

Using State Space Exploration to Determine How Gene Regulatory Networks Constrain Mutation Order in Cancer Evolution

Matthew A. Clarke, Steven Woodhouse*, Nir Piterman, Benjamin A. Hall and Jasmin Fisher

Abstract Cancer develops via the progressive accumulation of somatic mutations, which subvert the normal operation of the gene regulatory network of the cell. However, little is known about the order in which mutations are acquired in successful clones. A particular sequence of mutations may confer an early selective advantage to a clone by increasing survival or proliferation, or lead to negative selection by triggering cell death. The space of allowed sequences of mutations is therefore constrained by the gene regulatory network. Here, we introduce a methodology for the systematic exploration of the effect of every possible sequence of oncogenic mutations in a cancer cell modelled as a qualitative network. Our method uses attractor identification using binary decision diagrams and can be applied to both synchronous and asynchronous systems. We demonstrate our method using a recently developed model of ER-negative breast cancer. We show that there are differing levels of constraint in the order of mutations for different combinations of oncogenes, and that the effects of ErbB2/HER2 over-expression depends on the preceding mutations.

Matthew A. Clarke

Department of Biochemistry, University of Cambridge, UK e-mail: mac86@cam.ac.uk

Steven Woodhouse

Department of Genetics, University of Pennsylvania, USA e-mail: steven.woodhouse@pennmedicine.upenn.edu

Nir Piterman

Department of Computer Science, University of Leicester, UK e-mail: nir.piterman@leicester.ac.uk

Benjamin A. Hall

MRC Cancer Unit, University of Cambridge, UK e-mail: bh418@mrc-cu.cam.ac.uk

Jasmin Fisher

Department of Biochemistry, University of Cambridge, UK e-mail: jf416@cam.ac.uk and Microsoft Research Cambridge, UK e-mail: jasmin.fisher@microsoft.com

* Work done while author was a Post Doc Researcher at Microsoft Research Cambridge, UK.

1 Introduction

Multicellular organisms, in particular long-lived ones such as mammals, must assure the survival and procreation of the organism as a whole over the individual survival of a single cell. As such, they have evolved robust mechanisms to control proliferation and to detect and remove damaged or malfunctioning cells [12]. Consequently, no single mutation is believed to suffice to elicit tumour formation, rather cells require several key changes in order to escape these controls, out-compete healthy cells, and adapt to the new environment that emerges as a tumour forms [34]. These controls, and the different characteristics that are favoured at different stages of tumour evolution, will restrict which mutations are selected for and observed at different stages. This results in an optimal order in which mutations must be acquired for some cancers to be successful, as observed in colorectal cancer [2] or pancreatic ductal adenocarcinoma [52].

This evolution is not linear; while fitter sub-clones out-compete their neighbours they do not necessarily dominate and many tumours become highly heterogeneous, with competition [32] and cooperation [64] between clones. This poses a problem for treatment, with the degree of heterogeneity being a predictor of poor prognosis [44] and serving as a pool of possible adaptations to therapy. Heterogeneity therefore leads to the rapid emergence of acquired resistance [8], particularly in the case of targeted therapy [15].

Despite the heterogeneity of mature tumours, they largely originate from a single cell, even in cases of strong environmental factors such as smoking, or where there is a germline mutation that causes a predisposition [23, 61]. Testimony to the efficiency of the body's tumour surveillance mechanisms, this suggests that even with heterogeneity there will be shared mutations, with early mutations likely to be the most prevalent among tumour sub-clones [66]. As these are also likely to be driver mutations, treatments targeting these may be more broadly effective against the entire tumour [63]. Furthermore, existing treatment may preferentially affect late, sub-clonal mutations. For example, Sun *et al.* [63] observe that in acute myeloid leukaemia, cells carrying early mutations persist and may play a part in relapse, with similar results from Ding *et al.* [24]. This suggests that targeting early, founder mutations may help avoid this pool of early mutations and reduce rates of relapse. Finally, which mutations occur early may constrain the future behaviour of a tumour, as seen in studies of myeloproliferative neoplasms [49] and pancreatic cancer [59]. In order to take advantage of this commonality underlying tumour heterogeneity, we must be able to understand the constraints on the acquisition of tumour-promoting mutations as the tumour progresses. We can then predict which mutations are likely to occur early in tumour development, and so are best for treatment or early detection.

Existing studies have tried to shed light on the temporal evolution of tumours by comparing metastases to primary tumours [41, 55], analysing sequencing data from primary tumours for genetic recombinations [33, 47, 67], comparing regions within biopsies [69] and single-cell sequencing [29]. However, these studies either rely on metastases occurring late in tumour development, which may not always be the case [31], have very few time-points within tumour evolution, or must rely

on reconstruction of a history that has been distorted by selective sweeps over the history of the tumour. This, coupled with difficulty in obtaining biopsy data for early, asymptomatic, stages of tumour evolution, makes identification of the early behaviour of the tumour more difficult [62]. In this study, we present a method to study how positive and negative selection influences the evolution of tumours, and so determines the order of mutations, based on the gene regulatory network. This builds upon a biological understanding of the gene regulatory network of the cell to model the defining constraints on early tumour development and so better map out this stage using *in silico* models.

We model the gene regulatory network of the cell as a discrete qualitative network [54]; an extension of Boolean networks in which each node can take a finite number of values, corresponding to, for example, under-, normal and over-expression of a gene. This is necessary to capture the complexity of the regulatory network. For example, the c-myc oncogene elicits distinct responses at each of these levels [46]. This and similar phenomena cannot be captured in a Boolean framework. We can then model the evolution of cancer in the cell by changing the behaviour of nodes affected by mutation. Using a discrete system rather than a continuous mathematical model allows us to explore the entire state space of the system, and give definitive answers as to how the cell responds to each mutation acquired as the tumour evolves. By assigning cellular phenotypes to nodes that integrate the activity of the rest of the network, we are able to tie phenotypes to qualitative network attractors. We then identify all such attractors under all possible combinations of mutations.

We use the open source BioModelAnalyzer tool (BMA²), which we extend with support for attractor identification. Previously, BMA supported stability analysis [18] and Linear Temporal Logic (LTL) model checking [17] for synchronous qualitative networks. With our modifications it now supports full attractor analysis for both synchronous and asynchronous qualitative networks. This expands BMA to non-deterministic and non-stabilising systems such as development and stem cell fate specification.

As a proof-of-concept, we apply this to a recently developed model of the gene regulatory network of ER-negative breast cancer cells (currently in submission). Breast cancer is the second most common cancer worldwide at 11.9% of cancer diagnoses in 2012, and the cause of over 500,000 deaths per year around the world [27]. There has been a decline in breast cancer mortality in the UK of around 30% between the late 1980s and 2006 [4], likely as a result of a combination of the introduction of screening [39] and tamoxifen therapy [9]. However, of the major breast cancer sub-types, both basal and triple negative (TNBC) breast cancers are frequently estrogen-negative (ER-) and so not treatable by tamoxifen. These sub-types have higher mortality [60] and have fewer treatment options [38]. In order to better understand the development of such tumours, and so improve treatment, a better understanding of their evolution is required. We show that our new method can be used to explore the evolution of breast cancer based on an understanding of

² <http://biomodelanalyzer.org/>

the underlying gene regulatory network, and suggest how this can be built upon and applied to other cancers in the future.

2 Qualitative Networks

In this section we formally define qualitative networks, which are a framework for modelling gene regulatory networks. In a qualitative network, we have a variable for each gene or protein we wish to model. We also have variables that correspond to overall cellular behaviour, such as proliferation and apoptosis, which integrate the signals from the rest of the network. These give a read out of the effect on the observable behaviours of the cell, the phenotype, due to its genetic regulatory network being in a particular attractor. Each variable ranges over a finite range, such as 0–2, and has a user-defined target function that defines how its value is influenced by other genes. The system is updated over discrete time steps by following the logic of the target functions and updating the values of each variable. We may update variables using a synchronous or asynchronous update scheme.

A *qualitative network* (QN)[54], $Q(V, T, N)$, of granularity $N + 1$ consists of a set of *variables*: $V = \{v_1, v_2, \dots, v_n\}$, together with a set of *target functions*: $T = \{T_1, T_2, \dots, T_n\}$. A state of the system is a finite map $s : V \rightarrow \{0, 1, \dots, N\}$. For each variable $v_i \in V$ we have target function $T_i \in T$ associated with it: $T_i : \{0, 1, \dots, N\}^n \rightarrow \{0, 1, \dots, N\}$.³ Target functions direct the execution of the network: given state $s = (d_1, d_2, \dots, d_n)$, the *next value* of variable v_i is:

$$d'_i = \begin{cases} d_i + 1 & d_i < T_i(s) \text{ and } d_i < N, \\ d_i - 1 & d_i > T_i(s) \text{ and } d_i > 0, \\ d_i & \text{otherwise} \end{cases} \quad (1)$$

We say that a target function T_i is *enabled* at state s if $d'_i \neq d_i$, i.e. the next value of v_i is not equal to its current value.

In a *synchronous qualitative network* all variables are updated at each time step by firing all target functions using synchronous parallelism. That is, from state $s = (d_1, d_2, \dots, d_n)$, the *successor state* is $s' = (d'_1, d'_2, \dots, d'_n)$. A synchronous qualitative network has a deterministic dynamics – given an initial state, all future states are uniquely determined.

In an *asynchronous qualitative network*, we instead update variables using asynchronous parallelism. State $s' = (d_1, d_2, \dots, d'_i, \dots, d_n)$ is a *successor* of state $s = (d_1, d_2, \dots, d_i, \dots, d_n)$ if T_i is enabled. That is, we get to the next state s' , by non-deterministically selecting an enabled target function T_i and updating the value of the associated variable v_i . If no target function is enabled, for convenience we define the successor state to be $s' = s$. That is, the system remains in its current, stable, state, where it will remain.

³ For simplicity, here we assume that all variables have the same range $\{0, \dots, N\}$. Our implementation supports individual ranges for variables.

A concrete example of a qualitative network is the breast cancer model used in the case study of this paper. This qualitative network models the signal transduction network of a single mammary cell, where inputs from the environment are represented as special variables. Additionally, special variables like Proliferation represent the phenotypic read-out of the model as a function of intracellular variables. In this model, Beta-Catenin is represented as a variable which ranges from 0 (unexpressed) to 2 (highly expressed), and with a target function $2 - \text{GSK3_Axin_APC_CK1}$, which represents how its abundance in the cell is controlled by degradation by the GSK3-Axin-APC-CK1 complex. To model a mutation which leads to the constitutive activation of Beta-Catenin we would modify its target function to the constant 2. To model a knock-out we would modify the target function to 0. By applying such changes in target function to variables which represent genes observed to mutate in breast cancer, we can model the transition from a healthy cell to a cancerous cell.

A QN defines a transition system $T = (S, R)$, where S is the set of all $(N + 1)^n$ possible states of the network and $R \subseteq S \times S$ is the (synchronous or asynchronous) successor relation. Since a qualitative network has a finite number of possible states, any execution eventually converges to either a *fixpoint attractor* or a cycle of states, called a *loop attractor*. Formally, an attractor is a set of states $A \subseteq S$ such that for all $s \in A$, we have that $F(s, T) = A$, where $F(s, T)$ is the set of states reachable from s in T .

Attractors represent the long-term behaviour of the biological system. A qualitative network will have the same set of fixpoint attractors whether it is executed synchronously or asynchronously, but the two different execution types lead to different types of loop attractors [30]. The two modes of execution also change whether an attractor is reachable from a given initial state. The choice of whether to use the synchronous or asynchronous semantics when building a model depends on the characteristics of the biological system under consideration. For stochastic and non-stabilising systems such as stem cell differentiation, different potential interleavings of gene expression events may need to be explored, motivating the asynchronous semantics. For systems that are expected to exhibit stability the simpler synchronous semantics may be more appropriate, such as in the case of the development of breast cancer presented in this paper. We discuss identification of attractors of qualitative networks in the next section.

3 Identifying attractors in QNs using binary decision diagrams

Garg *et al.* [30] introduce an algorithm for identifying attractors in synchronous and asynchronous Boolean networks. From a high-level point of view, this algorithm works by choosing an arbitrary state of the network, and from this point simulating forward until the system falls into an attractor. This represents one possible long term-behaviour of the model. All states that reach this attractor are then eliminated by running all possible backwards executions from this point, doing so in an efficient

manner by using a data structure called a binary decision diagram. This process is repeated until all states are explored, and therefore all attractors are found. Here, we review this algorithm and extend it to qualitative networks for the first time.

3.1 Binary Decision Diagrams

A Reduced Ordered Binary Decision Diagram (ROBDD, or simply BDD) is a rooted directed acyclic graph with one or two terminal nodes of out-degree zero labelled 0 and 1, and with all other nodes having out-degree two and labelled with a variable x [11, 40]. A BDD is ordered — variables always occur in the same order along any path from root to terminal, and reduced — the left and right branches of a node cannot lead to the same node, and there are no two distinct nodes n and n' with isomorphic subgraphs.

A BDD is essentially a compressed representation of the truth table of a Boolean function, with each path to a root node representing an evaluation of the function. A left branch at a variable represents an assignment of 0 to that variable, and a right branch represents an assignment of 1. The value of the function evaluation is given by the terminal node that the path ends at. Given a fixed ordering for the variables of a function, there is one unique BDD representation for that function. Given two BDDs, we can combine them using any combination of Boolean operators (e.g., \vee , \wedge , or \neg). An example BDD for the majority function $(x_1 \wedge x_2) \vee (x_2 \wedge x_3) \vee (x_3 \wedge x_1)$ is shown in Figure 1.

BDDs allow for the efficient representation and manipulation of sets of objects and relations over sets of objects. In many practical cases a BDD representation allows the symbolic exploration of combinatorial spaces that would be intractable to represent and explore explicitly, although in the worst case a BDD still requires exponential space to represent all the solutions to a function.

3.2 Attractor identification algorithm

The algorithm of Garg *et al.* works by manipulating a BDD representing the successor relation of the network and BDDs representing sets of states. Starting from an arbitrary initial state s , the algorithm explores all states that are reachable from s , and all states that can reach s , by iteratively applying the successor relation forwards

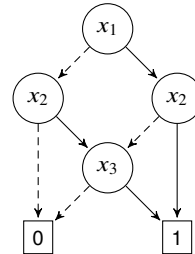


Fig. 1: In our approach to identifying attractors of qualitative networks, we represent Boolean functions using binary decision diagrams. Above, a binary decision diagram for the majority function over three variables $(x_1 \wedge x_2) \vee (x_2 \wedge x_3) \vee (x_3 \wedge x_1)$ is shown [40].

```

function ALLATTRACTORS(transitionBDD, statesToRemove)
  S ← ¬ statesToRemove
  REMOVEINVALIDBITCOMBINATIONS(S)
  attractors ← ∅
  while S ≠ false do
    s ← RANDOMSTATE(S)
    s ← RUNFORWARDS(transitionBDD, s)
    fr ← FORWARDREACHABLESTATES(transitionBDD, s)
    br ← BACKWARDREACHABLESTATES(transitionBDD, s)
    if fr ∧ ¬ br = false then
      attractors ← attractors ∪ {fr}
    end if
    S ← S ∧ ¬ (s ∨ br)
  end while
  return attractors
end function

function SYNCATTRACTORS(transitionBDD_sync)
  fix ← FIXPOINTS(transitionBDD_sync)
  br ← fix ∨ BACKWARDREACHABLESTATES(transitionBDD_sync, fix)
  loops ← ALLATTRACTORS(transitionBDD_sync, br)
  return fix, loops
end function

function ASYNCATTRACTORS(transitionBDD_sync, transitionBDD_async)
  fix ← FIXPOINTS(transitionBDD_sync)
  br ← fix ∨ BACKWARDREACHABLESTATES(transitionBDD_sync, fix)
  syncLoops ← ALLATTRACTORS(transitionBDD_sync, br)
  sharedAttractors ← fix
  asyncLoops ← ∅
  for loop ∈ syncLoops do
    if ISASYNCLOOP(loop) then
      sharedAttractors ← sharedAttractors ∨ loop
      asyncLoops ← asyncLoops ∪ {loop}
    end if
  end for
  br ← sharedAttractors
  br ← br ∨ BACKWARDREACHABLESTATES(transitionBDD_async, br)
  asyncLoops ← asyncLoops ∪ {ALLATTRACTORS(transitionBDD_async, br)}
  return fix, asyncLoops
end function

```

Fig. 2: Pseudocode for the attractor finding algorithm.

and backwards. An attractor has been found if the forward reachable states are contained in the backward reachable states. The explored states are removed from the state space and the process is repeated from another arbitrary state, until the entire state space has been explored. Our variant of this algorithm is shown in Figure 2.

Following [30], when analysing an asynchronous network, we first analyse the corresponding simpler synchronous network, and retain any loop attractors that are also valid attractors of the asynchronous system. As an optimisation to the original algorithm of Garg *et al.*, we find all fixpoint attractors in one step by intersecting the (synchronous) successor relation with the identity function.

To represent a set of states of a qualitative network as a binary decision diagram we use a binary encoding, introducing for each variable v_i of granularity N in the QN $\lceil \log_2 N \rceil$ binary variables, and ruling out invalid bit combinations which result when N is not a power of 2 (line 3 of the algorithm in Figure 2, and also in the fixpoints computation).

To represent each target function, we build a generalised truth table by evaluating the function at each possible combination of input values, and then encode this table as a binary decision diagram. To constrain the ranges of variables that need to be considered in order to build this table our implementation relies on the decreasing reachability sets algorithm of [17]. For models that contain target functions with a large number of input variables, encoding a symbolic representation of the function would be more appropriate. We then encode the successor relation of a synchronous or asynchronous network as a BDD following [30]. We implemented this algorithm in C++ and incorporated it into the open source tool BMA⁴. We make an implementation of the methodology outlined in this paper available through the command line interface of BMA and associated source code. However, in the future we plan to extend access to the web-based tool initially as part of the Natural Language Interface [1]⁵, and eventually tools in the GUI.

4 Exploring order of mutations

We wish to explore the effect of all possible combinations of mutations to a set of oncogenic genes, in order to determine if particular sequences of consecutive mutations are predicted to be more favourable to cancer progression than others. We make the simplification of assuming mutations are independent events, and that the system reaches an attractor between mutations. For example, if the cancer under consideration has mutations A , B and C , we consider the qualitative network model with all possible combinations of these three mutations (Figure 3). To model the mutation of a gene, we modify the associated target function to be a constant value. We describe this process more formally below.

⁴ <https://github.com/Microsoft/BioModelAnalyzer>

⁵ Development builds of this interface can be accessed as a chat-bot in Skype (<https://join.skype.com/bot/711331aa-e775-49be-b99d-6c42fc25f6d2>) or Telegram (<https://telegram.me/BioModelAnalyzerBot>)

We can represent a particular set of mutations M to a subset of variables $V' \subseteq V$ in a QN $Q(V, T, N)$ as a set of variable/value pairs:

$$M = \{(v_i, n_i) \mid v_i \in V'\} \quad (2)$$

where each v_i denotes a gene that is mutated, and $n_i \in \{0, \dots, N\}$ represents whether the gene is knocked out to value 0 or is constitutively expressed at some value > 0 . We model all mutations as maximally effective (either completely inhibiting or activating a variable). Different mutations may have different degrees of effect, for example point mutations at different locations within a gene may impair protein folding in different ways, but as most of the variables in the network range from 0-2 there is limited scope to explore this. However, a qualitative network model can incorporate higher granularities, so this could be a future extension of the model, but here we assume the most oncogenic mutation.

The modified QN given after application of these mutations is obtained by substituting the target function for each $v_i \in V'$ by a constant function, $T'_i(s) = n_i$:

$$Q'(V, \{T'_i \mid v_i \in V'\} \cup \{T_i \mid v_i \in V \setminus V'\}, N) \quad (3)$$

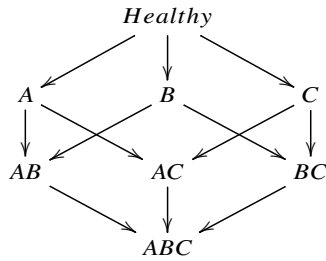


Fig. 3: To explore the order of mutations ABC, we apply them in all possible orders by constructing a powerset as above.

To consider all possible sequences of mutations leading to a set of observed mutations M , we construct the powerset 2^M and analyse all modified models resulting from applying each set of mutations $m \in 2^M$, starting with the healthy state in which no mutations are applied other than to model the presence of growth factors. (Figure 3). We compute the attractors for each of these models, and can then construct plots visualising levels of variables judged to be representative of the health of the cell,

such as those controlling proliferation or apoptosis.

Charting paths through these plots allows us to determine favourable and unfavourable sequences of mutations. For example, in a model with two oncogenes A and B , if applying mutations $\{A\}$ and $\{A, B\}$ both result in modified models with high levels of proliferation and low levels of apoptosis, while applying mutation $\{B\}$ results in a modified model with a low level of proliferation and a high level of apoptosis, this implies that it is favourable for the cancer for gene A to be mutated before gene B . When constructing these plots we consider only attractors that are reachable from a parent node. For example, for the model with mutations $\{A, B\}$, we consider only those attractors that are now reachable with the additional mutation, starting from states in the attractors of the model with mutation $\{A\}$, or from states in the attractors of the model with mutation $\{B\}$. This assumes that there is enough time to reach an attractor between mutation events. We have written R scripts that implement this methodology, available at <https://github.com/JFisherLab/MutationOrder>.

Model	Variables	Mutations	Synchronous time (s)	Asynchronous time (s)
Lung cancer	46	4	16	275
AML	42	5	37	37
CML	54	6	28	28
Breast cancer	75	7	57	57

Fig. 4: We assessed the practicality of our method by applying it to four qualitative network cancer cell models, with differing numbers of variables and mutations. Execution time is shown above.

5 Results

5.1 Benchmarks

To assess the practicality of our method, we applied it to four models of gene regulatory networks involved in cancers of different types: a lung cancer model (unpublished), a model of acute myeloid leukemia [58], a model of chronic myeloid leukemia [16], and a model of ER-negative breast cancer (in submission). These results are shown in Figure 4. All computations were performed on a Windows 10 PC with an Intel Xeon E5 @ 3.70Ghz and 16GB of RAM.

Note that if the model has only fixpoint attractors (no loops) under a particular set of mutations when considering the network synchronously, then there are the same attractors in the asynchronous case. The breast cancer, AML and CML leukemia models all have only fixpoint attractors and therefore there is no time difference when running the algorithm for either case. The lung cancer model, on the other hand, has a complex loop attractor and so different behaviour depending on whether it is being analysed synchronously or asynchronously. This leads to the much longer running time for the algorithm when executed in asynchronous mode.

5.2 Application to a model of ER-negative breast cancer

As a proof-of-concept application of our methodology, we turn to a recently developed synchronous qualitative network model of ER-negative breast cancer (in submission). To find mutations known to lead to cancer, we use the mutational profiles of common breast cancer cell lines [35]. We focus on ER-negative cell lines as the main driver genes are in our existing breast cancer model. The phenotypes we consider are proliferation and apoptosis, so we do not include mutations that affect parts of the genetic regulatory network not in the model, for example genes involved in invasiveness and differentiation such as E-cadherin. These phenotypes are represented in the model by their own variables representing proliferation and apoptosis, which integrate signals from terminally downstream genes and proteins in their respective pathways. For different attractors, these will have different levels

of activity indicating the predicted behaviour of the cell. The change in the activity in these variables is visualised for the different cell lines below.

5.2.1 SKBR3 demonstrates tightly constrained path to cancer

SKBR3, which is an ER-negative basal epithelial cell line [36], shows an example in which the model predicts that the order of mutations can be strongly constrained by apoptosis (Fig. 5b). There is only one path from the healthy state that minimises apoptosis while the cancer acquires the mutations shown in the SKBR3 cell line. Our method suggests that the level of apoptosis, due to the action of c-myc in the network model, is sufficient to strongly constrain the order of mutations. c-myc is a strong trigger of apoptosis, mainly via p53, perhaps as a fail-safe for its potent proliferative effects [25]. This means that in our model acquisition of an oncogenic c-myc mutation is most beneficial dependent upon several earlier mutations removing these safeguards, and we predict this encourages late mutation. This is as observed in breast cancer, where in many cases c-myc is amplified, but this appears to be associated with the later stages of tumour progression [13, 19, 3, 50]. In the case of the SKBR3 cell line, the path to breast cancer is particularly tightly constrained, compared to the other cell lines we studied (Section 5.2.2), as there is only one optimal order for these mutations to emerge.

Apoptosis remains lower in the model if p53 mutates early. This is as expected if we assume that the order of mutations is more strongly constrained by negative selection, in our model due to apoptosis, than positive selection by proliferation. This contradicts some models of tumour development in colorectal [26] and pancreatic cancer [37] but is consistent with experimental data for breast cancer [69, 56, 47] where p53 is often mutated early, as well as other cancers [14].

ErbB2 (a.k.a HER2) has a more contextual effect in the model, with it increasing apoptosis if it is the first mutation, but offering some protection from the downsides of c-myc activation, through the PI3K pathway, if mutated after c-myc and p53. It has been observed that when Ras, which is one of the key effectors of ErbB2 driven signalling, is mutated alone, it is a less effective driver of tumourigenesis than when it is combined with a mutation in c-myc, and vice-versa [42]. This is consistent with the model prediction in that a mutation in ErbB2, which will activate Ras, raises apoptosis, and so is detrimental to the cancer, when it mutates before c-myc (top left edge in Fig. 5b), but is beneficial by lowering apoptosis when it mutates after c-myc and p53 (bottom right edge in Fig. 5b), and so recapitulates this cooperation. Such contextual behaviour will be important to map out as it will be a determinant of what mutations are selected for in response to therapy. Similarly, if ErbB2 mutation is beneficial or detrimental at different stages of tumour evolution, this suggests that the order of mutation might influence whether a ErbB2 mutation is acquired at all. Here we restrict ourselves to considering combinations of oncogenes which are observed in one cell line, but this could be explored by considering together many possible oncogenes, and seeing whether ErbB2 mutations are likely to be selected for or against at different points, and whether this leads to diverging paths to cancer.

Proliferation is a weaker constraint for this cell line with different paths offering similar increases. Mutations which lead to high proliferation triggering higher apoptosis due to the coupling of the cell cycle to tumour suppression via p53, until this is mitigated by other mutations (Fig. 5a).

5.2.2 Acquisition of mutations is constrained to varying degrees

While there is one optimal order in which the mutations exhibited by the SKBR3 cell line can be acquired, this is not true for other sets of mutations observed in breast cancer cell lines. For example, SUM159PT [28] (Fig. 6) shows 2 equally optimal paths to cancer, whereas BT549 shows even looser ordering (Fig. 6). The fact there is

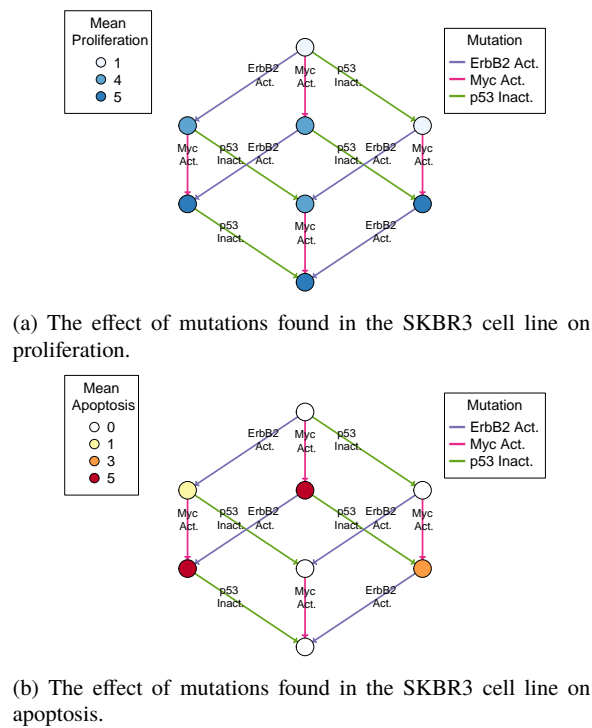


Fig. 5: How proliferation and apoptosis change as a tumour acquires the mutations seen in the SKBR3 cell line. The colour of the vertices corresponds to the level of proliferation (blue) and apoptosis (red). Edge colour and label corresponds to the mutation applied at each step, moving from top (healthy cell) to bottom (all mutations). Act. means that a node has been set to maximum activity, Inact. means a node has been set to 0 activity.



(a) The effect of mutations found in the SUM159PT cell line on apoptosis shows more flexibility than SKBR3 cell line.

(b) The effect of mutations found in the BT549 cell line on apoptosis show the order can vary more compared to SUM159PT, despite both being triple-negative cell lines.

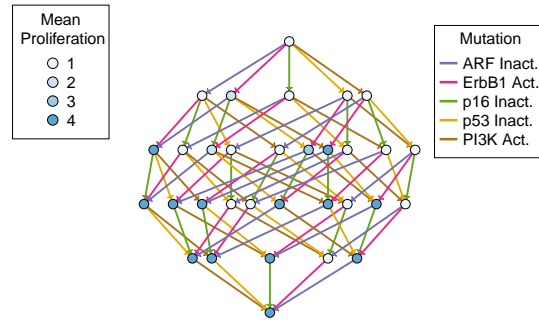
Fig. 6: Comparison of degree of constraint by apoptosis in different cell lines with an equal number of total mutations. (Left) The effect of mutations found in the SUM159PT cell line on apoptosis shows more flexibility than SKBR3 cell line. (Right) The effect of mutations found in the BT549 cell line on apoptosis show how the order is more flexible compared to SUM159PT, despite both being triple-negative cell lines.

such variance in the degree of constraint, even for phenotypically similar cell lines, implies some flexibility in the acquisition of mutations, but with underlying patterns.

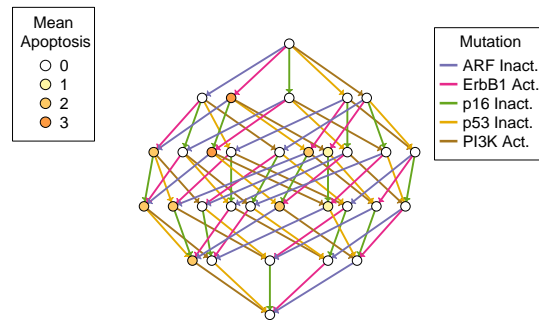
5.2.3 Difference in constraints from proliferation and apoptosis

While activation of tumour suppressor mechanisms places a strong negative selection pressure on tumour evolution, there must also be selection for increased proliferation. Strong constraints on oncogenes by apoptosis favour lineages in which *c-myc*, *ErbB1* and *ErbB2* mutate late. In the case of BT20 [43] however, there are many paths to cancer which are unconstrained by apoptosis (Fig. 7b). Nevertheless, there are still only a few paths which can be taken to maximise proliferation while also avoiding apoptosis (Fig. 7a).

However, the path selected for may change as during growth different phenotypes will be the limiting factor. In early stages, many mutations are often acquired by a small number of cells, implying avoidance of tumour suppression mechanisms is a prerequisite to the acquisition of proliferative mutations. Once escape from safeguards is achieved, then rapid growth can and does occur [63, 61]. As the tumour develops, other traits may take priority over such rapid growth, such as angiogenesis to support the tumour. As competition for space becomes a factor, higher apoptosis enabling a rapid turnover of cells may become tolerated.



(a) The effect of mutations found in the BT20 cell line on proliferation.



(b) The effect of mutations found in the BT20 cell line on apoptosis.

Fig. 7: Comparison of the role played in restriction of order of mutations in the BT20 cell line by proliferation vs apoptosis. There are only two attractors of the model which maximises possible proliferation for this cell line after 3 mutations, without apoptosis, out of 10 possible attractors at this stage of evolution. This means there are only 4 of a possible 21 minimal apoptosis paths that lead to maximum proliferative benefit.

5.2.4 Cell lines show redundancy in mutations

We would expect the optimum state to emerge only with the full complement of mutations for each cell line. This is true in the case of SKBR3, where proliferation is maximised and apoptosis minimised only in the final combination of all the mutations observed in the cell line which apply to our model. However, this is not true for all cell lines, for example BT20 (Fig. 7) gains no further benefit from mutation of p16 and p14ARF if p53 has been inactivated. Nevertheless, while the full path is not selected for, neither is it selected against by apoptosis. This implies that this is partly

a reflection of benefits not accounted for in our model. We only consider proliferation and apoptosis, but a cancer must also acquire other traits in order to be successful enough to be a candidate for a cell line. For example, the impact of our mutations on invasiveness [68], differentiation [65], angiogenesis [7], and immune response [21] will also exert selective pressures. However, these can be taken into account with further refinements to the models input into this method. It is also likely that the constraints of a cell line *in vitro* are different to those of a cancer *in vivo*, which may require extra mutation that a cancer would not benefit from.

Conversely, there may be variation within the cell line, and cells may need only a subset of all the mutations attributed to a cell line in order to be successful. If there are multiple viable endpoints within the set of possible mutations, then this method has the potential to predict which early mutations results in the choice of the final combination of mutations. Similarly, it has been observed that the same combination of mutations may result in different phenotypes, dependent upon the order in which they were acquired [49, 59]. We did not observe this in the case of our ER-negative breast cancer model, but the method is capable of finding such cases, as the attractors under a combination of mutations depend upon the initial states, which in turn depends upon the order of the previous mutations.

In these cases, the model could allow the use of early mutations as prognostic biomarkers. This could allow prediction of the future evolution of a tumour from early mutations and planning of therapy to account for this, perhaps to influence this evolution, as discussed by Basanta *et al.* [6].

6 Future Work

Our new method to explore the constraints on and effects of the order of mutations in the emergence of cancer could be extended to consider more phenotypes. However, this requires further assumptions about how to balance the priorities of a tumour, for example, how important is invasiveness to successful tumourigenesis compared to proliferation at different stages. Our breast cancer model also currently focuses on mutations which disrupt growth signalling in a cell-autonomous manner. External signals factors can be included in the network as variables just as genes and proteins, and so future models could explore the contribution of non-cell autonomous disruption to such signalling, and the spatial factors of cancer initiation, as reviewed in the case of colorectal cancer by van den Brink *et al* [10].

Our method assumes that all cells start on a level playing field in the same healthy state. However, even in healthy tissue there is an accumulation of neutral mutations [45], and so a wide variety of starting points exist. Similarly, while it is true that the vast majority of tumours will share a single founder cell, the landscape as the system transitions from a pre-malignant lesion to an established cancer, as described by Hanahan and Weinberg [34] will likely display neutral or only weakly selected mutations, resulting in a diverse and branching system of sub-clones. In order to explore this more fully, more diverse initial states can be considered, as well as

investigation of population effects and the effect of competition and cooperation between clones, with implications for early detection and treatment of cancer. We also currently use the method to reconstruct the history of known tumours, but it could also be used to predict how a larger set of mutations interacts.

A bottleneck in our methodology is the expansion of the 2^M mutated models. While we have not found this to be a problem with models of the scale we have considered so far, this step will become computationally expensive when considering a large number of mutated genes. A promising approach to deal with this limitation is to consider an extension of qualitative networks which have explicit support for dynamic modifications to target functions, akin to switching gene regulatory networks [57]. In principle this would allow us to consider all mutated models symbolically, in one step.

Analysis of the results will similarly become more complex as the number of mutations increases, and also if more cellular behaviours than proliferation and apoptosis are considered. As the number of combinations of mutations, M , increases, the number of paths between them increases as $M!$, therefore finding the best paths manually becomes infeasible. In the future it will be necessary to automate the finding of optimal paths based on scoring of the level of the different phenotypes the cancer passes through, for example penalising those orderings of mutations which pass through states with high apoptosis.

We currently assume that all mutations occur as single events, and with sufficient time for the system to find an attractor between mutations. Relaxing this assumption may be necessary in later stages of the cancer as chromosomal instability increases, or in the case of mass chromosomal rearrangement events such as chromothripsis [48, 20], chromoplexy [5] or kataegis [22].

Finally, while our model broadly agrees with the existing literature on the order of mutations in breast cancer, further experimental testing of our predictions is possible. The evolution of tumours has been studied in organoid systems, which more closely model the conditions a developing cancer would face *in vivo*, [51]. Alternatively, individual mutations can now be applied sequentially using emerging techniques such as CRISPR-Cas9 [53]. This would allow experiments guided by our model to closely recapitulate the early stages of tumour evolution.

7 Conclusions

We present a new method to find how behaviours such as proliferation and apoptosis constrain the order in which mutations are acquired during the emergence of cancer. This required the use of binary decision diagrams, which allow the efficient exploration of all possible attractors in qualitative networks, from any set of initial states, simulated both synchronously and asynchronously. Further, we demonstrate in a model of ER-negative breast cancer that different paths to cancer have varying degrees of constraint placed upon them by apoptosis and proliferation. *c-myc* places an especially strong bound on the paths that can be taken to tumourigenesis, while

this order depends on the context of other mutations, such as ErbB2. Our study could be expanded to provide a biological explanation, and be used as the basis for prediction of patterns in tumour evolution.

References

- [1] Ahmed Z, Benque D, Berezin S, Dahl ACE, Fisher J, Hall BA, Ishtiaq S, Nanavati J, Piterman N, Riechert M, Skoblov N (2017) Bringing ltl model checking to biologists. In: Bouajjani A, Monniaux D (eds) *Verification, Model Checking, and Abstract Interpretation*, Springer International Publishing, Cham, pp 1–13, DOI 10.1007/978-3-319-52234-0_1
- [2] Arends JW (2000) Molecular interactions in the Vogelstein model of colorectal carcinoma. *Journal of Pathology* 190(4):412–416, DOI 10.1002/(SICI)1096-9896(200003)190:4<412::AID-PATH533>3.0.CO;2-P
- [3] Aulmann S, Adler N, Rom J, Helmchen B, Schirmacher P, Sinn HP (2006) C-Myc Amplifications in Primary Breast Carcinomas and Their Local Recurrences. *Journal of Clinical Pathology* 59(4):424–428, DOI 10.1136/jcp.2005.029264
- [4] Autier P, Boniol M, La Vecchia C, LaVecchia C, Vatten L, Gavin A, Héry C, Heanue M (2010) Disparities in breast cancer mortality trends between 30 European countries: retrospective trend analysis of WHO mortality database. *BMJ (Clinical research ed)* 341:c3620, DOI 10.1136/bmj.c3620
- [5] Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, Park K, Kitabayashi N, MacDonald TY, Ghandi M, Van Allen E, Kryukov GV, Sboner A, Theurillat JP, Soong TD, Nickerson E, Auclair D, Tewari A, Beltran H, Onofrio RC, Boysen G, Guiducci C, Barbieri CE, Cibulskis K, Sivachenko A, Carter SL, Saksena G, Voet D, Ramos AH, Winckler W, Cipicchio M, Ardlie K, Kantoff PW, Berger MF, Gabriel SB, Golub TR, Meyerson M, Lander ES, Elemento O, Getz G, Demichelis F, Rubin MA, Garraway LA (2013) Punctuated evolution of prostate cancer genomes. *Cell* 153(3):666–677, DOI 10.1016/j.cell.2013.03.021
- [6] Basanta D, Gatenby RA, Anderson ARA (2012) Exploiting evolution to treat drug resistance: Combination therapy and the double bind. *Molecular Pharmaceutics* 9(4):914–921, DOI 10.1021/mp200458e
- [7] Blagosklonny MV, An WG, Romanova LY, Trepel J, Fojo T, Neckers L (1998) p53 inhibits hypoxia-inducible factor-stimulated transcription. *The Journal of biological chemistry* 273(20):11995–8
- [8] Bozic I, Reiter JG, Allen B, Antal T, Chatterjee K, Shah P, Moon YS, Yaqubie A, Kelly N, Le DT, Lipson EJ, Chapman PB, Diaz LA, Vogelstein B, Nowak MA (2013) Evolutionary dynamics of cancer in response to targeted combination therapy. *eLife* 2:e00747, DOI 10.7554/eLife.00747

- [9] Bray F, McCarron P, Parkin DM (2004) The changing global patterns of female breast cancer incidence and mortality. *Breast Cancer Research* 6(6):229, DOI 10.1186/bcr932
- [10] van den Brink GR, Offerhaus GJ (2007) The Morphogenetic Code and Colon Cancer Development. *Cancer Cell* 11(2):109–117, DOI 10.1016/j.ccr.2007.01.003
- [11] Bryant RE (1986) Graph-Based Algorithms for Boolean Function Manipulation. *IEEE Transactions on Computers* 100(8):677–691, DOI 10.1109/TC.1986.1676819
- [12] Casás-Selves M, DeGregori J (2011) How Cancer Shapes Evolution and How Evolution Shapes Cancer. *Evolution: Education and Outreach* 4(4):624–634, DOI 10.1007/s12052-011-0373-y, NIHMS150003
- [13] Chen Y, Olopade OI (2008) MYC in breast tumor progression. *Expert Review of Anticancer Therapy* 8(10):1689–1698, DOI 10.1586/14737140.8.10.1689
- [14] Chin L, Artandi SE, Shen Q, Tam A, Lee SL, Gottlieb GJ, Greider CW, DePinho RA (1999) P53 Deficiency Rescues the Adverse Effects of Telomere Loss and Cooperates With Telomere Dysfunction To Accelerate Carcinogenesis. *Cell* 97(4):527–538, DOI 10.1016/S0092-8674(00)80762-X
- [15] Chong CR, Jänne PA (2013) The quest to overcome resistance to EGFR-targeted therapies in cancer. *Nature medicine* 19(11):1389–400, DOI 10.1038/nm.3388, NIHMS150003
- [16] Chuang R, Hall BA, Benque D, Cook B, Ishtiaq S, Piterman N, Taylor A, Vardi M, Koschmieder S, Gottgens B, Fisher J (2015) Drug Target Optimization in Chronic Myeloid Leukemia Using Innovative Computational Platform. *Scientific Reports* 5:8190, DOI 10.1038/srep08190
- [17] Claessen K, Fisher J, Ishtiaq S, Piterman N, Wang Q (2013) Model-Checking Signal Transduction Networks through Decreasing Reachability Sets. In: *Computer Aided Verification*, Springer, pp 85–100
- [18] Cook B, Fisher J, Krepska E, Piterman N (2011) Proving Stabilization of Biological Systems. In: *VMCAI*, Springer, vol 11, pp 134–149, DOI doi:10.1007/978-3-642-18275-4_11
- [19] Corzo C, Corominas JM, Tusquets I, Salido M, Bellet M, Fabregat X, Serrano S, SolÀ F (2006) The myc oncogene in breast cancer progression: from benign epithelium to invasive carcinoma. *Cancer Genetics and Cytogenetics* 165(2):151 – 156, DOI <https://doi.org/10.1016/j.cancergencyto.2005.08.013>
- [20] Crasta K, Ganem NJ, Dagher R, Lantermann AB, Ivanova EV, Pan Y, Nezi L, Protopopov A, Chowdhury D, Pellman D (2012) DNA breaks and chromosome pulverization from errors in mitosis. *Nature* 482(7383):53–58, DOI 10.1038/nature10802
- [21] Cui Y, Guo G (2016) Immunomodulatory function of the tumor suppressor p53 in host immune response and the tumor microenvironment. *International Journal of Molecular Sciences* 17(11), DOI 10.3390/ijms17111942
- [22] D’Antonio M, Tamayo P, Mesirov JP, Frazer KA (2016) Kataegis Expression Signature in Breast Cancer Is Associated with Late Onset, Better

- Prognosis, and Higher HER2 Levels. *Cell Reports* 16(3):672–683, DOI 10.1016/j.celrep.2016.06.026
- [23] Davis A, Gao R, Navin N (2017) Tumor evolution: Linear, branching, neutral or punctuated? *Biochimica et Biophysica Acta - Reviews on Cancer* 1867(2):151–161, DOI 10.1016/j.bbcan.2017.01.003
- [24] Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, Ritchey JK, Young MA, Lamprecht T, McLellan MD, McMichael JF, Wallis JW, Lu C, Shen D, Harris CC, Dooling DJ, Fulton RS, Fulton LL, Chen K, Schmidt H, Kalicki-Veizer J, Magrini VJ, Cook L, McGrath SD, Vickery TL, Wendl MC, Heath S, Watson MA, Link DC, Tomasson MH, Shannon WD, Payton JE, Kulkarni S, Westervelt P, Walter MJ, Graubert TA, Mardis ER, Wilson RK, DiPersio JF (2012) Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* 481(7382):506–510, DOI 10.1038/nature10738
- [25] Evan GI, Wyllie AH, Gilbert CS, Littlewood TD, Land H, Brooks M, Waters CM, Penn LZ, Hancock DC (1992) Induction of apoptosis in fibroblasts by c-myc protein. *Cell* 69(1):119–28
- [26] Fearon ER, Vogelstein B (1990) A genetic model for colorectal tumorigenesis. *Cell* 61(5):759–767, DOI 10.1016/0092-8674(90)90186-I
- [27] Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F (2015) Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer* 136(5):E359–E386, DOI 10.1002/ijc.29210, arXiv:1011.1669v3
- [28] Flanagan L, Van Weelden K, Ammerman C, Ethier SP, Welsh J (1999) SUM-159PT cells: a novel estrogen independent human breast cancer model system. *Breast cancer research and treatment* 58(3):193–204
- [29] Gao R, Davis A, McDonald TO, Sei E, Shi X, Wang Y, Tsai PC, Casasent A, Waters J, Zhang H, Meric-Bernstam F, Michor F, Navin NE (2016) Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nature genetics* 48(10):1119–30, DOI 10.1038/ng.3641, 15334406
- [30] Garg A, Di Cara A, Xenarios I, Mendoza L, De Micheli G (2008) Synchronous versus asynchronous modeling of gene regulatory networks. *Bioinformatics* 24(17):1917–1925, DOI 10.1093/bioinformatics/btn336
- [31] Gray JW (2003) Evidence emerges for early metastasis and parallel evolution of primary and metastatic tumors. *Cancer Cell* 4(1):4–6, DOI 10.1016/S1535-6108(03)00167-3
- [32] Greaves M, Maley CC (2012) Clonal evolution in cancer. *Nature* 481(7381):306–13, DOI 10.1038/nature10762
- [33] Greenman CD, Pleasance ED, Newman S, Yang F, Fu B, Nik-Zainal S, Jones D, Lau KW, Carter N, Edwards PA, Futreal PA, Stratton MR, Campbell PJ (2012) Estimation of rearrangement phylogeny for cancer genomes. *Genome Research* 22(2):346–361, DOI 10.1101/gr.118414.110
- [34] Hanahan D, Weinberg RA (2011) Hallmarks of cancer: The next generation. *Cell* 144(5):646–674, DOI 10.1016/j.cell.2011.02.013, 0208024

- [35] Hollestelle A, Nagel JHA, Smid M, Lam S, Elstrodt F, Wasielewski M, Ng SS, French PJ, Peeters JK, Rozendaal MJ, Riaz M, Koopman DG, Ten Hagen TLM, De Leeuw BHCGM, Zwarthoff EC, Teunisse A, Van Der Spek PJ, Klijn JGM, Dinjens WNM, Ethier SP, Clevers H, Jochemsen AG, Den Bakker MA, Foekens JA, Martens JWM, Schutte M (2010) Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. *Breast Cancer Research and Treatment* 121(1):53–64, DOI 10.1007/s10549-009-0460-8
- [36] Holliday DL, Speirs V (2011) Choosing the right cell line for breast cancer research. *Breast cancer research : BCR* 13:215, DOI 10.1186/bcr2889
- [37] Hruban RH, Goggins M, Parsons J, Kern SE (2000) Progression model for pancreatic cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* 6(8):2969–72
- [38] Hurvitz S, Mead M (2015) Triple-negative breast cancer. *Current Opinion in Obstetrics and Gynecology* p 1, DOI 10.1097/GCO.0000000000000239
- [39] Independent UK Panel on Breast Cancer Screening (2012) The benefits and harms of breast cancer screening: an independent review. *Lancet (London, England)* 380(9855):1778–86, DOI 10.1016/S0140-6736(12)61611-0
- [40] Knuth DE (2009) *The Art of Computer Programming: Bitwise Tricks & Techniques; Binary Decision Diagrams, Volume 4, Fascicle 1*. Addison-Wesley
- [41] Kuukasjärvi T, Karhu R, Tanner M, Kähkönen M, Schäffer A, Nupponen N, Pennanen S, Kallioniemi A, Kallioniemi OP, Isola J (1997) Genetic heterogeneity and clonal evolution underlying development of asynchronous metastasis in human breast cancer. *Cancer research* 57(8):1597–604
- [42] Land H, Chen AC, Morgenstern JP, Parada LF, Weinberg RA (1986) Behavior of myc and ras oncogenes in transformation of rat embryo fibroblasts. *Molecular and Cellular Biology* 6(6):1917–1925
- [43] Lasfargues EY, Ozzello L (1958) Cultivation of human breast carcinomas. *Journal of the National Cancer Institute* 21(6):1131–47
- [44] Maley CC, Galipeau PC, Finley JC, Wongsurawat VJ, Li X, Sanchez CA, Paulson TG, Blount PL, Risques RA, Rabinovitch PS, Reid BJ (2006) Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nature genetics* 38(4):468–73, DOI 10.1038/ng1768
- [45] Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S, Wedge DC, Fullam A, Alexandrov LB, Tubio JM, Stebbings L, Menzies A, Widaa S, Stratton MR, Jones PH, Campbell PJ (2015) High burden and pervasive positive selection of somatic mutations in normal human skin (Supplement). *Science* 348(6237):880–886, DOI 10.1126/science.aaa6806
- [46] Murphy DJ, Junttila MR, Pouyet L, Karnezis A, Shchors K, Bui DA, Brown-Swigart L, Johnson L, Evan GI (2008) Distinct thresholds govern Myc’s biological output in vivo. *Cancer cell* 14(6):447–57, DOI 10.1016/j.ccr.2008.10.018
- [47] Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, Raine K, Jones D, Marshall J, Ramakrishna M, Shlien A, Cooke SL, Hinton J, Menzies A, Stebbings LA, Leroy C, Jia M, Rance R, Mudie LJ, Gamble SJ, Stephens PJ, McLaren S, Tarpey PS, Papaemmanuil E, Davies HR, Varela I, McBride DJ, Bignell GR, Leung K, Butler AP, Teague JW, Martin S, Jönsson

- G, Mariani O, Boyault S, Miron P, Fatima A, Langerød A, Aparicio SAJR, Tutt A, Sieuwerts AM, Borg Å, Thomas G, Salomon AV, Richardson AL, Børresen-Dale AL, Futreal PA, Stratton MR, Campbell PJ, Breast Cancer Working Group of the International Cancer Genome Consortium (2012) The life history of 21 breast cancers. *Cell* 149(5):994–1007, DOI 10.1016/j.cell.2012.04.023
- [48] Notta F, Chan-Seng-Yue M, Lemire M, Li Y, Wilson GW, Connor AA, Denroche RE, Liang SB, Brown AM, Kim JC, Wang T, Simpson JT, Beck T, Borgida A, Buchner N, Chadwick D, Hafezi-Bakhtiari S, Dick JE, Heisler L, Hollingsworth MA, Ibrahimov E, Jang GH, Johns J, Jorgensen LG, Law C, Ludkovski O, Lungu I, Ng K, Pasternack D, Petersen GM, Shlush LI, Timms L, Tsao MS, Wilson JM, Yung CK, Zogopoulos G, Bartlett JM, Alexandrov LB, Real FX, Cleary SP, Roehrl MH, McPherson JD, Stein LD, Hudson TJ, Campbell PJ, Gallinger S (2016) A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature* 538(7625):378–382, DOI 10.1038/nature19823, NIHMS150003
- [49] Ortmann CA, Kent DG, Nangalia J, Silber Y, Wedge DC, Grinfeld J, Baxter EJ, Massie CE, Papaemmanuil E, Menon S, Godfrey AL, Dimitropoulou D, Guglielmelli P, Bellosillo B, Besses C, Döhner K, Harrison CN, Vassiliou GS, Vannucchi A, Campbell PJ, Green AR (2015) Effect of Mutation Order on Myeloproliferative Neoplasms. *The New England journal of medicine* 372(7):601–612, DOI 10.1056/NEJMoa1412098, *N Engl J Med* 2011;365:981–92. Copyright
- [50] Robanus-Maandag EC, Bosch CAJ, Kristel PM, Hart AAM, Faneyte IF, Nederlof PM, Peterse JL, van de Vijver MJ (2003) Association of C-MYC amplification with progression from the in situ to the invasive stage in C-MYC-amplified breast carcinomas. *Journal of Pathology* 201(1):75–82, DOI 10.1002/path.1385
- [51] Roerink SF, Sasaki N, Lee-Six H, Young MD, Alexandrov LB, Behjati S, Mitchell TJ, Grossmann S, Lightfoot H, Egan DA, Pronk A, Smakman N, van Gorp J, Anderson E, Gamble SJ, Alder C, van de Wetering M, Campbell PJ, Stratton MR, Clevers H (2018) Intra-tumour diversification in colorectal cancer at the single-cell level. *Nature* DOI 10.1038/s41586-018-0024-3
- [52] Rowley M, Ohashi A, Mondal G, Mills L, Yang L, Zhang L, Sundsbak R, Shapiro V, Muders MH, Smyrk T, Couch FJ (2011) Inactivation of Brca2 promotes Trp53-associated but inhibits KrasG12D-dependent pancreatic cancer development in mice. *Gastroenterology* 140(4):1303–1313.e1–3, DOI 10.1053/j.gastro.2010.12.039
- [53] Sánchez-Rivera FJ, Jacks T (2015) Applications of the CRISPR-Cas9 system in cancer biology. *Nature reviews Cancer* 15(7):387–95, DOI 10.1038/nrc3950
- [54] Schaub MA, Henzinger TA, Fisher J (2007) Qualitative networks: a symbolic approach to analyze biological signaling networks. *BMC systems biology* 1:4, DOI 10.1186/1752-0509-1-4
- [55] Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A, Delaney A, Gelmon K, Guliany R, Senz J, Steidl C, Holt RA, Jones S, Sun M, Leung G, Moore R, Severson T, Taylor GA, Teschendorff AE, Tse K, Turashvili G, Varhol R, Warren RL, Watson P, Zhao Y, Caldas C, Huntsman D, Hirst

- M, Marra MA, Aparicio S (2009) Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* 461(7265):809–813, DOI 10.1038/nature08489
- [56] Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, Turashvili G, Ding J, Tse K, Haffari G, Bashashati A, Prentice LM, Khattra J, Burleigh A, Yap D, Bernard V, McPherson A, Shumansky K, Crisan A, Giuliany R, Heravi-Moussavi A, Rosner J, Lai D, Birol I, Varhol R, Tam A, Dhalla N, Zeng T, Ma K, Chan SK, Griffith M, Moradian A, Cheng SWG, Morin GB, Watson P, Gelmon K, Chia S, Chin SF, Curtis C, Rueda OM, Pharoah PD, Damaraju S, Mackey J, Hoon K, Harkins T, Tadigotla V, Sigaroudinia M, Gascard P, Tlsty T, Costello JF, Meyer IM, Eaves CJ, Wasserman WW, Jones S, Huntsman D, Hirst M, Caldas C, Marra MA, Aparicio S (2012) The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* 486(7403):395–399, DOI 10.1038/nature10933, NIHMS150003
- [57] Shavit Y, Yordanov B, Dunn SJ, Wintersteiger CM, Hamadi Y, Kugler H (2015) Switching Gene Regulatory Networks. In: *International Conference on Information Processing in Cells and Tissues*, Springer, Cham, pp 131–144, DOI 10.1007/978-3-319-23108-2_11
- [58] Silverbush D, Grosskurth S, Wang D, Powell F, Gottgens B, Dry J, Fisher J (2017) Cell-Specific Computational Modeling of the PIM Pathway in Acute Myeloid Leukemia. *Cancer research* 77(4):827–838, DOI 10.1158/0008-5472.CAN-16-1578
- [59] Skoulidis F, Cassidy LD, Pisupati V, Jonasson JG, Bjarnason H, Eyfjord JE, Karreth FA, Lim M, Barber LM, Clatworthy SA, Davies SE, Olive KP, Tuveson DA, Venkitaraman AR (2010) Germline Brca2 Heterozygosity Promotes KrasG12D -Driven carcinogenesis in a murine model of familial pancreatic cancer. *Cancer Cell* 18(5):499–509, DOI 10.1016/j.ccr.2010.10.015
- [60] Sotiriou C, Neo SYY, McShane LM, Korn EL, Long PM, Jazaeri A, Martiat P, Fox SB, Harris AL, Liu ET (2003) Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proceedings of the National Academy of Sciences of the United States of America* 100(18):10393–10398, DOI 10.1073/pnas.1732912100
- [61] Sottoriva A, Kang H, Ma Z, Graham TA, Salomon MP, Zhao J, Marjoram P, Siegmund K, Press MF, Shibata D, Curtis C (2015) A big bang model of human colorectal tumor growth. *Nature Genetics* 47(3):209–216, DOI 10.1038/ng.3214, 15334406
- [62] Sprouffske K, Pepper JW, Maley CC (2011) Accurate Reconstruction of the Temporal Order of Mutations in Neoplastic Progression. *Cancer Prevention Research* 4(7):1135–1144, DOI 10.1158/1940-6207.CAPR-10-0374
- [63] Sun QY, Ding LW, Tan KT, Chien W, Mayakonda A, Lin DC, Loh XY, Xiao JF, Meggendorfer M, Alpermann T, Garg M, Lim SL, Madan V, Hattori N, Nagata Y, Miyano S, Yeoh AEJ, Hou HA, Jiang YY, Takao S, Liu LZ, Tan SZ, Lill M, Hayashi M, Kinoshita A, Kantarjian HM, Kornblau SM, Ogawa S, Haferlach T, Yang H, Koefler HP (2017) Ordering of mutations in acute myeloid leukemia

- with partial tandem duplication of MLL (MLL-PTD). *Leukemia* 31(1):1–10, DOI 10.1038/leu.2016.160
- [64] Tabassum DP, Polyak K (2015) Tumorigenesis: it takes a village. *Nature Reviews Cancer* 15(8):473–483, DOI 10.1038/nrc3971
- [65] Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126(4):663–76, DOI 10.1016/j.cell.2006.07.024
- [66] Turajlic S, McGranahan N, Swanton C (2015) Inferring mutational timing and reconstructing tumour evolutionary histories. *Biochimica et Biophysica Acta - Reviews on Cancer* 1855(2):264–275, DOI 10.1016/j.bbcan.2015.03.005
- [67] Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, Chen K, Scheet P, Vattathil S, Liang H, Multani A, Zhang H, Zhao R, Michor F, Meric-Bernstam F, Navin NE (2014) Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* 512(7513):155–160, DOI 10.1038/nature13600, NIHMS150003
- [68] Wu X, Sun L, Wang X, Su P, Li Z, Zhang C, Wang Y, Gao P, Ma R (2016) Breast Cancer Invasion and Metastasis by mPR α Through the PI3K/Akt Signaling Pathway. *Pathology oncology research : POR* 22(3):471–6, DOI 10.1007/s12253-015-0023-8
- [69] Yates LR, Gerstung M, Knappskog S, Desmedt C, Gundem G, Van Loo P, Aas T, Alexandrov LB, Larsimont D, Davies H, Li Y, Ju YS, Ramakrishna M, Haugland HK, Lilleng PK, Nik-Zainal S, McLaren S, Butler A, Martin S, Glodzik D, Menzies A, Raine K, Hinton J, Jones D, Mudie LJ, Jiang B, Vincent D, Greene-Colozzi A, Adnet PY, Fatima A, Maetens M, Ignatiadis M, Stratton MR, Sotiriou C, Richardson AL, Lønning PE, Wedge DC, Campbell PJ (2015) Subclonal diversification of primary breast cancer revealed by multi-region sequencing. *Nature Medicine* 21(7):751–759, DOI 10.1038/nm.3886, arXiv:1408.1149