Article

A Structural Characterisation of the Mitogen-Activated Protein Kinase Network in Cancer

Evangelos Chatzaroulas [1,†] , Vytenis Sliogeris [1,†], Pedro Victori [2], Francesca M. Buffa [2,*,‡],
Sotiris Moschoyiannis [1,*,‡] and Roman Bauer [1,*,‡]

1   School of Computer Science and Electronic Engineering, University of Surrey, Guildford GU2 7XH, UK;
    e.chatzaroulas@surrey.ac.uk (E.C.); v.sliogeris@surrey.ac.uk (V.S.)
2   Computational Biology and Integrative Genomics Lab, Department of Oncology, Medical Sciences Division,
    University of Oxford, Oxford OX3 7DQ, UK; pedro.victori@oncology.ox.ac.uk
*   Correspondence: francesca.buffa@oncology.ox.ac.uk (F.M.B.); s.moschoyiannis@surrey.ac.uk (S.M.);
    r.bauer@surrey.ac.uk (R.B.)
†   These authors contributed equally to this work and should be considered joint first authors.
‡   These authors contributed equally to this work and should be considered joint senior authors.

Abstract: Gene regulatory networks represent collections of regulators that interact with each other and with other molecules to govern gene expression. Biological signalling networks model how signals are transmitted and how activities are coordinated in the cell. The study of the structure of such networks in complex diseases such as cancer can provide insights into how they function, and consequently, suggest suitable treatment approaches. Here, we explored such topological characteristics in the example of a mitogen-activated protein kinase (MAPK) signalling network derived from published studies in cancer. We employed well-established techniques to conduct network analyses, and collected information on gene function as obtained from large-scale public databases. This allowed us to map topological and functional relationships, and build hypotheses on this network's functional consequences. In particular, we find that the topology of this MAPK network is highly non-random, modular and robust. Moreover, analysis of the network's structure indicates the presence of organisational features of cancer hallmarks, expressed in an asymmetrical manner across communities of the network. Finally, our results indicate that the organisation of this network renders it problematic to use treatment approaches that focus on a single target. Our analysis suggests that multi-target attacks in a well-orchestrated manner are required to alter how the network functions. Overall, we propose that complex network analyses combined with pharmacological insights will help inform on future treatment strategies, exploiting structural vulnerabilities of signalling and regulatory networks in cancer.

Keywords: cancer; hallmarks; complex networks; network analysis

1. Introduction

It is estimated that each year, approximately 20 million people worldwide are afflicted with cancer, and 10 million lose their lives to it [1]. In most countries, cancer is the first or second leading cause of death before the age of 70. Despite significant biomedical advances in the last decades, and although there exist treatments that have proven highly effective in certain types of cancer, survival rates in many cancer types remain low. Hence, new treatments are needed, as well as methods to improve existing treatment strategies. To this end, a better understanding of the origins and progression of cancer would be of immense benefit.

Cancer is a highly complex phenomenon, involving numerous genetic factors, and processes across various spatial and temporal scales, as well as interactions across cells and the tumour microenvironment. A well-supported conceptual framework for understanding the workings of cancer was proposed by Hanahan and Weinberg in their seminal study "The

hallmarks of cancer" [2]. In this study, the authors proposed that cancer entails six essential alterations in cell physiology that lead to malignant growth, namely: self-sufficiency in growth signals, insensitivity to anti-growth signals, evading apoptosis, limitless replicative potential, sustained angiogenesis, tissue invasion and metastasis. In 2011, this hypothesis was extended by two additional hallmarks [3]—reprogramming energy metabolism and evading immune response.

However, it is currently unclear whether these hallmarks are based on individual cell behaviour, or are emerging phenomena only at the tissue level. Indeed, despite vast amounts of experimental data having been collected since then, it remains an open question as to how cellular pathways and genetic factors can give rise to such hallmarks. In particular, it is unknown how the symmetry-breaking dynamics [4] necessary for cancer development are controlled by cellular pathways and signalling. In other words, we lack a mechanistic understanding of how the capabilities necessary for the progression of cancer arise.

Computational analysis and modelling can help better understand the origins [5], progression [6] as well as treatment options [7] of cancer. The latter includes recent proposals of artificial intelligence (AI) based approaches to targeted therapeutics, directed at the control of Boolean network models [8] of signalling and regulatory networks, namely rule-based machine learning [9], reinforcement learning [10,11], and deep reinforcement learning [12] to address larger networks. We note that the Boolean paradigm allows the modelling of cellular functioning through attractor theory [13], where attractors model phenotypes [14], without the need for estimating a large number of parameters. To realise the promise of such computational approaches, it is imperative to apply them to networks that faithfully capture the origins and progression of cancer.

In this paper, we make initial steps towards mechanistically mapping signalling and regulatory networks to the hallmarks of cancer. To this end, we analyse a literature-based Boolean model of one of the arguably best-established cancer networks, namely the mitogen-activated protein kinase (MAPK) pathway, which is deregulated in numerous cancer types [15,16]. We employ complex network analysis [17] tools to characterise the topological organisation, as well as incorporating well-established information on biomolecular function. Overall, we present a connectomic analysis of a core network in cancer, providing insights into potential genetic interactions governing the progression of cancer. Notably, our results have implications in the context of treatment strategies to interrupt progression.

2. Methods

In this section, we present the key methods that allow a computational approach to the connectomic characterisation of the MAPK network in the sequel.

2.1. Complex Network Analysis

We start by listing a number of measures, or network metrics, found in complex networks [17] that are central to the subsequent study and topological analysis of the MAPK network.

To compute the complex network measurements, the publicly available Python library NetworkX [18] was used with minor modifications, which we open-sourced at https://github.com/UoS-PLCCN/grn-metrics (accessed on the 11 May 2022).

All computations were made on a standard laptop with an Intel Core i7 Processor and 32 GB of RAM.

Communities. Community structure is one of the most studied features of networked systems [19]. Communities in a network are usually described as groups of densely connected nodes with sparse connections to the nodes of other groups.

Communities in our study were identified using the Clauset–Newman–Moore greedy modularity maxim [20], and were implemented using NetworkX's implementation of the function `greedy_modularity_communities()`. Greedy modularity maximization begins

with each node in its own community, then joins the pair of communities that most increases the modularity metric until no such pair exists.

Modularity. With the communities identified using greedy modularity maximization, we can additionally compute the resulting modularity of this particular partition of the graph. This will naturally be the maximal modularity value.

Given a set of n communities C, Clauset *et al* in [20] proved that modularity can be computed as

$$Q = \sum_{c \in C} \left[\frac{L_c}{|E|} - \gamma \left(\frac{k_c}{2|E|} \right)^2 \right] \tag{1}$$

where L_c is the number of intra-community links, k_c the sum of degrees within the community and γ the resolution parameter.

A high modularity value naturally implies that the network naturally lends itself to division into modules. We use the pre-computed communities along with NetworkX's `modularity` function to calculate it.

Small-world coefficient. Next, we are interested in degrees of small-worldness of the network under study [21]. A small-world network is more clustered with a smaller characteristic path length than degree-preserved random networks. In other words, most nodes can be reached from every other node by a small number of hops or steps. The small-world coefficient that determines how intense this effect is in a network refers to the ratio of clustering coefficient and characteristic path length, which are normalized relative to those of the random networks.

The small-world coefficient in our study was calculated using a modified implementation from NetworkX's `sigma` function from the `smallworld` module. Specifically, the original implementation uses a random reference graph, which we deem insufficient, and thus modify the function to use Erdos–Renyi graphs instead.

Clustering coefficient. The global clustering coefficient measures the clustering present in the entire network, as opposed to localization around a single node's neighbourhood.

It is defined as the total number of triangles present in the network divided by the number of all possible triangles that the network could have [22]. Given a network with n vertices and t triangles, the global clustering coefficient can be computed as $C = \binom{n}{3}^{-1} t$. We use NetworkX's `triangles` function to calculate t.

Characteristic path length. The characteristic path length of a graph $G = (V, E)$, also known as the average path length, is the the average distance between any two pair of nodes in the graph, $l(G) = \sum_{u,v \in V} \frac{d(u,v)}{|V|(|V|-1)}$, where $d(u,v)$ is the shortest path between u and v. A smaller characteristic path length is linked to a higher efficiency in the transfer of information in the graph [23].

We compute the characteristic path length using NetworkX's `average_shortest_path_length` function.

Erdos–Renyi graphs. A $G_{n,p} = (V, E)$ Erdos–Renyi graph [24] is a random graph with $|V| = n$ vertices, where the presence of each edge $(u, v) \; \forall u, v \in V$ is determined to be independent and identically distributed with probability p. By the law of large numbers, as the number of nodes in such random graph tends to infinity, the number of generated edges approaches $\binom{n}{2} p$. The likelihood of generating a network G of n vertices and m edges is given by

$$P(G) = p^m (1-p)^{\binom{n}{2} - m} \tag{2}$$

We compute Erdos–Renyi graphs according to the procedure outlined in Algorithm 1. We use $n = 49$, $m_{initial} = 97$, which is the number of nodes and edges in the network that is the object of our study, and $p = \frac{1}{n} \approx 0.02$. All other random samples—e.g., during the operations that guarantee that the result graph is connected—are from a uniform probability distribution.

Rich-club coefficient. Afterwards, we turn our attention to the rich-club coefficient of a complex network [25], a feature which describes how well connected a set of hub nodes

are to one another and has been shown to influence structural and functional characteristics of networks, including topology, the efficiency of paths and distribution of load [26]. Intuitively, a subnetwork with only rich-club nodes should have more connections than a random network with the same degree and edge distributions.

Algorithm 1 ER graph generation algorithm.

Input: n, $m_{initial}$, p
Output: $G_{n,p}$

$\quad V \leftarrow \{x_1, x_2, \ldots, x_n\}$ $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ ▷ n nodes
$\quad E \leftarrow \varnothing$
$\quad$ **while** $|E| \neq m_{initial}$ **do**
$\qquad u, v \overset{i.i.d\ p}{\sim} V$
$\qquad E \leftarrow E \cup \{(u, v)\}$
$\quad$ **end while**
$\quad$ **while** $G_{n,p}$ not connected **do**
$\qquad e_{discard} \sim E$
$\qquad E \leftarrow E \setminus \{e_{discard}\}$
$\qquad$ c $\leftarrow$ connected_components(E) $\qquad\qquad$ ▷ the function is provided from NetworkX
$\qquad c_1 \leftarrow \varnothing, c_2 \leftarrow \varnothing$
$\qquad$ **while** $c1 = c2$ **do**
$\qquad\quad c_1, c_2 \sim c$
$\qquad$ **end while**
$\qquad u \sim c_1, v \sim c_2$
$\qquad E \leftarrow E \cup \{(u, v)\}$
$\quad$ **end while**

For each degree k in a complex network, given the amount of vertices $N_{>k}$ with degree larger than k, and the amount of edges $E_{>k}$ present between said vertices, the rich-club coefficient is defined as:

$$\phi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k} - 1)} \tag{3}$$

Intuitively, this is the ratio of the amount of possible edges between nodes with degree larger than k ($N_{>k}(N_{>k} - 1)/2$) and the actual amount of edges present in said nodes.

In our study, similar to [27] we normalize the above value with the rich-club coefficient value of a random network, $\phi_{rand}(k)$. As our null-model, we use random networks with the same number of edges and degree distribution as our original graph, generated according to the Maslov and Sneppen (MS) procedure [28]. Specifically, we use the original network as a starting point, then perform $100 * m$ degree-preserving edge swaps, where m is the number of edges in the network.

Unlike [27], however, our $\phi_{rand}(k)$ is the average of $\phi(k)$ for 1000 random networks. To this end, we modified NetworkX's the rich_club_coefficient implementation.

It is worth noting that given the denominator in Equation (3), the rich-club coefficient is only defined for degrees k such that $N_{>k} > 1$. In other words, we can only calculate the coefficient for degrees k such that the subgraph with vertices of degree higher than k has at least two nodes. Thus, given the degree distribution in the MAPK network, which will be further discussed in Section 3.2.1, we can only compute the coefficient for degrees $k = 0, \ldots, 8$, since $N_{>9} = 1$ and $N_{>11} = 0$. Specifically, we compute RCC For $k = 2, \ldots, 8$, as its value for $k = 0$ and $k = 1$ is generally not informative.

2.2. Mapping to Hallmarks of Cancer

Gene ontology (GO) IDs were manually associated with individual hallmarks of cancer (see Table S1). Subsequently, the sets of genes captured by the communities that were identified using the above-mentioned techniques were input to the BiNGO [29],

a Cytoscape plugin that has the ability to analytically infer the over-represented gene ontology biological process terms present in a network, or any of its subgraphs. We applied BiNGO to each identified community configured with a *p*-value threshold of 0.05, using hypergeometric statistical testing and the Benjamini and Hochberg false discovery rate (FDR) correction, to compute statistically significant GO IDs.

The top ten GO IDs that were linked to hallmarks of cancer (using Table S1) were then counted. A community was associated with a hallmark if at least two GO IDs linked to the same hallmark. This allowed us to associate each community to at least two hallmarks.

To further quantify how strongly the communities associate with their hallmarks, we computed a measure of prominence (MOP). The MOP of a hallmark for a given community is defined as the ratio between the number of GO IDs linked to that hallmark in that community, divided by the average number of GO IDs linked to that hallmark across all six communities, i.e., it is given by

$$\frac{N_{C_i}(j)}{\overline{N_C}(j)} \tag{4}$$

where $N_{C_i}(j)$ is the number of GO IDs in community i linked to hallmark j, and $\overline{N_C}(j)$ denotes the average number of GO IDs for hallmark j. Hence, the MOP quantifies how much the occurrence of a community's associated hallmark differs across all communities.

To quantify statistical significance, we conducted for each community and hallmark a one-sided binomial test using the Matlab function "myBinomTest()" implemented by Matthew Nelson [30]. The *p*-values for each community and its dominant hallmarks were calculated using Mathworks Matlab version R2020b.

3. Results

In this section, we highlight the results of applying the methods outlined in the previous section to the mitogen-activated protein kinase (MAPK) network.

3.1. The MAPK Network in Cancer

In this study, we analysed a model of the MAPK network (Figure 1) which is a complex molecular network comprising strongly intertwined signalling pathways [31]. Existing studies extensively report the signalling cascades to be involved in cancer deregulations. However, the precise underlying mechanisms remain elusive. Attempts to better understand the interplay between initial conditions and response include [32] with simulations performed using the *microC* framework.

In this work, we draw upon network analysis methods typically found in complex networks literature to elucidate such mechanisms by means of mapping the observed signalling relationships to cellular physiology. More specifically, we focus on a MAPK network built around the ERK, JNK and p38 cascades, including 232 transcription factors and other proteins and 167 interactions [16]. Grieco et al. [16] built a dynamical model for the network, aiming to study the response of the MAPK network to stimuli, such as growth factors and stress, which lead to the activation of MAPKs and downstream targets, and eventually affect cell fate.

This network was extended to include pathways activated under hypoxia, namely HIF and VEGFA signalling, as previously published by Voukantsis et al. [32]. We used this published version of the network, simply removing the input and output nodes which represents perturbation and cellular states (e.g., proliferation, apoptosis, growth arrest) rather than genes and their interactions. This network is a Boolean network, where all nodes can have one of two states, inactive or active. The nodes represent genes, and in some cases, the proteins encoded by them and their downstream targets, including transcription factors. The interactions between nodes are governed by Boolean rules that represent the relationships of activation and inhibition between genes/proteins.

The network was built based on the available experimental data, where every interaction between nodes is based on one or more published works, as explained by Grieco et al. Despite the fact that this MAPK network was initially applied to bladder cancer, its

main features are generalizable, due to the fact that the reaction map upon which it was built was assembled using data from different cell types in both human and mice [16]. Moreover, we chose this network because the MAPK cascade involvement in cancer is very well established across different tissues including bladder, breast, and particularly triple negative breast cancers, lung and several others. This network plays a big role in cancer deregulation and balance between cell proliferation and cell death, and it is involved in a series of major cellular processes, all of which make it a perfect target to study the hallmarks of cancer from a network analysis perspective.

Overall, the MAPK network we analyse is the result of numerous experiments where the cellular responses to various signalling stimuli were assessed. It describes the signalling cascades observed, including their directionality. Figure 1 visualises this network consisting of 49 nodes and 97 edges.
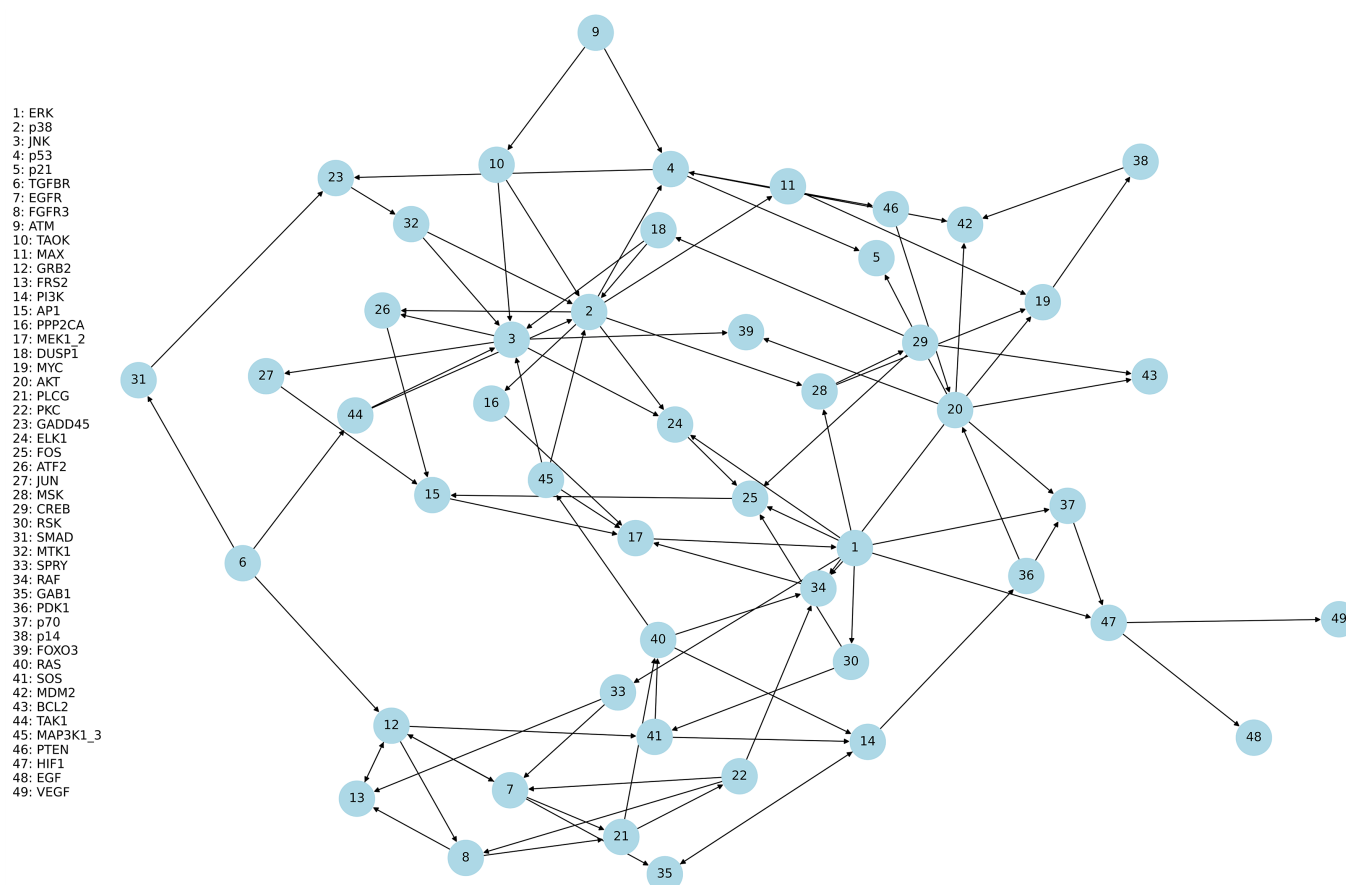

1: ERK
2: p38
3: JNK
4: p53
5: p21
6: TGFBR
7: EGFR
8: FGFR3
9: ATM
10: TAOK
11: MAX
12: GRB2
13: FRS2
14: PI3K
15: AP1
16: PPP2CA
17: MEK1_2
18: DUSP1
19: MYC
20: AKT
21: PLCG
22: PKC
23: GADD45
24: ELK1
25: FOS
26: ATF2
27: JUN
28: MSK
29: CREB
30: RSK
31: SMAD
32: MTK1
33: SPRY
34: RAF
35: GAB1
36: PDK1
37: p70
38: p14
39: FOXO3
40: RAS
41: SOS
42: MDM2
43: BCL2
44: TAK1
45: MAP3K1_3
46: PTEN
47: HIF1
48: EGF
49: VEGF

Figure 1. The Mitogen-Activated Protein Kinase (MAPK) network in cancer, arguably one of the best-established biological networks in cancer, supported by numerous experimental results.

3.2. Topological Organisation of the MAPK Network

Most biological networks are highly non-random. To investigate whether the MAPK network also has distinctive features, we first analysed its topological organisation.

3.2.1. Degree Distribution

First, we consider the number of connections of each node to other nodes in the network, and the probability distribution of this number across the network, also known as its *degree distribution*.

The degree distribution shown in Figure 2 shows that the network structure is non-regular and, like most biological networks, is not in accordance with a so-called Erdos–Renyi (ER) network topology [33]. In particular, the network comprises nodes that are

connected much more than expected from ER network topologies, i.e., hubs [34]. In agreement with the literature, these hubs (e.g., p38, ERK or JNK) have been shown to play central roles in cancer development [35–37].
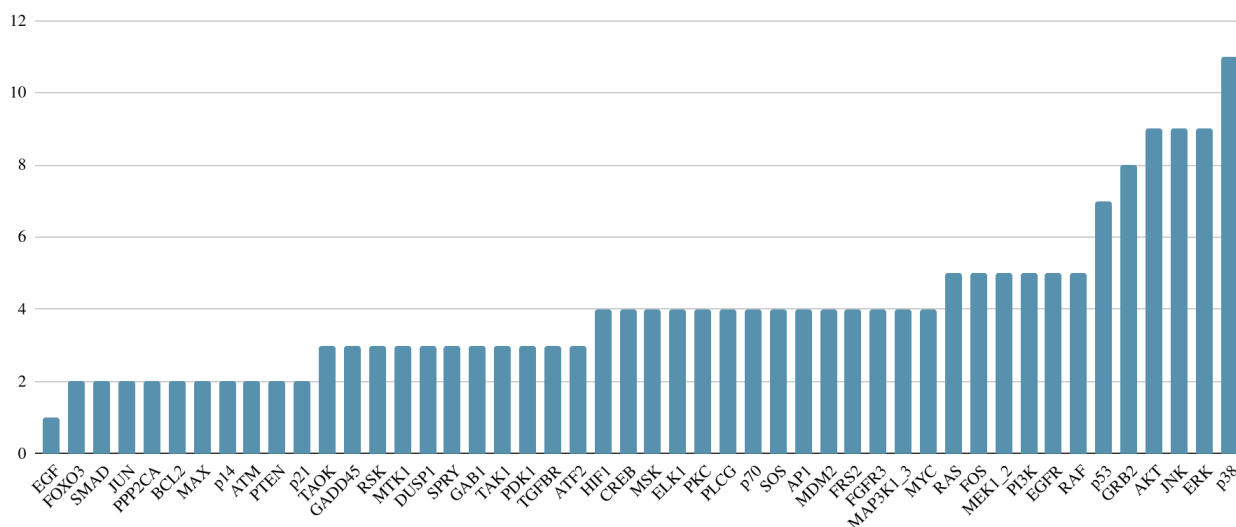
Node Degrees

Figure 2. Degree distribution of the MAPK network. In accordance with well-established literature, highly connected genes (e.g., p38 and ERK) are known to play critical roles in the development of cancer [38,39].

3.2.2. Community Detection

Next, we consider the fundamental complex network feature of "communities". These indicate groups of nodes that are characterised by the presence of numerous connections between them, as compared to fewer links to nodes in other communities [40]. To study whether the MAPK network comprises communities of cross-regulatory genes, we employed the community detection algorithm of Clauset et al. [20].

Indeed, we find that the MAPK network as modelled here comprises a structure that distinguishes itself from random reference networks by the presence of such communities. To this end, Figure 3 visualises the community structure of the MAPK network. In particular, we can observe six such separable communities of nodes. This indicates that the signalling pathways dynamics play out via sets of genes that form communities.

This is in accordance with the organisation of other signalling and gene regulatory networks [41], presenting a general principle for the structure of many biological networks. The genes of the individual communities are listed in Table 1.

Table 1. Identified communities of the MAPK network in cancer.

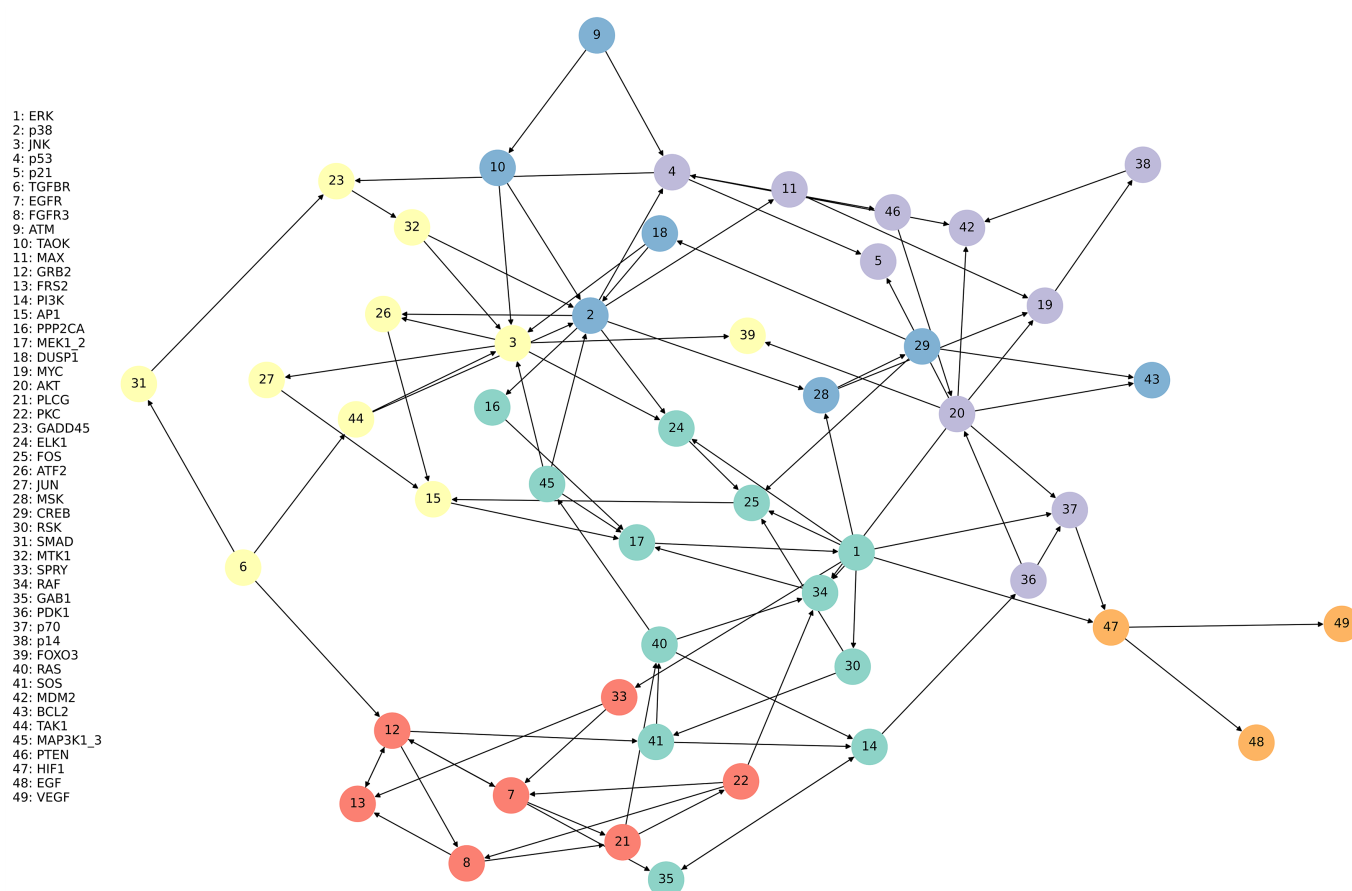| Community	Genes
C1	PPP2CA, MAP2K1, PI3K, RASA1, RAF, GAB1, FOS, RSK MAP3K13, SOS1, ELK1, ERK
C2	GADD45, SMAD2, ATF2, AP1, JUN, TAK1, JNK, MTK1 FOXO3, TGFBR1
C3	P14, RPS6KB1, MAX, MYC, MDM2, PTEN, AKT, P21 PDK1, P54
C4	SPRYD1, FRS2, GRB2, PRKCA, PLCG1, FGFR3, EGFR
C5	CREB, P38, MSK, DUSP1, TAOK1, BCL2, ATM
C6	EGF, VEGF, HIF1

1: ERK
2: p38
3: JNK
4: p53
5: p21
6: TGFBR
7: EGFR
8: FGFR3
9: ATM
10: TAOK
11: MAX
12: GRB2
13: FRS2
14: PI3K
15: AP1
16: PPP2CA
17: MEK1_2
18: DUSP1
19: MYC
20: AKT
21: PLCG
22: PKC
23: GADD45
24: ELK1
25: FOS
26: ATF2
27: JUN
28: MSK
29: CREB
30: RSK
31: SMAD
32: MTK1
33: SPRY
34: RAF
35: GAB1
36: PDK1
37: p70
38: p14
39: FOXO3
40: RAS
41: SOS
42: MDM2
43: BCL2
44: TAK1
45: MAP3K1_3
46: PTEN
47: HIF1
48: EGF
49: VEGF

Figure 3. Community structure of the MAPK network; the six communities that were detected are highlighted in different colours.

3.2.3. Complex Network Measures

In addition to the community structure, we also computed quantifiable, fundamental complex network features. Table 2 shows several well-established network measures, such as characteristic path, modularity, clustering coefficient, and small-worldness (recall Section 2). These measurements further demonstrate the highly non-random topology of the MAPK network. In particular, the characteristic path length denotes the average distance between any two pair of nodes in the network. We find that this value is significantly higher in the MAPK network as compared to matched ER networks. This indicates that the signalling in the MAPK network is surprisingly disconnected. Moreover, the modularity is higher than expected from a comparable, random network. This suggests that the MAPK network is made up of "cliques" of strongly interacting nodes, in accordance with the presence of communities described above. No statistically different organisation with regard to the clustering coefficient was observed, which measures the prevalence of triangular connectivity structure. This is related but different from the previously assessed modularity. Moreover, no non-random small-world organisation was observed, further confirming that communication in the MAPK network is not biased towards efficient communication.

Table 2. Complex network measures of the MAPK network. To study the topological organisation, well-established measures were computed. The analysis confirms a highly non-regular organisation with regard to the characteristic path length and the modularity, as compared to regular ER networks and randomised networks of an identical degree distribution. No significant difference is observed with regard to the clustering-coefficient or the small-worldness.

Measure	Value	ER Network Value	Randomized Network Value
char path	3.928	2.969 (std: 0.073)	2.941 (std: 0.060)
modularity	0.472	0.431 (std: 0.023)	0.425 (std: 0.020)
clustering coefficient	0.001	0.001 (std: 0.03)	0.001 (std: 0.0002)
small-worldness	0.993	0.986 (std: 0.298)	1.173 (std: 0.305)

We also analysed the rich-club organisation, which quantifies the tendency for high-degree nodes to connect with one another. Table 3 shows the rich-club coefficient values for degrees $2, \ldots, 8$ found in the MAPK network (recall the discussion in Section 2). The rich-club coefficients are always lower than 1, with low values for $k = 5$ and 0 for $k = 6, \ldots, 8$. This demonstrates the absence of a rich-club organisation in the MAPK connectome, and indicates that there are fewer links between the highest-degree nodes (hubs) than expected based on random reference networks [34]. The associated Z-scores and p-values indicate that these rich-club coefficient values for high degrees are statistically significant.

Table 3. Rich-club coefficient value for each degree found in the network, normalized using 1000 randomly generated reference networks with the same degree distribution as the original network as reference. The rich-club coefficients at high degrees are lower than expected for reference networks, indicating that the MAPK network's high-degree nodes avoid connections among each other.

Degree	Normalized Value	Z-Score	*p*-Value
2	0.981	−1.325	0.092
3	0.956	−0.936	0.175
4	0.602	−2.342	0.009
5	0.189	−2.754	0.003
6, 7, 8	0.0	−2.401	0.008

3.3. Relationship with Hallmarks of Cancer

We investigated whether the six communities identified via topological analysis (recall Table 1) are related to some of the hallmarks (see also Table 4) by applying BiNGO analysis to each one in order to infer statistically over-represented GO IDs. The resulting gene ontology biological processes identified indicate that individual communities of the network implicate functional annotations that are linked to specific hallmarks of cancer (see Table 5).

Moreover, we quantified the prominence of the hallmarks within a given community, indicating strong partitioning across all communities. To quantify this prominence, we created a "measure of prominence" (MOP). To compute the MOP, we counted for each community the number of genes that were associated with a hallmark of cancer. To this end, the ten most significantly ranked genes associated with hallmarks were identified from the community of genes generated by the BiNGO analysis (output files are available as Supplementary Material). This was done by associating the GO terms of the genes with individual hallmarks (Table S1). We found that the MOPs vary strongly across the communities, and the communities associate with individual hallmarks in a statistically significant way. In other words, individual communities express hallmarks in a highly non-random fashion.

Table 4. The eight hallmarks of cancer.

Hallmark	Shorthand
Self-sufficiency in growth signals	H1
Insensitivity to anti-growth signals	H2
Evading apoptosis	H3
Limitless replicative potential	H4
Sustained angiogenesis	H5
Tissue invasion and metastasis	H6
Reprogramming energy metabolism	H7
Evading immune response	H8

Table 5. Identified communities of the MAPK network in cancer entail network modules in accordance with specific hallmarks of cancer. Individual network communities are associated with subsets of hallmarks. The (relative) prominence of the hallmarks in individual communities is quantified using the MOP value. $\star$ indicates that the probability of observing the frequency of associations with a given hallmark is statistically significant ($p < 0.05$ for one-sided binomial test). Overall, the network topology reflects a compartmentalised and modular organisation with regard to the enabling mechanisms of cancer.

Community	Example GO Descriptions	Hallmarks (MOP)
C1	reg. of apoptosis, phosphorus metabolic process, chemotaxis	H3 (1.5), H6 (4.5$\star$), H7 (1.13)
C2	regulation of apoptosis, pos. reg. of metabolic process	H3 (1.5), H7 (2.63$\star$)
C3	response to extracellular stimulus, reg. of cell cycle reg. of cell cycle arrest	H2 (6.0), H4 (6.0$\star$)
C4	embryonic organ development, tissue development, phosphate metabolic process	H1 (4.8$\star$), H7 (0.75)
C5	reg. of apoptosis reg. of programmed cell death, phosphate metabolic process	H3 (3.0$\star$), H7 (1.5$\star$)
C6	reg. of cell proliferation pos. reg. of angiogenesis primitive heopoiesis	H1(1.2), H5 (5.15$\star$)

4. Discussion and Concluding Remarks

In this work, we made steps towards characterising the topology of the MAPK network, with a central role in cancer, and mapping its organisation to cellular physiology. To this end, we demonstrated that the MAPK network connectome exhibits topological features that are in accordance with existing data. We analysed the network features and related these to physiological behaviours. We found that the structure of the MAPK network renders it highly robust to modifications by external stimuli to small numbers of nodes. In particular, the characteristic path length (recall Table 2) is significantly higher than in comparable reference networks. This entails that signals between two nodes tend to require the crossing of many nodes. This indicates challenges from a clinical perspective, where the controllability of network dynamics is a desired feature. In other words, the MAPK network topology is better suited for interventions where multiple nodes are controlled.

Moreover, a rich-club organisation is lacking (recall Table 3). Rich-club organisation has been demonstrated in numerous biological networks, including neural networks [34]

and brain networks [42], where its presence increases the robustness of the network function to breakdowns of individual nodes. Again, the lack of such a rich-club organisation has clinically relevant implications. The nodes with the highest number of links tend to avoid each other, and so it is unlikely that controlling single nodes can have a large impact on the entire network dynamics. Hence, our analysis suggests that there exist "sweet-spots" comprising multiple nodes for disrupting the regulatory dynamics of cancer progression. It remains to be investigated how such a clinical interference could be realised, taking into account the impact of drugs on gene expression, in combination with knowledge of the underlying connectome.

We also provided a mapping of genetic communities or cliques to hallmarks of cancer. This is of clinical relevance because the removal of a single hallmark could be sufficient to significantly alter the fatality of cancer, as indicated by the difference of one hallmark between benign and malignant cancers [43]. Although the relationship between the hallmarks of cancer and cellular function are still debated [44], they constitute an established conceptual framework for understanding major cellular dynamics in cancer. However, it is currently unclear how these hallmarks are grounded in biological networks. This is a crucial question, because gaining insights into how the hallmarks of cancers are enabled by cell-intrinsic dynamics would revolutionise oncology: a causal understanding for the origins and progression of cancer would help identify suitable treatment options, potentially circumventing side-effects and increasing the likelihood of a cure.

Indeed, previous works study potential treatment options that exploit the impact of drugs on cellular pathways with relevance to hallmarks of cancer [45]. Our analysis confirms that the MAPK network is organised in accordance with the hallmarks of cancer. This agrees with the origin of this network, which was modelled from a collection of several cancer studies. In particular, our results show that the identified communities are prominently associated with subsets of hallmarks of cancer. This suggests that the conditions necessary for the progression of cancer are distributed in a modular and complementary fashion. This result is in agreement with the study of [41], where functional modules in breast cancer also display a compartmentalisation of specific hallmarks of cancer.

Of note, in our analysis, none of the communities was as strongly associated with the hallmark of "evading immune response" (H8). This is expected, as this network was modelled mainly using cell line experiments, which do not include this element.

Although this study is limited to a specific MAPK signalling model, it provides a demonstration of analysis of structural characterization of biological networks, and what such an approach can offer. Our analysis highlights that MAPK signalling is organised in a highly non-random and non-symmetrical way. The topological structure of this network exhibits a long characteristic path length, and the absence of rich-club organisation. Both features render the network dynamics robust to external perturbations where only a single node is affected. Furthermore, our results provide a mapping of genetic communities or cliques to hallmarks of cancer. The complex network-based analysis indicates that processes critical for the progression of cancer are leveraging a modular network structure, rendering it resistant to localised intervention. Hence, an orchestrated attack on multiple genes may be necessary to interrupt a given hallmark during cancer progression.

Cancer computational models, such as the one we present here, have long proven extremely useful for fast generation and testing of biological hypotheses, even in cases where the biological knowledge is incomplete (as it is often the case with genetic regulation) by using qualitative methods. They can point to probable avenues of treatment, and help provide mechanistic explanations to processes that were obscured to current experimental techniques. When combined with experimental validation, they remain an essential tool in oncology [46].

In future, we anticipate more uptake of computational approaches as used here, in combination with experimental results, to formulate quantitative models with high explanatory power. For instance, the study of Dessauges et al. [47] aligns well with our work in that it highlights the potential of the coordinated targeting of the MAPK network at

multiple locations to better impact oncogenic signalling. The employment of computational and complex network modelling can be used to incorporate such experimental evidence from different sources into a single, integrated framework for hypotheses generation and validation. Along those lines, future clinical applications are likely to benefit from taking into account the topological and functional organisation of cellular signalling cascades, including the MAPK network, in order to identify "sweet spots" for treatments. In particular, hub nodes and connectors between communities are promising candidates.

This leads to a key problem: identifying control nodes, or a set of *drivers* [48], which exert stronger influence, compared to other nodes in regulatory processes. Recent results in the structural controllability of complex networks [49] and their actionable application to network control (e.g., see [50]) may be pertinent to identifying *sweet spots*, which can then be used as the starting points for steering the dynamics of cellular signalling cascades in a computationally tractable manner.

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [CrossRef] [PubMed]
2. Hanahan, D.; Weinberg, R.A. The hallmarks of cancer. *Cell* **2000**, *100*, 57–70. [CrossRef]
3. Hanahan, D.; Weinberg, R.A. Hallmarks of cancer: The next generation. *Cell* **2011**, *144*, 646–674. [CrossRef] [PubMed]
4. Frost, J.; Pienta, K.; Coffey, D. Symmetry and symmetry breaking in cancer: A foundational approach to the cancer problem. *Oncotarget* **2018**, *9*, 11429. [CrossRef]
5. Bauer, R.; Kaiser, M.; Stoll, E. A computational model incorporating neural stem cell dynamics reproduces glioma incidence across the lifespan in the human population. *PLoS ONE* **2014**, *9*, e111219. [CrossRef]
6. de Montigny, J.; Iosif, A.; Breitwieser, L.; Manca, M.; Bauer, R.; Vavourakis, V. An in silico hybrid continuum-/agent-based procedure to modelling cancer development: Interrogating the interplay amongst glioma invasion, vascularity and necrosis. *Methods* **2021**, *185*, 94–104. [CrossRef]

7.   Axenie, C.; Bauer, R.; Martínez, M.R. The Multiple Dimensions of Networks in Cancer: A Perspective. *Symmetry* **2021**, *13*, 1559. [CrossRef]
8.   Albert, R.; Othmer, H.G. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster. *J. Theor. Biol.* **2003**, *223*, 1–18. [CrossRef]
9.   Karlsen, M.R.; Moschoyiannis, S. Evolution of control with learning classifier systems. *Appl. Netw. Sci.* **2018**, *3*, 30. [CrossRef]
10.   Papagiannis, G.; Moschoyiannis, S. Learning to Control Random Boolean Networks: A Deep Reinforcement Learning Approach. In *Complex Networks 2019*; Studies in Computational Intelligence; Springer: Cham, Swltzerland, 2019; Volume 881, pp. 721–734.
11.   Acernese, A.; Yerudkar, A.; Glielmo, L.; Del Vecchio, C. Reinforcement Learning Approach to Feedback Stabilization Problem of Probabilistic Boolean Control Networks. *IEEE Control Syst. Lett.* **2021**, *5*, 337–342. [CrossRef]
12.   Papagiannis, G.; Moschoyiannis, S. Deep Reinforcement Learning for Control of Probabilistic Boolean Networks. In *Complex Networks 2020*; Springer: Cham, Switzerland, 2020; Volume 944, pp. 361–371.
13.   Huang, S.; Ernberg, I.; Kauffman, S. Cancer attractors: A systems view of tumors from a gene network dynamics and developmental perspective. *Semin. Cell Dev. Biol.* **2009**, *20*, 869–876. [CrossRef] [PubMed]
14.   Jaeger, J.; Monk, N. Bioattractors: Dynamical systems theory and the evolution of regulatory processes. *J. Physiol.* **2014**, *592*, 2267–2281. [CrossRef] [PubMed]
15.   Dhillon, A.S.; Hagan, S.; Rath, O.; Kolch, W. MAP kinase signalling pathways in cancer. *Oncogene* **2007**, *26*, 3279–3290. [CrossRef] [PubMed]
16.   Grieco, L.; Calzone, L.; Bernard-Pierrot, I.; Radvanyi, F.; Kahn-Perles, B.; Thieffry, D. Integrative modelling of the influence of MAPK network on cancer cell fate decision. *PLoS Comput. Biol.* **2013**, *9*, e1003286. [CrossRef]
17.   Costa, L.d.; Rodrigues, F.A.; Travieso, G.; Boas, P.R.V. Characterization of complex networks: A survey of measurements. *Adv. Phys.* **2007**, *56*, 167–242. [CrossRef]
18.   Hagberg, A.; Swart, P.; Chult, D.S. *Exploring Network Structure, Dynamics, and Function Using NetworkX*; Technical Report; Los Alamos National Lab. (LANL): Los Alamos, NM, USA, 2008.
19.   Cherifi, H.; Palla, G.; Szymanski, B.; Lu, X. On community structure in complex networks: Challenges and opportunities. *Appl. Netw. Sci.* **2019**, *4*, 117. [CrossRef]
20.   Clauset, A.; Newman, M.E.; Moore, C. Finding community structure in very large networks. *Phys. Rev. E* **2004**, *70*, 066111. [CrossRef]
21.   Watts, D.; Strogatz, S. Collective dynamics of 'small-world' networks. *Nature* **1998**, *393*, 440–442. [CrossRef]
22.   Luce, R.D.; Perry, A.D. A method of matrix analysis of group structure. *Psychometrika* **1949**, *14*, 95–116. [CrossRef]
23.   Latora, V.; Marchiori, M. Efficient behavior of small-world networks. *Phys. Rev. Lett.* **2001**, *87*, 198701. [CrossRef]
24.   Erdos, P.; Rényi, A. On Random Graphs I. *Publ. Math.* **1959**, *6*, 290–297.
25.   Zhou, S.; Mondragón, R.J. The rich-club phenomenon in the Internet topology. *IEEE Commun. Lett.* **2004**, *8*, 180–182. [CrossRef]
26.   Csigi, M.; Kőrösi, A.; Bíró, J.; Heszberger, Z.; Malkov, Y.; Gulyás, A. Geometric explanation of the rich-club phenomenon in complex networks. *Sci. Rep.* **2017**, *7*, 1730. [CrossRef] [PubMed]
27.   Colizza, V.; Flammini, A.; Serrano, M.A.; Vespignani, A. Detecting rich-club ordering in complex networks. *Nat. Phys.* **2006**, *2*, 110–115. [CrossRef]
28.   Maslov, S.; Sneppen, K. Specificity and stability in topology of protein networks. *Science* **2002**, *296*, 910–913. [CrossRef]
29.   Maere, S.; Heymans, K.; Kuiper, M. BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics* **2005**, *21*, 3448–3449. [CrossRef]
30.   Nelson, M. mybinomtest. 2022. Available online: https://www.mathworks.com/matlabcentral/fileexchange/24813-mybinomtest-s-n-p-sided (accessed on 27 April 2022).
31.   Krishn, M.; Narang, H. The complexity of mitogen-activated protein kinases (MAPKs) made simple. *Cell. Mol. Life Sci.* **2008**, *65*, 3525–3544. [CrossRef]
32.   Voukantsis, D.; Kahn, K.; Hadley, M.; Wilson, R.; Buffa, F.M. Modeling genotypes in their microenvironment to predict single- and multi-cellular behavior. *GigaScience* **2019**, *8*, giz010. [CrossRef]
33.   Erdos, P.; Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci* **1960**, *5*, 17–60.
34.   Bauer, R.; Kaiser, M. Nonlinear growth: An origin of hub organization in complex networks. *R. Soc. Open Sci.* **2017**, *4*, 160691. [CrossRef]
35.   Wagner, E.F.; Nebreda, Á.R. Signal integration by JNK and p38 MAPK pathways in cancer development. *Nat. Rev. Cancer* **2009**, *9*, 537–549. [CrossRef] [PubMed]
36.   Bubici, C.; Papa, S. JNK signalling in cancer: In need of new, smarter therapeutic targets. *Br. J. Pharmacol.* **2014**, *171*, 24–37. [CrossRef] [PubMed]
37.   Maik-Rachline, G.; Hacohen-Lev-Ran, A.; Seger, R. Nuclear ERK: Mechanism of translocation, substrates, and role in cancer. *Int. J. Mol. Sci.* **2019**, *20*, 1194. [CrossRef] [PubMed]
38.   Martínez-Limón, A.; Joaquin, M.; Caballero, M.; Posas, F.; de Nadal, E. The p38 pathway: From biology to cancer therapy. *Int. J. Mol. Sci.* **2020**, *21*, 1913. [CrossRef]
39.   Liu, F.; Yang, X.; Geng, M.; Huang, M. Targeting ERK, an Achilles' Heel of the MAPK pathway, in cancer therapy. *Acta Pharm. Sin. B* **2018**, *8*, 552–562. [CrossRef]
40.   Newman, M.E. The structure and function of complex networks. *SIAM Rev.* **2003**, *45*, 167–256. [CrossRef]

41. de Anda-Jáuregui, G.; Alcalá-Corona, S.A.; Espinal-Enríquez, J. Functional and transcriptional connectivity of communities in breast cancer co-expression networks. *Appl. Netw. Sci.* **2019**, *4*, 22. [CrossRef]
42. Kim, D.J.; Min, B.K. Rich-club in the brain's macrostructure: Insights from graph theoretical analysis. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 1761–1773. [CrossRef]
43. Lazebnik, Y. What are the hallmarks of cancer? *Nat. Rev. Cancer* **2010**, *10*, 232–233. [CrossRef]
44. Fouad, Y.A.; Aanei, C. Revisiting the hallmarks of cancer. *Am. J. Cancer Res.* **2017**, *7*, 1016.
45. Buckley, A.M.; Lynam-Lennon, N.; O'Neill, H.; O'Sullivan, J. Targeting hallmarks of cancer to enhance radiosensitivity in gastrointestinal cancers. *Nat. Rev. Gastroenterol. Hepatol.* **2020**, *17*, 298–313. [CrossRef] [PubMed]
46. Victori, P.; Buffa, F.M. The many faces of mathematical modelling in oncology. *Br. J. Radiol.* **2018**, *92*, 20180856. [CrossRef] [PubMed]
47. Dessauges, C.; Mikelson, J.; Dobrzyński, M.; Jacques, M.A.; Frismantiene, A.; Gagliardi, P.A.; Khammash, M.; Pertz, O. Optogenetic actuator/ERK biosensor circuits identify MAPK network nodes that shape ERK dynamics. *bioRxiv* **2022**. [CrossRef]
48. Gibbs, D.L.; Shmulevich, I. Solving the influence maximization problem reveals regulatory organization of the yeast cell cycle. *PLoS Comput. Biol.* **2017**, *13*, e1005591. [CrossRef]
49. Liu, Y.Y.; Slotine, J.J.; Barabási, A.L. Controllability of complex networks. *Nature* **2011**, *473*, 167. [CrossRef]
50. Moschoyiannis, S.; Elia, N.; Penn, A.; Lloyd, D.J.B.; Knight, C. A web-based tool for identifying strategic intervention points in complex systems. Proc. Games for the Synthesis of Complex Systems (CASSTING @ ETAPS). *EPTCS* **2016**, *220*, 39–52. [CrossRef]