

Dynamics-based data science in biology

Jifan Shi ¹, Kazuyuki Aihara¹ and Luonan Chen ^{2,3,4,5,*}

Life science has long been a rich subject for research, and continues to develop at high speed. One of the major aims of life science is to study the mechanisms of various biological processes on the basis of biological big-data. Many statistics-based methods have been proposed to catch the essence by mining such data, including the popular category classification, variables regression, group clustering, statistical comparison, dimensionality reduction, and component analysis. However, these mainly elucidate static features or steady behavior of living organisms because of a lack of temporal data. A biological system is inherently dynamic, and with increasingly accumulated time-series data, there is a need for dynamics-based approaches based on physical and biological laws to reveal dynamic features or complex behavior of biological systems [1]. In this perspective, we review three dynamics-based data science approaches for studying dynamical bio-processes: namely, dynamical network biomarkers (DNB), landscapes of differentiation dynamics (LDD) and autorenservoir neural networks (ARNN). They are all data-driven or model-free approaches but based on the theoretical frameworks of nonlinear dynamics, that is, ordinary differential equations (ODE), partial differential equations (PDE) and artificial neural networks (NN), respectively. Figure 1A and B illustrates dynamical bio-processes and their omics data in biomedical fields, while Fig. 1C summarizes the three approaches of dynamics-based data science, which serve as typical

examples for studying biological systems from a data-driven dynamical perspective.

(i) **Providing early warning signals of pre-diseases/tipping-points by bifurcation theory.** The rapid development of high-throughput technology is allowing measurement of omics data to become more precise and less time-consuming. Identifying the tipping-point/pre-disease state is an urgent task for individuals in precision and prevention medicine. The DNB theory [2] is a dynamics-based data-driven approach to quantify critical states [1] during disease progression, as shown in the first column of Fig. 1C. A so-called pre-disease state (between normal and disease states) just before the disease state is defined as the state just before the bifurcation point of the dynamic system (ODE), which can be detected only by data based on three statistical conditions: (1) average standard deviations of DNB-members drastically increase; (2) Pearson correlation coefficients (PCCs) between DNB-members drastically increase in absolute values; (3) PCCs between DNB- and non-DNB-members drastically decrease. The three conditions are generically derived by exploring the low-dimensional feature of the center manifold just before the bifurcation of a fixed point (equilibrium) attractor in a nonlinear dynamical system. DNB not only quantitatively detects the ‘Wei-Bing’/pre-disease

state or predicts the imminent disease state for predictive medicine, but also provides early warning signals of other critical transitions, such as influenza outbreaks or even pandemics [2,3].

(ii) **Quantifying cell potential landscape by diffusion map theory with divergence theorem.** Single-cell sequencing data provide us the opportunity to study heterogeneity of tissues and capture features of different stages in cell differentiation processes. Various methods have been proposed to construct the differentiation tree, identify cells’ pseudo-time trajectory of differentiation, and study the transition rate between different cell-states [4,5]. In contrast to the traditional distance-based and entropy-based methods, LDD describes the cell differentiation as a non-equilibrium system by considering birth and death of cells on the basis of a source-sink Fokker-Planck equation (PDE) [6,7], as shown in the second column of Fig. 1C. By exploring diffusion map theory and divergence theorem, LDD can numerically estimate the potential of each cell cluster only with scRNA-seq data. In particular, LDD can not only identify the stem cell cluster with the highest pluripotency but also construct Waddington’s potential [8], without specific prior knowledge on stem cells and cell flow rates, which are required by the previous algorithms.

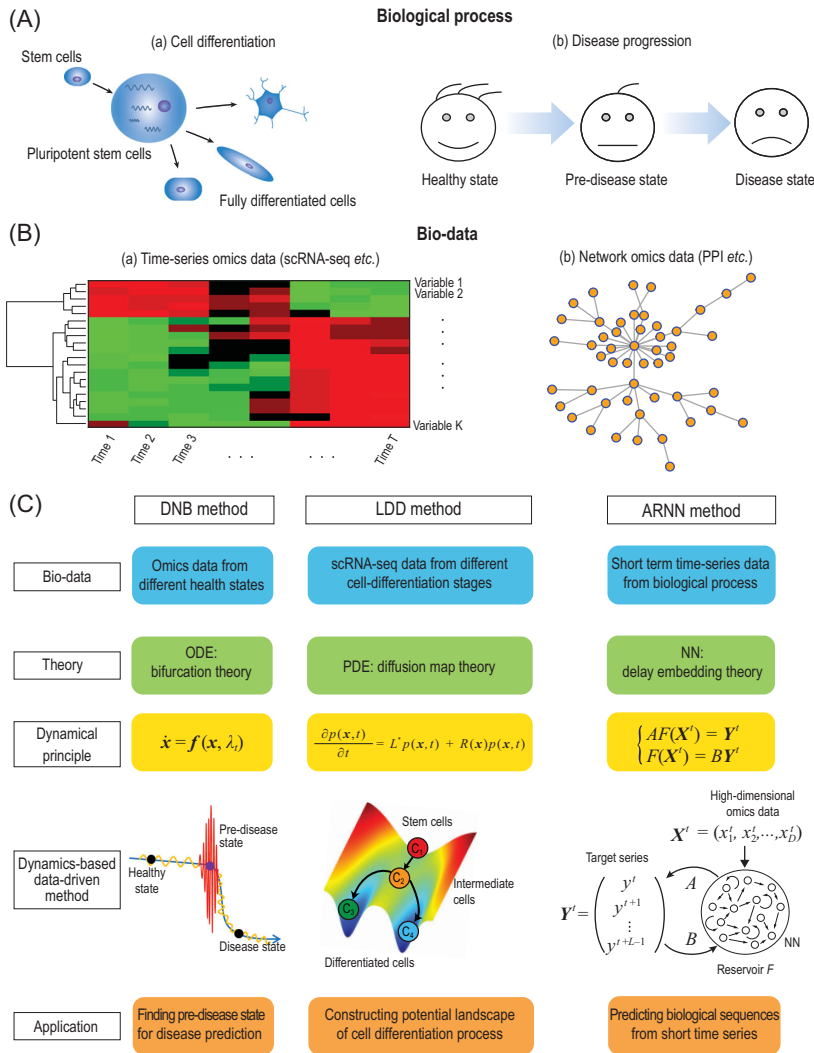


Figure 1. Dynamics-based data science approaches. (A) An illustration for a dynamic biological process, which usually includes a series of stages or states. (B) An illustration for omics-data extracted from experiments. (C) Three examples of dynamics-based data science approaches. The first column is the dynamic network biomarker (DNB) framework. DNB provides early warning signals for pre-disease/tipping-point detection from omics data of different health states, based on the bifurcation theory with ordinary differential equations (ODE). The second column is the landscape of a differentiation dynamics (LDD) framework, which distinguishes different cell types, provides pseudo-time trajectory for cell clusters (C_1 , C_2 , C_3 and C_4), and constructs a potential landscape for the cell differentiation process from single-cell RNA-sequencing (scRNA-seq) data during a cell differentiation process, based on the diffusion map theory of partial differential equations (PDE) named Fokker-Planck equations with source-sink terms. The third column is the autoreservoir neural network (ARNN) framework, which is able to predict short time-series from high-dimensional data by spatiotemporal information transformation (STI) and also significantly save computing resources, based on the delay-embedding and generalized embedding theorems of dynamic systems.

(iii) **Predicting time-series from short-term high-dimensional data by delay embedding theory.** Predicting future values from a short-term time-series is a difficult task but a hot topic. To overcome the small sample size problem, spatiotemporal

information transformation (STI) equations with nonlinear functions were derived based on the Takens' delay embedding theory for a nonlinear dynamic system [9] and its generalization [10]. ARNN [11] utilizes the STI equations by taking

advantage of the NN's ability to approximate any complicated nonlinear function, from the short-term high-dimensional data, as described in the third column of Fig. 1C. The STI equations represent the essential dynamics from the data, which equivalently enlarge the sample size and thus make the prediction accurate. In particular, different from the traditional reservoir structure with external dynamics [12], ARNN adopts nonlinear dynamics of a target system to efficiently transform the high-dimension spatial data into the low-dimension temporal data of a target variable, that is, future values or prediction [11]. ARNN not only could reduce the consumption of computing resources, but also avoids overfitting caused by much fewer training parameters. ARNN has been shown to have superior performances in predicting short-term time-series, such as gene expressions, daily cardiovascular disease admissions, wind speed, sea-level pressure, etc.

The efficiency of dynamical-based data science approaches on biological data has been demonstrated by the three methods above, which all show strong power in solving biological questions and are complementary to traditional statistics-based data science approaches. In addition, dynamical-based data science approaches can be applied to dynamical causality detection by exploring continuity of the cross mapping function between the observed variables. From a methodological viewpoint, we can summarize how to build a dynamics-based data-driven approach for studying biological dynamics as follows.

- (i) It generally starts from basic laws which a biological process obeys, e.g. proper dynamical equations for describing its time evolution. Note that how to narrow down to an appropriate dynamical model, e.g. ODE, PDE or NN, depends on the specific situation or prior information of the problem under study.
- (ii) Then, we need to derive the generic or essential statistic features, that

can be characterized by data, from such dynamical equations or models.

- (iii) With such features characterized by data, we can quantify various dynamical processes of biological systems based only on the measured data in a fully model-free manner.

Taken together, we conclude that the principles and advantages of dynamics-based data-driven approaches are explicable, quantifiable and generalizable. ‘Explicable’ indicates that every term in the dynamics-based data-driven approaches has its physical or biological sense. ‘Quantifiable’ ensures that the system can be measured by objective criteria and is comparable by quantitative indicators. ‘Generalizable’ says that the method can be improved by adding new factors or be generalized to other systems by proper modifications. In particular, dynamics-based data science approaches exploit the essential features of dynamical systems in terms of data, e.g. strong fluctuations near a bifurcation point, low-dimensionality of a center manifold or an attractor, and phase-space reconstruction from a single variable by delay embedding theorem, and thus are able to provide different or additional information to the traditional approaches, i.e. statistics-based data science approaches.

We believe that dynamical-based data science approaches will play an important role in systematic research in biology and medicine in future.

FUNDING

This work was supported by the National Key R&D Program of China (2017YFA0505500), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB38040400), the Japan Science and Technology Agency Moonshot R&D (JPMJMS2021), the Japan Agency for Medical Research and Development (JP20dm0307009), the Japan Society for the Promotion of Science KAKENHI (JP15H05707), and the National Natural Science Foundation of China (31930022, 31771476 and 12026608).

Conflict of interest statement. None declared.

Jifan Shi ¹, Kazuyuki Aihara¹ and Luonan Chen ^{2,3,4,5,*}

¹International Research Center for Neurointelligence, The University of Tokyo Institutes for Advanced Study, The University of Tokyo, Japan; ²Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, China; ³Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, China; ⁴School of Life Science and Technology, ShanghaiTech University, China and ⁵Key Laboratory of Systems Biology, Hangzhou Institute for Advanced Study, University of

Chinese Academy of Sciences, Chinese Academy of Sciences, China

*Corresponding author.

E-mail: lnchen@sibs.ac.cn

REFERENCES

- Scheffer M. *Nature* 2010; **467**: 411–2.
- Chen L, Liu R and Liu ZP *et al. Sci Rep* 2012; **2**: 342.
- Liu X, Chang X and Leng S *et al. Natl Sci Rev* 2019; **6**: 775–85.
- Schiebinger G, Shu J and Tabaka M *et al. Cell* 2019; **176**: 928–43.
- Saelens W, Cannoodt R and Todorov H *et al. Nat Biotechnol* 2019; **37**: 547–54.
- Weinreb C, Wolock S and Tusi BK *et al. Proc Natl Acad Sci USA* 2018; **115**: E2467–76.
- Shi J, Li T and Chen L *et al. PLoS Comput Biol* 2019; **15**: e1007488.
- Lang AH, Li H and Collins JJ *et al. PLoS Comput Biol* 2014; **10**: e1003734.
- Sauer T, Yorke JA and Casdagli M. *J Stat Phys* 1991; **65**: 579–616.
- Deyle ER and Sugihara G. *PLoS One* 2011; **6**: e18295.
- Chen P, Liu R and Aihara K *et al. Nat Commun* 2020; **11**: 4568.
- Lukoševičius M and Jaeger H. *Comput Sci Rev* 2009; **3**: 127–49.

National Science Review

8: nwab029, 2021

doi: [10.1093/nsr/nwab029](https://doi.org/10.1093/nsr/nwab029)

Advance access publication 12 February 2021

IMMUNOLOGY

Hunting field: insights on distribution pattern of bacteria and immune cells in solid tumors

Yang Li^{1,2,†}, Yuqi Wang^{1,†}, Xuefei Li^{1,2,*} and Chenli Liu^{1,2,*}

Bacterial cancer therapy, which was first applied in the clinic in 1868, has regained attention owing to the recent progress made in synthetic biology. Considering their easily manipulated genomes, preferential accumulation in tumors, and penetration abilities, bacteria have shown great therapeutic potential in tumor treatment. During

treatment, it has been found that bacteria in tumors lead to corresponding changes in the abundance as well as locations of a variety of cells and substances, especially immune cells, forming a unique distribution pattern. This has been suggested to contribute to the therapeutic effect of bacterial cancer therapy [1].

Generally, one to three days after the administration of *Salmonella*, *Clostridium*, *Escherichia* or *Pseudomonas* in mouse models, a relatively stable distribution pattern of bacteria and immune cells can be observed in tumors [2–4]. The stable distribution pattern (Fig. 1A) shares a common feature: bacteria mainly colonize the necrotic region of the tumor,

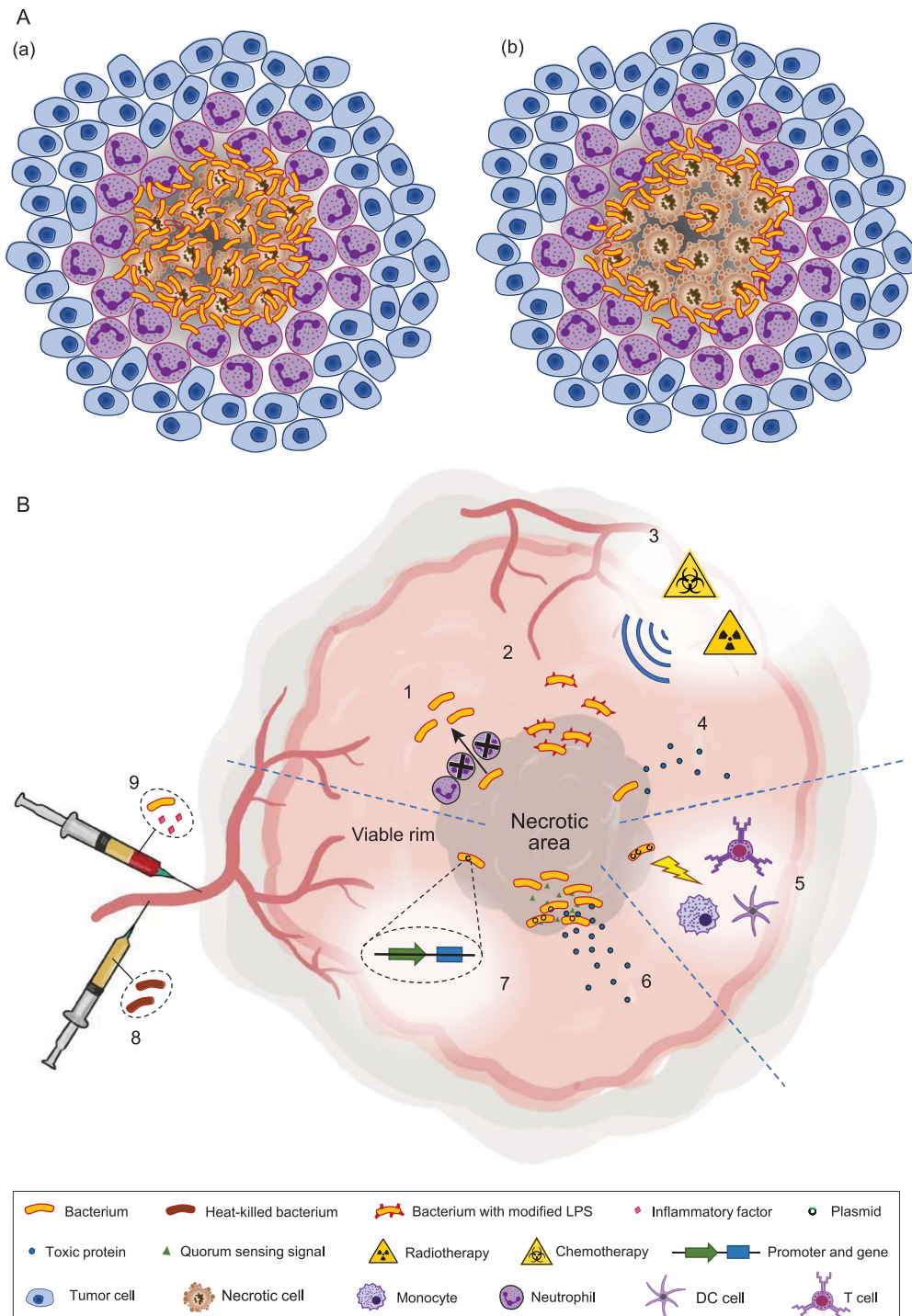


Figure 1. (A) The distribution pattern of bacteria in a tumor. Bacteria mainly colonize the necrotic/hypoxia area of a tumor, which can be divided into two modes: (a) evenly distributing throughout the necrotic area or (b) accumulating in the hypoxic area with a few colonies deeper in the necrotic area. Neutrophils gather around the necrotic area and form a ring to surround the bacteria, which sometimes overlaps with part of the bacterial area. (B) Proposals for optimizing bacterial therapy. The methods can be divided into three aspects. One is to target the viable area in the tumor by preventing the formation of the neutrophils ring (1), modifying the bacteria to have the potential of escaping the confinement of the neutrophil ring (2), combining bacteria with chemotherapy and radiotherapy (3) or designing bacteria to secrete drug proteins which could spread to the viable rim (4). Another is to introduce plasmids expressing tumor antigens/cytokines/immunostimulators/immunosuppressive/checkpoint inhibitor or other proteins with immunomodulatory activities into bacteria to enhance the bacterial stimulation on the immune system (5). The third is to reduce the side effects elicited by the bacteria and ensure safety by controlling the synthesis and secretion of toxic proteins or immuno-regulatory factors specifically within the tumor tissue by quorum-sensing system (6) or tumor-specific promoter (7), pre-exposing mice to heat-killed bacteria (8) or co-injecting the attenuated bacteria with inflammatory factors (9).

with neutrophils forming a ring-like structure surrounding the area of bacteria. Two modes of bacterial distribution are observed: an even distribution throughout the necrotic area (Fig. 1Aa), or accumulation in the hypoxic area in close proximity to the necrotic region, with a few colonies deeper in the necrotic area (Fig. 1Ab) [2,3,5–9].

The development of the intra-tumoral distribution pattern is a dynamic process with interactions among bacteria, tumor cells and the immune system. After transport into tumor tissues through the blood stream, bacteria colonize tumors, while simultaneously the concentrations of immune factors increase as well. This results in the formation and expansion of necrotic areas in tumors, while a rim of viable tumor cells is left on the periphery. Tumors grow and create an immune-privileged environment that protects cancer cells from being easily found and killed by immune cells. After the entry of bacteria, however, the ‘peace’ is broken. Like a hunting process, bacteria are the ‘rabbits’ running into the flush forest where they can hide well, and the innate immune cells are the ‘dogs’ chasing behind. When entering the forest, rabbits and dogs wake up the sleeping ‘tigers’ (adaptive immune cells) and the other dogs. Then they find that there are also many ‘sheep’ (cancer cells) hiding there, and more tigers may come into the forest. Therefore, tigers and dogs start to hunt both the sheep and the rabbits, forming a busy and crowded ‘hunting field’.

In the ‘hunting field’, both dogs and tigers contribute to the death of sheep. The invasion of bacteria in a tumor can promote the infiltration of a large number of innate as well as adaptive immune cells. Importantly, the exhausted effector immune cells can be re-activated. There is a possibility that specific types of bacteria have similar antigens with cancer cells, which can help to stimulate the immune cells that can recognize neoantigens. Furthermore, the death of cancer cells caused by bacteria and innate immune cells may expose the neoantigens to the adaptive immune system, which could further enhance the cancer-specific killing. Therefore, when, where and how the bacteria interact with the

immune system impacts the effectiveness of therapy.

This hunting field not only reflects the state of the tumor during treatment, but also affects the curative potential of bacterial treatment. Based on the distribution pattern, we can analyze and utilize the colonization mode of bacteria and overcome known limitations to optimize therapy. The recurrence of tumors after bacterial treatment is due to the proliferation of tumor cells within the viable tumor area, especially for large tumors [10]. This suggests taking additional measures to target the rim of viable cells (Fig. 1B1–4). Enhancing the antitumor effect of the immune system can be another potential method (Fig. 1B5), e.g. by increasing infiltration or anti-neoplastic activity of immune cells induced by bacteria. It is also crucial to balance toxicity and efficacy of bacterial therapy. Measures should be applied to reduce the side effects of bacteria to alleviate the harm on normal tissues, ensuring safety (Fig. 1B6–9).

Nevertheless, some questions require further exploration. This spatial pattern can be important for the survival of bacteria, but whether this prolonged existence of bacteria in tumors helps or limits the therapeutic effects still awaits an answer. It remains unclear why some genetically engineered bacteria show poor therapeutic effect and fail to induce a similar spatial pattern in a tumor. It is vital to know which abstracted appendages or gene products of bacteria contribute to the formation of a spatial pattern. In addition, although the immunology of the tumor microenvironment has been intensively studied, the behavior and effect of immune cells after bacteria have entered the tumor tissue are still obscure. Which immune cells or immune factors have crucial impacts on therapeutic effects remains to be elucidated.

Currently, the primary methods used in studies of intra-tumoral patterns are immunofluorescence and immunohistochemical staining of tumor sections. Therefore, only static images and snapshots of the spatiotemporal evolution can be captured. The continuous and dynamic characterization of different components in the whole tumor, before and after bacterial treatment, is lacking.

We should note that even for primary tumors, the spatiotemporal evolution of the tumor microenvironment is still an active area of study. How bacteria trigger the required cancer-killing by immune cells will be an important focus of future study. To achieve this goal, standardized and quantitative data acquisition and analysis are required. Specifically, we need to quantify and understand how bacteria help to attract and activate corresponding immune cells and how the immune cells interact with cancer cells. Since the interactions among bacteria and the tumor microenvironment are complex, mathematical models can be helpful for explorations of the detailed mechanisms underlying the pattern evolution. More importantly, such exploration will be helpful for indicating the potential directions of strain modifications and possible treatment strategies. The challenges are not limited to bacteria engineering, and mechanistic studies on the spatiotemporal evolution of patterns will shed light on the rational engineering of the tumor microenvironment for a safe and effective therapy.

FUNDING

This work was financially supported by the National Key R&D Program of China (2018YFA0902701), the Strategic Priority Research Program of Chinese Academy of Sciences (XDB29050501), the National Natural Science Foundation of China (32025022, 3201101136 and 32000886), and the Shenzhen Grants (JCYJ20170818164139781, KQTD2015033117210153).

Conflict of interest statement. None declared.

Yang Li^{1,2,†}, Yuqi Wang^{1,†}, Xuefei Li^{1,2,*} and Chenli Liu^{1,2,*}

¹CAS Key Laboratory for Quantitative Engineering Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China and ²University of Chinese Academy of Sciences, China

*Corresponding authors. E-mails: cl.liu@siat.ac.cn; xuefei.li@siat.ac.cn

[†]Equally contributed to this work.

REFERENCES

1. Zhou S, Gravekamp C and Bermudes D *et al.* *Nat Rev Cancer* 2018; **18**: 727–43.

2. Weibel S, Stritzker J and Eck M *et al. Cell Microbiol* 2008; **10**: 1235–48.
3. Ganai S, Arenas RB and Sauer JP *et al. Cancer Gene Ther* 2011; **18**: 457–66.
4. Forbes NS. *Nat Rev Cancer* 2010; **10**: 785–94.
5. Jiang SN, Park SH and Lee HJ *et al. Mol Ther* 2013; **21**: 1985–95.
6. Ryan RM, Green J and Williams PJ *et al. Gene Ther* 2009; **16**: 329–39.
7. Crull K, Rohde M and Westphal K *et al. Cell Microbiol* 2011; **13**: 1223–33.
8. Westphal K, Leschner S and Jablonska J *et al. Cancer Res* 2008; **68**: 2952–60.
9. Beckman RA, Schemmann GS and Yeang C-H. *Proc Natl Acad Sci USA* 2012; **109**: 14586–91.
10. Chen F, Zang Z and Chen Z *et al. Biomaterials* 2019; **214**: 119226.

National Science Review
8: nwab023, 2021
doi: 10.1093/nsr/nwab023
Advance access publication 6 February 2021