

## Research



**Cite this article:** Yuan R *et al.* 2017 Beyond cancer genes: colorectal cancer as robust intrinsic states formed by molecular interactions. *Open Biol.* **7**: 170169. <http://dx.doi.org/10.1098/rsob.170169>

Received: 10 July 2017

Accepted: 6 October 2017

**Subject Area:**  
systems biology

**Keywords:**  
colorectal cancer, systems biology, robust dynamical states, endogenous molecular–cellular network, stochastic nonlinear dynamics

### Authors for correspondence:

Ping Ao

e-mail: [aoping@sjtu.edu.cn](mailto:aoping@sjtu.edu.cn)

Xiaomei Zhu

e-mail: [xiaomeizhu@yahoo.com](mailto:xiaomeizhu@yahoo.com)

<sup>†</sup>Present address: Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA.

Electronic supplementary material is available online at <http://dx.doi.org/10.6084/m9.figshare.c.3911893>.

# Beyond cancer genes: colorectal cancer as robust intrinsic states formed by molecular interactions

Ruoshi Yuan<sup>1,†</sup>, Suzhan Zhang<sup>2,3</sup>, Jiekai Yu<sup>2,3</sup>, Yanqin Huang<sup>2,3</sup>, Demin Lu<sup>2,3</sup>, Runtan Cheng<sup>1</sup>, Sui Huang<sup>4</sup>, Ping Ao<sup>1,5</sup>, Shu Zheng<sup>2,3</sup>, Leroy Hood<sup>4</sup> and Xiaomei Zhu<sup>1,5</sup>

<sup>1</sup>Key Laboratory of Systems Biomedicine, Ministry of Education, Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China

<sup>2</sup>Key Laboratory of Cancer Prevention and Intervention, Chinese Ministry of Education, Key Laboratory of Molecular Biology in Medical Sciences, Hangzhou, Zhejiang Province 310009, People's Republic of China

<sup>3</sup>Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310009, People's Republic of China

<sup>4</sup>Institute for Systems Biology, 401 Terry Ave. N., Seattle, WA 98109-5234, USA

<sup>5</sup>Shanghai Center of Quantitative Life Sciences, Shanghai University, Shanghai 200444, People's Republic of China

**id** RY, 0000-0002-9508-7326; PA, 0000-0001-7109-628X; XZ, 0000-0002-3674-3565

Colorectal cancer (CRC) has complex pathological features that defy the linear-additive reasoning prevailing in current biomedicine studies. In pursuing a mechanistic understanding behind such complexity, we constructed a core molecular–cellular interaction network underlying CRC and investigated its nonlinear dynamical properties. The hypothesis and modelling method has been developed previously and tested in various cancer studies. The network dynamics reveal a landscape of several attractive basins corresponding to both normal intestinal phenotype and robust tumour subtypes, identified by their different molecular signatures. Comparison between the modelling results and gene expression profiles from patients collected at the second affiliated hospital of Zhejiang University is presented as validation. The numerical ‘driving’ experiment suggests that CRC pathogenesis may depend on pathways involved in gastrointestinal track development and molecules associated with mesenchymal lineage differentiation, such as Stat5, BMP, retinoic acid signalling pathways, Runx and Hox transcription families. We show that the multi-faceted response to immune stimulation and therapies, as well as different carcinogenesis and metastasis routes, can be straightforwardly understood and analysed under such a framework.

## 1. Introduction

Colorectal cancer (CRC) is a leading cause of cancer deaths in the USA [1] and around the world [2,3]. The current research on combined molecular-targeting agents [4], the need for better risk models to incorporate genetic, lifestyle and environmental effects [5], and the development for early detection method [6] all require a better understanding at the molecular level. Ideally, a causal and quantitative model may be used as a ‘dry-experiment’ platform to recapitulate carcinogenesis and metastasis routes, and to test efficacy of drug combinations. Towards such a goal, we investigated CRC under the framework that cancer is a robust state(s) evolutionarily formed from the underpinning endogenous molecular–cellular interaction network [7,8]. The method has been developed and tested in recent years [9–12]. The model construction and analysis are presented in this work. We show that the model indeed captures the essential features of CRC complexity. The nonlinearity of the interaction network dynamics is directly responsible for the formation of the robust CRC subtypes and their differential responses to intervention.

CRC is a clinically heterogeneous disease. About 40% of patients with CRC develop liver metastasis at the time of presentation, with approximately 20% presenting as synchronous metastasis and the remaining 20% as metachronous metastasis [13–15]. Classification schemes are researched to associate clinical outcome with both genetic and phenotypic biomarkers, such as KRAS/BRAF mutations [16], overall mutation frequency [17], mesenchymal character [18], methylation [19] and gene expression profiles [20]. For example, patients who differ in microsatellite instability, which accounts for 7–20% of all patients, appear to have a different prognosis and benefit from chemotherapy [18]. However, these tumour classification schemes have intrinsic difficulties in themselves, such as the choice and comparison across and within different patient sets used in association studies [21]. In addition, the classification schemes do not directly translate molecular heterogeneity of CRC into the disease progression and response to therapies mechanistically. A better understanding for CRC heterogeneity is still needed. Nevertheless, a remarkable finding of a possible robust classification system, by analysing six independent classification systems and their convergence into four consensus molecular subtypes [22], suggests that CRC may have intrinsic robust subtypes regardless of approaching angles. This observation of robust molecular subtypes is consistent with our hypothesis that cancer corresponds to intrinsic robust states formed by an endogenous molecular–cellular interaction network [7].

The hypothesis of an endogenous molecular–cellular interaction network for cancer has been tested in hepatocellular [9], prostate [10] and gastric cancers [11], and acute promyelocytic leukaemia [12]. The major assumption is that the collective effects of interactions among molecular/cellular agents lead to robust states including normal physiological states and cancer subtypes. Carcinogenesis may be understood as a transition from a normal state to cancer state(s). The effect of mutations is naturally incorporated [23]. The idea of collective emergence in biology goes back to Waddington and Delbruck, followed by Monod and Jacob, Hinshelwood and Kauffman [24–29]. High-dimensional attractor states in gene expression patterns and their functional robustness have meanwhile been experimentally demonstrated [30,31].

Genetic and epigenetic studies have identified dozens of genes and molecular pathways, which could possibly have causal relation to CRC. Large-scale molecular profiling, on the other hand, has identified thousands of molecules differentially expressed in CRC and its progression [32,33]. As constructing a network that includes all the genes known to be differentially expressed in CRC and their regulatory molecules is not practical and may not be necessary to provide a picture for the mechanism, we focused on the ones which may be casual to CRC pathology. In addition to well-known molecules and pathways in carcinogenesis, such as p53, Myc, Wnt and NF- $\kappa$ B [34–36], we included molecules important in embryonic development, notably of the gastrointestinal (GI) track and mesenchymal lineage. The feedbacks from intra- and extracellular signalling as well as other molecular interactions are collected from molecular biology and biochemical low-throughput experiments designed to be valid under normal physiological conditions.

The network dynamics simulation reveals the robust states, their transition paths and the associated drives by the orchestrated activities of key regulatory molecules. The calculated robust heterogeneity is validated by the reported phenotypes in the literature and microarray experiments. The implication

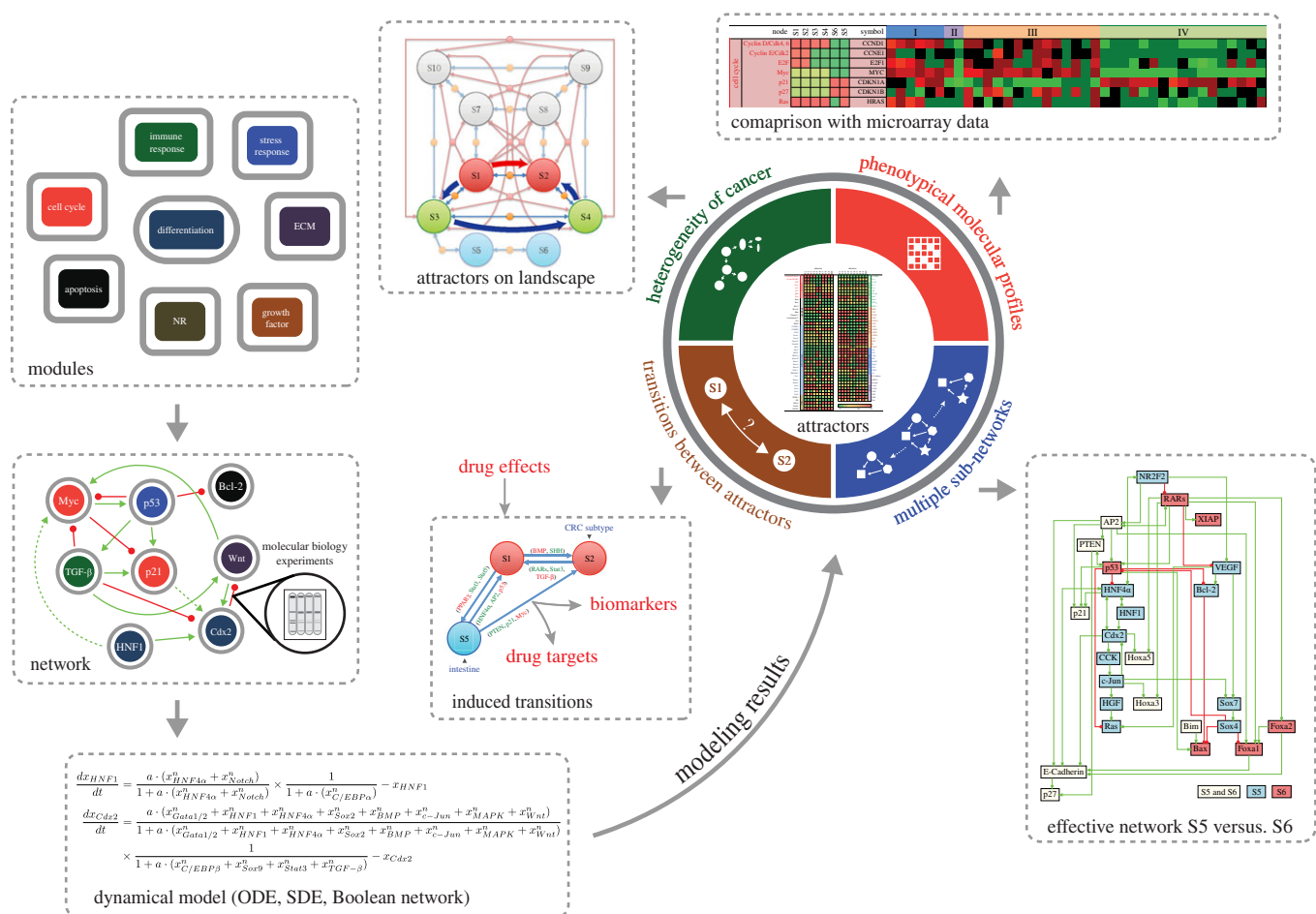
of such a roadmap in carcinogenesis and metastasis potential is considered in the Discussion section. The intrinsic landscape of CRC may also shed new light on the debate about the stimulating/inhibitory effect of immune response to cancer. Future clinical and biological studies are needed to fully validate these predictions and explore for therapeutic purposes.

## 2. Results

### 2.1. Molecular–cellular network model for colorectal cancer

The molecular–cellular network we constructed is aimed to be a consolidated and simplified core network for studying CRC development (figure 1). Some of the nodes in the network represent pathways, such as Notch, SHH and Wnt signalling pathways. Cascading effects within these pathways are included in the same manner. Similarly, EGFR, VEGF represent pathway activities, not the expression level of these proteins. Growth factor feedback loops are assumed to be self-contained within the tumour tissue. For example, although VEGF receptors are primarily expressed on endothelial cells, tumour cells might also express them [37]. Production of cytokines is not assumed to be from tumour cells but from the tumour tissue, and emphasis in defining interactions is on receptors/ligands and crosstalk between signalling pathways. Importantly, the model involves cell–cell interactions because of the feedback from cytokines and extracellular matrix. To keep the model simple, only representative molecules of given functional modules are included. For example, for cell cycle regulation, only Cyclin E and Cyclin D are included. The nodes and edges of the CRC network that represent molecules, pathways and their interactions are listed in electronic supplementary material, table S1. Cell cycle, apoptosis, growth factor signalling, differentiation, immune response, stress response, angiogenesis and metastasis are incorporated in the model. The interactions among molecules are obtained from the literature, with preference given to interactions with solid biochemical basis. The constructed network is presented in electronic supplementary material, table S1, with detailed references in electronic supplementary material, table S2.

The network construction is hypothesis-driven in several respects. First, a number of important molecules and pathways identified in previous genetic and molecular biological studies (such as APC and KRAS), with a causal relationship to CRC development, are included as nodes in the network. Second, we assumed that there are intrinsic molecular feedback loops of these genes or molecules. We looked into developmental programmes for these feedback loops. There are studies showing that overactive pathways can lead to pathological conditions. For example, the Wnt pathway, which is active in the developmental process, induced a programme of intestinal genes in the developing lung [38]. Third, we hypothesized that feedback loops mediated by immune response also might have a developmental link in haematopoiesis. Development of leucocyte is orchestrated by the comprehensive regulation of inflammation-related molecular programmes. In addition, physiological inflammation in response to external stimulation dissolves after withdrawal of stimuli, while the developmental programme presumably requires self-sustained feedbacks to quarantine



**Figure 1.** Schematics of endogenous molecular–cellular network construction and modelling. We started with a minimal core network representing regulation of basic cellular functions, such as cell cycle, apoptosis and stress response, similar to previous cancer models [9–12]. Molecules and molecular pathways specific for GI track development and functions, such as transcription factors Cdx2, HNF1, glucocorticoids signalling pathways, were added to the minimal core network. The molecular interactions were collected from the literature, with priority given to those verified by molecular biology experiments. Feedback loops related to inflammation and hematopoiesis were also included. Dynamical system equations (described in electronic supplementary material) were used to compute the attractor states generated by the defined network structure, as well as saddle points for spontaneous transitions between attractors. Random parameter tests were performed to demonstrate robustness of the obtained results. Comparison of gene activity profiles predicted by the attractors with microarray data validated the modelling. Specifically, CRC subtypes as well as normal intestinal phenotype corresponded to the attractors of network dynamics.

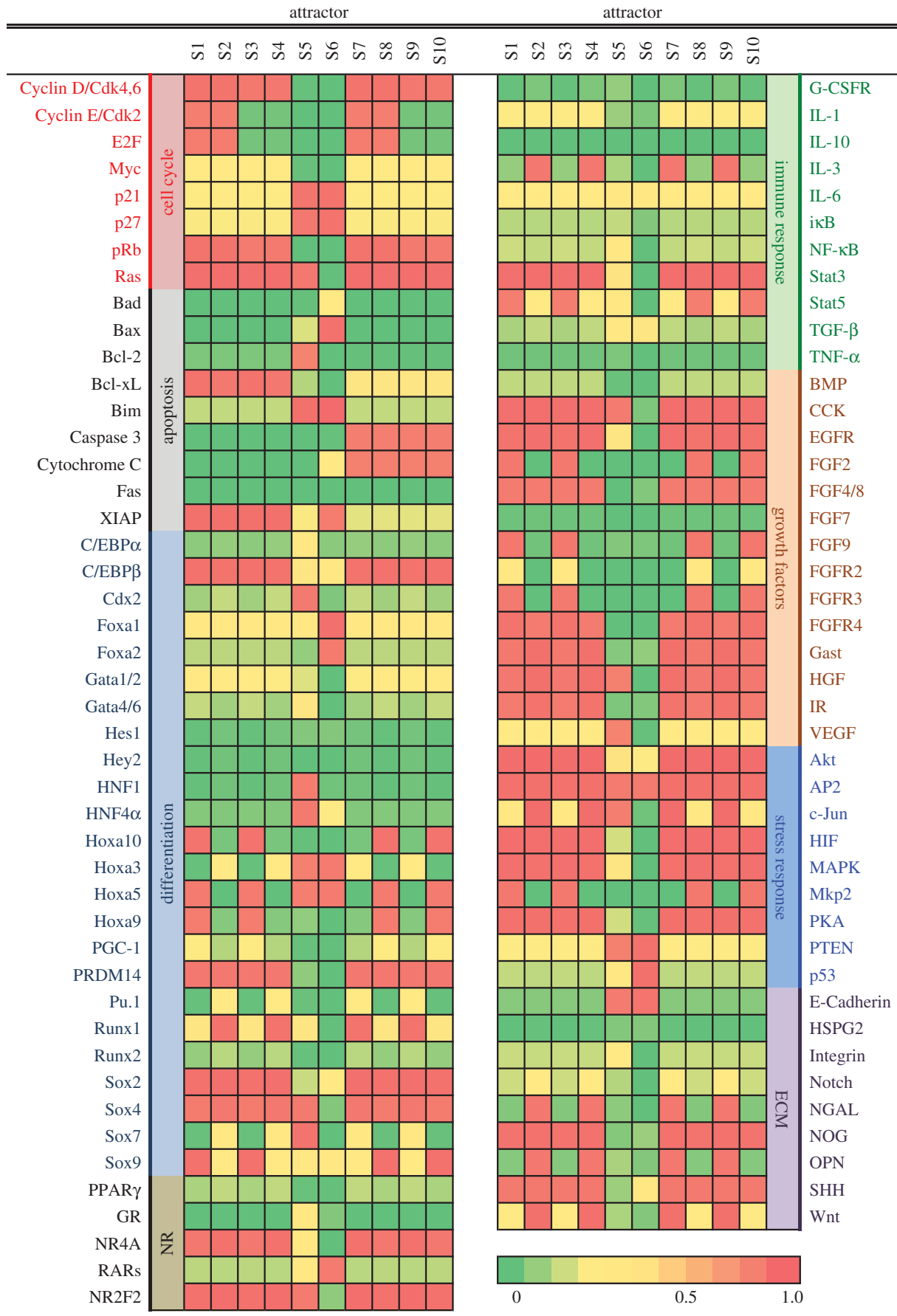
robustness. More details can be found in the electronic supplementary material.

### 2.2. From network model to phenotypes: robust states

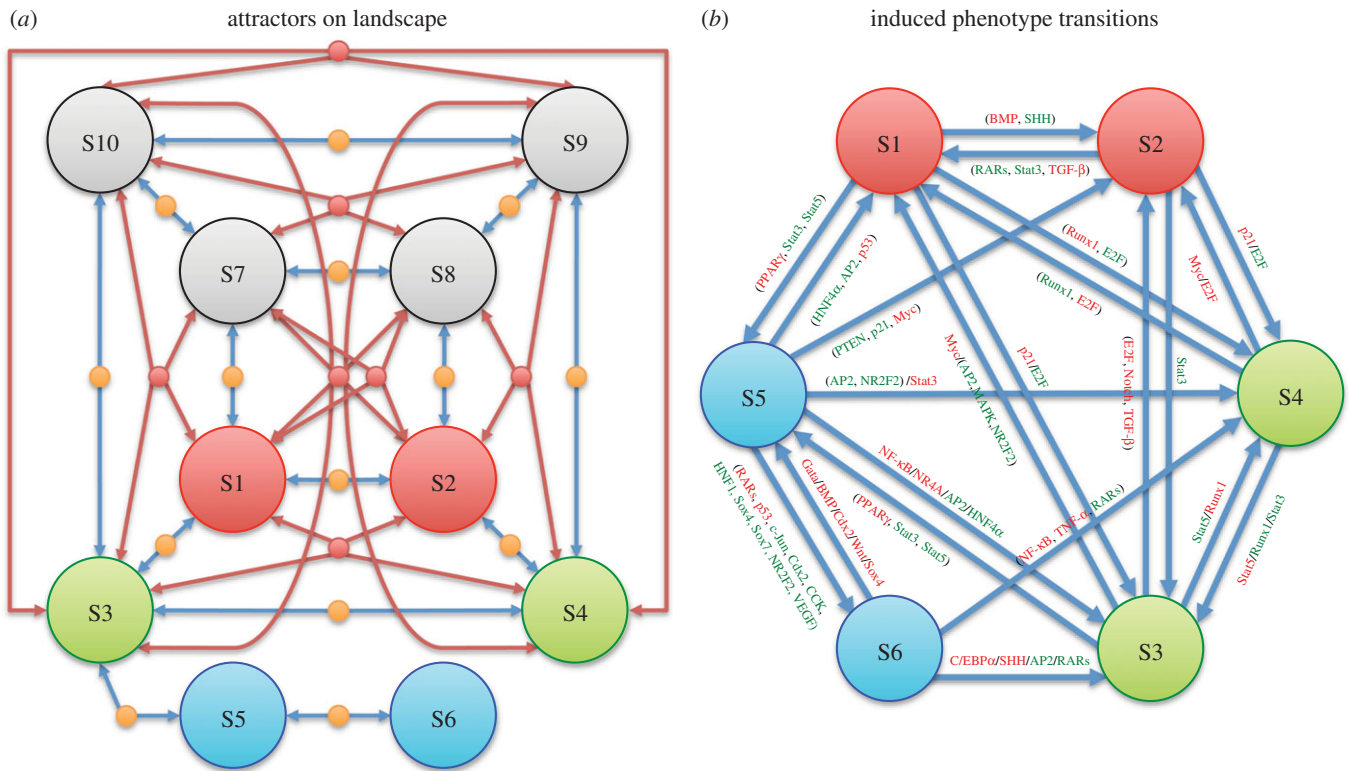
In a molecular network, due to the interactions in the network that include feedbacks only for a limited number of combinations, these interactions are ‘balanced’, that is, do not exert any driving force to change the network state. Mathematically, these ‘balanced states’ correspond to fixed-point solutions of a nonlinear dynamical system [39]. If a fixed point is also stable under perturbation, it is an attractor of the system. These attractors are more likely to occur and have a longer residence time in real biological systems with the presence of noise. Saddle points or other fixed points that are partially (i.e. in a subset of state space directions) stable can serve as passes for stochastic transitions between attractors. The attractors are interpreted as cell phenotypes because of their stability, which affords homeostatic robustness to the specific molecular profile that determines phenotype [9,10,40]. We obtained 10 attractors from the network constructed (figure 2), representing the predicted phenotypes. These attractors are robust over a large range

of parameters. Among them, attractors S7–S10 display molecular profiles that resemble apoptotic states. Attractors S1 and S2 are proliferation phenotypes, epitomized by upregulation of E2F, Cyclin D/E. By contrast, attractors S5 and S6 represent differentiated phenotypes. Attractors S3 and S4 appear to encode cell cycle arrested state, but otherwise resemble the proliferation attractors S1 and S2.

Among the two growth states with predicted active cell cycle activity (attractors S1 and S2 in figure 2), one seems to represent growth factor-stimulated proliferation, characterized by active FGF, HGF, IGF, EGF signalling, consistent with epithelia-to-mesenchymal transition and endoderm organ formation (attractor S1 in figure 2) [41], while the other appears more typical of CRC: active Wnt pathway [42], high expression of Runx1 and osteopontin [43]. These two growth states might both contribute to cancer progression because they interconvert into each other relatively easily. The growth-factor-stimulated state (attractor S1) is inducible from attractor S2 by reduction in Stat3 pathway (figure 3). By contrast, the typical CRC-like state (attractor S2) is not inducible from attractor S1 by inflammation alone. Instead, attractor S2 is induced by both inflammatory and immune suppression from attractor S1 by increasing TNF-α and TGF-β. There are different ways



**Figure 2.** The molecular profiles of the attractors in the dynamical system model of the CRC network (see also electronic supplementary material, table S1). The corresponding equations are listed in electronic supplementary material. Note that attractors S1 and S2, which correspond to proliferating, differ mainly in Stat5 activities. This difference may influence metastasis. Attractors S3 and S4 correspond to non-proliferating states but are otherwise similar to S1 and S2. S5 represents the normal intestine phenotype. S6 resembles a differentiated phenotype with a secretory signature. S7–S10 map into apoptotic states.



S1	proliferating state	Runx1 ↓	OPN ↓	Stat3 ↑	Stat5 ↑	E2F ↑	p21 ↓	p53 ↓
S2	proliferating state	Runx1 ↑	OPN ↑	Stat3 ↑	Stat5 ↓	E2F ↑	p21 ↓	p53 ↓
S3	non-proliferating, S1 like	Runx1 ↓	OPN ↓	Stat3 ↑	Stat5 ↑	E2F ↓	p21 ↓	p53 ↓
S4	non-proliferating, S2 like	Runx1 ↑	OPN ↑	Stat3 ↑	Stat5 ↓	E2F ↓	p21 ↓	p53 ↓
S5	differentiated state	Runx1 ↓	OPN ↓	Stat3 ↓	Stat5 ↓	HNF1 ↑	Cdx2 ↑	p53 ↓
S6	differentiated state	Runx1 ↓	OPN ↓	Stat3 ↓	Stat5 ↓	HNF1 ↓	Cdx2 ↓	p53 ↑
S7	apoptotic state		similar to S2		caspase 3 ↑		cytochrome C ↑	
S8	apoptotic state		similar to S1		caspase 3 ↑		cytochrome C ↑	
S9	apoptotic state		similar to S4		caspase 3 ↑		cytochrome C ↑	
S10	apoptotic state		similar to S3		caspase 3 ↑		cytochrome C ↑	

**Figure 3.** (a) Spontaneous transitions between the attractors characterized by saddle/unstable fixed points. In addition to attractors, the dynamical model also contains fixed points of different types, including saddles and other unstable fixed points. These points usually play the role of passes for spontaneous transitions between the attractors. S1–S10 represent the attractors, as shown in figure 2. The saddle/unstable fixed points are denoted by small dots. The flows of cell states from saddle/unstable fixed points to the attractors are represented by arrowed lines. (b) Predicted switching between these attractors triggered by induction (perturbation of gene activity). Multiple paths for transitions between any two attractors, representing (tumour) cell type conversions (only selected are shown). Inducers in the same brackets must be operated simultaneously to induce a switch. The slash ‘/’ represents different paths triggered by different inducers. Red/green represents upregulation/down-regulation. For clarity, the corresponding phenotypes of these attractors are also listed. Attractor S5 is normal intestine-like, while all the other attractors might contribute to CRC. Since attractors S3 and S4 are similar to S1 and S3, there are essentially three attractors contributing to CRC subtypes: S1, S2 and S6.

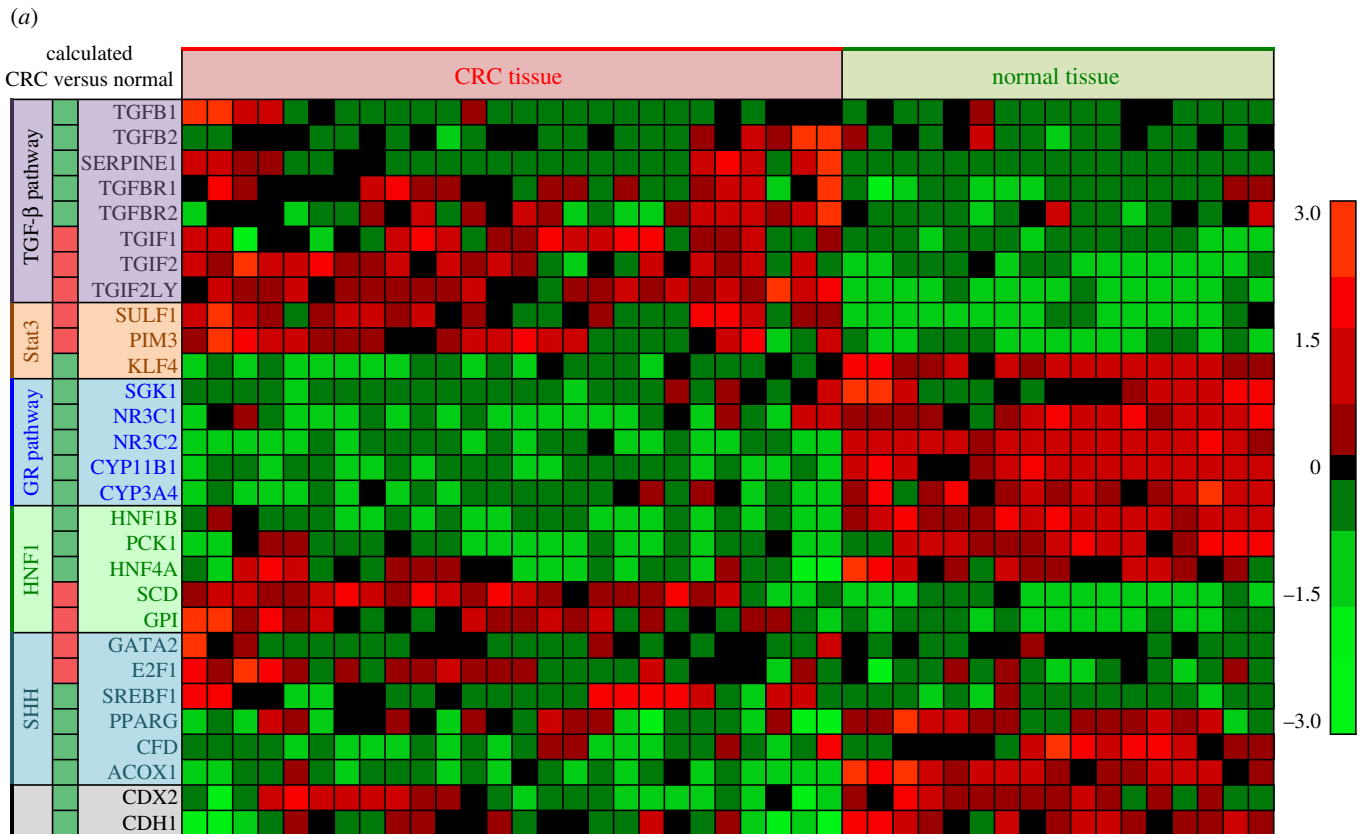
to interpret the possible phenotypes represented by attractors S1 and S2, but it is apparent that they have characteristics of both an aberrant developmental process and inflammation.

The differentiated states (attractor S5) represent an intestine-like phenotype, characterized by activated Cdx2 and HNF1. Both genes are indispensable for intestinal function [44]. Attractor S5 also has higher GATA4/6 level, consistent with an intestinal epithelial cell phenotype [44]. Attractor S6 differs from S5 by being Foxa1,2/SHH/Sox-positive. It might represent stomach or intestinal secretory cell lineage [11,45–47]. For these attractors, the active network nodes (with a high expression level) can be identified, as shown in electronic supplementary material, figure S1. This is a way to visualize the effective subnetwork specific for each attractor. The active nodes for the normal intestine-like state (attractor S5) are compared with those for S1, S2 and S6 separately in electronic

supplementary material, figure S1. The network topology indeed suggests that attractors have their own separate positive feedback loops. Also, attractors have mutual suppressing relations, as shown in the graph, although not in the simple form as seen in direct node-to-node switches.

### 2.3. Comparison with observations I: common features of colorectal cancer

CRC is a disease that exhibits inter-individual variation both in pathology and in its response to treatment. The model predicts multiple fixed points. This multi-stability opens the possibility that cells in a tumour occupy distinct attractors, and possibly interconvert between them—giving rise to mixtures of cell phenotypes—which would explain the widely



**Figure 4.** Comparison between computed results and microarray data. The microarray profiles were obtained from the second affiliated hospital of Zhejiang University for a total of 17 normal tissues and 26 CRC tissues of different patients. Every column is the profile of a computed attractor, the profile of relative gene activity between CRC and normal, as indicated, or a microarray profile from CRC tissue. Comparison of the predicted attractors profiles with observed microarray data show common features as listed in (a) and (b). The cell cycle module is shown in (b). The list of genes for (b) were obtained from Theilgaard-Mönch *et al.* [104]. (c) A broad comparison of the model prediction and observation. Normal tissues are in group IV. Group I are patients showing attractor S6 signature, group II patients showing some normal like attractor S5 signature, and group III patients showing a signature of attractor S1 and S2, but not S6. References used for annotations are [105–110]. A full list of data and references is available in electronic supplementary material, file S1. (Parts (b) and (c) shown on following pages.)

observed non-genetic cellular heterogeneity of tumour tissues [48,49]. However, obviously transcriptome analyses of tumours [50] still can identify subtypes of CRC, perhaps because some attractor dominates. Thus, we first identified common features in CRC profiles derived from modelling and compared them with clinical measurements.

Attractor S5 (figure 2) represents normal intestinal phenotype. If the intestinal cells were trapped in other attractors, they would be pathological, although not necessarily CRC. We assume that cancer tissue is a heterogeneous combination of attractors S1–S4 and S6, then compare with real tumours. Figure 2 shows that HNF1, HNF4, Cdx2, TGF- $\beta$  pathway, E-cadherin and glucocorticoid pathway characterize normal intestine. They are unambiguously downregulated in every possible abnormal state in the computed attractors S1–S4 and S6. The microarray profile from patients obtained from the second affiliated hospital of Zhejiang University showed downregulation of HNF1B expression in CRC patients. As expected [51], HNF1B transcript downregulation accompanied a decrease in gluconeogenesis (manifest in PCK1, phosphoenolpyruvate carboxykinase), upregulated glycolysis (GPI, glucose-6-phosphate isomerase) and lipogenesis (SCD, stearoyl-CoA desaturase). Microarray results for Cdx2 and HNF4A were not conclusive. TGF- $\beta$  signalling impairment is consistent with upregulation of negative regulators TGIFs [52,53]. Glucocorticoid receptor NR3C1 and mineralocorticoid receptor NR3C2 were both downregulated in CRC patients. Glucocorticoids target genes CYP3A4 [54],

SGK1 [55] and CYP11B1 (steroid 11- $\beta$ -hydroxylase), which are required for synthesis of glucocorticoids [56], were downregulated in CRC patients. The model predicts upregulation of Stat3 in attractors S1–S4. Expression of several Stat3 target genes, SULF1, PIM3 and KLF4 [57] was altered accordingly, as shown in figure 4a. In addition to transcripts, the kinase activities of the Akt and MAPK pathways were shown to be active in CRC stem cells [58], consistent with both attractors S1 and S2. These experimental results show that network modelling results are consistent with observation. One application for these common features (shown in figure 4a,b) predicted by models is to use them as molecular signature in prognosis and diagnosis. Unlike signatures defined by brute-force statistical analysis of data [50,59], attractor-informed patterns capture the natural constraints imposed by the underlying regulatory network.

## 2.4. Comparison with observations II: heterogeneous features of colorectal cancer

The computed attractors S1–S4 and S6 (figure 2) together with their combination define a set of subtypes that the network model predicts. There are two proliferating attractors S1 and S2. Stat3 and Stat5 pathways are both active in attractor S1, while only Stat3 is active in attractor S2. Runx1, interleukin signalling, Wnt pathways and NR4A are more active in attractor S2, while Hox genes and FGF signalling are more active in S1

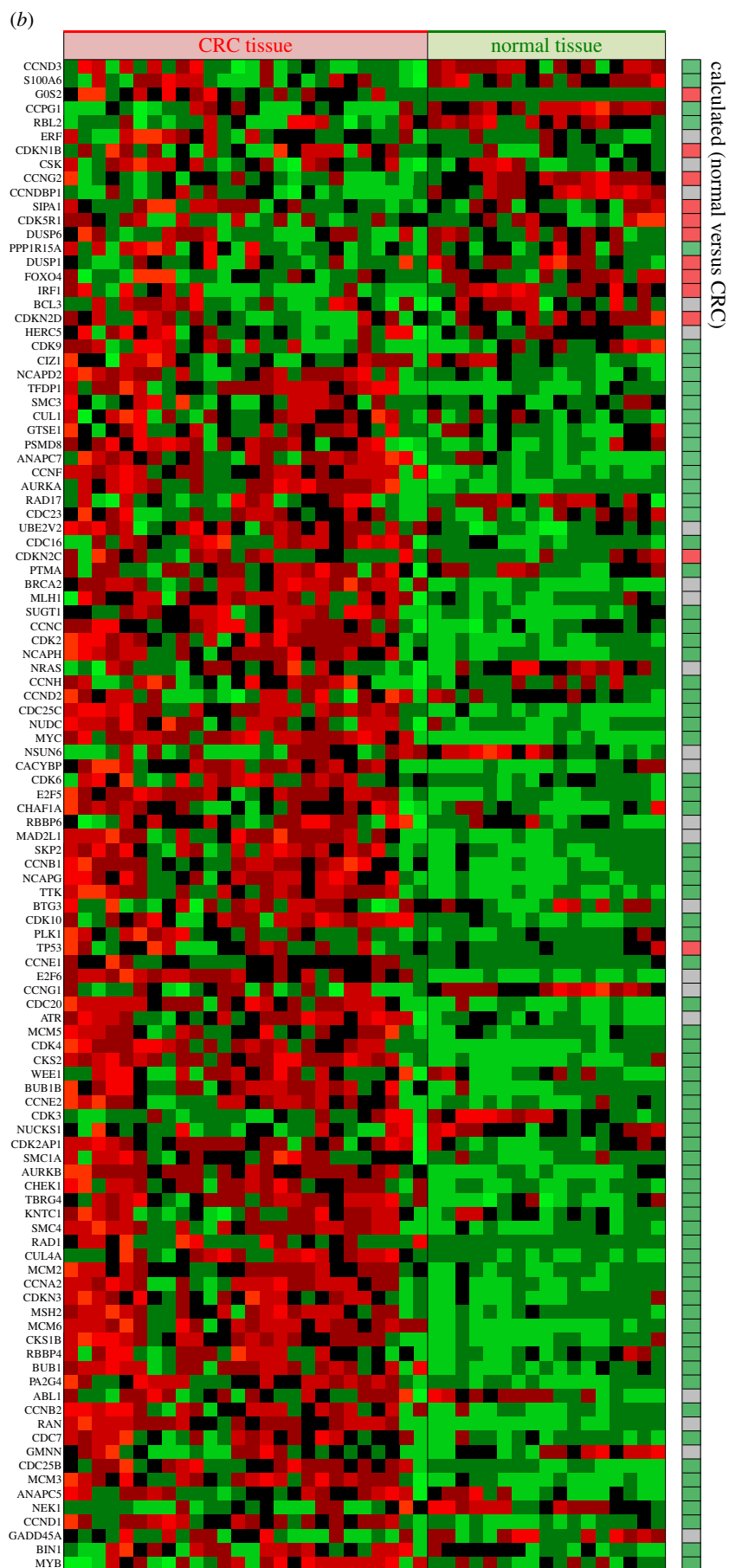


Figure 4. (Continued.)

attractor. Attractor S2 also predicts upregulated OPN, which favours metastasis [60]. When comparing patients' mRNA profiles, we can determine which attractor the profile of an

individual tumour is closer to. For example, patient 6 (in electronic supplementary material, file 2) is close to attractor S2, with high OPN, Runx1 and low Hoxa10. Compared with



Figure 4. (Continued.)

patient 6, patient 9 is closer to attractor S1, with low OPN, Runx1 and high Hoxa10. Attractor S6 resembles a differentiated state with possible secretory signature. There are

reports of mucinous CRCs [61–63], which express Sox2, MUC5AC, MUC2, MUC5B and p21 (CDKN1A). In the microarray data (in electronic supplementary material, file 2), several



tumours (of patients 2, 3, 4, 5, 7, 15) exhibit a molecular profile resembling that of S6, characterized by upregulated p53 or p27 and upregulated mucins. However, there is a difference between attractor S6 and these patients' profiles, in that the apparent Stat3 activity is high in these patients, as suggested by elevated SULF1, PIM3 and KLF4. It is now well documented that most (if not all) tumours, including CRC, contain multiple cell types, even among the isogenic cells [58]. Such phenotypic heterogeneity is best explained by multi-stability [48,49]. One possibility is that these patients' tumours may contain cell types other than S6 that would map into attractor S1 or S2. These p53- or p27-positive patients are particularly interesting because they have a worse clinical outcome.

## 2.5. Transition between robust states

Mathematically, in addition to attractors, the network dynamics also predict saddles and other unstable fixed points. However, these states are not locally stable. When perturbed, they will flow to different attractors, driven by molecular interactions [9,10,64,65]. The saddle points have straightforward dynamical interpretation as the point of passage of the 'least effort' path for spontaneous transitions between attractors driven by noise [66–68]. Such paths are shown in figure 3a. Attractors S1–S4 are connected through multiple saddles and other unstable fixed points to each other. Therefore, it is likely that cells interconvert due to external perturbations or internal fluctuations. The exception is attractor S6, which is not connected to proliferating states represented by S1 and S2. As S6 resembles a non-proliferative and differentiated state, it has to coexist with either S1 or S2 to allow for tumour growth. This heterogeneity of growth behaviour fits the biological framework of cancer stem cells that drive tumour growth. Among the possible CRC attractors S1–S4, attractor S3 is connected by a saddle point to the normal intestine-like attractor S5. This would suggest the possibility that carcinogenesis could be initiated by a transition from the normal intestine-cell-type attractor S5–S3 through the saddle point.

Transition from one attractor to another in the model can also be enforced by resetting the values of a few selected nodes of the network, which in reality maps into upregulating or downregulating of the corresponding molecules by external induction. Starting from normal intestine-like attractor S5, if Myc is upregulated, and p21 and PTEN downregulated simultaneously, this would cause a switch to the proliferative state S2, with typical CRC molecular profile. In the model, when NF- $\kappa$ B is numerically upregulated in the normal intestine-like attractor S5, the cell will switch to attractor S3, consistent with the notion of inflammation as promoter of carcinogenesis. A switch from attractor S3–S1 is caused by activation of Myc. Computed trajectories for selected attractor switching are shown in figure 3b. Consistent with the modelling results that attractor S6 is not as closely connected through saddles or other unstable fixed points to the rest of the attractors, inducing S6 from S5 is more difficult.

## 2.6. Validation of the modelling results

We use standard hypothesis testing for the validation.

*Comparison between modelling results and clinical data:*

*Vector (a).* Extracting results from clinical data. We use the average gene expression value of the microarray data from CRC (26 samples) versus the average value of those from

normal tissue (17 samples) to generate a *vector* (a column matrix) with 1 and  $-1$  elements. The rule is listed below: if the expression level of a gene from the CRC data over that from the normal tissue is larger than 1 plus a threshold value, we consider this gene is upregulated in CRC and vice versa. In case the upregulation or downregulation is not determined by data, we remove the gene from the list, because we cannot distinguish the two possibilities that (i) the upregulation or downregulation of the gene is not shown due to inaccurate measurements, and (ii) the gene expression level is similar in CRC and normal tissue.

$$\begin{cases} \frac{\text{CRC}}{\text{Normal}} > 1 + \text{threshold}, & 1 \text{ (upregulated in CRC)} \\ \frac{\text{Normal}}{\text{CRC}} > 1 + \text{threshold}, & -1 \text{ (downregulated in CRC)} \\ \text{Otherwise,} & 0 \text{ (undetermined by data, removed)}. \end{cases}$$

*Vector (b).* The *vector* (with  $-1, 1, 0$  elements) for the modelling results can be similarly generated without removing zeros, because zeros mean that the modelling results predict a similar gene expression level. We use the same *threshold* for the two *vectors* (a) and (b). We could then compare the two *vectors* directly and compute an accuracy rate for the consistency (see electronic supplementary material, file 3 for details). For a 5% threshold value, the accuracy is about 73% with a total of 78 genes, which is close to the accuracy of microarray data based on our previous experience [8,10–12,23]. We also compute the accuracy with a varying threshold value in figure 5b. It demonstrates the robustness of our results.

*Test 1: Comparison with random guess.* We generate 1 million random vectors with uniformly distributed 1 and  $-1$ . For better matching with the data, we do not include zeros here. The distribution of the accuracy is shown in electronic supplementary material, figure S3. The test shows that our modelling results have  $p < 0.00001$ .

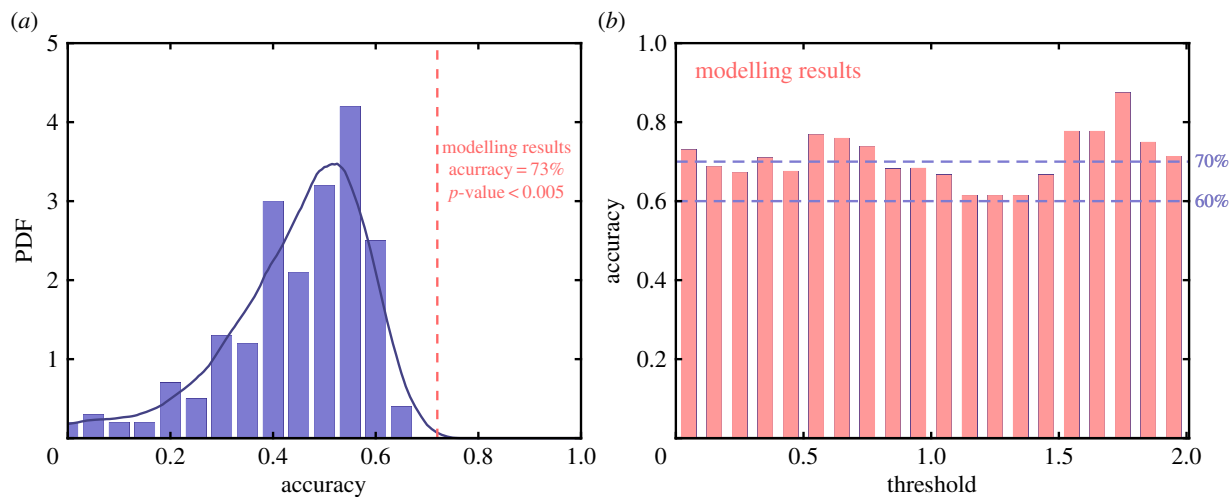
*Test 2: Comparison with randomly rewiring networks.* The null hypothesis is that the accuracy of the results from the CRC network model versus clinical data is not significant compared with results generated from randomly rewiring networks.

For randomly rewiring networks that have the same total number of interactions but are randomly rewired the connections and interaction types (inhibition or activation), we compute the accuracy rate distribution shown in figure 5a. For each randomly rewiring network, we compute its dynamical stable states and the optimal two-state combination (assigned to cancer and normal separately) compared with the data. The estimated distribution from 200 networks shows that the  $p$ -value is less than 0.005, such that the null hypothesis should be rejected.

## 3. Discussion

### 3.1. Clinical implications: colorectal cancer carcinogenesis, metastasis and treatment

CRC development may take years and involve multiple risk factors [3,69–71]. Intermediate steps have been observed. The molecular network constructed here exhibits both normal intestine-like attractors and proliferating CRC-like attractors. The predicted possibility for switching from normal intestinal to CRC states may be relevant to carcinogenesis summarized in figure 3. The results are consistent with the heterogeneous



**Figure 5.** Validation of the modelling results through comparison with randomly rewiring networks. (a) Distribution of consistency with clinical data for a group of 200 randomly rewiring networks. Our model has 73% accuracy, which is significantly larger than randomly rewiring networks with  $p < 0.005$ . (b) The influence of the threshold parameter in the comparison with the clinical data is not significant. The details of the comparison are provided in the electronic supplementary material, file S3.

nature of carcinogenesis and the role of inflammation and stress [72,73]. The network model naturally allows for a multitude of paths for the transition to the cancerous attractors [74], many of which are associated with an inflammatory programme. Specifically, the model predicts that tumour attractors can be reached through changes in Myc, p21 or PTEN, or activation of NF- $\kappa$ B as intermediate steps. In addition, increased stress and p53, in combination with loss of retinoic inducible transcription factor AP2 and nuclear factor HNF4, may also lead to switching to CRC, from normal intestine-like attractor S5–S1.

The network supports several attractors that might associate with distinct phenotypes; in principle, how they contribute to metastasis can be discussed. For instance, the proliferating attractors S1 and S2 differ in Stat5 pathway activities. Stat5 signalling has been shown to have an independent effect on cell morphology via regulation on the expression of adhesion molecules [75,76]. Impaired Stat5 signalling leads to non-healing wounds in the intestine [76]. The loss of intestinal barrier due to Stat5 dysfunction might contribute to metastasis in the phenotypes given by the network. Another metastasis-related protein, PRL-3, is a downstream target of Stat5 and Stat3 [77,78]. Note that all the network-supported phenotypes of CRC have metastasis potential. The Runx1 and OPN high/low phenotype might favour metastasis through loss of intestinal barrier/activate PRL-3.

Transition from CRC to normal intestine state, on the other hand, might be relevant to the prevention and treatment of CRC. There are three major types of CRC, represented by the type with positive p53 or p27 (S6), and the OPN- and Runx1-positive (S2/S4) and-negative (S1/S3) types. For the OPN- and Runx1-negative phenotype, switching to the normal intestinal phenotype requires a suppression of the inflammation programme, which could be achieved through a transition from S3 to S5 by reducing Stat3 and Stat5. For the OPN- and Runx1-positive phenotype, switching to the normal intestinal phenotype would require first a switching to the OPN- and Runx1-negative phenotype, from S2 to S1, through suppression of RARs signalling and simultaneous anti-inflammation (reducing Stat3, increasing TGF- $\beta$  signalling). The p53/p27-positive CRC type could be converted to the OPN- and Runx1-positive phenotype via S4, by promoting inflammation (NF- $\kappa$ B and TNF- $\alpha$ ) and suppressing retinoic acid signalling RARs.

The theoretical base under immunotherapies is the cancer immunosurveillance hypothesis [79]. However, experimental evidence leads to various conclusions, ranging from that the immune system naturally protects against cancer [80] to that the immune reaction is almost always stimulatory to the tumour's growth [81]. The current focus on the differential responses to immunotherapy is usually about the state of the immune system, such as the T cell population and types [82,83]. Our results show that different intrinsic robust cancer subtypes may respond to immune cells, such as T cells and macrophages, differently (figure 3b). Since the focus of the model is on tumour cells, the calculated different responses are due to the network-wide regulation over NF- $\kappa$ B and Stat3/5 inside tumour cells. If the internal status of a tumour does not favour NF- $\kappa$ B activation, the remaining effect of cytokines is left as a growth factor. Therefore, the calculated likely routes for transitions between normal and cancer states, corresponding to the genesis and development of cancer and transition among subtypes, have non-identical dependence on immune activation.

The network for CRC presented in this work is essentially a core decision-making network for intestinal development and function. Signal transduction and transcriptional regulation are two main molecular interaction types included. There are no particular cancer-specific genes or cancer pathways in the network. As consequences of the network wiring diagram, its dynamics generate multiple attractors, which are robust states and correspond to both normal intestinal phenotype and CRC subtypes. For each of the robust state, a subnetwork of the active nodes thus represents 'cancer network' or 'normal network' separately, although they are part of the same endogenous network. Interestingly, the nodes active in the normal and CRC-like attractors have mutual inhibiting effects. The concept of multiple attractors emanating from a single network naturally suggests that healthy and abnormal cell states are 'alternative sides of the same coin'—a feature not manifest in the traditional notion of specific linear pathways implicated in individual 'hallmarks' of cancer [84,85]. Here, we tested the cancer attractor concept, which has a solid foundation in a formal framework, but typically has been articulated in generic terms. We link it for the first time to specific biological observations, both in terms of the network topology (by modelling experimentally tested molecular interactions) and network

dynamics (by comparing measured transcriptomes with predicted attractors). We show high consistency in that our network model predicted features of CRC including its diversity among individual patients.

## 4. Material and methods

### 4.1. Network construction

The construction of the network followed an incremental procedure aimed at capturing the essential features of CRC while keeping the network to the core level. To start with, we first built a minimal cellular model, with cell cycle, apoptosis, growth factor signalling, immune response and stress response modules, similar to previous works [9–12]. Molecules and molecular pathways well established in cancer, such as PTEN, p53, Myc, Ras and NF- $\kappa$ B, were in this minimal model. Then we added several additional modules specific to CRC. First, Cdx2 [86] and Notch signalling pathway [87] were included for their role in intestinal differentiation. Cdx2, HNF1 and Gata family were added for their indispensable roles for intestinal function [44]. Wnt [88,89], SHH and BMP signalling pathways [90–92] were also included for mesoderm development. Wnt signalling is active in posterior endoderm in early mouse development. Wnt induces Cdx2, a cell fate switch of intestinal differentiation [93]. Glucocorticoid pathway was included for its effect on GI mucosa [94] and PPAR $\gamma$  for inhibiting intestinal inflammation [95,96]. We explored inflammation beyond NF- $\kappa$ B pathway, into haematopoiesis because of many shared transcription factors in CRC and leukaemia. One possible feedback link is Runx1, which is often over-expressed in CRC. Runx1 is both a gastric gene and indispensable in haematopoiesis, and associates with the myeloid leukaemia [97–99]. A number of leukaemia-related genes and molecular signalling pathways (Hox family, Pu.1, IL3) were also added for their possible role in immune response. Finally, angiogenesis and metastasis modules were included in the model.

### 4.2. Network modelling

The dynamical model for the CRC network was constructed based on a framework of the endogenous molecular–cellular network proposed previously [7,8]. Endogenous network models have been built for hepatocellular [9], prostate [10] and gastric cancers [11], as well as acute promyelocytic leukaemia [12]. As detailed knowledge about the interactions between the network nodes (e.g. *in vivo* parameters) is usually absent, a coarse-grained modelling of the network dynamics was employed. The interactions between the network nodes (agents) were described by activation/inhibition using differential equations [9–12], as a refinement of coarser descriptions such as the threshold Boolean network [100,101].

The dynamical equations for the concentration/activity of an agent  $x_i$  under the influence of other agents  $x = (x_1, x_2, \dots, x_n)$  take the following form:

$$\frac{dx_i}{dt} = f_i(x) - \frac{x_i}{\tau_i}.$$

This expression consists of a generation term and a degradation term, which is consistent with the chemical rate equation. The degradation time is set to be unity for all

agents (i.e.  $\tau_i = 1$ ). The underlying assumption is that the baseline decay times for the agents are of the same order of magnitude. Although protein degradation time may vary significantly, most of the agents in the network for the core endogenous interactions operate at the regulatory level relevant to cell proliferation and death. It has been shown that the lifetimes of more than 4000 proteins in a non-dividing mammalian cell are narrowly distributed, with a median value of 35.5 h [102]. No structural proteins with a long lifetime are included in the model. For the signal transduction processes, the signalling pathways are treated as a whole until downstream transcription takes effect, avoiding the possible short response time of the initial stages. The active regulations of protein degradation, such as the regulation of PTEN stability, are included as activation/inhibition in the network [103].

Here, we use a nonlinear Hill function to implement the coarse-grained activation and inhibition. For example (see electronic supplementary material, table S1), HNF1 is activated by HNF4  $\alpha$  and Notch, and inhibited by C/EBP  $\alpha$ ; the equation has the form

$$\frac{dx_{\text{HNF1}}}{dt} = \frac{a(x_{\text{HNF4}\alpha}^n + x_{\text{Notch}}^n)}{1 + a(x_{\text{HNF4}\alpha}^n + x_{\text{Notch}}^n)} \times \frac{1}{1 + ax_{\text{C/EBP}\alpha}^n} - x_{\text{HNF1}}.$$

In general, if an agent  $x$  is activated by agents  $y_1, y_2, \dots, y_k$  and inhibited by agents  $z_1, z_2, \dots, z_l$ , we have

$$\frac{dx}{dt} = \frac{a \sum_{i=1}^k y_i^n}{1 + a \sum_{i=1}^k y_i^n} \times \frac{1}{1 + a \sum_{i=1}^l z_i^n} - x,$$

where two parameters  $n$  and  $a$  can be tuned in the equation. Note that our differential equation modelling results are consistent with those from the parameter-free Boolean dynamics (see electronic supplementary material, table S6), such that the model is intrinsically not sensitive to parameters. Our large-scale random parameter tests also validate this point (see electronic supplementary material, figure S2).

The value for each agent is normalized from zero to one. The dynamics have a conservation property: they will keep the value of each agent normalized (between zero and one). A full list of 89 equations can be viewed in the electronic supplementary material. The equations can generate several local attractors with biological significance. In this CRC model, we obtained 10 attractors, as well as 14 saddle points and seven other unstable fixed points (shown in electronic supplementary material, table S3 and table S7 separately). A series of random searches crossing several magnitudes in the number of random initial points are executed. It turns out that no new attractors can be found, which means the results may cover the major attractors that are biologically stable (with a relatively large attractive basin) in the state space.

To validate the modelling results: (i) we use both Newton's iteration method and Euler's method separately to obtain attractors from the differential equations (electronic supplementary material, table S3); (ii) we made a series of random parameter tests (see electronic supplementary material, figure S2) that allow independent random Hill coefficient  $n$  and prefactor  $a$  for each interaction (total of 525 interactions), and the obtained results are almost invariant; (iii) we used alternative forms of equations and the attractors are reproduced (electronic supplementary material, tables S4 and S5); and (iv) we computed the Boolean network model and found the result supports nine of the 10 attractors

obtained from the differential equations (see electronic supplementary material, table S6).

### 4.3. Sample collection and microarray experiments

Colorectal tumour and normal tissue specimens from 26 CRC patients with clinical outcome and three healthy individuals were collected from the second affiliated hospital of Zhejiang University. All samples were concurrently analysed using Affymetrix Human Genome U133 Plus 2.0 microarrays. The extraction of RNA was performed following standard protocols provided by the manufacturers.

### 4.4. Microarray data analyses

For gene expression analysis, tumour and adjacent normal tissues were investigated using an Affymetrix Human Genome U133 Plus 2.0 microarray. Data were acquired by GENECHIP operating software v. 1.4. After quality checks, raw intensity

data were processed by quantile normalization with robust multi-array average to remove systematic bias using Affymetrix EXPRESSION CONSOLE v. 1.12. The complete table of microarray data with recognized gene symbol and function is provided in electronic supplementary material, file S2.

**Ethics.** The Ethics Committee of the Second Affiliated Hospital, Zhejiang University School of Medicine approved this study. Written informed consent for participation was obtained from each patient.

**Data accessibility.** All data of this work are available in the electronic supplementary material.

**Authors' contributions.** R.Y., X.Z. and P.A. constructed endogenous network, performed data and model analyses; Su.Z., J.Y., Y.H., D.L. and Sh.Z. collected and analysed data from patients; X.Z., R.Y., P.A., Su.Z., Sh.Z. and L.H. conceived the study; all authors wrote the manuscript.

**Competing interests.** We declare we have no competing interests.

**Funding.** This work was supported in part by the Natural Science Foundation of China No. NSFC91329301 and No. NSFC91529306; and by the grants from the State Key Laboratory of Oncogenes and Related Genes (No. 90-10-11).

## References

- US Cancer Statistics Working Group. 2014 *United States cancer statistics: 1999–2010 incidence and mortality web-based report*. Atlanta, GA: Department of Health and Human Services, Centers for Disease Control and Prevention.
- Yang L, Parkin DM, Li LD, Chen YD, Bray F. 2004 Estimation and projection of the national profile of cancer mortality in China: 1991–2005. *Br. J. Cancer* **90**, 2157–2166. (doi:10.1038/sj.bjc.6601813)
- Sung JJ *et al.* 2008 Asia Pacific consensus recommendations for colorectal cancer screening. *Gut* **57**, 1166–1176. (doi:10.1136/gut.2007.146316)
- Heinemann V, Douillard JY, Ducreux M, Peeters M. 2013 Targeted therapy in metastatic colorectal cancer—an example of personalised medicine in action. *Cancer Treat. Rev.* **39**, 592–601. (doi:10.1016/j.ctrv.2012.12.011)
- Chan AT, Giovannucci EL. 2010 Primary prevention of colorectal cancer. *Gastroenterology* **138**, 2029–2043. (doi:10.1053/j.gastro.2010.01.057)
- Zamanianazodi M, Rezaeitavirani M, Hasanzadeh H, Rahmati RS, Dalilan S. 2015 Introducing biomarker panel in esophageal, gastric, and colon cancers; a proteomic approach. *Gastroenterol. Hepatol. Bed Bench* **8**, 6–18.
- 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)
- Yuan R, Zhu X, Wang G, Li S, Ao P. 2017 Cancer as robust intrinsic state shaped by evolution: a key issues review. *Rep. Prog. Phys.* **80**, 042701. (doi:10.1088/1361-6633/aa538e)
- Wang G-W, Zhu X-M, Gu J-R, Ao P. 2014 Quantitative implementation of the endogenous molecular-cellular network hypothesis in hepatocellular carcinoma. *Interface Focus* **4**, 20130064. (doi:10.1098/rsfs.2013.0064)
- Zhu X, Yuan R, Hood L, Ao P. 2015 Endogenous molecular-cellular hierarchical modeling of prostate carcinogenesis uncovers robust structure. *Prog. Biophys. Mol. Biol.* **117**, 30–42. (doi:10.1016/j.pbiomolbio.2015.01.004)
- Li S, Zhu X, Liu B, Wang G, Ao P. 2015 Endogenous molecular network reveals two mechanisms of heterogeneity within gastric cancer. *Oncotarget* **6**, 13 607–13 627. (doi:10.18632/oncotarget.3633)
- Yuan R, Zhu X, Radich JP, Ao P. 2016 From molecular interaction to acute promyelocytic leukemia: calculating leukemogenesis and remission from endogenous molecular-cellular network. *Sci. Rep.* **6**, 24307. (doi:10.1038/srep24307)
- Manfredi S, Lepage C, Hatem C, Coatmeur O, Faivre J, Bouvier A-M. 2006 Epidemiology and management of liver metastases from colorectal cancer. *Ann. Surg.* **244**, 254–259. (doi:10.1097/01.sla.0000217629.94941.cf)
- Leporrier J, Maurel J, Chiche L, Bara S, Segol P, Launoy G. 2006 A population-based study of the incidence, management and prognosis of hepatic metastases from colorectal cancer. *Br. J. Surg.* **93**, 465–474. (doi:10.1002/bjs.5278)
- Cummings LC, Payes JD, Cooper GS. 2007 Survival after hepatic resection in metastatic colorectal cancer: a population-based study. *Cancer* **109**, 718–726. (doi:10.1002/cncr.22448)
- Yeh JJ *et al.* 2009 KRAS/BRAF mutation status and ERK1/2 activation as biomarkers for MEK1/2 inhibitor therapy in colorectal cancer. *Mol. Cancer Ther.* **8**, 834. (doi:10.1158/1535-7163.MCT-08-0972)
- Sammalkorpi H, Alhopuro P, Lehtonen R, Tuimala J, Mecklin J-P, Jarvinen HJ, Jiricny J, Karhu A, Aaltonen LA. 2007 Background mutation frequency in microsatellite-unstable colorectal cancer. *Cancer Res.* **67**, 5691–5698. (doi:10.1158/0008-5472.CAN-06-4314)
- Roepman P *et al.* 2014 Colorectal cancer intrinsic subtypes predict chemotherapy benefit, deficient mismatch repair and epithelial-to-mesenchymal transition. *Int. J. Cancer* **134**, 552–562. (doi:10.1002/ijc.28387)
- Ogino S. 2009 CpG island methylator phenotype, microsatellite instability, BRAF mutation and clinical outcome in colon cancer. *Gut* **58**, 90–96. (doi:10.1136/gut.2008.155473)
- Marisa L *et al.* 2013 Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med.* **10**, e1001453. (doi:10.1371/journal.pmed.1001453)
- Sztupinski Z, Györfi B. 2016 Colon cancer subtypes: concordance, effect on survival and selection of the most representative preclinical models. *Sci. Rep.* **6**, 37169. (doi:10.1038/srep37169)
- Guinney J *et al.* 2015 The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356. (doi:10.1038/nm.3967)
- 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. (doi:10.1098/rsif.2015.1115)
- Waddington CH. 1957 *The strategy of the genes: a discussion of some aspects of theoretical biology*. New York, NY: Macmillan.
- Waddington CH. 1977 *Tools for thought: how to understand and apply the latest scientific techniques of problem solving*. St Albans, UK: Paladin.
- Delbrück M. 1949 *Unités biologiques douées de continuité génétique Colloques Internationaux du centre national de la Recherche Scientifique*, pp 33–35. Paris, France: CNRS.
- Monod J, Jacob F. 1961 General conclusions: teleonomic mechanisms in cellular metabolism, growth, and differentiation. In *Cold Spring Harbor Symposia on Quantitative Biology* (ed. L Frisch), pp. 389–401. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

28. Dean ACR, Hinshelwood C. 1963 Integration of cell reactions. *Nature* **199**, 7–11. (doi:10.1038/199007a0)
29. Kauffman SA. 1969 Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.* **22**, 437–467. (doi:10.1016/0022-5193(69)90015-0)
30. Huang S, Eichler G, Bar-Yam Y, Ingber DE. 2005 Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.* **94**, 128701. (doi:10.1103/PhysRevLett.94.128701)
31. Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S. 2008 Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453**, 544–547. (doi:10.1038/nature06965)
32. Shih W, Chetty R, Tsao M-S. 2005 Expression profiling by microarrays in colorectal cancer. *Oncol. Rep.* **13**, 517–524.
33. Ahmed S, Banerjee A, Hands RE, Bustin S, Dorudi S. 2005 Microarray profiling of colorectal cancer in Bangladeshi patients. *Colorectal Dis.* **7**, 571–575. (doi:10.1111/j.1463-1318.2005.00818.x)
34. Vogelstein B, Kinzler KW. 2004 Cancer genes and the pathways they control. *Nat. Med.* **10**, 789–799. (doi:10.1038/nm1087)
35. Wood LD *et al.* 2007 The genomic landscapes of human breast and colorectal cancers. *Science* **318**, 1108–1113. (doi:10.1126/science.1145720)
36. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. 2013 Cancer genome landscapes. *Science* **339**, 1546–1558. (doi:10.1126/science.1235122)
37. Simiantonaki N, Taxeidis M, Jayasinghe C, Kirkpatrick CJ. 2007 Epithelial expression of VEGF receptors in colorectal carcinomas and their relationship to metastatic status. *Anticancer Res.* **27**, 3245–3250.
38. Okubo T, Hogan BL. 2004 Hyperactive Wnt signaling changes the developmental potential of embryonic lung endoderm. *J. Biol.* **3**, 11. (doi:10.1186/jbiol3)
39. Hirsch MW, Smale S. 1974 *Differential equations, dynamical systems, and linear algebra*. San Diego, CA: Academic Press.
40. Zhu X-M, Yin L, Hood L, Ao P. 2004 Calculating biological behaviors of epigenetic states in the phage lambda life cycle. *Funct. Integr. Genomics* **4**, 188–195. (doi:10.1007/s10142-003-0095-5)
41. Zorn AM, Wells JM. 2009 Vertebrate endoderm development and organ formation. *Annu. Rev. Cell Dev. Biol.* **25**, 221. (doi:10.1146/annurev.cellbio.042308.113344)
42. Lin AY *et al.* 2011 Comparative profiling of primary colorectal carcinomas and liver metastases identifies LEF1 as a prognostic biomarker. *PLoS ONE* **6**, e16636. (doi:10.1371/journal.pone.0016636)
43. Zhao M, Liang F, Zhang B, Yan W, Zhang J. 2015 The impact of osteopontin on prognosis and clinicopathology of colorectal cancer patients: a systematic meta-analysis. *Sci. Rep.* **5**, 12713. (doi:10.1038/srep12713)
44. Bosse T, Fialkovich JJ, Piaseckij CM, Beuling E, Broekman H, Grand RJ, Montgomery RK, Krasinski SD. 2007 Gata4 and Hnf1alpha are partially required for the expression of specific intestinal genes during development. *Am. J. Physiol. Gastrointest. Liver Physiol.* **292**, G1302–G1314. (doi:10.1152/ajpgi.00418.2006)
45. Ye DZ, Kaestner KH. 2009 Foxa1 and Foxa2 control the differentiation of goblet and enteroendocrine L- and D-cells in mice. *Gastroenterology* **137**, 2052–2062. (doi:10.1053/j.gastro.2009.08.059)
46. Bastide P *et al.* 2007 Sox9 regulates cell proliferation and is required for Paneth cell differentiation in the intestinal epithelium. *J. Cell Biol.* **178**, 635–648. (doi:10.1083/jcb.200704152)
47. Mori-Akiyama Y, van den Born M, van Es JH, Hamilton SR, Adams HP, Zhang J, Clevers H, de Crombrughe B. 2007 SOX9 is required for the differentiation of paneth cells in the intestinal epithelium. *Gastroenterology* **133**, 539–546. (doi:10.1053/j.gastro.2007.05.020)
48. Brock A, Chang H, Huang S. 2009 Non-genetic heterogeneity—a mutation-independent driving force for the somatic evolution of tumours. *Nat. Rev. Genet.* **10**, 336–342. (doi:10.1038/nrg2556)
49. Marusyk A, Almendro V, Polyak K. 2012 Intra-tumour heterogeneity: a looking glass for cancer? *Nat. Rev. Cancer* **12**, 323–334. (doi:10.1038/nrc3261)
50. Cancer Genome Atlas N. 2012 Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337. (doi:10.1038/nature11252)
51. Rebouissou S, Imbeaud S, Balabaud C, Boulanger V, Bertrand-Michel J, Tercé F, Auffray C, Bioulac-Sage P, Zucman-Rossi J. 2007 HNF1 $\alpha$  inactivation promotes lipogenesis in human hepatocellular adenoma independently of SREBP-1 and carbohydrate-response element-binding protein (ChREBP) activation. *J. Biol. Chem.* **282**, 14 437–14 446. (doi:10.1074/jbc.M610725200)
52. Tateossian H, Morse S, Parker A, Mburu P, Warr N, Acevedo-Arozena A, Cheeseman M, Wells S, Brown SDM. 2013 Otitis media in the Tgfr knockout mouse implicates TGFbeta signalling in chronic middle ear inflammatory disease. *Hum. Mol. Genet.* **22**, 2553–2565. (doi:10.1093/hmg/ddt103)
53. Hu Y, Yu H, Shaw G, Renfree MB, Pask AJ. 2011 Differential roles of TGIF family genes in mammalian reproduction. *BMC Dev. Biol.* **11**, 58. (doi:10.1186/1471-213X-11-58)
54. Cooper BW, Cho TM, Thompson PM, Wallace AD. 2008 Phthalate induction of CYP3A4 is dependent on glucocorticoid regulation of PXR expression. *Toxicol. Sci.* **103**, 268–277. (doi:10.1093/toxsci/kfn047)
55. Webster MK, Goya L, Ge Y, Maiyar A, Firestone G. 1993 Characterization of sgk, a novel member of the serine/threonine protein kinase gene family which is transcriptionally induced by glucocorticoids and serum. *Mol. Cell. Biol.* **13**, 2031–2040. (doi:10.1128/MCB.13.4.2031)
56. Kawamoto T *et al.* 1992 Role of steroid 11beta-hydroxylase and steroid 18-hydroxylase in the biosynthesis of glucocorticoids and mineralocorticoids in humans. *Proc. Natl Acad. Sci. USA* **89**, 1458–1462. (doi:10.1073/pnas.89.4.1458)
57. Bourillot PY, Aksoy I, Schreiber V, Wianny F, Schulz H, Hummel O, Hubner N, Savatier P. 2009 Novel STAT3 target genes exert distinct roles in the inhibition of mesoderm and endoderm differentiation in cooperation with Nanog. *Stem Cells* **27**, 1760–1771. (doi:10.1002/stem.110)
58. Wang YK, Zhu YL, Qiu FM, Zhang T, Chen ZG, Zheng S, Huang J. 2010 Activation of Akt and MAPK pathways enhances the tumorigenicity of CD133<sup>+</sup> primary colon cancer cells. *Carcinogenesis* **31**, 1376–1380. (doi:10.1093/carcin/bgq120)
59. Khirade MF, Lal G, Bapat SA. 2015 Derivation of a fifteen gene prognostic panel for six cancers. *Sci. Rep.* **5**, 13248. (doi:10.1038/srep13248)
60. Paschos KA, Majeed AW, Bird NC. 2014 Natural history of hepatic metastases from colorectal cancer—pathobiological pathways with clinical significance. *World J. Gastroenterol.* **20**, 3719. (doi:10.3748/wjg.v20.i14.3719)
61. Park ET, Gum JR, Kakar S, Kwon SW, Deng G, Kim YS. 2008 Aberrant expression of SOX2 upregulates MUC5AC gastric foveolar mucin in mucinous cancers of the colorectum and related lesions. *Int. J. Cancer* **122**, 1253–1260. (doi:10.1002/ijc.23225)
62. Byrd JC, Bresalier RS. 2004 Mucins and mucin binding proteins in colorectal cancer. *Cancer Metastasis Rev.* **23**, 77–99. (doi:10.1023/A:1025815113599)
63. Kim DH, Kim JW, Cho JH, Baek SH, Kakar S, Kim GE, Sleisenger MH, Kim YS. 2005 Expression of mucin core proteins, trefoil factors, APC and p21 in subsets of colorectal polyps and cancers suggests a distinct pathway of pathogenesis of mucinous carcinoma of the colorectum. *Int. J. Cancer* **27**, 957–964. (doi:10.3892/ijo.27.4.957)
64. Huang S, Guo YP, May G, Enver T. 2007 Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev. Biol.* **305**, 695–713. (doi:10.1016/j.ydbio.2007.02.036)
65. Huang S, Kauffman S. 2013 How to escape the cancer attractor: rationale and limitations of multi-target drugs. *Semin. Cancer Biol.* **23**, 270–278. (doi:10.1016/j.semcancer.2013.06.003)
66. Freidlin MI, Wentzell AD. 2012 *Random perturbations of dynamical systems*, 3rd edn. Berlin, Germany: Springer.
67. Ao P. 2005 Laws in Darwinian evolutionary theory. *Phys. Life Rev.* **2**, 117–156. (doi:10.1016/j.pprev.2005.03.002)
68. Zhou JX, Aliyu MDS, Aurell E, Huang S. 2012 Quasi-potential landscape in complex multi-stable systems. *J. R. Soc. Interface* **9**, 3539–3553. (doi:10.1098/rsif.2012.0434)
69. Wang X, Lei T, Ma X. 2001 Colon cancer risk factors in Jiashan county, Zhejiang province, the highest incidence area in China. *Chin. J. Oncol.* **23**, 480–482.
70. Van Faassen A, Tangerman A, Bueno-de-Mesquita BH. 2004 Serum bile acids and risk factors for colorectal cancer. *Br. J. Cancer* **90**, 632–634. (doi:10.1038/sj.bjc.6601608)
71. Slattery ML, Fitzpatrick F. 2009 Convergence of hormones, inflammation, and energy-related

- factors: a novel pathway of cancer etiology. *Cancer Prev. Res.* **2**, 922–930. (doi:10.1158/1940-6207.CAPR-08-0191)
72. Trinchieri G. 2012 Cancer and inflammation: an old intuition with rapidly evolving new concepts. *Annu. Rev. Immunol.* **30**, 677–706. (doi:10.1146/annurev-immunol-020711-075008)
73. Fordyce CA, Patten KT, Fessenden TB, DeFilippis R, Hwang ES, Zhao J, Tlsty TD. 2012 Cell-extrinsic consequences of epithelial stress: activation of protumorigenic tissue phenotypes. *Breast Cancer Res.* **14**, R155. (doi:10.1186/bcr3368)
74. Huang S. 2011 On the intrinsic inevitability of cancer: from foetal to fatal attraction. *Semin. Cancer Biol.* **21**, 183–199. (doi:10.1016/j.semcancer.2011.05.003)
75. Miyoshi K *et al.* 2001 Signal transducer and activator of transcription (Stat) 5 controls the proliferation and differentiation of mammary alveolar epithelium. *J. Cell Biol.* **155**, 531–542. (doi:10.1083/jcb.200107065)
76. Gilbert S, Zhang R, Denson L, Moriggi R, Steinbrecher K, Shroyer N, Lin J, Han X. 2012 Enterocyte STAT5 promotes mucosal wound healing via suppression of myosin light chain kinase-mediated loss of barrier function and inflammation. *EMBO Mol. Med.* **4**, 109–124. (doi:10.1002/emmm.201100192)
77. Rios P, Li X, Köhn M. 2013 Molecular mechanisms of the PRL phosphatases. *FEBS J.* **280**, 505–524. (doi:10.1111/j.1742-4658.2012.08565.x)
78. Zhou J *et al.* 2011 PRL-3, a metastasis associated tyrosine phosphatase, is involved in FLT3-ITD signaling and implicated in anti-AML therapy. *PLoS ONE* **6**, e19798. (doi:10.1371/journal.pone.0019798)
79. Burnet FM. 1970 The concept of immunological surveillance. *Prog. Exp. Tumor. Res.* **13**, 1–27. (doi:10.1159/000386035)
80. Corthay A. 2014 Does the immune system naturally protect against cancer? *Front. Immunol.* **5**, 197. (doi:10.3389/fimmu.2014.00197)
81. Prehn RT. 2010 The initial immune reaction to a new tumor antigen is always stimulatory and probably necessary for the tumor's growth. *Clin. Dev. Immunol.* **2010**, 5. (doi:10.1155/2010/851728)
82. Disis ML. 2010 Immune regulation of cancer. *J. Clin. Oncol.* **28**, 4531–4538. (doi:10.1200/JCO.2009.27.2146)
83. Draghiciu O, Nijman HW, Daemen T. 2011 From tumor immunosuppression to eradication: targeting homing and activity of immune effector cells to tumors. *Clin. Dev. Immunol.* **2011**, 439053. (doi:10.1155/2011/439053)
84. Hanahan D, Weinberg RA. 2011 Hallmarks of cancer: the next generation. *Cell* **144**, 646–674. (doi:10.1016/j.cell.2011.02.013)
85. Breitkreutz D, Hlatky L, Rietman E, Tuszynski JA. 2012 Molecular signaling network complexity is correlated with cancer patient survivability. *Proc. Natl Acad. Sci. USA* **109**, 9209–9212. (doi:10.1073/pnas.1201416109)
86. Sherwood RI, Maehr R, Mazzoni EO, Melton DA. 2011 Wnt signaling specifies and patterns intestinal endoderm. *Mech. Dev.* **128**, 387–400. (doi:10.1016/j.mod.2011.07.005)
87. Stanger BZ, Datar R, Murtaugh LC, Melton DA. 2005 Direct regulation of intestinal fate by Notch. *Proc. Natl Acad. Sci. USA* **102**, 12 443–12 448. (doi:10.1073/pnas.0505690102)
88. Theodosiou NA, Tabin CJ. 2003 Wnt signaling during development of the gastrointestinal tract. *Dev. Biol.* **259**, 258–271. (doi:10.1016/S0012-1606(03)00185-4)
89. Korinek V, Barker N, Moerer P, van Donselaar E, Huls G, Peters PJ, Clevers H. 1998 Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nat. Genet.* **19**, 379–383. (doi:10.1038/1270)
90. Roberts DJ. 2000 Molecular mechanisms of development of the gastrointestinal tract. *Dev. Dyn.* **219**, 109–120. (doi:10.1002/1097-0177(2000)9999:9999<::AID-DVDY1047>3.3.CO;2-Y)
91. Keding M, Simon-Assmann P, Bouziges F, Arnold C, Alexandra E, Haffen K. 1990 Smooth muscle actin expression during rat gut development and induction in fetal skin fibroblastic cells associated with intestinal embryonic epithelium. *Differentiation* **43**, 87–97. (doi:10.1111/j.1432-0436.1990.tb00434.x)
92. Apelqvist A, Ahlgren U, Edlund H. 1997 Sonic hedgehog directs specialised mesoderm differentiation in the intestine and pancreas. *Curr. Biol.* **7**, 801–804. (doi:10.1016/S0960-9822(06)00340-X)
93. Gao N, White P, Kaestner KH. 2009 Establishment of intestinal identity and epithelial–mesenchymal signaling by Cdx2. *Dev. Cell* **16**, 588–599. (doi:10.1016/j.devcel.2009.02.010)
94. Quaroni A, Tian JQ, Göke M, Podolsky DK. 1999 Glucocorticoids have pleiotropic effects on small intestinal crypt cells. *Am. J. Physiol.* **277**, G1027–G1040.
95. Su CG *et al.* 1999 A novel therapy for colitis utilizing PPAR-gamma ligands to inhibit the epithelial inflammatory response. *J. Clin. Invest.* **104**, 383. (doi:10.1172/JCI7145)
96. Dubuquoy L, Jansson EA, Deeb S, Rakotobe S, Karoui M, Colombel J-F, Auwerx J, Pettersson S, Desreumaux P. 2003 Impaired expression of peroxisome proliferator-activated receptor gamma in ulcerative colitis. *Gastroenterology* **124**, 1265–1276. (doi:10.1016/S0016-5085(03)00271-3)
97. Pradhan MP, Prasad NK, Palakal MJ. 2012 A systems biology approach to the global analysis of transcription factors in colorectal cancer. *BMC Cancer* **12**, 331. (doi:10.1186/1471-2407-12-331)
98. Scheitz CJF, Lee TS, McDermitt DJ, Tumber T. 2012 Defining a tissue stem cell-driven Runx1/Stat3 signalling axis in epithelial cancer. *EMBO J.* **31**, 4124–4139. (doi:10.1038/emboj.2012.270)
99. Levanon D *et al.* 2001 Spatial and temporal expression pattern of Runx3 (Aml2) and Runx1 (Aml1) indicates non-redundant functions during mouse embryogenesis. *Mech. Dev.* **109**, 413–417. (doi:10.1016/S0925-4773(01)00537-8)
100. Bornholdt S. 2008 Boolean network models of cellular regulation: prospects and limitations. *J. R. Soc. Interface* **5**(Suppl. 1), S85–S94. (doi:10.1098/rsif.2008.0132.focus)
101. Kauffman S, Peterson C, Samuelsson B, Troein C. 2003 Random Boolean network models and the yeast transcriptional network. *Proc. Natl Acad. Sci. USA* **100**, 14 796–14 799. (doi:10.1073/pnas.2036429100)
102. Cambridge SB, Gnäd F, Nguyen C, Bermejo JL, Krüger M, Mann M. 2011 Systems-wide proteomic analysis in mammalian cells reveals conserved, functional protein turnover. *J. Proteome Res.* **10**, 5275–5284. (doi:10.1021/pr101183k)
103. Chen Z *et al.* 2005 Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature* **436**, 725–730. (doi:10.1038/nature03918)
104. Theilgaard-Mönch K, Jacobsen LC, Borup R, Rasmussen T, Bjerregaard MD, Nielsen FC, Cowland JB, Borregaard N. 2005 The transcriptional program of terminal granulocytic differentiation. *Blood* **105**, 1785–1796. (doi:10.1182/blood-2004-08-3346)
105. Hollnagel A, Oehlmann V, Heymer J, Rütter U, Nordheim A. 1999 Id genes are direct targets of bone morphogenetic protein induction in embryonic stem cells. *J. Biol. Chem.* **274**, 19 838–19 845. (doi:10.1074/jbc.274.28.19838)
106. Dai X *et al.* 2013 Requirement for integrin-linked kinase in neural crest migration and differentiation and outflow tract morphogenesis. *BMC Biol.* **11**, 107. (doi:10.1186/1741-7007-11-107)
107. Bhatia B, Hsieh M, Kenney AM, Nahle Z. 2011 Mitogenic sonic hedgehog signaling drives E2F1-dependent lipogenesis in progenitor cells and medulloblastoma. *Oncogene* **30**, 410–422. (doi:10.1038/onc.2010.454)
108. Craven SE, Lim KC, Ye W, Engel JD, de Sauvage F, Rosenthal A. 2004 Gata2 specifies serotonergic neurons downstream of sonic hedgehog. *Development* **131**, 1165–1173. (doi:10.1242/dev.01024)
109. Wang X, Shi M, Li H, Wang N. 2011 Hedgehog signaling pathway and adipogenesis. *Chin. J. Cell Biol.* **8**, 017.
110. Suh JM, Gao X, McKay J, McKay R, Salo Z, Graff JM. 2006 Hedgehog signaling plays a conserved role in inhibiting fat formation. *Cell. Metab.* **3**, 25–34. (doi:10.1016/j.cmet.2005.11.012)