INTERFACE

rsif.royalsocietypublishing.org

Research

Cite this article: Wang G, Su H, Yu H, Yuan R, Zhu X, Ao P. 2016 Endogenous network states predict gain or loss of functions for genetic mutations in hepatocellular carcinoma. *J. R. Soc. Interface* **13**: 20151115.
http://dx.doi.org/10.1098/rsif.2015.1115

Received: 30 December 2015
Accepted: 3 February 2016

Subject Category:
Life Sciences–Mathematics interface

Subject Areas:
computational biology, systems biology, medical physics

Keywords:
genetic mutations, endogenous network theory, dynamical systems, stable state

Author for correspondence:
Ping Ao
e-mail: aoping@sjtu.edu.cn

Electronic supplementary material is available at http://dx.doi.org/10.1098/rsif.2015.1115 or via http://rsif.royalsocietypublishing.org.

THE ROYAL SOCIETY PUBLISHING

Endogenous network states predict gain or loss of functions for genetic mutations in hepatocellular carcinoma

Gaowei Wang[1], Hang Su[1], Helin Yu[1], Ruoshi Yuan[1], Xiaomei Zhu[2] and Ping Ao[1,3]

[1]Ministry of Education Key Laboratory of Systems Biomedicine, Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai, China
[2]GenMath, Seattle, WA, USA
[3]State Key Laboratory for Oncogenes and Related Genes, Shanghai Cancer Institute, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Cancers have been typically characterized by genetic mutations. Patterns of such mutations have traditionally been analysed by *posteriori* statistical association approaches. One may ponder the possibility of *a priori* determination of any mutation regularity. Here by exploring biological processes implied in a mechanistic theory recently developed (the endogenous molecular–cellular network theory), we found that the features of genetic mutations in cancers may be predicted without any prior knowledge of mutation propensities. With hepatocellular carcinoma (HCC) as an example, we found that the normal hepatocyte and cancerous hepatocyte can be represented by robust stable states of one single endogenous network. These stable states, specified by distinct patterns of expressions or activities of proteins in the network, provide means to directly identify a set of most probable genetic mutations and their effects in HCC. As the key proteins and main interactions in the network are conserved through cell types in an organism, similar mutational features may also be found in other cancers. This analysis yielded straightforward and testable predictions on accumulated and preferred mutation spectra in normal tissue. The validation of predicted cancer state mutation patterns demonstrates the usefulness and potential of a causal dynamical framework to understand and predict genetic mutations in cancer.

1. Background

Cancer is often characterized by abundant genetic mutations and, in fact, is assumed to be disease owing to these mutations [1–5]. Work from several cancer genomics studies, such as The Cancer Genome Atlas (TCGA), has provided the complete DNA sequence of large numbers of cancers [6–8]. Detailed analyses of such big data revealed that the mutation frequency of each gene is usually different. In some cases, there is even no mutation for the known oncogenes or tumour-suppressing genes [9]. Nevertheless, key features of genetic mutations, including within the top mutated genes in a large number of cancer samples, have been documented [10]. Recent studies have shown that these key features were influenced by factors such as chromatin organization [11], selection pressures [12], carcinogens [13] and environmental conditions [14]. However, these factors are inadequate to completely explain the key features of genetic mutations in cancer [15,16]. An understanding of the underlying cause of these key features should provide us with a deeper understanding of cancer genesis [17]. Here, using hepatocellular carcinoma (HCC) as an example, we examined factors that could influence the key features of genetic mutation from a theoretical point of view, and the predictions were compared against experiments.

The 'endogenous molecular–cellular network' theory states that functions of complex biological processes can be understood via an analysis of the

network dynamics [18,19]. Such molecular–cellular networks have been shaped by evolution and its core structure, and main properties are conserved through both species and biological processes [20–22]. Specifically, the salient properties include that (i) biological systems are built by a set of essential modules and cross-talks between these modules [23]; (ii) each module can be further specified by a set of key molecular agents, namely proteins [24] and (iii) key proteins and causal biochemical interactions between them via well-documented gene regulation and signalling transduction form a network [24]. Given the dominancy of feedback among these proteins, the network is closed and autonomous [25]. Finally, a quantitative description of the endogenous network consists of a set of coupled differential equations that can generate many local stable states with obvious or non-obvious biological functions [18,19,26–30].

In this work, based on the core endogenous molecular–cellular network for hepatocytes [18,19,31], we found that the normal hepatocyte and cancerous hepatocyte can be represented by two robust stable states of the core network. Relationship between genetic mutations and biological function can be analysed in this context. Here, we focused on the stable states of the network, as these are the most robust part of the network. Particular genetic mutations that were associated with the cancerous hepatocyte stable state and confer selective advantages to establish and maintain cancerous hepatocyte stable states were then predicted to have higher mutation frequency in HCC. Subsequent inspection of experimental data demonstrated agreement with this prediction: the results reproduced key features of the top mutated genes in HCC. As most of the proteins and interactions in the core network are conserved in many other cell types, we predicted and validated that the key features of the genetic mutations in HCC will be found in other cancers. Moreover, this analysis yielded straightforward and testable predictions, including that there are accumulated and preferred mutation spectra in normal tissue. Overall, this work illustrates the usefulness of network-level analyses as a means to predict and understand genetic mutations in cancers.

2. Results

2.1. Normal and cancerous hepatocytes are robust stable states of the core endogenous network

A set of essential modules, including the cell cycle, apoptosis, metabolism, hepatocyte-specific function, cell adhesion, immune response and angiogenesis, were selected to capture the essential features of hepatocytes [32]. According to the current understanding of these modules at a molecular level, each module was simplified and specified by a set of proteins that play key roles in each module (table 1) [24,32]. Interactions between the proteins, activation/upregulation or inhibition/downregulation, were summarized from well-documented gene regulation and signalling transduction. These proteins and their causal interactions formed the core endogenous molecular–cellular network of the hepatocyte, which integrated our current understanding of hepatocyte behaviour at a molecular level (figure 1 and electronic supplementary material, table S1). The core endogenous network of hepatocytes included 37 proteins and 114 interactions. It should be emphasized that these proteins were selected based on their normal biological functions without any prior knowledge of their mutation propensity in cancer.

Table 1. Key proteins of essential modules and pathways. The 37 agents, proteins here, distributed among eight modules and pathways.

essential modules and pathways	key proteins
cell cycle	Rb, E2F, cyclin D-CDK4/6, cyclin E-CDK2
apoptosis	Casp3, Bad, Casp8, Casp9/Cytc, IAP, Bcl-2/xL, Bax, Fas
metabolism	HIF, Akt, PTEN
hepatocyte-specific function	HNF4α, C/EBPα, Foxa2
cell adhesion	snail, slug, E-cadherin, integrin
immune response	NF-κB (KC), IL-1, IL-6, TNFα
angiogenesis	VEGF
pathways	EGF, Ras, Erk, Myc, TGF-β, NF-κB, P53, β-catenin, GSK-3β, P21

A quantitative description of such networks provides a way to reveal emergent properties that are difficult to understand by intuitive reasoning alone [33]. Therefore, the network was described quantitatively using a set of coarse-grained ordinary differential equations (electronic supplementary material, Methods). Because of the lack of information on kinetic parameters, the dynamical equations are doubly normalized to reduce such dependence [34]: the maximum strength of each interaction is normalized to 1, and the maximum expression level or activity of each gene or protein is normalized to 1. While this normalization procedure makes the model lose some detailed descriptive power, it gains some advantages. For example, the relative expression level of each gene becomes a useful and predictive indicator, to be explored below. Algorithms were designed to calculate stable states of the differential equations (electronic supplementary material, Methods). By using different equation forms and different parameter values in the equations (electronic supplementary material, table S2), four robust stable states which were specified by distinct activation/expression levels of the proteins in the network were obtained (electronic supplementary material, table S3–S12). Again, it should be stressed that those states are new predictions, because their descriptions are not in the biological data used to construct the network.

Theoretically, these robust stable states may represent obvious or non-obvious biological functions. When comparing with clinical observation and experimental data, we found that two robust stable states from the model reproduced the main known features of the normal hepatocyte and cancerous hepatocyte at the modular and molecular level (electronic supplementary material, tables S13 and S14, more detailed information in our previous work [31]). As the core endogenous network was constructed using knowledge obtained independently and the stable states of the complex nonlinear dynamical systems were not obvious before analysis, it is striking that a single network can

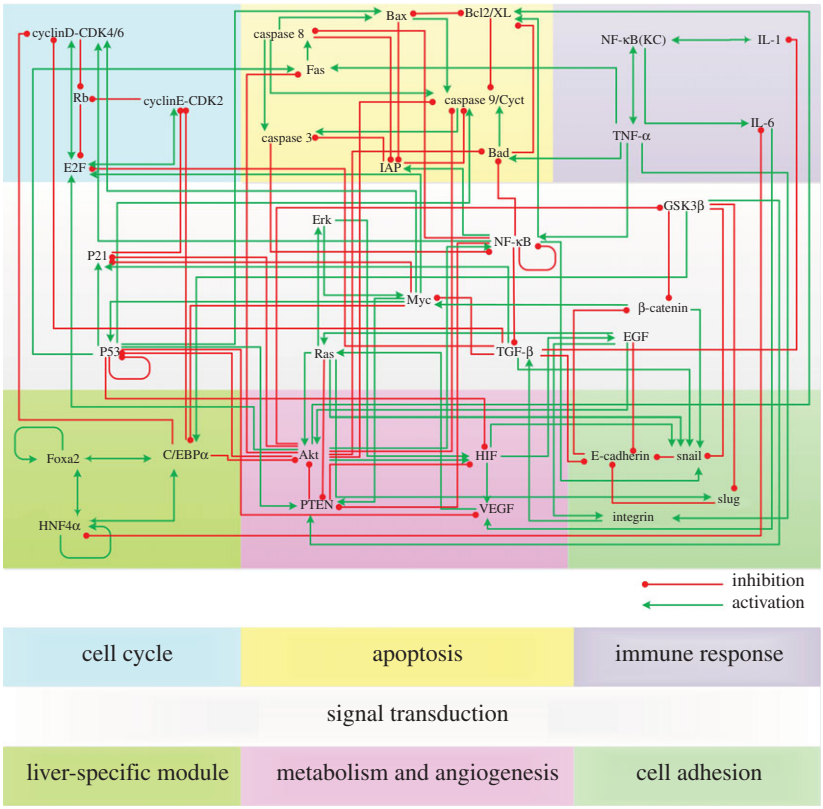

Figure 1. The core endogenous molecular–cellular network of hepatocytes. Cell cycle, apoptosis, metabolism, hepatocyte-specific function, cell adhesion, immune response and angiogenesis modules were selected to capture the essential features of hepatocytes. Each module was simplified and specified by a set of key proteins. Interactions between the proteins, activation or inhibition, were summarized from well-documented and conserved gene regulatory network and signaling transduction pathways. The core endogenous network of hepatocytes included 37 key proteins and 114 interactions. (Online version in colour.)

reproduce key features of both normal and cancerous hepatocytes, among other predictions.

2.2. Prediction of top mutated genes in cancers

It has been known that dynamical equations can be used to predict the effect of mutations. For example, a detailed analysis of mutations against experiments has been performed for the core regulatory network of phage lambda genetic switch [35]. Here, we examined the molecular-level details of these two stable states not at this dynamical level, but at the relative expression level, which is less sensitive to kinetic parameters. We first characterized the activity of each protein in a given stable state as activated or inactivated by setting a threshold: if the activity of a protein is greater than this threshold, we identified the protein as active, and if lower, the protein was identified as inactive. Activated proteins in each stable state were then highlighted to reveal subnetworks that are expected to play key roles in the establishment and maintenance of the stable state. Because the activity of each protein was normalized to range from 0 (minimal activation) to 1 (full activation), the threshold can be selected within a reasonable range (from 0.4 to 0.6). We found that these thresholds will not affect the main conclusions. Figure 2 showed the subnetworks of the normal hepatocyte and cancerous hepatocyte when we set the threshold to 0.4 (electronic supplementary material, table S3).

These subnetworks then provided means to directly identify a few key features of the genetic mutation patterns in HCC. Biologically, genetic mutations can have varying effects on the function of protein. Mutations that confer enhanced activity were defined as gain-of-function mutations, whereas those that reduce or abolish protein function were defined as loss-of-function mutations. It was known that some random mutations in cancers were selected for and accumulated in response to phenotypic consequences [36–38]. We reasoned that proteins which were inactive in the normal hepatocyte stable state and activate in the cancerous hepatocyte stable state could have a higher probability of undergoing gain-of-function mutations in cancers, as gain-of-function of this gene can adapt to the cancerous hepatocyte stable state and confer selective advantages to establish and maintain the cancerous hepatocyte stable state. For example, as shown in electronic supplementary material, table S3, the normalized activity of cyclin D-CDK4 in the normal hepatocyte stable state and cancerous hepatocyte stable state was 0.0000 and 0.8399, respectively. Accordingly, cyclin D-CDK4 was inactive in the normal hepatocyte stable state and active in the cancerous hepatocyte stable state, and so cyclin D-CDK4 was identified to have a higher probability of undergoing a gain-of-function mutation in HCC. Similarly, proteins that were active in the normal hepatocyte stable state and inactivate in the cancerous hepatocyte stable state were expected to have a higher probability of undergoing loss-of-function mutations in HCC. For example, Rb was active in the normal hepatocyte stable state and inactive in the cancerous hepatocyte stable state, and so Rb was identified to have a higher probability of undergoing a loss-of-function mutation in HCC. In this way, we identified probable genetic mutations in HCC (table 2). It should be noted that there were

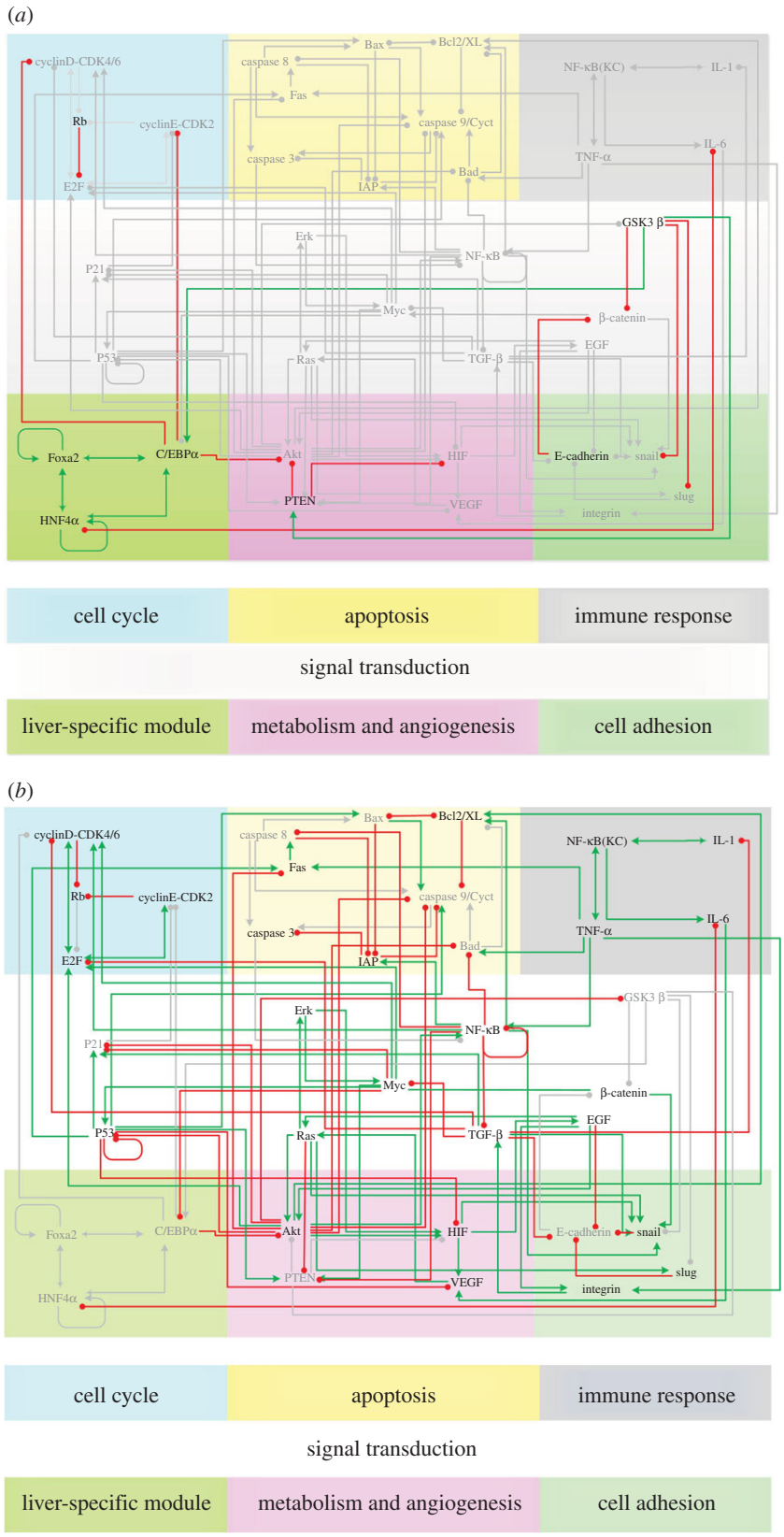

Figure 2. Subnetworks of normal hepatocyte and cancerous hepatocyte stable states. Activated proteins and interactions are highlighted in bold in the normal hepatocyte stable state (*a*) and the cancerous hepatocyte stable state (*b*) to form different subnetworks. (Online version in colour.)

also six proteins whose activity did not significantly differ between the normal hepatocyte stable states and cancerous hepatocyte stable states (table 2). For example, Bax was inactive in both the normal hepatocyte stable state and cancerous hepatocyte stable state. Using this approach, we could not decide the probable mutation of these genes.

2.3. Model prediction reproduces key features of cancer genetic mutation data

We next compared these predicted mutated genes with well-documented genetic mutation data from the Catalogue of Somatic Mutations in Cancer (COSMIC) [39]. Given the

Table 2. Prediction of top mutated genes in cancerous hepatocytes from the model. There are 24 proteins that were predicted to have gain-of-function mutations and seven proteins that have loss-of-function mutations. The present approach could not decide the mutation type of the reminding six genes (—).

gene name	model result
cyclin D-CDK4/6	gain-of-function
cyclin E-CDK2	gain-of-function
Rb	loss-of-function
E2F	gain-of-function
C/EBPα	loss-of-function
Foxa2	loss-of-function
HNF4α	loss-of-function
Fas	gain-of-function
Bcl-2/xL	gain-of-function
IAP	gain-of-function
Bax	—
Bad	—
Casp 9/Cytc	—
Casp 8	—
Casp 3	—
E-cadherin	loss-of-function
snail	gain-of-function
slug	gain-of-function
integrin	gain-of-function
Akt	gain-of-function
PTEN	loss-of-function
HIF	gain-of-function
TNFα	gain-of-function
IKK- NF-κB	gain-of-function
IKK-NF-κB(KC)	gain-of-function
IL-1	gain-of-function
IL-6/stat	gain-of-function
EGF	gain-of-function
VEGF	gain-of-function
Ras	gain-of-function
ERK	gain-of-function
GSK3β	loss-of-function
β-catenin	gain-of-function
Myc	gain-of-function
P53	gain-of-function
TGF-β	gain-of-function
P21	—

heterogeneity of mutations in different HCC patients, and even different regions or cells of the same HCC patient, a set of 20 top mutated genes in a large number of HCC samples was chosen to reflect key features of genetic mutation data. Four genes (TP53 (P53), CTNNB1 (β-catenin), RB1, PTEN) can be linked to proteins in the core network directly, whereas seven genes (AXIN1, CDKN2A, PIK3CA, HNF1A, ATM, CREBBP and IL6ST) can be linked to proteins in the core network indirectly, as they have well-defined relationships with proteins in the core network according to the Kyoto Encyclopaedia of Genes and Genomes (KEGG; relationships are summarized in electronic supplementary material, table S16). These 11 proteins that can be linked to proteins in the core network directly or indirectly were classified into category 1 in table 3; the remaining nine proteins which have not been considered in the present network were classified into category 2 in table 3. Overall, we found that 11 of the top 20 mutated genes in HCC were included in our theoretical analysis that was based on a 37 node network constructed with a totally different set of experimental data.

As most of the proteins and interactions in the core network of hepatocytes are conserved in other cell types, the results obtained in the present analysis of HCC are expected to be applicable to other cancers. As such, we also investigated the top 20 mutated genes in biliary tract cancer, bone cancer, breast cancer, central nervous cancer, eye cancer, prostate cancer, skin cancer, small intestine cancer, soft tissue cancer, stomach cancer [10]. Genes that can be linked to proteins in the core network directly or indirectly were classified into category 1 in table 3 (relationships are summarized in electronic supplementary material, table S17). Proteins that were not considered in the present network were classified into category 2 in table 3. Strikingly, we found that 10, 13, 9, 10, 13, 12, 11, 13, 13 and 13 of the top 20 mutated genes in these different cancers can be explained by the present core network of the hepatocyte, respectively. Thus, the inherent core network structure, including the relative activity of the constituent proteins, appears to capture common features also relevant to the development of cancer in other cell types. It should be emphasized that these proteins were selected without any prior knowledge of their mutation propensity in cancer.

Moreover, the present model generated the types of aberration, gain-of-function or loss-of-function, of these probable mutational genes (table 2). We noted that experiments have identified the type of aberration of some mutations in HCC [40] and of these, six proteins were in the core network and five agreed with model predictions (table 4). A similar summary of the type of aberration of some mutations in other cancer types [41] reveals that eight proteins were in the core network and seven agreed with model predictions (table 4). This overall agreement further supports the significant potential of this analysis to predict genetic mutations in cancer. However, it should be mentioned that there is one disagreement between the model and the literature. It was well known that p53 has a loss-of-function mutation in many cancers, whereas p53 was predicted to have a gain-of-function mutation in our model (table 2). This disagreement may be owing to one of two sources. First, given the heterogeneous nature of cancer mutations and the current incomplete network, experimental results of certain genes which do not appear to fit this model are expected. Second, the aforementioned results were obtained with a threshold as 0.4. We found that p53 was one of three genes whose results are sensitive to threshold values (electronic supplementary material, table S15). Thus, the behaviour of p53 may be more complex than presently believed, as some recent studies have suggested [42].

2.4. Further testable predictions

Our analysis also affords two additional intriguing and testable predictions. First, our model suggests that there are mutations

Table 3. Classification of the top 20 mutated genes in different cancers. The first column denotes different cancer types. Genes in category 1 can be linked to proteins in the core network directly or indirectly. Genes in category 2 are not considered in the present network. The last column denotes the coverage rate.

cancer type	category 1	category 2	
HCC	TP53, CTNNB1, AXIN1, CDKN2A, PIK3CA, HNF1A, ATM, CREBBP, RB1, IL6ST, PTEN	TERT, ARID1A, ARID2, KMT2C, NFE2L2, KMT2D, PTPRB, TSC2, SMARCA4	11/20
biliary tract	TP53, KRAS, CDKN2A, SMAD4, AXIN1, CTNNB1, PIK3CA, BRAF, CDH1, PTEN	MLL3, BAP1, IDH1, AR1D1A, PBRM1, TERT, FBXW7, RNF43, IDH2, GNAS	10/20
bone	TP53 ,CDKN2A, RB1, CTNNB1, BRAF, AKT1, APC, KIT, NRAS, FGFR2, KRAS, HRAS, EGFR	GNAS, IDH1, PTCH1, IDH2, SMARCB1, CDC73, SMO	13/20
breast	PIK3CA, TP53, CDH1, PTEN, AKT1, RB1, ATM, NF1, APC	MLL3, GATA3, ARID1A, MED12, KMT2D, RUNX1, AKAP9, MAP2K4, UBR5, MYH9, BRCA1	9/20
central nervous	TP53, PTEN, CDKN2A, CTNNB1, EGFR, BRAF, PIK3CA, NF1, RB1, PIK3R1	IDH1, TERT, SMARCB1, H3F3A, ATRX, CIC, CHEK2, PTCH1, KMT2D, SMARCA4	10/20
eye	GNA11, RB1, BRAF, TP53, PTEN, KIT, NRAS, CDKN2A KRAS, EGFR, PDGFRA, MET, CTNNB1	GNAD, BAP1, SF3B1, TERT, BCOR, FBXW7, DICER1	13/20
prostate	TP53, PTEN, KRAS, EGFR, CTNNb1, HRAS, ATM, APC, RB1, TRRAP, PIK3CA, BRAF	MLL3, FOXA1, KMT2D, MLL, MED12, AKAP9, MLLT3, KDM6A	12/20
skin	BRAF, TP53, CDKN2A, FGFR3, NRAS, HRAS, PTEN, PIK3CA, KIT, CTNNB1, MAP2K1	TERT, PTCH1, CYLD, ARID2, ROS1, MLL3, NF, RAC1, GNAQ	11/20
intestine	KRAS, TP53, APC, SMAD4, CTNNB1, PIK3CA, BRAF, EGFR, KDR, NRAS, ATM, CDKN2A, PDGFRA	GNAS, MEN1, FBXW7, PTPN11, STK11, SMARCB1, ERBB2,	13/20
soft tissue	KIT, CTNNB1, VHL, PDGFRA, TP53, CDKN2A, APC, KRAS, HRAS, NRAS, PTEN, PIK3CA, BRAF	SMARCB1, NF2, MED12, NF, TERT, MEN1, GNAS	13/20
stomach	TP53, CDH1, APC, PIK3CA, CTNNB1, KRAS, TRRAP, CDKN2A, AXIN1, PTEN, EGFR, PDGFRA, KDR	ARID1A, MSH6, FBXW7, RNF43, NSD1, ERBB2, GNAS,	13/20

Table 4. Main gain-of-function or loss-of-function mutations in HCC and other cancer types. There are six and eight well-documented gain-of-function or loss-of-function mutations in HCC [40] and other cancer types [41], respectively. Gain-of-function mutations are marked in bold and loss-of-function mutations underlined. P53 does not agree with model prediction and is given in italics.

HCC	cancers
PTEN	**Cyclin D-CDK4/6**
IL-6/stat3	Rb
VEGF	C/EBPα
Ras	E-cadherin
β-catenin	PTEN
P53	**EGF**
	Ras
	P53

that can confer selective advantages to establish and maintain the normal hepatocyte phenotype: a preferred mutation spectrum in the normal hepatocyte (electronic supplementary material, table S18). There is some evidence showing that cells and tissues can maintain their normal phenotype in the face of myriad mutated genes [43]. It should be biologically interesting to determine whether there would be such a mutation spectrum in normal liver. Second, while we have used this model to predict mutations in cancer successfully as shown by above analysis, it showed that the normal hepatocyte and cancerous hepatocyte are endogenous stable states of one single endogenous network. This indicated that there are cancers, especially at the early stage, which can take place without major genetic alterations such as the well-documented oncogenes and tumour suppressor genes. Indeed, there is evidence supporting the existence of mutation-free cancer from different standpoints [9,44]. If firmly established, the cancer genesis and procession is mechanistically completely different from that of the cancer mutation theory.

3. Discussion

In the present core endogenous network of hepatocytes, three transcription factors were used to capture the key features of hepatocytes and 34 genes/proteins were used to capture essential modules and pathways in cells, such as the cell cycle, apoptosis, metabolism, cell adhesion, immune response and angiogenesis (table 1). These 37 genes/proteins and the interactions between them are usually conserved in different human cell types [24]. Normal hepatocytes and cancerous hepatocytes were found to be represented by two robust stable states of the core network. We wish to point out that this network has been greatly simplified. For example, there are many more functional modules than the modules used in the model, other molecular–cellular agents such as microRNA and metabolites were not considered, and oscillatory activities such as cell cycle processes were simplified by a binary switch. It is expected that a more extensive network will reproduce more of these detailed features through the inclusion of more modules and proteins.

The probable genetic mutations in HCC were predicted from an analysis of the two stable states, showing good agreement with key features of the genetic mutations data of HCC. We

focused here on the stable states of the network, as these are the most robust part of the network. We should note that the expressions or activities of proteins could be affected by multiple factors that are different from genetic alterations such as epigenetics and feedback regulations. At an initial stage, we predicated mutations from protein activities in cancer versus normal stable states in a deterministic context. Other important details such as the order of mutations or cooperative effects of mutations could be incorporated in further modelling. These issues require an explicit inclusion of stochastic effects, where the potential energy landscape can be used to explore more detailed issues [35,45]. Nonetheless, without any *a priori* knowledge of genetic mutation propensity in HCC, our results show that such network-level analyses are indeed a powerful approach to enable the prediction and a better understanding of genetic mutations in HCC. The core endogenous network was also approximated by the Boolean dynamics to check the internal consistency of modelling. We found that the probable genetic mutations predicted from Boolean analysis were consistent with results obtained from coarse-grained ordinary differential equations in table 2 at the present stage of modelling. We should note that Boolean analysis cannot be used to explore some detailed issues such as the results in electronic supplementary material, table S15. As most of the proteins and interactions in the core network of hepatocytes are conserved in other cell types, the results obtained in the present analysis of HCC are expected to be applicable to other cancers. Therefore, the observation that there are common features of genetic mutations in different cancers is evident, likely reflecting the conserved nature of the core network. This analysis also yielded straightforward and testable predictions, including that there are mutation-free cancers and a preferred mutation spectrum in normal tissue. Overall, this work illustrates the usefulness of network-level analyses as a means to predict and understand genetic mutations in cancers.

The implications of this work indicate that cancer is more complex and less determined by mutations than previously thought. During the past several decades, the prevailing cancer somatic mutation theory assumed that cancer is caused by genetic alterations. However, the present model shows that normal and cancerous hepatocytes are endogenous stable states of a single endogenous network, which indicates that cancers can develop by mechanisms other than genetic alterations. In its strongest form, it indicates that there are cancers, especially at early stage, that can arise without major genetic alterations. The development of cancer can be compared with the development of a multicellular organism [24]. The genome is statistically identical in all cells; the cells differ not because they contain different genetic information, but because they express different sets of genes and with different expression levels.

4. Methods

4.1. Construction of core endogenous network of hepatocytes

A set of essential modules, including the cell cycle, apoptosis, metabolism, hepatocyte-specific functions, cell adhesion, immune response and angiogenesis, were selected to capture the key features of the hepatocyte. Each module is simplified and specified by a set of key proteins at a molecular level (table 1). These proteins and their causal interactions (electronic supplementary material, table S1) form the core endogenous network of the hepatocyte (figure 1).

4.2. Quantitative description and analysis of the core endogenous network of hepatocytes

The core endogenous network of hepatocytes (figure 1) was quantified by a set of coarse-grained ordinary differential equations in a standard way (detailed description and full equation in electronic supplementary material, Methods). In this modelling framework, we focused on the properties that are determined by the logical and topological architecture of the network, rather than the precise form of the equation and parameter values. Therefore, the microscopic details underlying the forms of equation and parameter values are not required to be overly precise. We used two independent algorithms to calculate the stable states of the nonlinear dynamical system (detailed description in electronic supplementary material, Methods), and obtained consistent results from the independent algorithms.

Data accessibility. The top 20 genetic mutations in different cancers (table 3) were obtained from COSMIC (http://cancer.sanger.ac.uk/cosmic/). In table 4, well-documented genetic mutations in HCC are obtained from a recent review of HCC [40], and the mutations summarized from other cancer are obtained from a recent review [41]. The relationship between the top 20 mutated proteins and proteins in the core network are obtained from KEGG (http://www.kegg.jp/).

Authors' contributions. G.W. conceived the study, carried out the analysis, drafted the manuscript; H.S., H.Y., R.Y., X.Z. participated in data analysis, helped draft the manuscript; P.A. conceived the study, coordinated the study and drafted the manuscript. All authors gave final approval for publication.

Competing interests. We declare no competing interests.

Funding. This work was supported in part by the National Basic Research Programme of China no. 2010CB529200 (P.A.), Natural Science Foundation of China no. NFSC91029738 (P.A.).

Acknowledgements. We thank Site Li and Junqiang Wang for comments and discussions, Dr Daniel Czajkowsky for critical reading and editing the manuscript.

References

1. Boveri T, Boveri M. 1929 *The origin of malignant tumors*. Baltimore, MD: Williams and Wilkins Company.
2. Nowell P. 1976 The clonal evolution of tumor cell populations. *Science* **194**, 23–28. (doi:10.1126/science.959840)
3. Lengauer C, Kinzler KW, Vogelstein B. 1998 Genetic instabilities in human cancers. *Nature* **396**, 643–649. (doi:10.1038/25292)
4. Hanahan D, Weinberg RA. 2011 Hallmarks of cancer: the next generation. *Cell* **144**, 646–674. (doi:10.1016/j.cell.2011.02.013)
5. Greaves M, Maley CC. 2012 Clonal evolution in cancer. *Nature* **481**, 306–313. (doi:10.1038/nature10762)
6. Weinstein JN *et al.* 2013 The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* **45**, 1113–1120. (doi:10.1038/ng.2764)
7. Burrell RA, McGranahan N, Bartek J, Swanton C. 2013 The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501**, 338–345. (doi:10.1038/nature12625)
8. Totoki Y *et al.* 2014 Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. *Nat. Genet.* **46**, 1267–1273. (doi:10.1038/ng.3126)
9. Versteeg R. 2014 Cancer: tumours outside the mutation box. *Nature* **506**, 438–439. (doi:10.1038/nature13061)

10. Forbes SA *et al.* 2009 COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Res.* **38**(Suppl. 1), D652–D657. (doi:10.1093/nar/gkp995)
11. Schuster-Böckler B, Lehner B. 2012 Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488**, 504–507. (doi:10.1038/nature11273)
12. Pollard KS *et al.* 2006 Forces shaping the fastest evolving regions in the human genome. *PLoS Genet.* **2**, e168. (doi:10.1371/journal.pgen.0020168)
13. Westcott PMK *et al.* 2014 The mutational landscapes of genetic and chemical models of *Kras*-driven lung cancer. *Nature* **517**, 489–492. (doi:10.1038/nature13898)
14. Yun J *et al.* 2009 Glucose deprivation contributes to the development of *KRAS* pathway mutations in tumor cells. *Science* **325**, 1555–1559. (doi:10.1126/science.1174229)
15. Shendure J, Akey JM. 2015 The origins, determinants, and consequences of human mutations. *Science* **349**, 1478–1483. (doi:10.1126/science.aaa9119)
16. Martincorena I, Campbell PJ. 2015 Somatic mutation in cancer and normal cells. *Science* **349**, 1483–1489. (doi:10.1126/science.aab4082)
17. Stratton MR. 2011 Exploring the genomes of cancer cells: progress and promise. *Science* **331**, 1553–1558. (doi:10.1126/science.1204040)
18. Ao P, Galas D, Hood L, Zhu X. 2008 Cancer as robust intrinsic state of endogenous molecular–cellular network shaped by evolution. *Med. Hypotheses* **70**, 678–684. (doi:10.1016/j.mehy.2007.03.043)
19. Wang GW, Zhu X, Hood L, Ao P. 2013 From phage lambda to human cancer: endogenous molecular–cellular network hypothesis. *Quant. Biol.* **1**, 32–49. (doi:10.1007/s40484-013-0007-1)
20. Waddington CH. 1957 The strategy of the genes: a discussion of some aspects of theoretical biology. London, UK: George Allen & Unwin, Ltd.
21. Wright S. 1982 The shifting balance theory and macroevolution. *Annu. Rev. Genet.* **16**, 1–20. (doi:10.1146/annurev.ge.16.120182.000245)
22. Haldane JBS. 1990 *The causes of evolution*. Princeton, NJ: Princeton University Press.
23. Hartwell LH, Hopfield JJ, Leibler S, Murray AW. 1999 From molecular to modular cell biology. *Nature* **402**(6761 Suppl.), C47–C52. (doi:10.1038/35011540)
24. Alberts B *et al.* 2007 *Molecular biology of the cell*, 5th edn. New York, NY: Garland Science Taylor & Francis Group.
25. Noble D. 2008 Claude Bernard, the first systems biologist, and the future of physiology. *Exp. Physiol.* **93**, 16–26. (doi:10.1113/expphysiol.2007.038695)
26. Kauffman S. 1993 *The origins of order: self organization and selection in evolution*. Oxford, UK: Oxford University Press.
27. Bar-Yam Y, Harmon D, de Bivort B. 2009 Attractors and democratic dynamics. *Science* **323**, 1016–1017. (doi:10.1126/science.1163225)
28. Huang S, Ernberg I, Kauffman S. 2009 Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Semin. Cell Dev. Biol.* **20**, 869–876. (doi:10.1016/j.semcdb.2009.07.003)
29. Zhang J *et al.* 2014 TGF-β-induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. *Sci. Signal.* **7**, ra91. (doi:10.1126/scisignal.2005304)
30. Davila-Velderrain J, Martínez-García J, Alvarez-Buylla ER. 2015 Modeling the epigenetic attractors landscape: towards a post-genomic mechanistic understanding of development. *Front. Genet.* **6**, 160. (doi:10.3389/fgene.2015.00160)
31. Wang GW *et al.* 2014 Quantitative implementation of the endogenous molecular–cellular network hypothesis in hepatocellular carcinoma. *Interface Focus* **4**.
32. Monga SPS. 2010 *Molecular pathology of liver diseases*. Berlin, Germany: Springer.
33. Tyson JJ, Chen KC, Novak B. 2003 Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* **15**, 221–231. (doi:10.1016/S0955-0674(03)00017-6)
34. Ao P, Galas D, Hood L, Yin L, Zhu XM. 2010 Towards predictive stochastic dynamical modeling of cancer genesis and progression. *Interdiscip. Sci. Comput. Life Sci.* **2**, 140–144. (doi:10.1007/s12539-010-0072-3)
35. Zhu XM *et al.* 2004 Calculating biological behaviors of epigenetic states in the phage λ life cycle. *Funct. Integr. Genomics* **4**, 188–195. (doi:10.1007/s10142-003-0095-5)
36. Rosenberg SM. 2001 Evolving responsively: adaptive mutation. *Nat. Rev. Genet.* **2**, 504–515. (doi:10.1038/35080556)
37. Cairns J, Overbaugh J, Miller S. 1988 The origin of mutants. *Nature* **335**, 142–145. (doi:10.1038/335142a0)
38. Cairns J. 1998 Mutation and cancer: the antecedents to our studies of adaptive mutation. *Genetics* **148**, 1433–1440.
39. Bamford S *et al.* 2004 The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Brit. J. Cancer* **91**, 355–358. (doi:10.1038/sj.bjc.6601894)
40. Nault J-C, Zucman-Rossi J. 2014 Genetics of hepatocellular carcinoma: the next generation. *J. Hepatol.* **60**, 224–226. (doi:10.1016/j.jhep.2013.08.019)
41. Rahman N. 2014 Realizing the promise of cancer predisposition genes. *Nature* **505**, 302–308. (doi:10.1038/nature12981)
42. Prives CL. 2014 Abstract PL01-02: The two faces of p53: tumor suppressor and oncogene. *Cancer Res.* **74**(Suppl. 19), PL01-02-PL01-02.
43. Rubin H. 2006 What keeps cells in tissues behaving normally in the face of myriad mutations? *Bioessays* **28**, 515–524. (doi:10.1002/bies.20403)
44. Wang X *et al.* 2011 Residual embryonic cells as precursors of a Barrett's-like metaplasia. *Cell* **145**, 1023–1035. (doi:10.1016/j.cell.2011.05.026)
45. Yuan RS, Ao P. 2012 Beyond Itô versus Stratonovich. *J. Stat. Mech. Theory Exp.* **2012**, P07010. (doi:10.1088/1742-5468/2012/07/P07010)