Detection of markers for discrete phenotypes

Hannes Klarner*
Freie Universität Berlin
Berlin, Germany
hannes.klarner@fu-berlin.de

Elisa Tonello
Freie Universität Berlin
Berlin, Germany
elisa.tonello@fu-berlin.de

Laura Cifuentes
Freie Universität Berlin
Berlin, Germany
laura.cifuentes@fu-berlin.de

Florence Janody
Instituto de Patologia e Imunologia
Molecular
Porto, Portugal
fjanody@ipatimup.pt

Claudine Chaouiya
Institute of Mathematics of Marseille
Marseille, France
claudine.chaouiya@univ-amu.fr

Heike Siebert
Freie Universität Berlin
Berlin, Germany
heike.siebert@fu-berlin.de

ABSTRACT

Motivation: Capturing the molecular diversity of living cells is not straightforward. One approach is to measure molecular markers that serve as indicators of specific biological conditions or phenotypes. This is particularly relevant in modern medicine to provide precise diagnostics and pinpoint the best treatment for each patient. The challenge is to select a minimal set of markers whose activity patterns are in correspondence with the phenotypes of interest. **Results:** This article approaches the marker detection problem in the context of discrete phenotypes which arise, for example, from Boolean models of cellular networks. Mathematically this poses a combinatorial optimization problem with many answers. We propose a solution to this optimization problem that is based on the modelling language answer set programming (ASP). A case study of a death cell receptor network illustrates the methodology. **Discussion and code:** For code, discussions and reporting errors visit https://github.com/hklarner/detection_of_markers_for_discrete_phenotypes.

CCS CONCEPTS

• **Applied computing → Biological networks**; **Systems biology**; **Bioinformatics**.

KEYWORDS

Marker Detection, Discrete Phenotypes, Biological Networks

---

*Corresponding Author

---

1 INTRODUCTION

1.1 Biological background

In 2001, the National Institutes of Health's Biomarkers Definitions Working Group defined a biomarker as "a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention", see [8]. Among biomarkers, molecular biomarkers can disclose changes in gene copy number, gene mutations, alterations of the levels of gene or protein expression or variations of the functional activity of a component. They have a wide array of uses in a variety of fields, including medicine, developmental biology, and basic scientific research. In particular, the use of biomarkers has increased greatly in precision medicine to diagnose pathogenic processes, to assess patient prognosis, to select the best treatment for each individual, as well as to evaluate risk of developing diseases. However, the clinical translation of novel molecular biomarkers predicting disease susceptibility, progression or treatment outcome remains poor, in particular for complex diseases, such as Alzheimer and related forms of dementia, cancer, diabetes and heart disease, with less than two approvals per year across all diseases among thousands of potential identified biomarkers. This lack of success can arise from experimental limitations of the biomarker discovery pipeline. In addition, because these diseases consist of various pathophysiological processes, resulting from multiple molecular and environmental factors, and exhibit variable disease course and response to therapy, a single biomarker rarely fulfills all necessary criteria for a comprehensive diagnostic or prognostic assessment, see [1]. There is therefore an unmet need for multi-marker discovery using biomarkers reflecting different pathophysiologies that could be used in routine clinical practice, see [3]. Analysis of data in terms of complex networks may help to identify these key biomarkers. This path requires a multidisciplinary approach, engaging researchers in numerous fields: biologists, informatics specialists, and mathematicians.

1.2 Mathematical background

In its essence the marker detection approach proposed in this article is one of combinatorial optimization. The input to this problem consists of a list of binary vectors whose coordinates designate the available molecular components, and whose values indicate the activity or abundance of the respective component. Each vector is an abstract representation of a state of the biological system. Usually

Hannes Klarner, Elisa Tonello, Laura Cifuentes, Florence Janody, Claudine Chaouiya, and Heike Siebert

these states will belong to the long-term behaviors of the system, the so-called attractors, but technically this is not a requirement.

We are aware of two distinct settings to which our work is applicable. First, such binarized states are the output of discretization algorithms that are frequently a preliminary step to the analysis of continous gene expression data. For an example in the context of time series data, see [5].

Second, they are the principal output of discrete models of molecular interaction networks. A prominent example of discrete models are Boolean interaction networks. These consist of Boolean variables that are governed by fixed update functions and some assumption about the number of variables that are permitted to change state simultaneously during a state transition. While reachability questions, the enumeration of all attractors and descriptions of the basins of attraction are currently limited to medium-sized models of about 50 components, see [10], the detection of steady states is feasible for large networks of hundreds of components and more, see for example [2]. In the context of this paper we assume that the steady states capture the phenotypes under investigation. For a survey on Boolean networks in systems biology, see for example [12, 14].

The overall challenge we are facing is therefore to nominate multi-element markers that are 1) minimal and therefore low cost, and 2) respect the practical limitations of available marker candidates. The relationship between the solutions of a marker detection problem and its underlying Boolean network model, if there is one, is also of interest.

1.3 Summary

The article is organized in five sections. Sec. 2 introduces the notation for Boolean networks and basic set theory required in this text. It is followed by Sec. 3 which defines phenotypes and marker types for steady states and introduces the notions of consistency and marker component equivalence. The methods section, Sec. 4, gives details on our ASP encoding of the marker detection problem. A case study which illustrates the full methodology for a death cell receptor model published in [4] is presented in Sec. 5. The discussion and outlook are given in Sec. 6.

2 DEFINITIONS AND NOTATION

A *Boolean network* of n components is an n-dimensional Boolean function $f = (f_1, \ldots, f_n)$ with $f_i : \mathbb{B}^n \to \mathbb{B}$ where $\mathbb{B} = \{0, 1\}$ are the Boolean values. A *state* is a n-dimensional Boolean vector $x = (x_1, \ldots, x_n)$. The synchronous update of a state x is the image of x under f, i.e., $f(x) = (f_1(x), \ldots, f_n(x))$. A *steady state* is a fixed point of f, that is, a state that satisfies $f(x) = x$. The *interaction graph* is a directed graph (V, A) where an arc $(i, j) \in A$ captures the dependence of f_j on the variable x_i.

Example 2.1 (toy problem). The toy network consists of six components with two positive feedback loops $f_1 = x_2, f_2 = x_1, f_3 = x_4, f_4 = x_3$ and two read-out nodes whose update functions are $f_5 = x_2 \wedge x_4, f_6 = x_2 \vee (\neg x_2 \wedge \neg x_3)$. It has four steady states $S = \{000001, 110001, 001100, 111111\}$.

The *power set* of a set A is denoted by $\mathcal{P}(A) = \{B \mid B \subseteq A\}$. Given two families of subsets $\mathbf{X}, \mathbf{Y} \in \mathcal{P}(\mathcal{P}(X))$, we denote by $\otimes$ the set-product $\mathbf{X} \otimes \mathbf{Y} := \{X \cup Y \mid X \in \mathbf{X}, Y \in \mathbf{Y}\}$. If $\mathbf{Z} = \mathbf{X} \otimes \mathbf{Y}$ we say

that $\mathbf{Y}, \mathbf{X}$ are *factors* of $\mathbf{Z}$. The set-product is often useful for giving concise descriptions of set families with many elements.

An equivalence relation $\sim$ on a set A induces a *partition* of A into k *blocks* $A_1, \ldots, A_k$ with $A_i \subseteq A$ and $A_i \cap A_j = \emptyset$ for all i, j and $A_1 \cup \cdots \cup A_k = A$. We call blocks of size one *trivial blocks*.

The number of elements of a set A is denoted by $|A|$.

3 MARKERS AND PHENOTYPES

The term phenotype refers to the classification of individual organisms by observable traits. The traits are determined, essentially, by the organism's genotype and the influence of environmental factors. More specifically, when studying mathematical models of biological systems, a phenotype usually refers to the expression pattern of specific components of interest under the assumption that the system has reached an attractor.

In this text we focus on the steady states of a network and define the phenotypes by partitioning the steady states according to given phenotype components. Let S be the steady states of a Boolean network and $P \subseteq V$ a set of *phenotype components* with $|P| = k$. A *phenotype* $p \in \mathbb{B}^k$ is the projection of a state $x \in S$ onto the phenotype components. We denote by $\pi_P : \mathbb{B}^n \to \mathbb{B}^k$ the *projection* of x onto P. For each phenotype p we define the *phenotype states* $\chi_P(S)$ by $\chi_p(S) := \{x \in S \mid \pi_P(x) = p\}$.

Analogously, we define *marker components*, or *markers* for short, as a subset $M \subseteq V$ and the *marker-type* as a state of M. Note that, mathematically, markers and phenotypes are therefore objects of the same type.

Example 3.1 (toy problem). Suppose the phenotype components are $P = \{4, 5\}$ and the marker components are $M = \{1, 2\}$, then there are three phenotypes $\pi_P(S) = \{00, 10, 11\}$ and two marker-types $\pi_M(S) = \{00, 11\}$.

3.1 Consistency between markers and phenotypes

In practice, we are given phenotype components and are asked to find marker components such that the phenotypes and the marker-types are consistent across all the steady states. That is, given only the marker-type of a steady state the phenotype must be uniquely deducible.

Definition 3.2 (consistency). The markers M are *consistent with* the phenotype components P on S iff $\forall x, x' \in S : \pi_M(x) = \pi_M(x') \implies \pi_P(x) = \pi_P(x')$.

Note that this definition allows a 1-to-many relationship between the phenotypes and the marker-types on S. In practice this may introduce an unnecessary complexity in the experimental setup. If we are interested in 1-to-1 relationships we need to demand in addition that P is consistent with M in which case we say that P and M are *equivalent*.

Usually, to reduce the cost of adding markers or of reading marker values, we are furthermore interested in subset minimal markers M and the *exclusion constraint* which limits the marker candidate space to $M \subseteq U$ with $U = V \setminus P$.

Example 3.3 (toy problem). Suppose we are interested in subset minimal markers and the exclusion constraint. The markers

$M = \{1, 2\}$ are not consistent with $P = \{4, 5\}$ because the steady states $x = 000001$, $y = 001100$ satisfy $\pi_P(x) = 00 \neq \pi_P(y) = 10$ but $\pi_M(x) = 00 = \pi_M(y)$. The markers $M' = \{1, 3\}$ are consistent with P but M' and P are not equivalent because $|\pi_P(S)| = 3$ and $|\pi_{M'}(S)| = 4$. Finally, the markers $M'' = \{3, 6\}$ are equivalent to P. While M'' is the unique marker set that is equivalent to P, there is a family of three consistent marker sets, namely $\{\{1, 3\}, \{2, 3\}, \{3, 6\}\}$.

3.2 Marker component equivalence

Marker detection problems often have hundreds of solutions. We have noticed that this is partly due to an equivalence relation on V that partitions V into blocks of interchangeable components. In a sense, components within the same block carry the same information with regards to steady state identification. Intuitively, two components are equivalent if their activities are locked in a fixed relationship across all steady states under consideration.

Definition 3.4 (Steady state correlation). Two components $i, j \in V$ are *correlated* across the steady states S if either $\forall s, s' \in S : s_i = s_j$ or $\forall s, s' \in S : s_i \neq s_j$. We write $i \sim_S j$.

The key observation is that correlated components are interchangeable in marker sets. That is, given $i, j \in V$ with $i \sim_S j$ and suppose that $i \in M$ for a marker set $M \subseteq V$. Then $M' := (M \setminus \{i\}) \cup \{j\}$ is another valid marker set and of the same consistency type.

Note that the steady state correlation is independent of the specification of phenotype components and therefore holds for all choices of P. In fact, the relation $\sim_S$ is an equivalence relation on V and correlated components hence form blocks of a partition of V.

Example 3.5 (toy problem). The blocks of the steady state correlation are $\mathbf{B} = \{\{1, 2\}, \{3, 4\}, \{5\}, \{6\}\}$, two are trivial and two are non-trivial. Since $M = \{2, 3\}$ is a marker set and $\{1, 2\} \in \mathbf{B}$, it follows by the observation above that $M' = \{1, 3\}$ is another valid marker set.

4 MATERIALS AND METHODS

4.1 Introduction to ASP

Answer set programming (ASP) is a particular form of logic programming. Mathematical problems are modelled in this language by stating rules over user-defined predicates. The answer sets of a program correspond to the solutions of the mathematical problem. ASP solvers like *clasp* [7] are capable of enumerating all subset minimal answer sets.

ASP has been applied to problems in bioinformatics before, see for example the computation of minimal interventions in logical signaling networks [9]. The ASP modelling language is well-suited for combinatorial problems and the *clasp* solver outperforms a brute force detection algorithm significantly. Encoding a mathematical problem in ASP and debugging can be tricky but small changes in the mathematical problem often correspond to small changes in the ASP program. In contrast, imperative programs often require large changes to adapt to variations of the mathematical problem.

The following quick introduction to ASP is given in terms of the input language *gringo* [6]. An *atom* is a function symbol followed by terms which are either constants or variables. Examples of atoms

are `f(c,10)` and `g(X)` where `c` and `10` are constants and uppercase letters, like `X`, denote first-order variables. A *rule* is a language construct that denotes an implication and consists of a *head*, the implicant, and a *body*, a conjunction of atoms followed by a full stop. The implication symbol is `:-` and means "if". An example is `f(c) :- g(1), g(10)` which makes `f(c)` true if both `g(1)` and `g(10)` are true. A rule without a body is called a *fact* and its head is unconditionally considered true. Facts are a means of including data into the program. An *answer set* consists of all atoms that are derivable by rules of the program.

For our encoding of the marker detection problem we need two more *gringo* language constructs. An *integrity constraint* is a rule without a head. It serves as a solution filter that removes answer sets that satisfy the body. An example is `:- f(a), g(10)`. A *choice rule* is a construct for stating that any subset of set of atoms may be considered factual, i.e. as fact. It is often used to define the candidate space for combinatorial problems. An example is `{f(1), f(5), f(12)}`. It states that any of the 8 subsets of the 3 atoms may be considered factual. Instead of enumeration, conditional statements using `:` are allowed and sometimes necessary. An example of a conditional choice rule is `{f(I): g(I), not h(I)}`. Here, `not h(I)` is the *classical negation* which evaluates to true for all values of `I` for which `h(I)` can not be derived. The count of true atoms in a choice rule can be limited above and below by adding the integer values: `3 {f(I): g(I), not h(I)} 10`.

4.2 Encoding of the marker detection problem

The first step is to encode the steady states. We use the predicate `x(I,J,K)` to encode the steady state data. Here, $0 \leq i < |S|$ is the identifier of a steady state, $j \in V$ is the index of a component and $k \in \mathbb{B}$ is the value of j in the ith steady state.

The following line numbers refer to Table 1. The predicate `forbidden(J)` states the indices $j \in V$ that are excluded from the marker detection problem. The phenotype components are declared by the predicate `p(J)` where $j \in V$. The marker component candidates are encoded by the predicate `m(J)` and defined via the choice rule in Line 14. Here, the underscores in the predicate `x` are place holders for which *clasp* allows any value.

Next we introduce the predicate `different_phenotype(S1,S2)` for steady states $s, s' \in S$ to be true if their phenotypes are different. Since the steady states are Boolean vectors we may simply enumerate the two cases under which s, s' are of different phenotype, namely if there is a phenotype component for which the expression in the steady states differs.

Note that we could instead introduce `equal_phenotype(S1,S2)`, but it is our observation that the "exists counterexample" formulation performs better than the "all values are equal" formulation. With the intention of increasing the computational efficiency we introduce a rule that states that the predicate is symmetric: `different_phenotype(S1,S2):-different_phenotype(S2,S1)`. The predicate `different_marker_type(S1,S2)` is defined analogously.

We use one integrity constraint to enforce the consistency of markers with phenotypes, see Line 27. An optional second integrity constraint may be added to achieve equivalence, see Line 30.

The full program for the toy problem is listed in Table 1.

Table 1: ASP encoding of marker detection for toy problem

```
1    % steady states
2    x(0,0,0). x(0,1,0). x(0,2,0). x(0,3,0). x(0,4,0). x(0,5,1).
3    x(1,0,1). x(1,1,1). x(1,2,0). x(1,3,0). x(1,4,0). x(1,5,1).
4    x(2,0,0). x(2,1,0). x(2,2,1). x(2,3,1). x(2,4,0). x(2,5,0).
5    x(3,0,1). x(3,1,1). x(3,2,1). x(3,3,1). x(3,4,1). x(3,5,1).
6
7    % phenotype components
8    p(4). p(5).
9
10   % exclusion contraint
11   forbidden(4). forbidden(5).
12
13   % marker components
14   1 m(C) : x(_,C,_), not forbidden(C).
15
16   % marker-types
17   different_marker_type(S1,S2) :- x(S1,C,1), x(S2,C,0), m(C).
18   different_marker_type(S1,S2) :- x(S1,C,0), x(S2,C,1), m(C).
19   different_marker_type(S1,S2) :- different_marker_type(S2,S1).
20
21   % phenotypes
22   different_phenotype(S1,S2) :- x(S1,C,0), x(S2,C,1), p(C).
23   different_phenotype(S1,S2) :- x(S1,C,1), x(S2,C,0), p(C).
24   different_phenotype(S1,S2) :- different_phenotype(S2,S1).
25
26   % consistency
27   :- different_phenotype(S1,S2), not different_marker_type(S1,S2).
28
29   % equivalence
30   :- different_marker_type(S1,S2), not different_phenotype(S1,S2).
31
32   % output
33   #show m/1.
```

5 RESULTS

5.1 Markers for a Boolean death receptor model

In this case study we compute markers for the death receptor model published in [4]. The network is a model of the cross talk between the *NFkB* pro-survival pathway, the *RIP1*-dependent necrosis pathway and the apoptosis pathway. The pathways are stimulated via the cytokines *TNF* and *FASL*, which are input components of the network, and the response is modelled via the phenotype components $P = \{Survival, NonACD, Apoptosis\}$.

The network consists of 28 components and has 27 steady states. The steady state detection is completed instantly on a personal computer using the python package *pyboolnet*, see [11]. There are four phenotypes with between six and eight steady states in the respective class. The steady state correlation detects seven non-trivial blocks. The blocks form tree or cascade like structures when mapped onto the interaction graph of the model, see Figure 1. Five blocks are representable by one of the following root components $\{apoptosome, CASP8, RIP1, MOMP, RIP1ub\}$, the other two blocks involve feedback loops.

We computed the subset minimal marker sets for four scenarios that are the result of all combinations of two binary parameters. The first parameter determines the consistency type between marker-type and phenotype. Its values are either *1-to-1* or *1-to-many*. The second determines whether the exclusion constraint for P is enabled. Table 2 lists the marker families by number of components required and parameter values. The marker detection is solved instantly on a personal computer.

Scenario 1: Subset minimality and equivalence required, exclusion constraint enabled. There are 126 markers, all of size 3. In
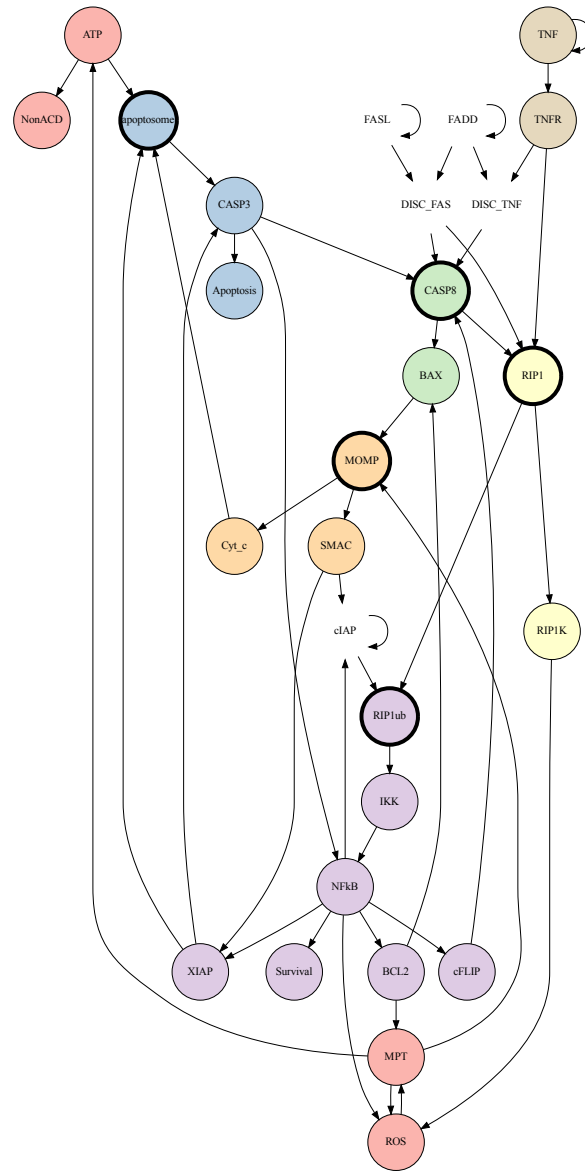

Figure 1: The blocks of the steady state correlation analysis of the death receptor model. Nodes without color belong to trivial blocks. Note that most blocks are tree-like. These have roots which can be seen as representatives of their blocks. They are highlighted with thick borders.

set-product notation they are

$$\{\{BCL2\}, \{IKK\}, \{NFkB\}, \{RIP1ub\}, \{XIAP\}, \{cFLIP\}\} \otimes \mathbf{L}_1$$

where $\mathbf{L}_1$ consists of 21 component pairs, including, for example $\{CASP3, MOMP\}$, $\{MPT, SMAC\}$, $\{MOMP, ROS\}$. An example of a valid marker set in this scenario is therefore $\{BCL2, CASP3, MOMP\}$.

Table 2: The number of markers of the death receptor model for four different parameter combinations.

# marker components	# marker sets	consistency type	exclusion constraint
3	126	1-to-1	active
3	231	1-to-1	inactive
3	216	1-to-many	active
4	54		
5	54		
3	369	1-to-many	inactive
4	82		
5	82		

Note that the first factor corresponds to the purple block in the steady state correlation analysis of Figure 1.

Scenario 2: Subset minimality and consistency required, exclusion constraint enabled. In this scenario we require only consistency between marker-types and phenotypes, not equivalence. There are 324 markers of sizes 3, 4, 5, and we denote the respective subset of equal sized markers by $\mathbf{X}^3, \mathbf{X}^4, \mathbf{X}^5$. In set-product notation they are

$$\mathbf{X}^3 = \{\{BCL2\}, \{IKK\}, \{NFkB\}, \{RIP1\}, \{RIP1K\}, \{RIP1ub\},$$
$$\{XIAP\}, \{cFLIP\}\} \otimes \mathbf{L}_2$$
$$\mathbf{X}^4 = \{\{DISC\_FAS\}\} \otimes \{\{TNF\}, \{TNFR\}\} \otimes \mathbf{L}_2$$
$$\mathbf{X}^5 = \{\{FADD\}\} \otimes \{\{FASL\}\} \otimes \{\{TNF\}, \{TNFR\}\} \otimes \mathbf{L}_2$$

where $\mathbf{L}_2$ consists of 27 pairs of components, including, for example $\{ATP, SMAC\}$, $\{CASP8, ROS\}$, $\{CASP3, Cyt\_c\}$. Note that the first factor of $\mathbf{X}^3$ is the union of the purple and yellow blocks in Figure 1. Similarly, the factors of $\mathbf{X}^4, \mathbf{X}^5$ are unions of the trivial blocks $FADD$ and $FASL$ and the brown block.

Examples of valid marker sets $\mathbf{Y}^3, \mathbf{Y}^4, \mathbf{Y}^5$ of sizes 3, 4 and 5 are therefore

$$\mathbf{Y}^3 = \{RIP1, ATP, SMAC\}$$
$$\mathbf{Y}^4 = \{DISC\_FAS, TNF, CASP8, ROS\}$$
$$\mathbf{Y}^5 = \{FADD, FASL, TNF, CASP3, Cyt\_c\}$$

6 DISCUSSION

The results in this paper give theoretical foundations to a central problem in the design of multi-marker experiments. We have introduced a notion of consistency between markers and phenotypes and demonstrated that the enumeration of all subset minimal consistent marker sets is instantaneous for medium-sized Boolean networks. Furthermore, the steady state correlation analysis leads to an interesting tool, the map of interchangeable marker components of Fig. 1. With it one may transform any valid marker set into other valid sets and detect, broadly speaking, clusters in the interaction graph that separate steady state information. It seems possible to

extend the underlying relation $\sim_S$ from V to $\mathcal{P}(V)$. This might be fruitful when further exploring the structure of marker families.

The next steps are to validate the theoretical results of this work experimentally in a follow-up study on a recently published model of epithelial-to-mesenchymal transition (EMT), see [13]. Here, selected markers will be used to experimentally distinguish between cells that acquire distinct phenotypes on the EMT continuum during cancer progression.

Also, the factorizations in Sec. 5 are of a heuristic nature, and we do not yet have rigorous mathematical statements with regards to minimal factorizations. The product description seems promising, as it reveals 1) the number of independent choices and 2) the number of options for each choice. Research in this direction may lead to some notion of marker building blocks.

REFERENCES

[1] Asier Antoranz, Theodore Sakellaropoulos, Julio Saez-Rodriguez, and Leonidas G Alexopoulos. 2017. Mechanism-based biomarker discovery. *Drug discovery today* 22, 8 (2017), 1209–1215.

[2] Julio Aracena, Luis Cabrera-Crot, and Lilian Salinas. 2021. Finding the fixed points of a Boolean network from a positive feedback vertex set. *Bioinformatics* 37, 8 (2021), 1148–1155.

[3] John R Ball, Christine M Micheel, et al. 2010. Evaluation of biomarkers and surrogate endpoints in chronic disease. (2010).

[4] Laurence Calzone, Laurent Tournier, Simon Fourquet, Denis Thieffry, Boris Zhivotovsky, Emmanuel Barillot, and Andrei Zinovyev. 2010. Mathematical modelling of cell-fate decision in response to death receptor engagement. *PLoS Comput Biol* 6, 3 (2010), e1000702.

[5] Elena S Dimitrova, M Paola Vera Licona, John McGee, and Reinhard Laubenbacher. 2010. Discretization of time series data. *Journal of Computational Biology* 17, 6 (2010), 853–868.

[6] Martin Gebser, Roland Kaminski, Benjamin Kaufmann, and Torsten Schaub. 2012. Answer set solving in practice. *Synthesis lectures on artificial intelligence and machine learning* 6, 3 (2012), 1–238.

[7] Martin Gebser, Benjamin Kaufmann, André Neumann, and Torsten Schaub. 2007. clasp: A conflict-driven answer set solver. In *International Conference on Logic Programming and Nonmonotonic Reasoning*. Springer, 260–265.

[8] Biomarkers Definitions Working Group, Arthur J Atkinson Jr, Wayne A Colburn, Victor G DeGruttola, David L DeMets, Gregory J Downing, Daniel F Hoth, John A Oates, Carl C Peck, Robert T Schooley, et al. 2001. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical pharmacology & therapeutics* 69, 3 (2001), 89–95.

[9] ROLAND KAMINSKI, TORSTEN SCHAUB, ANNE SIEGEL, and SANTIAGO VIDELA. 2013. Minimal intervention strategies in logical signaling networks with ASP. *Theory and Practice of Logic Programming* 13, 4-5 (2013), 675–690. https://doi.org/10.1017/S1471068413000422

[10] Hannes Klarner, Frederike Heinitz, Sarah Nee, and Heike Siebert. 2018. Basins of Attraction, Commitment Sets, and Phenotypes of Boolean Networks. *IEEE/ACM transactions on computational biology and bioinformatics* 17, 4 (2018), 1115–1124.

[11] Hannes Klarner, Adam Streck, and Heike Siebert. 2017. PyBoolNet: a python package for the generation, analysis and visualization of boolean networks. *Bioinformatics* 33, 5 (2017), 770–772.

[12] Julian D Schwab, Silke D Kühlwein, Nensi Ikonomi, Michael Kühl, and Hans A Kestler. 2020. Concepts in Boolean network modeling: What do they all mean? *Computational and structural biotechnology journal* 18 (2020), 571–582.

[13] Gianluca Selvaggio, Sara Canato, Archana Pawar, Pedro T Monteiro, Patrícia S Guerreiro, M Manuela Brás, Florence Janody, and Claudine Chaouiya. 2020. Hybrid epithelial–mesenchymal phenotypes are controlled by microenvironmental factors. *Cancer research* 80, 11 (2020), 2407–2420.

[14] Rui-Sheng Wang, Assieh Saadatpour, and Reka Albert. 2012. Boolean modeling in systems biology: an overview of methodology and applications. *Physical biology* 9, 5 (2012), 055001.