using statistics: spotting and avoiding them (2011). Available in: http://www.ma.utexas.edu/users/mks/statmistakes/errortypes.html

17. Freiman JA, Chalmers TC, Smith H, Kuebler RR. The Importance of Beta, the Type II Error and Sample Size in the Design and Interpretation of the Randomized Control Trial — Survey of 71 Negative Trials**.** N Engl J Med. 1978, 299:690-694.

18. Ma S, Huang J. Regularized ROC method for disease classification and biomarker selection with microarray data. Bioinformatics. 2005, 21: 4356-4362.

19. Smoot M, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics. 2011, 27: 431–432.