Introduction

Cancer is a malignant disease that involves dysregulated cellular processes in many aspects, including but not limited to abnormal proliferation and death, altered metabolism, tumour angiogenesis, loss of differentiation biomarkers, as well as invasion and metastasis to the surrounding and distant tissues\(^1\)\(^2\). Complex signalling cascades have been identified that are often substantially altered across the progression of the disease. However, a number of targeted therapeutics and immunotherapies have failed to deliver promising outcomes in clinical trials or in the actual practice. The reasons may include the loss of target generality due to tumour heterogeneity, the resistance from the cancer cell as a result of the cell’s robust bypassing of a particular signalling cascade, and other off-target effects\(^3\).

Aquaporins (AQPs) are a group of membrane channels that selectively permit water and some other small molecules like CO\(_2\) and urea. Since the first isolation of a 28kDa protein (CHIP28, later as AQP1) from human erythrocytes\(^4\), 13 aquaporins in total have been identified (AQP0-12) in human cells\(^5\). They distribute widely across almost every human cell type, but some localisation patterns do also exist. For instance, AQP4 is the most common aquaporin found in the astrocyte processes\(^6\), AQP2 on the apical membranes of the cells in the collecting duct, AQP5 on the luminal surfaces in pulmonary and colorectal epithelia\(^7\), and so forth. Studies from structural biology indicates that aquaporins are assembled as homo-tetramers, with each monomer containing a conserved NPA (asparagine-proline-alanine) motif that facilitates water selectivity. On the other hand, amino acids near the N or C terminals may be phosphorylated and thus may participate in signalling\(^5\).

There have been extensive studies since the late 1990s relating aquaporins to a number of stages and phenotypes in the progression of various cancers. A 1999 article\(^8\) examined the heterogeneous expression of AQP1 in some mammary carcinoma and glioblastoma cell lines, which was not found in normal brain/skin cell line controls. From then on, research has focused on the role of aquaporins in cancer due to and beyond their water permeability. Among them two notable groups have stood out: one group of studies\(^9\)\(^-\)\(^11\) have investigated the role of aquaporins in angiogenesis and the invasion of cancer cells, notably that of AQP1; the other group\(^12\)\(^-\)\(^15\) have looked into how aquaporins involve in a number of cell signalling pathways including Ras-MAPK and PI3K-Akt. Together these studies imply that aquaporins may be important players in the development and progression of cancer.

Computational modelling has become an indispensable tool for life scientists with its ever increasing importance in biomedical research. By integrating established knowledge with phenotypic data, we are able to extract useful information from large datasets, study molecular mechanisms behind the observations, as well as make reasonable predictions for a complex behaviour, all of which may further our understanding towards the basis and treatment of the disease. Two major approaches of modelling have arisen in practice. The
continuous (differential, or mathematical) approach uses a set of differential equations (or transfer functions if in the s-domain) to describe a certain biological system, like enzyme kinetics or the initiation and propagation of action potentials. Computational power is used to solve the equations, often non-linear and stiff; or to infer certain useful parameters with machine learning. On the other hand, the discrete (logical, or computational) approach assigns both discrete steps for time and discrete levels for the concentration, activity or phosphorylation of molecules. This encapsulates most of the qualitative features of biological interaction in a finite-state machine (FSM), hence is particularly useful in modelling networks like genetic regulation and cell signalling. Computational power is used to execute the FSM computer algorithm that mimics the biological phenomena.

This project arises from a series of experiments whose results seem incompatible with literature consensus. Riedel et al. discovered a sharp increase in Aqp1 transcription in lymph node stromal cells 4 days after tumour implantation into a murine lung, which, however, returned to its normal level after one further week when metastasis started to establish in the lymph nodes. Strikingly, p53-null murine cells transfected with Aqp1-siRNA showed similar or even increased level of proliferation in the assay (data not shown). The lack of understanding of aquaporin regulation in cells, and cancer cells in particular, has impeded us from providing a satisfactory explanation for the above observations. Therefore, the need became apparent for a model that is firmly based on our knowledge established in the literature, but is also able to explain the unprecedented behaviour in those experiments.

For the first time, a computational model of aquaporin regulation in cancer cells has been constructed in the form of a Qualitative Network based on the software BioModelAnalyzer (see Methods). The model connects some important aquaporins expressed in human cancer to important cellular phenotypes like proliferation, apoptosis, invasion and differentiation, via a number of pathways in cell signalling that are often dysregulated in a cancer cell. Based on over 60 publications, this model can not only satisfactorily reproduce the results reported in a discrete, qualitative manner, but also reconcile some seemingly incompatible phenotypes with research consensus by suggesting molecular mechanisms accountable. Novel predictions have also been made by mimicking real-life experimental conditions in the model.

**Methods**

**Qualitative Network and BioModelAnalyzer**

The logical modelling of a regulatory network in biology was pioneered by Stuart Kauffman and was developed extensively by René Thomas. A network model consists of three components: nodes, edges and target functions. A node (variable) represents a particular type of molecule, which has an integer value in a fixed range that represents its level of activity. At a particular step, the vector containing the values of all nodes is called a state. Edges are directed connections between nodes that can elicit either inhibitory or excitatory effect. Each node is also assigned a target function, which is an expression that takes inputs from upstream nodes. At every time step, the target function determines the value of the node at the next step.
If the range for every node is restricted to binary (0-1), the network is termed a *Boolean network*. It is able to encapsulate the essence of a biological regulatory process, and can be easy to analyse since Karnaugh maps can be used to simplify Boolean functions substantially. Nevertheless, it can be oversimplifying to model molecules with various activity levels as binary ON/OFF switches. In fact, it is a rule rather than an exception that a protein in different activity levels may exhibit distinct or sometimes even contrary behaviours.

A useful extension of the Boolean network is termed, first by Schaub et al., a *Qualitative Network* (QN), whose nodes are allowed to take ranges larger than [0,1], and whose target functions can go beyond logical operators and become algebraic. However, node values can only change by 1 *at most* at each time step (which effectively creates an inherent time delay; see Discussion). Thus for a node \( x \in [0, N] \) (\( x \in \mathbb{Z} \)), with a target function \( f(v_1, v_2, \ldots, v_i, \ldots) \) taking in input variables \( (v_1, v_2, \ldots, v_i, \ldots) \) at the current step, \( x \) at the next time step will become

\[
\begin{align*}
    x - 1 & \quad \text{if} \quad f < x \text{ and } x > 0 \\
    x + 1 & \quad \text{if} \quad f > x \text{ and } x < N \\
    x & \quad \text{if} \quad \begin{cases} 
        f = x, & \text{OR} \\
        f < x \text{ and } x = 0, & \text{OR} \\
        f > x \text{ and } x = N. & \text{OR}
    \end{cases}
\end{align*}
\]

More formal definitions of QN can be found in a number of articles, but there are a few issues that require further attention.

The first is about the *synchronicity* of updates across different variables. If, at each time step, we evaluate the target functions of all variables and update them simultaneously, the update will be *synchronous*. On the other hand, if only a proper subset of the variables is updated at a time, the update will be *asynchronous*. It should be carefully considered as to whether the synchronous or the asynchronous strategy should be adopted for a particular model. For systems with high concentrations of molecules that do not exhibit localisation, e.g. bacteria chemotaxis, it is safe to use synchronous updates. Whereas for systems that contain localised or low concentrations of molecules, like transcription regulation, asynchronous updates may be more accurate when describing the behaviour.

The second is about the *outcome* of network analysis. Since the number of states is finite and all executions are determinate, each execution must end in a cycle that falls within one of the following three endpoints:

(a) all executions end in a cycle of length 1 (i.e. in a fix point; the network stabilises);
(b) all executions end in a cycle of length greater than 1 (the network oscillates); or
(c) all executions end in some fix point, but there exists more than one (the network bifurcates).

Although all three outcomes may exist in the biological context, network stability is still favourable, as sound predictions can be made on the actual state of the cells. Instead of performing simulations that visit every possible state, efficient algorithms have been
developed to check the stability of a QN. One of the algorithms iteratively generate new lemmas based on existing knowledge of the model and shrink the variable ranges accordingly. This model checking console, together with a synchronous simulation strategy and a biologist-friendly interface, forms the basis of the software BioModelAnalyzer (BMA, http://biomodelanalyzer.org/). The BMA has been used previously to build models of signalling in chronic myeloid leukaemia (CML) and of the PIM signalling axis in acute myeloid leukaemia (AML), both of which successfully recapitulated experiment results in silico and suggested novel drug targets verified in the follow-up in vitro experiments.

**Model Building**

Workflows for developing a QN have been established in some articles, but in general it is an iterative process consisting of three main steps:

(a) establish the molecules (nodes) involved and draw a regulatory graph, which, of note, is the point at which the majority of cancer researchers would stop;
(b) discretise experiment results from the literature into state transition tables, and deduce target functions that yield the state transitions; and
(c) validate the model either by performing in vitro/in vivo experiments, or by reproducing in silico the experiments found in other publications; then refine the model accordingly.

In fact, the first two steps may be performed simultaneously as they both involve manual curation of literature. Once the model has been built and refined, novel predictions can be made by carrying out in silico experiments (e.g. knockout/overexpression of some nodes of interest), whose validation would be subject to real-life experiments. The workflow is summarised in Fig. 1.

In this project, the curation of literature is performed using discretisation tables (see Supplementary Note). For a given experiment, numerical levels are assigned to molecules of interest in order to represent the observations qualitatively (e.g. Western blot band intensities, cell count in assays, etc.). This essentially creates a local state transition table. After the curation of all publications available, the local tables may then be integrated into a global state transition table (Supplementary Tables S1-3) that encapsulates changes in every node in the model under different experimental inputs. Both tables are useful in that discretisation tables represent experiments in silico, whilst a global table provides a big picture of the changes in the overall cell behaviour. We then build target functions for the nodes, which are able to reproduce the state transitions and thus to capture the biological relationship. In the BMA, if no target function is specified for a node, the default action would be subtracting the average of inhibitory inputs from that of excitatory inputs.

Results

Building an initial model of AQP co-regulation in cancer

![Cellular Pathways](https://github.com/trinjacky/AQP-Modelling/)

**Figure 2.** Initial model of AQP co-regulation in cancer. ([https://github.com/trinjacky/AQP-Modelling/](https://github.com/trinjacky/AQP-Modelling/))
We generated an initial, primitive model from the curation of around 30 publications studying aquaporins or documenting basic cancer cell signalling (Fig. 2). The model connects 26 nodes to 4 cell phenotypes. Most nodes take the range [0, 2], other than the cascade of nodes from Ras to Fos whose values range from 0 to 3. As discussed, the ranges represent the proteins’ respective activity (phosphorylation) levels, with 0 representing low or no activity and 2/3 representing high or excessive activity. Note that range conversion takes place when a node takes input from a node of a different range (see Discussion). Cell behaviour is categorised by numerical levels of four phenotypes: proliferation, invasion/migration, differentiation and apoptosis.

Some important target functions are shown in Table 1, whilst other nodes either serve as inputs or are updated according to the default rule in the BMA (see Methods). The functions can be interpreted in a biological context. For example, proliferation phenotype is inhibited by the p53 and Rb proteins, whose effects are assigned weights of 60% and 40% respectively. For protein kinase A (PKA), the slight inhibition from AQ55 yields an overshoot before reaching steady state for both molecules in the simulation, which has also been documented in the literature25. Such overshoot behaviour is also captured in the regulation of AQP426, whose target function includes a factor of $(2 - \frac{3}{2}AQ51)$.

The initial model is able to capture most of the qualitative features from the literature. However, the model risks oversimplifying the cascades of cancer cell signalling, therefore puts the generality, validity and the robustness of the model into question. The complexity of AQ5 signalling implied by the primitive model also requires a better organisation of different components.

**The complete model**

Based on the initial QN model and a further curation of over 30 publications, a refined model was built (Fig. 3), incorporating a more complete picture of aquaporin signalling. Compared to the initial model, there have been a number of changes. First, we included AQ53, whose role in cancer had not been extensively studied but might be of importance.
Figure 3. Refined (complete) model of AQP co-regulation in cancer. (https://github.com/trinjacky/AQP-Modelling/)

Figure 4. A merged network from the interactome data accessed at the software Cytoscape [27], showing the three AQPs all interact with each other via CREB3.
given its permeability to glycerol, a potential source of energy; and to H$_2$O$_2$, a reactive oxygen species (ROS) that has a well-established role in signalling. Second, we refined the signalling axes, which provided more potential targets and extra robustness to the model. Not only did we add in receptor tyrosine kinases (RTK), CREB also came in given its role downstream of PKA and the sound evidence of CREB-AQP interaction from the interactome$^{27}$ (Fig. 4). Note that the node may actually represent a protein complex within which the labelled protein is of the greatest importance$^{24}$. Third, we organised the full regulatory graph into three modules: hypoxia and PKA regulation; the Ras-MAPK axis and the Akt-apoptosis cascades. Each of the modules takes limited inputs, undergoes individual signalling cascades, and performs crosstalk with other modules. Again, apart from the Ras-CREB cascade, all other nodes take the range [0,2].

We iteratively refined the target functions in the initial model and added in new features, which gave rise to a new set of functions as in Table 2. Although their biological interpretation becomes more complex, we may still dissect the functions into understandable components as in the Table. For an input variable $x \in [0,2]$, say, we can express its necessity (but not sufficiency) by multiplying a factor of ceil($x/2$) to the target function.

This complete model is able to recapitulate the expected behaviours of the initial model. Moreover, the validity is increased since we can reproduce some well-established findings in silico, for instance the knockout of RTK retards proliferation and invasion phenotypes (Supplementary Table S3). The following predictions from the in silico experiments are made with the refined model only.

![Figure 5](image-url)  
**Figure 5.** Phenotypic results of in silico experiments of individual AQP knockout in the context of p53-wt or p53-null. HIF1a is set to 1.

HIF1a has always been kept at the basal level (1) in these experiments, since it demonstrates the tendency of a ‘master switch’ (see Discussion). The state transitions are documented in Supplementary Tables S1-2.

**In silico experiments of AQP knockouts**

Now that a model has been constructed, we are able to perform in silico experiments to generate new predictions that have considerable biological implications and help direct future real-life experiments. Here we report the results of ‘knocking out’ individual aquaporins (i.e. by forcing its level to be 0), in the presence or absence of p53 (Fig. 5). No AQPs were knocked out in both controls. Note that
**Table 2.** Target functions for the complete model.

| Node   | Target Function                                                                 | Interpretation                                                                 |
|--------|---------------------------------------------------------------------------------|--------------------------------------------------------------------------------|
| PKA    | $cAMP \times \left( \frac{2 \times \text{avg}(HIF1\alpha, TGF\beta, cAMP)}{\text{cell}(\frac{1}{2} AQP5)} \right)$ | cAMP necessary and sufficient; when cAMP at basal level, other factors may regulate PKA |
| Src    | $PKA \times AQP5 \times \text{cell}(\frac{1}{2} FAK)$                           | All necessary; PKA and AQP5 are sufficient                                        |
| PTPs   | $2 - H_2O_2,\text{intra}$                                                       | Default level 2; intracellular $H_2O_2$ as antagonist                              |
| Nox2   | $\max(1, HIF1\alpha) \times \left( \frac{1}{2} \text{CXCR4} + \frac{1}{2} \text{RTK} \right)$ | HIF1\alpha sufficient; weighted effects of CXCR4 (50%), RTK & HIF1\alpha (each 25%) |
| AQP3   | $1 + \text{cell}(\frac{1}{2} \text{RTK}) - \frac{1}{2} \text{siAQP3}$           | RTK & AQP3 co-immunoprecipitates; siAQP3 cannot completely knock out AQP3         |
| AQP5   | $(2 - HIF1\alpha)(PKA + \text{floor}(\frac{1}{2} \text{RTK}))$                 | PKA necessary and sufficient; HIF1\alpha inhibits AQP5; high RTK(HER2) induces AQP5 & inhibition reverses it |
| RTK    | $2 - \text{cell}(\frac{1}{2} \text{PTPs})$                                     | EGFR & ErbB (HER2/3): Basal level 1; PTPs as antagonist                           |
| Ras    | $\text{RTK} \times \text{max}(1, \text{Src})$                                  | RTK necessary and sufficient; Src over-expression can trigger Ras                 |
| Rap    | $\text{cell}(\frac{1}{2} \text{RTK}) \times \text{Src}$                        | RTK necessary; Src necessary and sufficient                                        |
| CREB   | $\text{cell}(\frac{1}{2} \text{PKA}) \times (\frac{1}{2} \text{Akt} + \text{floor}(\frac{1}{2} \text{PKA})) \times \text{cell}(\text{avg}(\text{PKA}, \text{ERK}))$ | PKA necessary; Akt necessary but can be compensated by high PKA; averaging effect of PKA and ERK |
| CycD   | $\text{floor}(\frac{1}{2} (\text{cell}(3 \times \text{CREB} + 1) - \text{INK4A}))$ | CycD-CDK4 complex: round CREB to nearest integer; INK4A as antagonist            |
| AQP1   | $\text{cell}(\frac{1}{2} \text{CREB}) \times (2 - \max(\text{AQP4}_1, 1, 1 - \text{AQP4}_1))$ | CREB necessary (from interactome); AQP4 gives bell-like regulation                |
| AQP4   | $\text{cell}(2(1 - \frac{1}{2} \text{AQP1}) \times HIF1\alpha \times (p53 + \frac{1}{3} \text{CREB} + \frac{1}{3} HIF1\alpha - \frac{1}{2} \text{AQP1}))$ | HIF1\alpha necessary and sufficient; p53 sufficient; averaging effect of CREB, HIF1\alpha & AQP1; AQP1 as antagonist (but gives initial overshoot) |
| FAK    | $\max(1, \text{AQP1})$                                                          | FAK & AQP1 co-immunoprecipitates; basal level 1                                  |
| PI3K   | $\text{RTK} \times \text{Rap}$                                                 | Both necessary                                                                   |
| PTEN   | $2 - 2 \times H_2O_2, \text{intra}$                                             | Presence of intracellular $H_2O_2$ inhibits PTEN                                   |
| Bad    | $2 - \text{avg}(\text{Akt}, \text{PKA})$                                        | Default level 2; Akt/PKA as antagonist                                             |
| Bcl2   | $2 - \text{Bcl2}$                                                              | Default level 2; Bcl2 as antagonist                                               |
| Casp9  | $2 - \text{avg}(\text{Akt}, \text{Bcl2})$                                       | Default level 2; Akt/Bcl2 as antagonist                                            |
| p53    | $2 - \text{Mdm2}$                                                              | Default level 2; Mdm2 as antagonist                                               |
| AQP8_9 | $\text{PKA} \times (p53 + \text{floor}(\frac{1}{3} \text{PKA}))$               | PKA necessary and sufficient; p53 necessary but can be compensated by high PKA   |
| Casp3  | $\text{floor}(\frac{1}{2} \text{Casp9} + \frac{5}{3} \text{p53} + \text{cell}(\frac{1}{4} \text{TGF}\beta))$ | Weighted average of 40% Casp9 and 60% p53 (both necessary); but TGF\beta is sufficient |
| Apoptosis | $AVD \times \text{Casp3}$                                                      | Both are necessary                                                               |
| Proliferation | $\text{cell}(\frac{1}{3} \text{CycD}) \times (2 - \text{cell}(\frac{1}{3} \text{p53} + 1, 2 - \text{CycD}))$ | CycD necessary; weighted inhibition from p53 (60%) and Rb (2-CycD, 40%)           |
| Invasion | $\text{cell}(\frac{1}{3} \text{AQP1} + \frac{1}{3} \text{Src} - \frac{1}{2} \text{AQP4}_1) \times (\text{floor}(\frac{1}{4} \text{AQP1}) + \text{cell}(\frac{1}{4} \text{FAK}))$ | Averaging effect of AQP1, Src & AQP4; FAK necessary but can be compensated by high AQP1 |
| Differentiation | $\text{AQP4}$                                                                    | AQP4 is necessary and sufficient                                                  |
AQP1

From literature, AQP1 is highly related to the invasive and metastatic potential of cancer cells\textsuperscript{9,11,28,29}. It is overexpressed in over 10\% of the colorectal cancer and glioblastoma cell lines documented in the COSMIC database\textsuperscript{30} (http://cancer.sanger.ac.uk/). Mechanisms postulated include:

a) the cascade of local actin depolymerisation $\rightarrow$ AQP1-facilitated local expansion $\rightarrow$ actin re-polymerisation that stabilises the expansion\textsuperscript{31}; and
b) the ‘osmotic engine model’ which depicts the localisation of AQP1 at the lamellipodia, resulting in cell expansion at the lead and shrinkage at the back\textsuperscript{32}.

Therefore, it is appropriate to propose a direct relationship between AQP1 and invasion.

In both p53-wt and p53-null conditions, AQP1 knockout completely abolished the level of invasion. In the p53-null condition, the level of differentiation increased as a result of the increase in AQP4 level. Interestingly, the level of proliferation and apoptosis were unaltered compared to the controls in both conditions. Even though AQP1 may be connected to the two phenotypes via a FAK-dependent cascade\textsuperscript{28}, the model does not suggest a major role. In general, it agrees with our previous observation of a similar (or even increased; though not captured) level of proliferation in an AQP1-siRNA transfected p53-null cell line \textit{in vitro}.

AQP3

Studies have proposed the role of AQP3 in cancer signalling as a transporter for H\textsubscript{2}O\textsubscript{2} and glycerol. Although studies relating AQP3-facilitated glycerol transport to cancer seem contradictory (interestingly, because of the opposite arguments on the role of glycerol in cancer metabolism\textsuperscript{33,34}), the relationship is well-established between AQP3-dependent intracellular H\textsubscript{2}O\textsubscript{2} accumulation and its downstream signalling\textsuperscript{35-37}.

Here in our executable QN model, we could see a decreased level of both proliferation and invasion compared to the controls following the ‘transfection’ of AQP3-siRNA (although this does not completely deactivate the protein\textsuperscript{35}). In p53-wt condition apoptosis also went up. By examining the transition of states of other nodes, we could see that the decrease in intracellular H\textsubscript{2}O\textsubscript{2} level resulted in higher PTEN activity\textsuperscript{36}, thus inhibiting Akt survival pathways (CREB and p53 regulation). Further laboratory work therefore may also include the investigation of the levels of p53, Akt and CREB in an AQP3-related experiment.

AQP4

The majority of research relates AQP4 to the pathophysiology of oedema, particularly in the brain where the protein is abundant\textsuperscript{6,38}. Tumour-associated oedema has become an important aspect of study\textsuperscript{7,26,38}, but there has been no sufficient evidence of its protumorigenic effect. In contrast, a study of over 600 lung cancer samples has associated higher AQP4 expression with higher differentiation level and better prognosis\textsuperscript{39}. Surprisingly, this study also suggested a strong correlation between AQP1 and AQP4 (Pearson’s $r=0.82$), which is encapsulated in the model by mutual inhibition since they contribute to distinct phenotypes.
In the QN model, with p53 in place, the level of AQP4 was higher than AQP1, implying a differentiating, non-invasive behaviour; p53 knockout shifted the pattern to its opposite, which agreed with the literature\textsuperscript{38}. Knockout of AQP4 completely abolished differentiation activity regardless of p53, which seemed apparent as it was the only input to the phenotype node (see Discussion). Compared to both controls, other phenotype activities remained unchanged apart from a decrease in invasion in p53-null, due to a subsequent decrease in the level of AQP1.

**AQP5**

Studies have described the relationship between AQP5 and cancer cell proliferation and invasion. For a cell overexpressing AQP5, proliferation is increased via Ras-MAPK\textsuperscript{13} and/or PI3K-Akt\textsuperscript{14} axes, while invasion is related to Src\textsuperscript{12}. Both interactions concern the intracellular domain of the protein, therefore water permeability may not be relevant.

In the model, AQP5 knockout increased the level of PKA, because of the slight inhibition from AQP5 needed for its overshoot behaviour\textsuperscript{25} (see the Initial Model section). This resulted in a high level of CREB and hence of CycD, but whether p53 was present determined the eventual phenotypes. In the p53-wt condition, AQP5 knockout led to a decrease in proliferation and invasion compared to the control, as well as an increase in apoptosis. Interestingly, these were all seem to be the downstream effects of a low Src level. Decrease in Src activity directly affected the cell’s invasive potential, whilst a downregulation of Rap halted the Akt pathways and thus triggered high p53. Strikingly, when we knockout AQP5 and p53 altogether, the level of proliferation increased due to a high CycD level. Other phenotypes were unaltered with respect to the controls.

**AQP8 and AQP9**

The two aquaporins are considered pro-apoptotic, as two studies have documented a significant increase in their levels during the course of apoptosis\textsuperscript{40,41}. They are thought to be related to the apoptotic volume decrease (AVD) which is a necessary morphological change for the cell death to take place. In the model, AQP8/9 knockout directly inhibited AVD, hence maintaining the apoptosis phenotype at zero level.

**Discussion**

**Addressing the cell-line specificity**

When building the model, the aim has been the generation of a broad picture of AQP regulation and signalling in cancer cells, which is able to mimic the conversation between AQPS and different fundamental signalling cascades, and also that among AQPS themselves. Although AQP distribution can be distinct across different tissue types and hence can play a differential role in the signalling of cancers of different origins, we did not try to incorporate a cell-specific context in this model, given the fact that the literature relating AQPS to a specific cancer type is scarce. Nevertheless, from the in silico experiments we observed
varying behaviours when p53 was present or absent. This may capture some degree of cell-line specificity.

**The role of AQP5s and the ‘overfitting’ of literature**

While a number of molecular mechanisms have been postulated that relate AQP5s to different cancer phenotypes, careful consideration is required as to whether AQP5s have a critical, determining role; or are in fact markers that are possibly regulated by some common confounders. Such uncertainty is highlighted in the model, where differentiation phenotype simply follows the level of AQP4; where AQP5 can interfere strongly with Src signalling; and where HIF1a seems to be a ‘general switch’ that could alter the levels of other nodes dramatically. This again is possibly the consequence of the scarce literature we were able to find relating AQP5 to different signalling cascades and phenotypes. In order to accommodate the literature, we may risk overfitting by putting excessive importance into the AQP5s. Further studies in the area are highly expected, but researchers are advised to be cautious when suggesting a novel causal relationship between two signalling events.

**Target functions**

When developing target functions, a few issues have arisen that are worth discussing. First, the target functions are expected not only to reproduce the qualitative behaviour, but also to contain certain biological intuition. For instance, multiplication is used to describe the necessity of a molecule; addition is used to describe its sufficiency; and so on. Second, they are expected, in general, to be of simple forms. Most state transitions can be encapsulated using linear functions. Quadratic equations (and beyond) are unfavourable, as they can be simplified to the first degree in the discrete domain. For example, the function \((x - 1)^2\) in the domain \([0,1,2]\), which yields a bell-like regulation, can be simplified as \(\max(1 - x, x - 1)\). Third, one needs to take extra care when connecting nodes of different ranges. In the BMA, the values of inputs to a node will be adjusted to accommodate its range. The underlying biology is that the value of a node is regarded as a relative level of activity, such that the upper limit represents 100% activity of that node and the lower limit represents 0% activity. Therefore, for a node \(x \in [0,2]\), an input node \(y \in [0,3]\) will appear in the target function of \(x\) as \(y^\prime = \frac{2}{3}y\). This is the case in our QN model when taking inputs from the Ras-CREB cascade.

In general, we may be able to build, or even machine-learn complex target functions that perfectly follow the tables and publications, but such high level of abstraction could render biological interpretation much more difficult. We aim to capture the regulation algebraically but also in a simple, biologically-understandable manner.

**Conclusions**

Taken together, complex cellular signalling and the diverse roles membrane channels play in cancer result in the need for a computational model that is rapidly interpretable. By integrating various sources of primary literature and bioinformatics data, we have generated a qualitative network model that helps decipher the co-regulation of aquaporins in cancer cells. A workflow of literature curation and model construction has been
proposed; and by refining the target functions iteratively we are able to reproduce a number of documented observations in silico. Novel predictions have been made regarding specific aquaporin knockouts in the presence or absence of p53. In vivo or in vitro experiments that validate both the model behaviour and the predictions are highly anticipated. The issue of cell-line specificity may be further addressed to pave the way for potential applications in personalised cancer medicine. Even if the off-target effects and the poor specificity, given a high structural homology among aquaporins, might be prominent for their inhibitors, further exploration that improves the druggability of AQP5s is promised.

Reference

1 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. Cell 144, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).
2 Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. Cell 100, 57-70, doi:10.1016/s0092-8674(00)81683-9 (2000).
3 Kaelin, W. G., Jr. Common pitfalls in preclinical cancer target validation. Nat Rev Cancer 17, 425-440, doi:10.1038/nrc.2017.32 (2017).
4 Preston, G. M., Carroll, T. P., Guggino, W. B. & Agre, P. Appearance of Water Channels in Xenopus Oocytes Expressing Red Cell CHIP28 Protein. Science 256, 385-387, doi:10.1126/science.256.5055.385 (1992).
5 King, L. S., Kozono, D. & Agre, P. From structure to disease: the evolving tale of aquaporin biology. Nat Rev Mol Cell Biol 5, 687-698, doi:10.1038/nrm1469 (2004).
6 Papadopoulos, M. C. & Verkman, A. S. Aquaporin water channels in the nervous system. Nature Reviews Neuroscience 14, 265-277, doi:10.1038/nrn3468 (2013).
7 Verkman, A. S., Anderson, M. O. & Papadopoulos, M. C. Aquaporins: important but elusive drug targets. Nat Rev Drug Discov 13, 259-277, doi:10.1038/nrd4226 (2014).
8 Endo, M., Jain, R. K., Witwer, B. & Brown, D. Water channel (aquaporin 1) expression and distribution in mammary carcinomas and glioblastomas. Microvasc Res 58, 89-98, doi:10.1006/mvre.1999.2158 (1999).
9 Esteva-Font, C., Jin, B. J. & Verkman, A. S. Aquaporin-1 gene deletion reduces breast tumor growth and lung metastasis in tumor-producing MMTV-PyVT mice. FASEB J 28, 1446-1453, doi:10.1096/fj.13-245621 (2014).
10 Hu, J. & Verkman, A. S. Increased migration and metastatic potential of tumor cells expressing aquaporin water channels. FASEB J 20, 1892-1894, doi:10.1096/fj.06-5930fe (2006).
11 Saadoun, S., Papadopoulos, M. C., Hara-Chikuma, M. & Verkman, A. S. Impairment of angiogenesis and cell migration by targeted aquaporin-1 gene disruption. Nature 434, 786-792, doi:10.1038/nature03460 (2005).
12 Chae, Y. K. et al. Expression of aquaporin 5 (AQP5) promotes tumor invasion in human non small cell lung cancer. PLoS One 3, e2162, doi:10.1371/journal.pone.0002162 (2008).
13 Kang, S. K. et al. Role of human aquaporin 5 in colorectal carcinogenesis. Am J Pathol 173, 518-525, doi:10.2353/apjpath.2008.071198 (2008).
14 Chae, Y. K. et al. Human AQP5 plays a role in the progression of chronic myelogenous leukemia (CML). PLoS One 3, e2594, doi:10.1371/journal.pone.0002594 (2008).
15 Hoque, M. O. et al. Aquaporin 1 is overexpressed in lung cancer and stimulates NIH-3T3 cell proliferation and anchorage-independent growth. Am J Pathol 168, 1345-1353, doi:10.2353/apjpath.2006.050596 (2006).
16 Fisher, J. & Henzinger, T. A. Executable cell biology. Nat Biotechnol 25, 1239-1249, doi:10.1038/nbt1356 (2007).
17 Riedel, A., Shorthouse, D., Haas, L., Hall, B. A. & Shields, J. Tumor-induced stromal reprogramming drives lymph node transformation. Nat Immunol 17, 1118-1127, doi:10.1038/ni.3492 (2016).
18 Benque, D. et al. in Computer Aided Verification: 24th International Conference, CAV 2012, Berkeley, CA, USA, July 7-13, 2012 Proceedings (eds P. Madhusudan & Sanjit A. Seshia) 686-692 (Springer Berlin Heidelberg, 2012).
Thomas, R. & d'Ari, R. Biological feedback. (CRC press, 1990).

Schaub, M. A., Henzinger, T. A. & Fisher, J. Qualitative networks: a symbolic approach to analyze biological signaling networks. *BMC Syst Biol* 1, 4, doi:10.1186/1752-0509-1-4 (2007).

Silverbush, D. et al. Cell-Specific Computational Modeling of the PIM Pathway in Acute Myeloid Leukemia. *Cancer Res* 77, 827-838, doi:10.1158/0008-5472.CAN-16-1578 (2017).

Chuang, R. et al. Drug target optimization in chronic myeloid leukemia using innovative computational platform. *Sci Rep* 5, 8190, doi:10.1038/srep08190 (2015).

Cook, B., Fisher, K., Krepska, E. & Piterman, N. in Verification, Model Checking, and Abstract Interpretation: 12th International Conference, VMCAI 2011, Austin, TX, USA, January 23-25, 2011. *Proceedings* (eds Ranjit Jhala & David Schmidt) 134-149 (Springer Berlin Heidelberg, 2011).

Faure, A., Naldi, A., Chaouiya, C. & Thieffry, D. Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics* 22, e124-131, doi:10.1093/bioinformatics/btl210 (2006).

Yang, F., Kawedia, J. D. & Menon, A. G. Cyclic AMP regulates aquaporin 5 expression at both transcriptional and post-transcriptional levels through a protein kinase A pathway. *J Biol Chem* 278, 32173-32180, doi:10.1074/jbc.M305149200 (2003).

Mou, K. et al. AQP-4 in peritumoral edematous tissue is correlated with the degree of glioma and with expression of VEGF and HIF-alpha. *J Neurooncol* 100, 375-383, doi:10.1007/s11060-010-0205-x (2010).

Cline, M. S. et al. Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* 2, 2366-2382, doi:10.1038/nprot.2007.324 (2007).

Meng, F. et al. Aqp1 enhances migration of bone marrow mesenchymal stem cells through regulation of FAK and beta-catenin. *Stem Cells Dev* 23, 66-75, doi:10.1089/scd.2013.0185 (2014).

Monzani, E., Bazzotti, R., Perego, C. & La Porta, C. AQP1 is not only a water channel: it contributes to cell migration through Lin7/beta-catenin. *PLoS One* 4, e6167, doi:10.1371/journal.pone.0006167 (2009).

Forbes, S. A. et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res* 45, D777-D783, doi:10.1093/nar/gkw1121 (2017).

Papadopoulos, M. C., Saadoun, S. & Verkman, A. S. Aquaporins and cell migration. *Pflugers Arch* 456, 693-700, doi:10.1007/s00424-007-0357-5 (2008).

Stroka, K. M. et al. Water permeation drives tumor cell migration in confined microenvironments. *Cell* 157, 611-623, doi:10.1016/j.cell.2014.02.052 (2014).

Goldstein, I. et al. p53 promotes the expression of gluconeogenesis-related genes and enhances hepatic glucose production. *Cancer Metab* 1, 9, doi:10.1186/2049-3002-1-9 (2013).

Hara-Chikuma, M. & Verkman, A. S. Prevention of skin tumorigenesis and impairment of epidermal cell proliferation by targeted aquaporin-3 gene disruption. *Mol Cell Biol* 28, 326-332, doi:10.1128/MCB.01482-07 (2008).

Thiagarajah, J. R., Chang, J., Goettel, J. A., Verkman, A. S. & Lencer, W. I. Aquaporin-3 mediates hydrogen peroxide-dependent responses to environmental stress in colonic epithelia. *Proc Natl Acad Sci U S A* 114, 568-573, doi:10.1073/pnas.1612921114 (2017).

Satoooka, H. & Hara-Chikuma, M. Aquaporin-3 Controls Breast Cancer Cell Migration by Regulating Hydrogen Peroxide Transport and Its Downstream Cell Signaling. *Mol Cell Biol* 36, 1206-1218, doi:10.1128/MCB.00971-15 (2016).

Hara-Chikuma, M., Watanabe, S. & Satoooka, H. Involvement of aquaporin-3 in epidermal growth factor receptor signaling via hydrogen peroxide transport in cancer cells. *Biochem Biophys Res Commun* 471, 603-609, doi:10.1016/j.bbrc.2016.02.010 (2016).

Yan, J. H. et al. p53-induced uncoupling expression of aquaporin-4 and inwardly rectifying K+ 4.1 channels in cytotoxic edema after subarachnoid hemorrhage. *CNS Neurosci Ther* 18, 334-342, doi:10.1111/j.1755-5949.2012.0299.x (2012).

Warth, A. et al. Loss of aquaporin-4 expression and putative function in non-small cell lung cancer. *BMC Cancer* 11, 161, doi:10.1186/1471-2407-11-161 (2011).

Jablonski, E. M. et al. Decreased aquaporin expression leads to increased resistance to apoptosis in hepatocellular carcinoma. *Cancer Lett* 250, 36-46, doi:10.1016/j.canlet.2006.09.013 (2007).

Jessica Chen, M. et al. Water and ion channels: crucial in the initiation and progression of apoptosis in central nervous system? *Curr Neuropharmacol* 6, 102-116, doi:10.2174/157015908784533879 (2008).