Briefings in Bioinformatics, 2024, **25(6)**, bbae520
https://doi.org/10.1093/bib/bbae520
Problem Solving Protocol

Cancerous time estimation for interpreting the evolution of lung adenocarcinoma

Yourui Han[1,‡], Bolin Chen[1,2,‡,*], Jun Bian[3], Ruiming Kang[4], Xuequn Shang[1,2]

[1]School of Computer Science, Northwestern Polytechnical University, Xi'an 710012, China
[2]Key Laboratory of Big Data Storage and Management, Northwestern Polytechnical University, Ministry of Industry and Information Technology, Xi'an 710012, China
[3]Department of General Surgery, Xi'an Children's Hospital, Xi'an Jiaotong University Affiliated Children's Hospital, Xi'an 710003, China
[4]Rewise (Hangzhou) Information Technology Co., LTD, Hangzhou 310000, China

*Corresponding author: blchen@nwpu.edu.cn
‡Yourui Han and Bolin Chen contributed equally.

Abstract

The evolution of lung adenocarcinoma is accompanied by a multitude of gene mutations and dysfunctions, rendering its phenotypic state and evolutionary direction highly complex. To interpret the evolution of lung adenocarcinoma, various methods have been developed to elucidate the molecular pathogenesis and functional evolution processes. However, most of these methods are constrained by the absence of cancerous temporal information, and the challenges of heterogeneous characteristics. To handle these problems, in this study, a patient quasi-potential landscape method was proposed to estimate the cancerous time of phenotypic states' emergence during the evolutionary process. Subsequently, a total of 39 different oncogenetic paths were identified based on cancerous time and mutations, reflecting the molecular pathogenesis of the evolutionary process of lung adenocarcinoma. To interpret the evolution patterns of lung adenocarcinoma, three oncogenetic graphs were obtained as the common evolutionary patterns by merging the oncogenetic paths. Moreover, patients were evenly re-divided into early, middle, and late evolutionary stages according to cancerous time, and a feasible framework was developed to construct the functional evolution network of lung adenocarcinoma. A total of six significant functional evolution processes were identified from the functional evolution network based on the pathway enrichment analysis, which plays critical roles in understanding the development of lung adenocarcinoma.

Keywords: cancerous time estimation; cancer evolution; oncogenetic graph; functional evolution process; lung adenocarcinoma.

Introduction

Lung adenocarcinoma is one of the most common subtypes of lung cancer. Like other cancers, the occurrence and development of lung adenocarcinoma exhibit a complex and gradual evolutionary process, where cancer cells continually adapt to and transform their environment after experiencing accidental mutations [1–3]. This high adaptability of cancer cells makes the traits exhibited by patients strongly heterogeneous, accompanied by multiple distinct pathological stages along their evolutionary processes [4]. Exploring the evolutionary process of lung adenocarcinoma is conducive to studying the molecular pathogenesis and functional evolution processes, which also promote the development of diagnosis biomarkers and therapeutic methods.

To explore the molecular pathogenesis, some oncogenetic modeling methods have been developed for cross-sectional samples collected through biopsies of patients. These methods assume different individuals' genetic alteration profiles are independent observations from the same multivariate stochastic process, and construct directed graphical models that reflect the dependencies or causalities between genetic alterations among the patient population. Various conceivable forms of oncogenetic graphs have been attempted, such as in the forms of (i) a tree (e.g. oncotrees [5]); (ii) a forest structure (e.g. Mtreemix [6] and CAPRESE [7]); (iii) a directed acyclic graph (e.g. Conjunctive Bayesian Networks [8], DiProg [9], Bayesian Mutation Landscape [10], CAPRI [11], and TO-DAG [12]); or (iv) a general directed graph without any structural constraints (e.g. NAM [13] and Mutual Hazard Networks [14]). However, the biggest limitation of current oncogenetic models lies in their failure to explicitly incorporate time variables in the algorithm design, primarily due to the unknown regarding the progression time of sample. If the temporal information can be recovered, it would be crucial in the oncogenetic analysis as it can reflect the trends and patterns of the event accumulation process in cancer development, potentially enhancing the reliability of oncogenetic models [15].

To explore the functional evolution processes, a prevalent method, which is developed for samples' gene expression RNA-seq profile collected through biopsies from patients, is to identify differentially expressed genes between control and patient samples of different pathological stages, followed by the pathway enrichment analyses [16, 17]. However, many of those studies ignore the fact that different pathological stages of cancer are often highly different from each other. The mixture of those heterogeneous samples may lack the efficiency of identifying

Received: February 25, 2024. **Revised:** August 19, 2024. **Accepted:** October 2, 2024

the real differentially expressed genes, and loss the opportunity to analyze the dynamic evolution process of cancer [18].

To model the evolutionary process of tumors, the epigenetic landscape [19] is utilized to characterize the evolutionary process of tumor cells from cell lines [20]. Then, methods for engrafting xenografts that derived from patient tumors into model organisms are proposed [21–23], and the cancer fitness landscape [24] is adopted to characterize the fitness changes of clone composition during evolutionary process to predict evolutionary trajectories, perform drug screening, or design therapeutic windows [25, 26]. However, data from cell lines and xenografts lack the natural cellular environment and may have limited similarities with the disease progression observed in patients, since the tissue microenvironment significantly influences cancer evolution [27].

With the rapid development of high-throughput biological experiment, more and more lung adenocarcinoma genomic data have been accumulated, which provides the possibility to reconstruct the evolutionary process of lung adenocarcinoma from a patient data-driven perspective, and deals with the problems of temporal information missing, heterogeneous characteristics ignoring among distinct pathological stages, and the indirect insight into the evolutionary process. To overcome those challenges, in this study, a patient quasi-potential landscape was proposed to estimate the cancerous time as progression time, which can reflect the time of phenotypic states' emergence during the evolutionary process. On the one hand, cancerous time can be used directly in the oncogenetic analysis to identify the oncogenetic graph, which can reveal the molecular pathogenesis and as common evolutionary patterns to interpret the evolution of cancer. On the other hand, cancerous time can re-divided samples into evolutionary stages rather than pathological stages, which can reduce the impact of cancer heterogeneity and identify the real differentially expressed genes to analyze the dynamic evolution process of cancer.

The gene expression RNA-seq profile of the patient was assumed as independent observation from the same multivariate stochastic process. Using the assumption of branching evolution, the evolutionary process of lung adenocarcinoma is graphically described as a hillside terrain, where healthy people will be at the top of the terrain, and driven by genetic mutation and natural selection, they will move along different routes on the hillside to different phenotypic states of patients. The hillside-like topography is denoted as the patient quasi-potential landscape of lung adenocarcinoma evolutionary process, and each point (x, y, z) on this landscape is represented as a phenotypic state of lung adenocarcinoma. To construct this landscape, the layer-wise narrowing down process was first employed to obtain differentially expressed genes strongly related to survival (*DESGs*) as features of the patient sample. After this, similar samples were aggregated as the same phenotypic state, which are projected on the patient quasi-potential landscape by the self-organizing constraint map method and the quasi-potential entropy. The self-organizing constraint map method was adopted to determine the coordinate (x, y) of the sample, which is the "identity label" of the phenotypic state. And the quasi-potential entropy was defined to calculate the quasi-potential of the sample, and the coordinate z of the phenotypic state is calculated as the mean of all samples mapped to that phenotypic state. Considering that there are still some phenotypic states that do not correspond to the observed samples after such spatial mapping, k-diminishing nearest neighbor method and biharmonic spline interpolation method were adopted to restore these unsampled phenotypic states according to the information of neighbor phenotypic states that have been sampled. The workflow is illustrated in Figure 1(a)–(d).

Moreover, as the patient's progression time of cancer is unknown and the patient quasi-potential landscape vividly reflects the specific position of the patient in the evolutionary process, the cancerous time of the phenotypic state during evolutionary process is estimated as the Euclidean distance between the phenotypic state coordinates and the cluster center coordinates of healthy people in the landscape. The longer the cancerous time, the longer the patient's actual progression time may also be. To verify the effectiveness of the cancerous time, the molecular pathogenesis and function evolution process of lung adenocarcinoma were explored by cancerous time. Firstly, the deep search and tree merging method were utilized to identify oncogenetic graphs by sorting patient somatic mutation data according to cancerous time. The workflow is illustrated in Figure 1(e). Secondly, a feasible framework was developed to identify functional evolution processes according to cancerous time. The workflow is illustrated in Figure 1(f). The source code of our method can be accessed at "https://github.com/YR1108/Cancerous-Time-Estimation-for-Interpreting-the-Evolution-of-Lung-Adenocarcinoma.git".

Materials and methods

Data sources

In order to depict the patient quasi-potential landscape of lung adenocarcinoma evolutionary process, the gene expression dataset was obtained from TCGA-LUAD, which includes 503 cases and 560 samples. The number of samples of normal, stages I, II, III, and IV, is 57, 273, 120, 84 and 26, respectively, and the samples are denoted as $s_i, i = 1, \ldots, 560$. Among them, 498 samples have somatic mutation data. All these data were retrieved from the TCGA Data Portal (https://tcga-data.nci.nih.gov/tcga).

The construction of the patient quasi-potential landscape

Layer-wise narrowing down process

Gene expression profiles, considered as high-dimensional datasets, encompass a multitude of features; however, the majority of these features do not contribute to the cancerous process. In order to find a set of cancerous process related genes for constructing the patient quasi-potential landscape, a layer-wise narrowing down process is conducted to identify differentially expressed genes strongly related to survival, which is denoted as $DESGs = \{gene_1, gene_2, \ldots, gene_m\}$. The first layer is differentially expressed analysis, and the analysis of variance method is employed to obtain differentially expressed genes *DEGs*. The second layer is survival analysis to generate a set of survival-related genes *SGs*. The intersection of *DEGs* and *SGs* is the genes set of *DESGs*.

Notably, to overcome individual specificity and implicitly incorporate the temporal information contained within the pathological stages, the sample $s_i = (g_{i,1}, g_{i,2}, \ldots, g_{i,m})$ is processed by stage normalization, $g_{i,j}$ is the gene expression level of $gene_j$ in *DESGs*. The gene expression level of the sample s_i after normalization is denoted as $\tilde{s}_i = (\tilde{g}_{i,1}, \tilde{g}_{i,2}, \ldots, \tilde{g}_{i,m})$. (The details of *DESGs* and the stage normalization method are provided in Section 1.1 of the "Supplementary Document".)

Spatial mapping and quasi-potential calculating

As similar samples should be aggregated as the same phenotypic state to construct the patient quasi-potential landscape, the

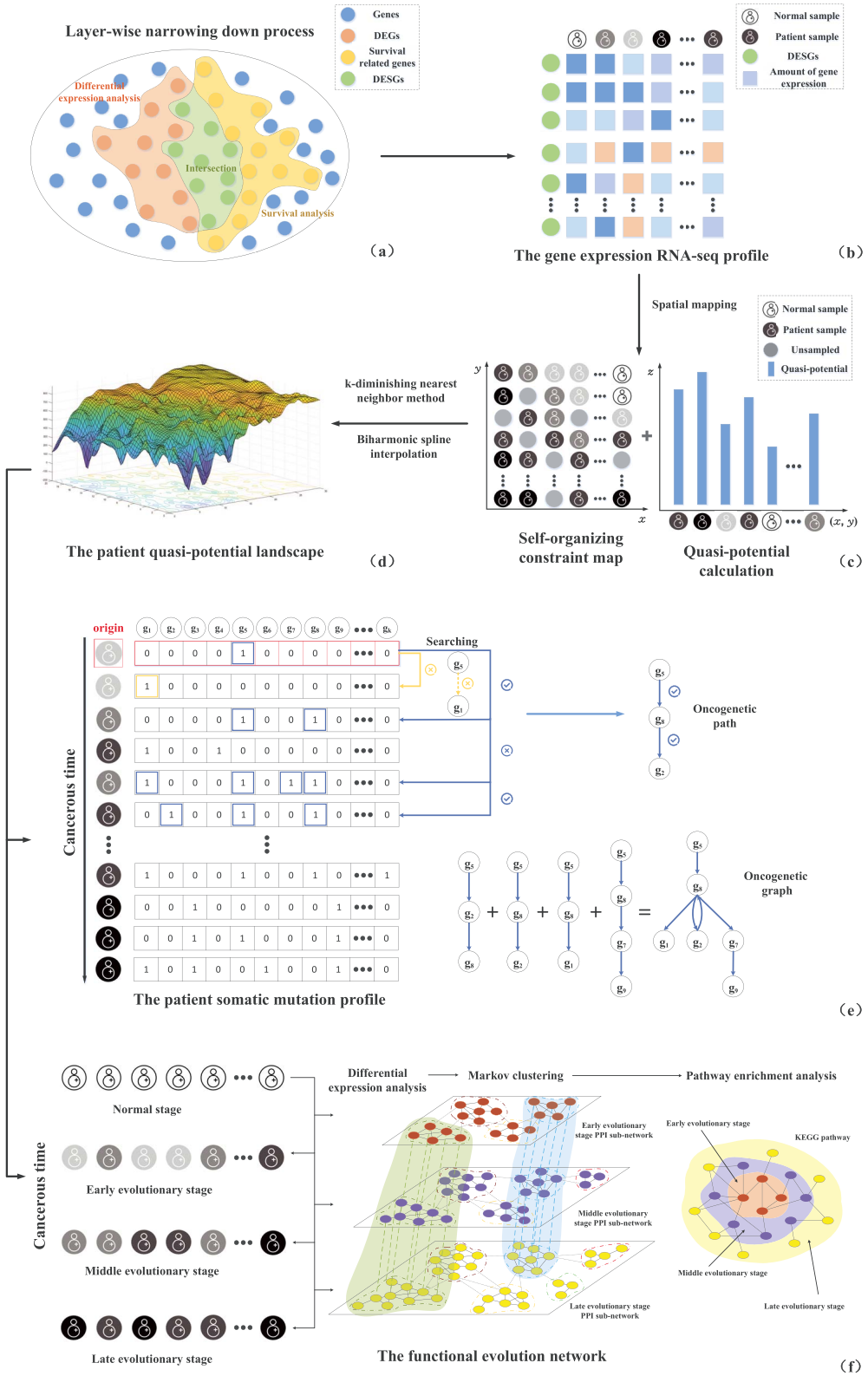

Figure 1. The overall framework of cancerous time for interpreting the evolution of lung adenocarcinoma. (a) The layer-wise narrowing down process was used to obtain *DESGs*; (b) The gene expression RNA-seq profile was chosen as the basic data to construct the landscape; (c) Spatial mapping and quasi-potential entropy: the self-organizing constraint map method was adopted to determine the coordinate (x, y) of the sample, which is the "identity labels" of the phenotypic state. And the quasi-potential entropy was defined to calculate the quasi-potential of the sample, and the coordinate z of phenotypic state is calculated as the mean of all samples mapped to that phenotypic state; (d) The k-diminishing nearest neighbor method and biharmonic spline interpolation method were adopted to restore these unsampled phenotypic states according to the information of neighborhood samples; (e) Sorting patient somatic mutation data with cancerous time, the deep search and tree merging method were utilized to identify oncogenetic graphs; (f) A feasible framework was developed to identify functional evolution processes according to cancerous time.

self-organizing constraint map method is designed to facilitate small-scale clustering, which achieves a balance between inter-sample variability and inter-sample heterogeneity.

The self-organizing constraint map method trains a neural network with grid structure based on unsupervised learning method. A neuron in the neural network is represented as a phenotypic state of lung adenocarcinoma, and the geographic coordinate (x, y) of the neuron is the "identity label" of the phenotypic state. The set of samples that map to the same neuron, i.e. to the same phenotypic state, are denoted as $D_{x,y} = \{\tilde{s}_1, \ldots, \tilde{s}_{n_{x,y}}\}$, and $n_{x,y}$ is the number of these samples.

Meanwhile, the quasi-potential entropy $e(\tilde{s}_i)$ is defined to compute the quasi-potential for each sample:

$$e(\tilde{s}_i) = M - \left(-\sum_{j=1}^{m} \tilde{g}_{i,j} * \log_2 \tilde{g}_{i,j}\right) = M + \sum_{j=1}^{m} \tilde{g}_{i,j} * \log_2 \tilde{g}_{i,j},$$

where M is a positive integer. Then the coordinate $z(x, y)$ of phenotypic state (x, y) is calculated as the mean of all samples mapped to that phenotypic state:

$$z(x, y) = \frac{1}{n_{x,y}} \sum_{i=1}^{n_{x,y}} e(\tilde{s}_i).$$

Thus, diverse samples will be mapped to distinct phenotypic states on the landscape, and their respective geographic coordinates (x, y, z) will be ascertained. It is noteworthy that this mapping does not constitute a bijection, as similar samples will be ascribed to the same phenotypic state. (The details of the self-organizing constraint map method are provided in Section 1.2 of the "Supplementary Document").

Reduction of 3D hillside topography

After spatial mapping and quasi-potential calculating, there are still some phenotypic states that do not correspond to the observed patient samples, i..e. $D_{x,y} = \emptyset$. Thus the k-diminishing nearest neighbor method is adopted to restore the coordinate z of these unsampled phenotypic states according to the information of neighbor phenotypic states that have been sampled. Then, the corner property is eliminated by biharmonic spline interpolation while maintaining smoothness and second-order continuity to generate the patient quasi-potential landscape. (The details of the k-diminishing nearest neighbor method and biharmonic spline interpolation are provided in Sections 1.3 and 1.4 of the "Supplementary Document", respectively).

The identification of oncogenetic graphs

The estimation of cancerous time

The cancerous time $T(\tilde{s}_i)$ of the sample $\tilde{s}_i \in D_{x,y}$ during cancer evolutionary process is estimated as the Euclidean distance $d_{ct}(\tilde{s}_i)$ between the phenotypic state's geographic coordinate (x, y, z) and the cluster center's geographic coordinate (x_{hc}, y_{hc}, z_{hc}) of healthy people in the landscape:

$$d_{ct}(\tilde{s}_i) = \|(x, y, z) - (x_{hc}, y_{hc}, z_{hc})\|.$$

To be more specific, as cancerous time $T_{\tilde{s}_i}$ is the pseudo time used to represent how long a sample has been cancerous, it is mainly used to compare the sequence of samples in the lung adenocarcinoma evolutionary process.

The selection of genes in oncogenetic graphs

To select genes for the generation of oncogenetic graphs, the CancerMapp pipeline [28] is adopted, which applies a statistical approach to identify significant changes in gene mutations along a progression model inferred from expression profiles. However, since the validity of cancerous time of lung adenocarcinoma evolutionary process needs to be verified, current clinical application of oncogenic driver gene mutations in lung adenocarcinoma is directly selected from the latest literature [29]. There are 17 genes selected: TP53, KRAS, EGFR, RET, ROS1, ALK, MET, ERBB2, BRAF, HRAS, RIT1, NRAS, NRG1, NRG3, NTRK1, NTRK2, and NTRK3.

The generation of oncogenetic graphs

Sorting patient somatic mutation data with cancerous time, the search origins are selected from the top three percent of samples. The oncogenetic paths are generated by searching in the direction of increasing cancerous time. For oncogenetic path reconstruction from somatic mutations made three assumptions: (i) no mutation occurs twice in the course of cancer evolution; (ii) no mutation is ever lost; and (iii) mutations are acquired consecutively over time in a gradual manner. In other words, the latter sample must contain all the mutations of the former sample, and the number of mutations can only be increased by one in the searching. Then tree merging method is utilized to merge oncogenetic paths for the generation of oncogenetic graphs.

The Identification of functional evolution processes

Division of evolutionary stage and differential expression analysis

Patients are evenly re-divided into early, middle, and late evolutionary stage through cancerous time. Differentially expressed genes between control samples and early, middle, and late evolutionary stage samples are screened using the limma algorithm [30]. We calculate *log2FC* of each gene between two groups. Only the genes that meet p-value < 0.05 and $|\ log2FC\ | \geq 3$ are considered as differentially expressed genes.

Functional evolution network between adjacent evolutionary stages

STRING (Search Tool for the Retrieval of Interacting Genes) database [31] is utilized to obtain the Protein-protein interactions (PPIs) between differentially expressed genes identified above. Three Protein-protein interaction (PPI) sub-networks are constructed by selecting PPIs scores larger than 0.4, and those sub-networks are then visualized in Cytoscape [32] (Supplementary Figs. 5–7). Identifying functional modules in a PPI network is important for understanding the functional evolution processes and complex molecular mechanisms of cancer. Markov clustering method [33] is employed to identify functional modules(criteria: inflation parameter=3) in three PPI sub-networks, respectively. The functional evolution network is constructed by taking individual modules as vertices. Modules of different stages of the PPI sub-networks may share the same differentially expressed genes, which makes it possible to calculate overlapping score between modules of adjacent evolutionary stages. The module relationship between adjacent evolutionary stages as edges of the functional evolution network.

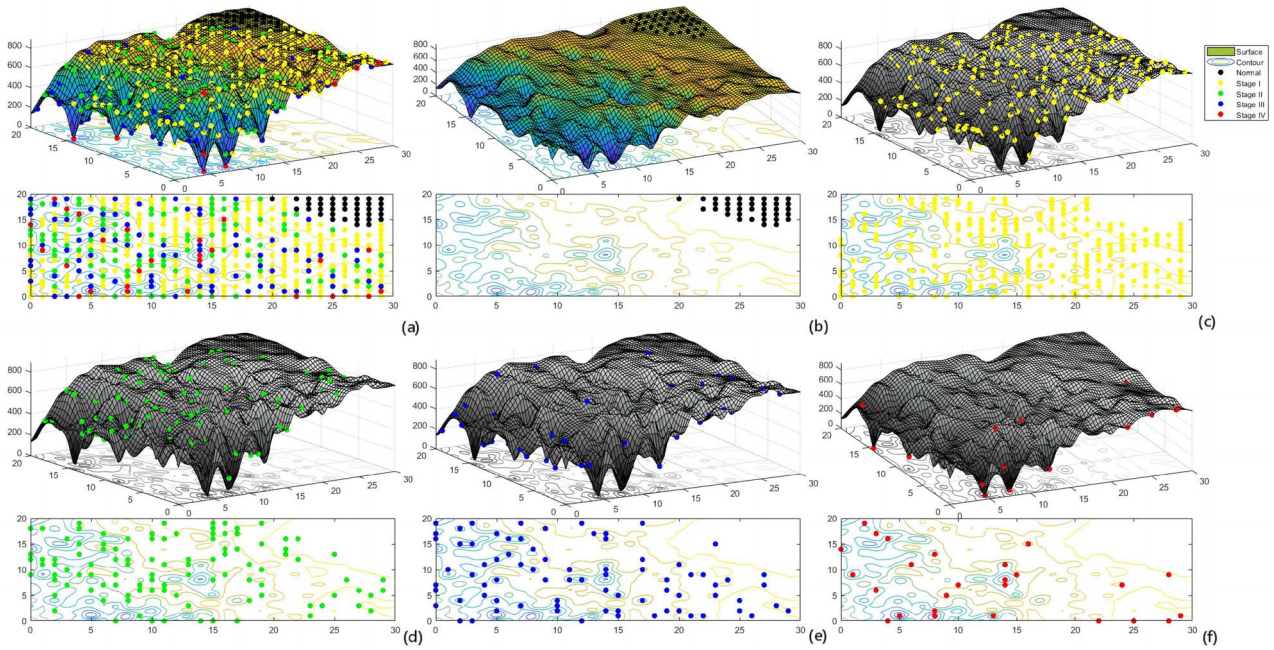

Figure 2. The patient quasi-potential landscape with stage information. (a) The patient quasi-potential landscape of lung adenocarcinoma evolutionary process and its contour with stage information; (b)–(f) The distribution of patients in the landscape and contour with different stage. In the following order: normal, stage I, stage II, stage III, and stage IV.

Pathway enrichment analysis of functional evolution processes

The significant functional evolution processes in the functional evolution network are screened by the weight of the edges. The DAVID [34] is used to conduct pathway enrichment analysis of function evolution processes using the KEGG pathways. Cancerous time is used in the division of evolutionary stages, which makes differentially expressed genes presentable in adjacent evolutionary stages and can be used to characterize the changes of significant KEGG pathways at evolutionary stages.

Experiments and results

The basic features in patient quasi-potential landscape of lung adenocarcinoma evolutionary process

The patient quasi-potential landscape of lung adenocarcinoma evolutionary process and its contour are shown in Figure 2(a), whose samples are annotated using cancer stage information. In view of general state, the landscape really depicts healthy people are at the top of the terrain, and driven by genetic mutation and natural selection, they will move along different routes on the hillside to different phenotypic states of patients. The direction of evolution is from high quasi-potential to the low quasi-potential.

In view of concrete state, the different stages are mapped to the landscape separately, which are shown in Figure 2(b)–(f). Figure 2(b) shows normal samples with high quasi-potential are clustered in the highest corner of the landscape. Figure 2(c) shows stage I patients are mostly distributed in and around the peaks of the landscape. Figure 2(d) shows stage II patients are mostly distributed on the mountainside of the landscape and close to the peaks. Figure 2(e) shows stage III patients are mostly distributed on the mountainside of the landscape and close to the valleys. Figure 2(f) shows stage IV patients are mostly scattered in the valleys of the landscape.

Moreover, the threshold 840 ± 5, 700 ± 5, and 600 ± 5 are selected to divide the landscape into normal region, high, middle, and low quasi-potential region (Supplementary Fig. 1 in "Supplementary Document"). It is not difficult to see that patients in the early stage are mostly distributed in the high quasi-potential and middle quasi-potential region, while patients in the late stage are mostly distributed in the low quasi-potential region. And normal sample is mostly distributed in the normal region, which is shown in the Table 1.

In addition to the stage information, the subtype information of lung adenocarcinoma patients is also matched to the landscape, which is shown in Figure 3. Also, it is not difficult to see that patients of terminal respiratory unit (TRU) subtype are mostly distributed in the high quasi-potential region, while patients of proximal inflammatory (PI) and proximal proliferative (PP) (non-TRU) subtypes are mostly distributed in the middle and low quasi-potential region, which is shown in the Table 1. When comparing the distribution of subtypes and stages in the patient quasi-potential landscape, most patients of the TRU subtype have a lower tumor stage and high quasi-potential, while both PI and PP (nonTRU) subtypes show lower quasi-potential. This is consistent with the observation that the majority of TRU subtype patients are in an earlier tumor stage and exhibit lower proliferation capacities, while both PI and PP (nonTRU) subtypes indicate higher proliferation capacities [35, 36].

The oncogenetic graphs in patient quasi-potential landscape of lung adenocarcinoma

Given that the true progression time of the cancer remains unknown, the survival time of the sample was utilized to evaluate the accuracy and validity of the cancerous time estimation. To be more specific, if the cancerous time is prolonged, indicative of a longer duration of cancer progression, then under normal circumstances, the survival time tends to be abbreviated. Conversely, a relatively extended survival time frequently implies

Table 1. Statistical table of sample distribution in different quasi-potential regions.

Region	Normal	Stage I	Stage II	Stage III	Stage IV	PI	PP	Non-TRU(PP+PI)	TRU
Normal region	**97.2%**	0.4%	0%	0%	0%	0%	0%	0%	0%
High quasi-potential	2.8%	**45.9%**	26.2%	21.8%	23.1%	10.1%	21.8%	15.3%	**76.8%**
Middle quasi-potential	0%	29.3%	32.0%	19.2%	7.7%	31.9%	21.8%	27.4%	20.3%
Low quasi-potential	0%	24.4%	**45.6%**	**59.0%**	**69.2%**	**58.0%**	**56.4%**	**57.3%**	2.9%

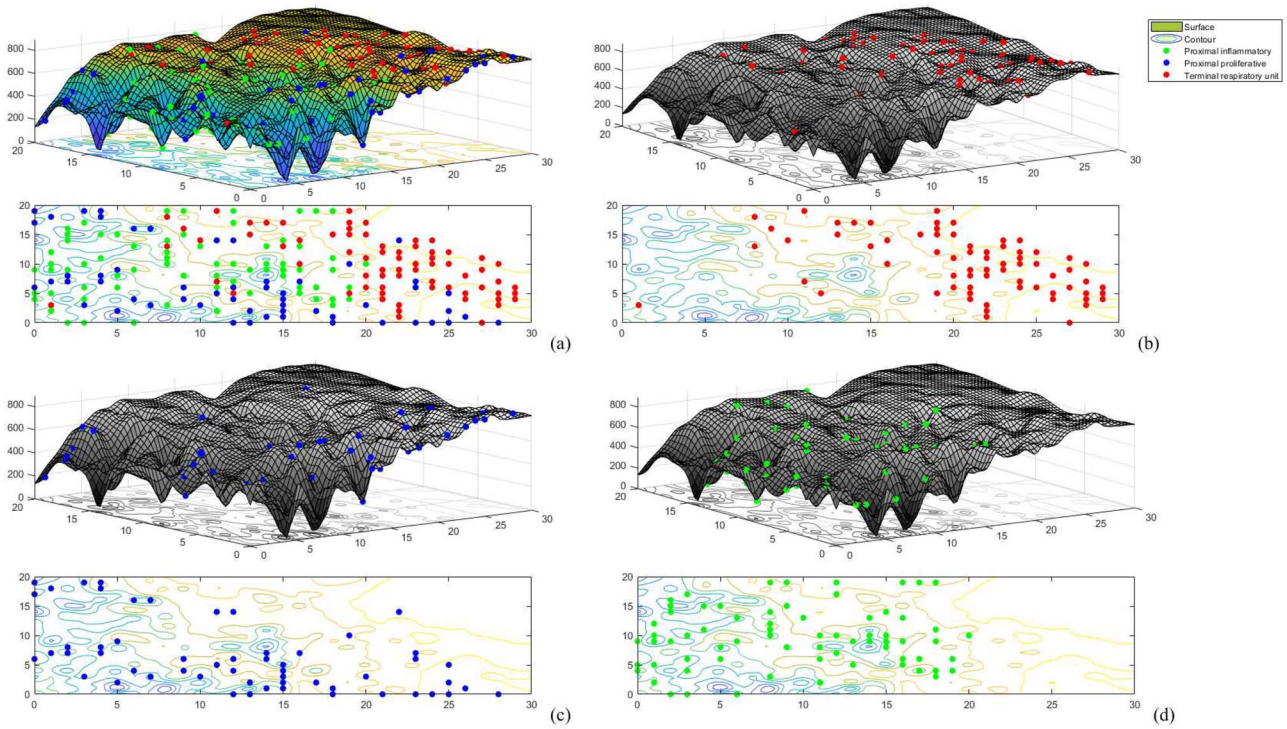

Figure 3. The patient quasi-potential landscape with subtype information. (a) The patient quasi-potential landscape of lung adenocarcinoma evolutionary process and its contour with subtype information; (b)–(d) The distribution of patients in the landscape and contour with different subtype. In the following order: TRU, PP, and PI.

a shorter cancerous time, suggesting that an inverse correlation should exist between survival time and cancerous time. The samples were sorted based on their cancerous time, utilizing a sampling interval of 4. Then the average cancerous time and survival time of the samples within each interval were calculated, followed by an analysis to investigate the correlation between these two variables (Figure 4(a)). The Pearson correlation coefficient r between the two variables is $r = -0.8086$, with a range of $r \geq -0.8$, indicating a strong negative correlation between survival time and cancerous time. And the $p\text{-}value = 0.0151$, demonstrating that this strong negative correlation is statistically significant.

Subsequently, in order to verify the robustness of cancerous time estimation, $\omega = 10\%, 20\%$ of the features were randomly selected to introduce perturbations for all patient samples. Perturbation is defined as the process of randomly adding or subtracting the selected feature by the proportion $\lambda = 10\%, 20\%, 30\%, 40\%$ of the maximum value of that feature across all samples. When $\lambda \leq 20\%$ and $\omega = 10\%, 20\%$, the correlation between cancerous time and survival time remains constant, as demonstrated in Figure 4(b). Furthermore, the permutation order of samples based on cancerous time does not undergo any change following perturbation, with a cosine similarity of 1 and a Pearson correlation coefficient of 1 observed between the cancerous time before and after perturbation, which is illustrated in the upper right subplot of Figure 4(b).

When $\lambda = 30\%$ and $\omega = 10\%, 20\%$, there were minor alterations observed in the cancerous time of the samples with lesser prevalence, which subsequently resulted in a slight variation in the correlation between cancerous time and survival time, which are shown in Figure 4(c) and Figure 4(d). As the degree of perturbation continues to escalate, a larger proportion of samples experience slight modifications in their cancerous time, nonetheless, the cancerous time and survival time still demonstrate a robust negative correlation(Figure 4(e) and Figure 4(f)). Furthermore, for each individual feature, a subset of samples is randomly selected and subjected to perturbation, ultimately revealing that alterations in the phenotypic state of the sample are observed exclusively when perturbations are made to a very limited number of key genes(Supplementary Fig. 2).

Sorting patient somatic mutation data with cancerous time, the deep search and tree merging methods were utilized to identify oncogenetic graphs. A total of 498 samples are measured for both gene expression RNA-seq profiles and somatic mutations. Among the top 3% samples in time sorting, EGFR, KRAS, and BRAF are chosen as origin to search the oncogenetic paths, and 39 different oncogenetic paths are searched, which is shown in Figure 5. To be more specific, there are a total of 473 oncogenetic

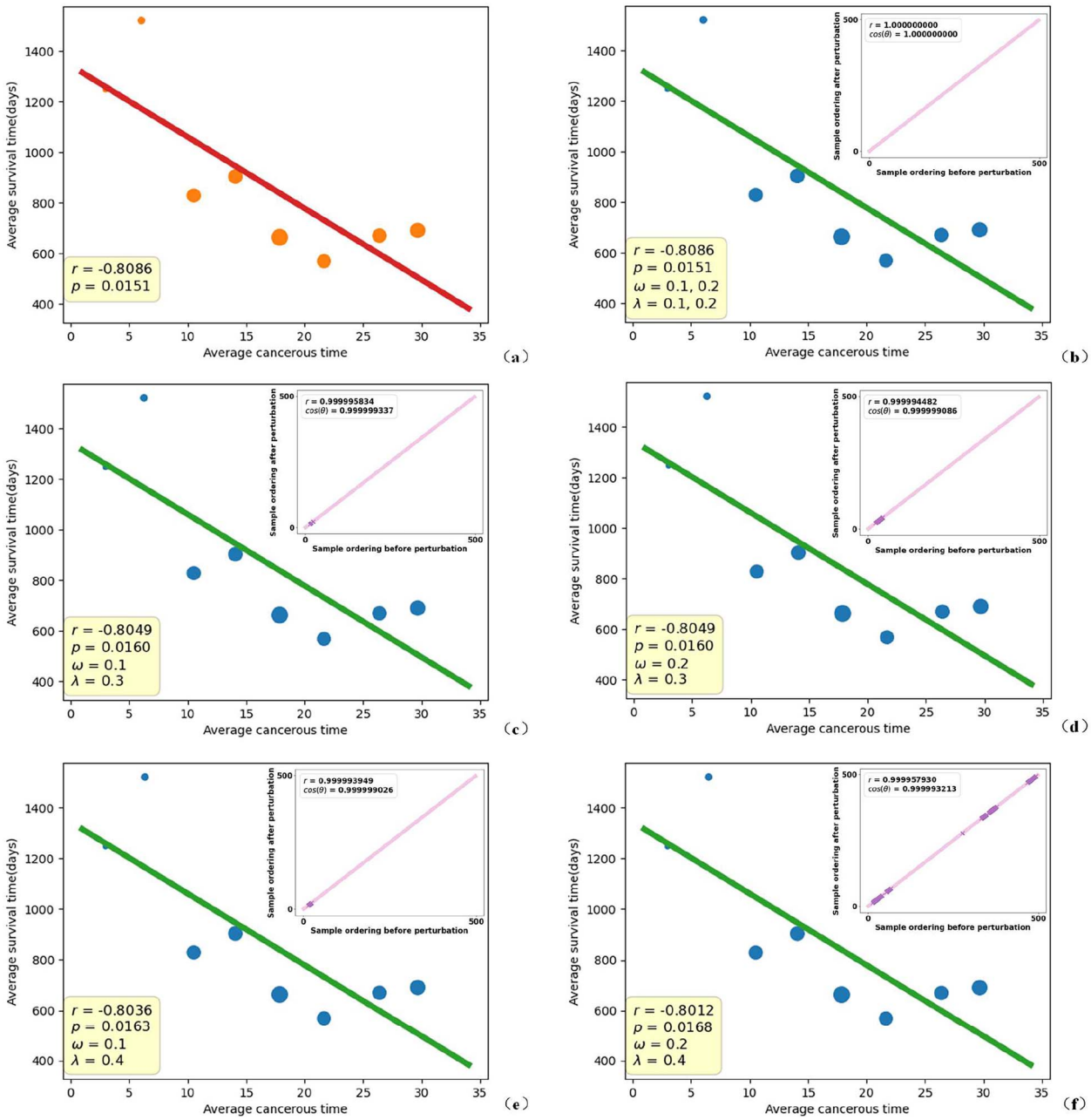

Figure 4. The correlation analysis between cancerous time and survival time. (a) The correlation analysis between cancerous time and survival time before the perturbation, the size of the discrete points in the figure is positively correlated with the number of samples within the corresponding sampling interval; (b)–(e) The correlation analysis between cancerous time and survival time after the different perturbations. The subplot embedded within each figure is employed to depict the variation in the permutation order of the samples, which is determined by the cancerous time, both prior to and subsequent to the perturbation. If the order remains unchanged, it is represented by a pink diagonal line, whereas if a sample in a specific order undergoes a change, a purple cross is plotted at the corresponding position on the diagonal line.

paths; some oncogenic paths are visualized in Supplementary Fig. 3. Among them, KRAS and EGFR oncogenetic paths accounted for the majority proportion, with 311 KRAS oncogenetic paths and 125 EGFR oncogenetic paths. In fact, the genetic landscape of lung adenocarcinoma is indeed dominated by mutations in KRAS and EGFR, with mutations in each of these genes individually accounting for up to 30% of patients with advanced-stage lung adenocarcinoma, while BRAF with mutations accounting for 5% of patients [29, 37].

Then, these oncogenetic paths were merged to generate three oncogenetic graphs of lung adenocarcinoma, which are shown in Figure 6. Three oncogenetic graphs illustrate three evolutionary patterns. The EGFR/BRAF/KRAS oncogenetic graph shows that the lung adenocarcinoma evolutionary process has an evolutionary pattern where normal people firstly have a mutation in EGFR/BRAF/KRAS, then progress to a more severe pathological state with EGFR/BRAF/KRAS amplification and mutation in TP53 or other genes. These three evolutionary patterns exhibit similarities to molecular pathogenesis through a stepwise continuum, thereby providing ancillary evidence for the rationality of cancerous time estimation. The EGFR-mutated atypical adenomatous hyperplasia progresses to adenocarcinoma *in situ*, followed

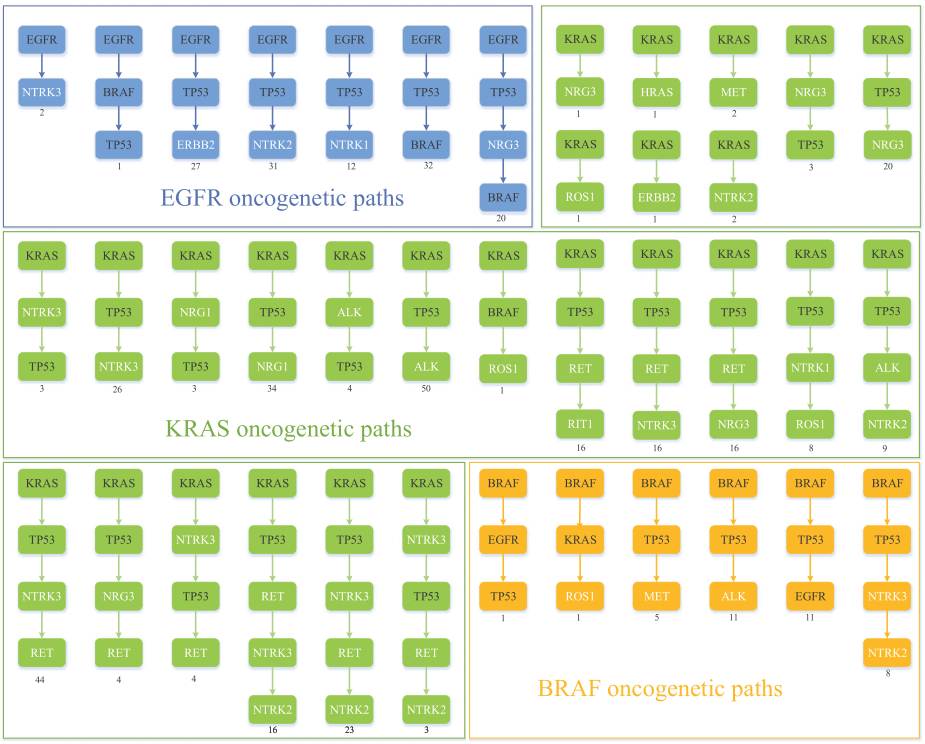

Figure 5. The oncogenetic paths under cancerous time. The number below the oncogenetic path represents the number of times the path appeared in the search.

by invasive adenocarcinoma after EGFR/TTF-1 amplification or the inactivation of TP53, while KRAS- or BRAF-mutated atypical adenomatous hyperplasia rarely progresses to adenocarcinoma *in situ* or invasive adenocarcinoma, but can do so after the inactivation of TP53 or other genomic events [38–40]. The molecular pathogenesis through a stepwise continuum is shown in Figure 7 from literature [40].

To further corroborate the availability of the patient quasi-potential landscape and assess the stability of the oncogenetic path modeling method, a subset constituting 90% of the patient samples augmented with Gaussian white noise was randomly selected to simulate new observational samples for analysis. Certainly, the mutation data of the original sample were utilized as the corresponding mutation data for the simulated sample. The neural network, trained through the self-organizing constraint map method, was utilized to ascertain the "identity label" of the phenotypic state corresponding to the sample within the simulated dataset. Subsequently, it was mapped onto the patient quasi-potential landscape to calculate the corresponding cancerous time, and generate the oncogenetic paths. Then the index $J(A,B) = \frac{2|A \bigcap B|}{|A|+|B|}$ was utilized to quantify the degree of similarity between two sets, and a comparison between the original oncogenetic paths and those based on simulated dataset is shown in Figure 8. Firstly, EGRF, BRAF, and KRAS consistently emerge as distinct mutations within the top three percentile of the simulated samples, rendering them viable candidates as initial genes for the exploration of oncogenic pathways. Secondly, it is observed that the majority categories of oncogenic paths derived from simulated dataset are encompassed within the set of original oncogenetic paths(Supplementary Fig. 4). Furthermore, the majority of oncogenic paths belonging to the new category are found to be subpath of the original oncogenetic path. For instance, KRAS → TP53 is the subpath of KRAS → TP53 → NRG3.

The functional evolution processes in patient quasi-potential landscape of lung adenocarcinoma

Based on the gene expression profile dataset and division of evolutionary stages, a total of 236, 523, and 878 differentially expressed genes were identified between control samples and early, middle, and late evolutionary stages, respectively, which are shown in Figure 9.

The corresponding differentially expressed gene sets are denoted as NE_DEGs, NM_DEGs, and NL_DEGs. A total of 222 differentially expressed genes are inherited from the early evolutionary stage to the middle evolutionary stage, and 481 differentially expressed genes are inherited from the middle evolutionary stage to the late evolutionary stage, which provides a basis for the analysis of functional evolution processes. Three PPI sub-networks were established based on three differentially expressed gene sets, respectively (Supplementary Figs 5–7).

A total of 25, 51, and 88 modules, which contain at least three proteins, were identified among three PPI networks obtained above. The number of overlapped genes between modules of adjacent evolutionary stages serves as the weights of the edges to construct the functional evolution network(Supplementary Fig. 8). There are seven prominent functional evolution processes in the network: NM1-NL1(Mod1), NE1-NM7-NL27(Mod2), NE2-NM5-NL12(Mod3), NE3-NM6-NL5(Mod4), NE4-NM14-NL18(Mod5), NE5-NM4-NL3(Mod6), and NE15-NM2-NL16(Mod7), whose sum of edges weight is greater than or equal to 10.

Mod1-Mod7 are enriched in a total of 30 pathways, as shown in the Table 2. To be more specific, KEGG pathways of Mod1 have great significance to the analysis of the functional evolution process of lung adenocarcinoma, and there are nine pathways whose $p\text{-}value \leq 0.05$. The pathway with the smallest $p\text{-}value$ is the cell cycle pathway, which is a fundamental process responsible for cell

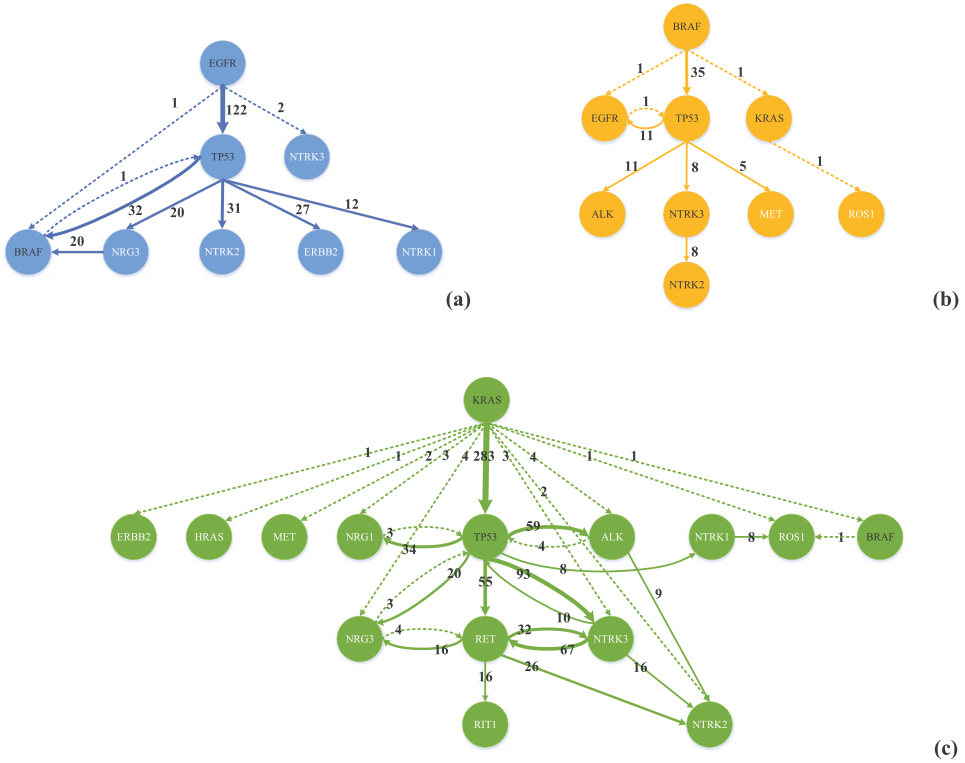

Figure 6. The oncogenetic graphs after oncogenetic paths merging. The dashed lines indicate less frequency of occurrence, the thickness of the solid line is positively related to the frequency of the edge. (a) The oncogenetic graph generated with EGFR as its origin; (b) The oncogenetic graph generated with BRAF as its origin; (c) The oncogenetic graph generated with KRAS as its origin.

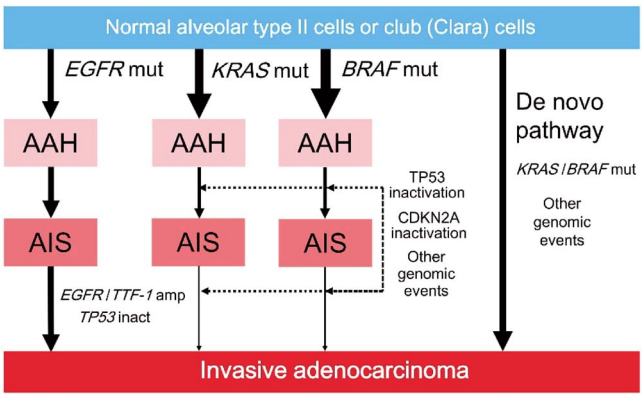

Figure 7. The molecular pathogenesis through a stepwise continuum. AAH develops from normal alveolar type II cells or club (Clara) cells after mutations in EGFR, KRAS, and BRAF. EGFR-mutated AAH progresses to AIS, followed by invasive adenocarcinoma after EGFR/TTF-1 amplification. KRAS- or BRAF-mutated AAH rarely progresses to AIS or invasive adenocarcinoma, but can do so after the inactivation of TP53/CDKN2A (p16) or other genomic events (dotted arrows). KRAS- or BRAF-mutated invasive adenocarcinoma may arise from normal lung epithelium via a *de novo* pathway that involves KRAS or BRAF mutations and other genomic events. The solid arrows indicate progression. AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma *in situ*; amp, amplification; inact, inactivation; mut, mutation. The figure and corresponding description are cited from literature [40].

growth, DNA replication, and cell division. Dysregulation of cell cycle control is commonly observed in cancer, including lung adenocarcinoma [41]. Genetic mutations in key cell cycle regulatory genes, such as TP53 (p53), Cyclin-dependent kinases (CDKs), and Cyclins, have been associated with lung adenocarcinoma [42, 43]. Cellular senescence is a state of irreversible growth arrest that may contribute to tumor suppression or tumor progression, depending on the context [44, 45]. Senescence-associated markers and regulators have been investigated in various cancers, including lung adenocarcinoma. Numerous investigations have indicated that p53 is a tumor suppressor gene [46] and plays a critical role in the development of tumors. It not only contributes to controlling cell cycle [46], but also to keeping chromosome stability [47] and mitochondrial genetic stability [48, 49].

In addition to Mod6, the KEGG pathways enriched in other Mods also play crucial roles in the development and progression of lung adenocarcinoma, e.g. the Wnt signaling pathway enriched in Mod3 regulates cell proliferation, differentiation, migration, genetic stability, and apoptosis in almost all tissues and organs [50, 51]. The PPAR signaling pathway enriched in Mod4 is known to play a crucial role in various biological processes, including energy metabolism, inflammation, and cell proliferation [52, 53]. The PI3K-Akt pathway enriched in Mod7 is one of the most common abnormal pathways in lung adenocarcinoma, which can promote aggressive cancer behaviors such as cell proliferation, invasion, metastasis, and angiogenesis, and prevents programmed cell death through the regulation of several downstream effectors [54–56]. The discussions of other pathways are provided in the "Supplementary Document."

Moreover, these pathways are dynamically changed with evolutionary stages of lung adenocarcinoma, some pathways are shown in Figure 10. These pathways of functional evolution processes are progressively activated or inhibited by evolving in a growth-like manner, which provides ancillary evidence for the rationality of cancerous time estimation. In general, the intersection of pathways enriched by different mods and differentially expressed genes(NE_DEGs, NM_DEGs, and NL_DEGs) all expanded with cancerous time, i.e. PI3K-Akt signaling pathway and Wnt signaling

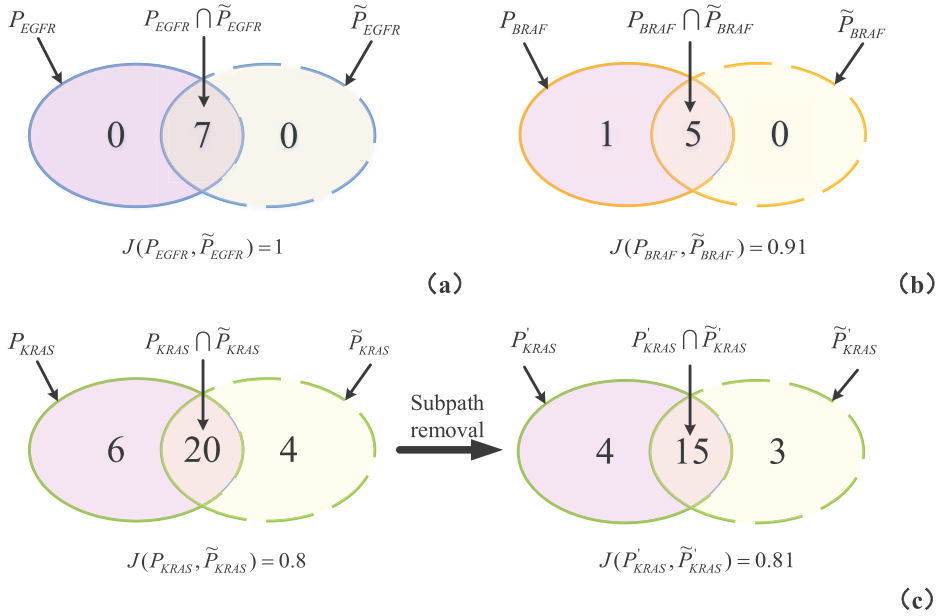

Figure 8. A comparison between the original oncogenetic paths and those based on simulated dataset. (a) A comparison between the original EGFR oncogenetic paths and those derived from simulated dataset; (b) A comparison between the original BRAF oncogenetic paths and those derived from simulated dataset; (c) A comparison between the original KRAS oncogenetic paths and those derived from simulated dataset. P_i denotes the set of original oncogenetic paths, and $\tilde{P}_i$ denotes the set of oncogenetic paths derived from simulated dataset, $i \in \{EGFR,BRAF,KRAS\}$. Specially, P'_i denotes the set of original oncogenetic paths following the removal of the subpath, while $\tilde{P}'_i$ denotes the set of oncogenetic paths derived from simulated dataset following the removal of the subpath.

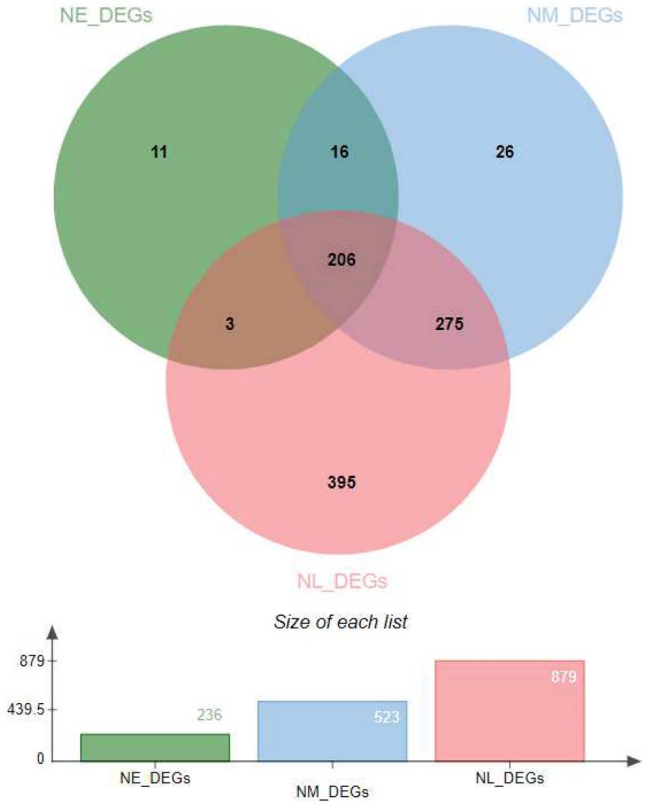

Figure 9. The differential expression analysis between control and different evolutionary stages.

pathway. In addition, the *log2FC* = 3, which is much larger than the commonly chosen cutoff of 1.5, is used to identify differentially expressed genes, suggesting that some pathways do not play a key role in early evolutionary stage, rather than not functioning in early evolutionary stage, i.e. p53 signaling pathway and cellular senescence pathway.

Similarly, the simulated dataset was used to further corroborate the availability of the patient quasi-potential landscape, and assess the stability of the feasible framework for identifying functional evolution processes. There are six prominent functional evolution processes (Simulated Mod1, Simulated Mod2, Simulated Mod3, Simulated Mod4, Simulated Mod5, Simulated Mod6) that are identified from the simulated dataset, whose sum of edges weight is greater than or equal to 10. The index *J* was utilized to quantify the degree of similarity between Mods and Simulated Mods based on genes, which is shown in Fig. 11(a). And the correspondence between Mods and Simulated Mods was determined by *J*, e.g. Mod1 corresponds to Simulated Mod1. It is noteworthy that Mod4 is unrecognized in the simulated dataset, presumably due to the fact that the dataset has been generated from only 90% of the original data, ultimately resulting in an insignificant sum of edge weights for this particular evolutionary path. Furthermore, the KEGG pathways identified by the Simulated Mods exhibit substantial consistency with those of the original dataset, with only minor variations observed in the number of gene enrichments and *p-value*, which is shown in Fig. 11(b).

Conclusions and future research directions

The patient quasi-potential landscape reconstructs the evolutionary process of cancer from a patient data-driven perspective, and accurately estimates the emergence time of cancerous phenotypes within this evolutionary process. This holds significant promise for exploring the accumulation of mutations under the constraints of cancerous time ordering. A total of three oncogenetic graphs are obtained as the common evolutionary patterns: EGFR, BRAF, and KRAS oncogenetic graphs,

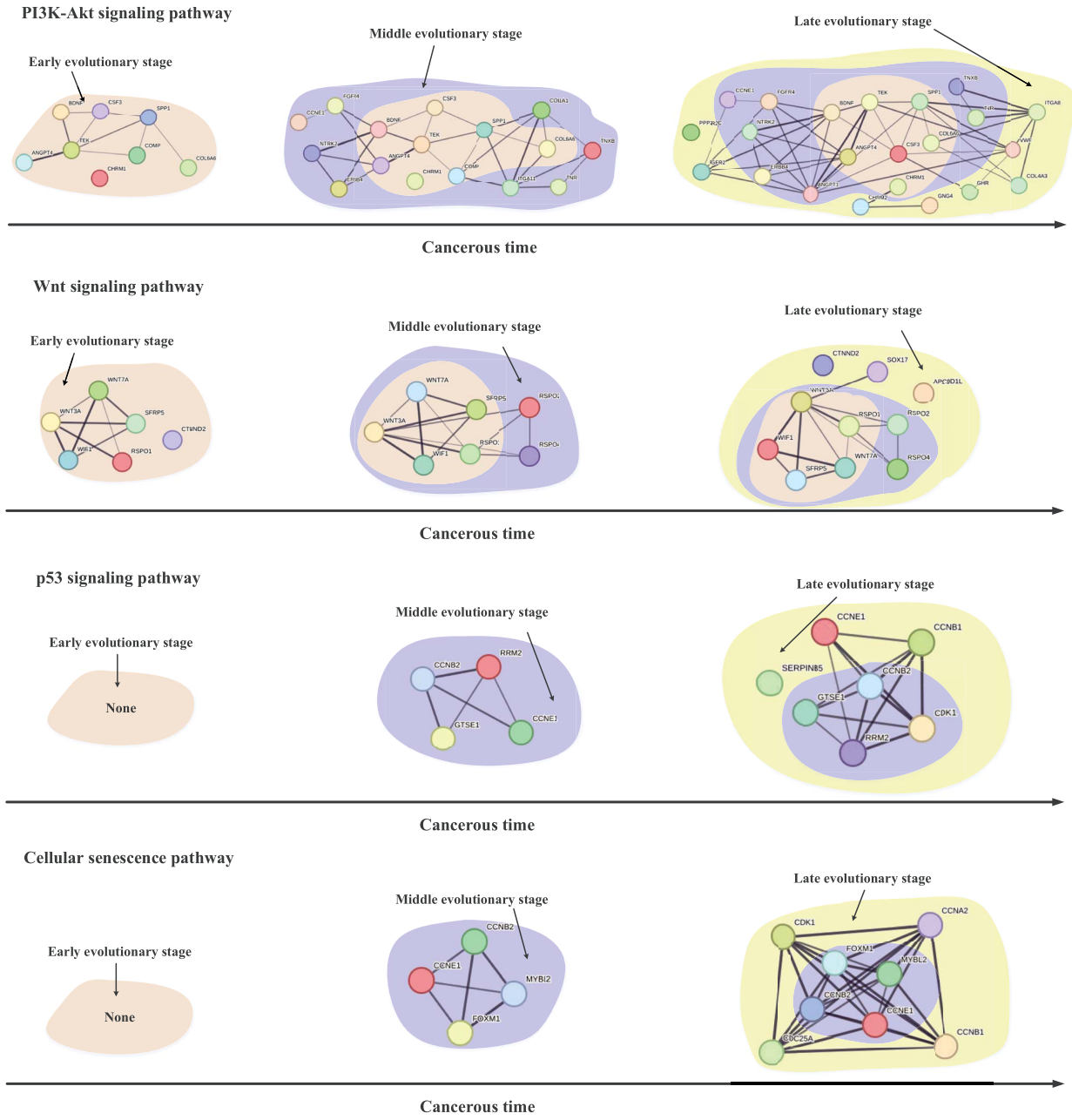

Figure 10. The pathways evolve in a growth-like manner with cancerous time.

which show that normal people firstly have a mutation in EGFR/BRAF/KRAS, then progress to a more severe pathological state with EGFR/BRAF/KRAS amplification and mutation in TP53 or other genes. And a total of six significant functional evolution processes are identified, which may play critical roles in the development of lung adenocarcinoma. And KEGG pathways of these evolution processes are progressively activated or inhibited by evolving in a growth-like manner.

However, the construction process of landscape is data-driven, i.e. the phenotypic state space is sampled based on clinical samples. When gathering data, the requirement that clinical samples should be representative of the entire evolutionary process must be considered. In addition, there are many other uses of the patient quasi-potential landscape that need to be explored in subsequent research, such as exploring the contributing factors of special terrain in the landscape and the key inflection point of cancer metastasis in the landscape, etc., which may offer potential strategies for diagnosis and treatment.

Key Points

- The patient quasi-potential landscape reconstructs the evolutionary process of cancer from the patient data-driven perspective vividly, and can deal with the indirect insight into the evolutionary process.
- Cancerous time deals with the problems of temporal information missing, and can be used directly in the oncogenetic analysis to identify the oncogenetic graph.
- Cancerous time deals with heterogeneous characteristics ignoring among distinct pathological stages, and can identify the real differentially expressed genes to analyze the dynamic functional evolution process.

Table 2. KEGG pathways of Mod1-Mod7.

Mod	Term	p	Count
Mod1	Cell cycle	9.1E-37	31
	Oocyte meiosis	2.60E-13	15
	Progesterone-mediated oocyte maturation	2.20E-09	11
	Motor proteins	9.60E-07	11
	Cellular senescence	1.40E-05	9
	Human T-cell leukemia virus 1 infectione	1.70E-04	9
	p53 signaling pathway	1.80E-04	6
	Homologous recombination	2.00E-04	5
	Fanconi anemia pathway	6.70E-03	4
	MicroRNAs in cancer	7.10E-02	6
Mod2	Complement and coagulation cascades	3.00E-02	2
Mod3	Wnt signaling pathway	4.20E-10	7
	Basal cell carcinoma	5.00E-02	2
	Melanogenesis	7.90E-02	2
Mod4	PPAR signaling pathway	1.60E-06	5
	Taurine and hypotaurine metabolism	2.00E-02	2
	Metabolic pathways	3.20E-02	6
	Butanoate metabolism	3.40E-02	2
	Alanine, aspartate and glutamate metabolism	4.60E-02	2
	Cholesterol metabolism	6.30E-02	2
Mod5	African trypanosomiasis	5.30E-05	3
	Malaria	9.80E-05	3
Mod6	Phagosome	8.00E-03	3
	Pertussis	6.08E-02	2
Mod7	Protein digestion and absorption	5.20E-08	6
	ECM-receptor interaction	2.10E-06	5
	Focal adhesion	5.60E-05	5
	Human papillomavirus infection	3.70E-04	5
	PI3K-Akt signaling pathway	5.00E-04	5
	Relaxin signaling pathway	9.20E-03	3

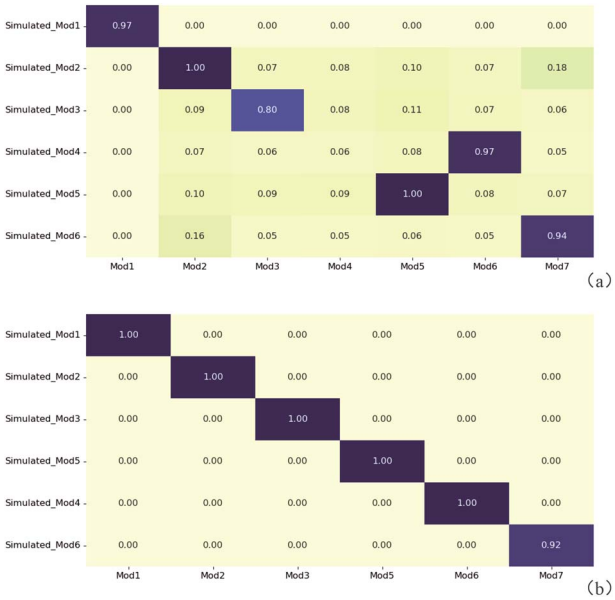

Figure 11. A comparison of functional evolution processes between original and simulated dataset. (a) A genes comparison of evolution paths between original and simulated dataset; (b) A KEGG pathways comparison of evolution paths between original and simulated dataset.

Acknowledgement statement

The results published here are in whole or part based upon data generated by The Cancer Genome Atlas managed by the NCI and NHGRI. Information about TCGA can be found at http://cancergenome.nih.gov.

Author contributions

B.C. initialized this study. Y.H. and B.C. discussed many times to finalize the work plan. X.S. and B.C. gave suggestions many times to modify this study. Y.H. conducted the numerical experiments and drafted the manuscript. Everyone read the manuscript and revised it, and agreed with the final version.

Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Funding

This work was supported by the National Key R&D Program of China under grant No. 2021YFA1000402, the National Natural Science Foundation of China under grant No. 61972320, and Xi'an municipal bureau of science and technology under grant No. 22YXYJ0057.

Conflict of Interest Statement

The authors declare that there is no conflict of interest.

Data availability statement

All data can be retrieved from the TCGA Data Portal (https://tcga-data.nci.nih.gov/tcga). All relevant data are within "https://github.com/YR1108/Cancerous-Time-Estimati\on-for-Interpreting-the-Evolution-of-Lung-Adenocarcinoma.git".

Code availability statement

The source code of our method can be accessed at "https://github.com/YR1108/Cancerous-Time-Estimation-for-Interpreting-the-Evolution-of-Lung-Adenocarcinoma.git".

References

1. Vendramin R, Litchfield K, Swanton C. Cancer evolution: Darwin and beyond. *EMBO J* 2021;**40**:e108389. https://doi.org/10.15252/embj.2021108389.
2. Szüts D. A fresh look at somatic mutations in cancer. *Science* 2022;**376**:351–2. https://doi.org/10.1126/science.abo7425.
3. Nowell PC. The clonal evolution of tumor cell populations. *Science* 1976;**194**:23–8. https://doi.org/10.1126/science.959840.
4. Greaves M, Maley CC. Clonal evolution in cancer. *Nature* 2012;**481**:306–13. https://doi.org/10.1038/nature10762.
5. Desper R, Jiang F, Kallioniemi OLLI-P. *et al.* Inferring tree models for oncogenesis from comparative genome hybridization data. *J Comput Biol* 1999;**6**:37–51. https://doi.org/10.1089/cmb.1999.6.37.
6. Beerenwinkel N, Rahnenführer J, Kaiser R. *et al.* Mtreemix: A software package for learning and using mixture models

of mutagenetic trees. *Bioinformatics* 2005;**21**:2106–7. https://doi.org/10.1093/bioinformatics/bti274.
7. Loohuis LO, Caravagna G, Graudenzi A. *et al.* Inferring tree causal models of cancer progression with probability raising. *PloS One* 2014;**9**:e108358. https://doi.org/10.1371/journal.pone.0108358.
8. Gerstung M, Baudis M, Moch H. *et al.* Quantifying cancer progression with conjunctive bayesian networks. *Bioinformatics* 2009;**25**:2809–15. https://doi.org/10.1093/bioinformatics/btp505.
9. Farahani HS, Lagergren J. Learning oncogenetic networks by reducing to mixed integer linear programming. *PloS One* 2013;**8**:e65773. https://doi.org/10.1371/journal.pone.0065773.
10. Misra N, Szczurek E, Vingron M. Inferring the paths of somatic evolution in cancer. *Bioinformatics* 2014;**30**:2456–63. https://doi.org/10.1093/bioinformatics/btu319.
11. Ramazzotti D, Caravagna G, Loohuis LO. *et al.* Capri: Efficient inference of cancer progression models from cross-sectional data. *Bioinformatics* 2015;**31**:3016–26. https://doi.org/10.1093/bioinformatics/btv296.
12. Lecca P, Casiraghi N, Demichelis F. Defining order and timing of mutations during cancer progression: The to-dag probabilistic graphical model. *Front Genet* 2015;**6**:309.
13. Hjelm M, Höglund M, Lagergren J. New probabilistic network models and algorithms for oncogenesis. *J Comput Biol* 2006;**13**:853–65. https://doi.org/10.1089/cmb.2006.13.853.
14. Schill R, Solbrig S, Wettig T. *et al.* Modelling cancer progression using mutual hazard networks. *Bioinformatics* 2020;**36**:241–9. https://doi.org/10.1093/bioinformatics/btz513.
15. Chen J. Timed hazard networks: Incorporating temporal difference for oncogenetic analysis. *PloS One* 2023;**18**:e0283004. https://doi.org/10.1371/journal.pone.0283004.
16. Chen B, Wang Y, Zhang J, *et al.* Specific feature recognition on group specific networks (SFR-GSN): a biomarker identification model for cancer stages. *Front Genet* 2024; **15**:1407072. https://doi.org/10.3389/fgene.2024.1407072.
17. Chen B, Wang T, Shang X. Identification and analysis of genes involved in stages of colon cancer. In: Huang DS, Jo KH, (eds.) *Intelligent Computing Theories and Application* . 2020, p. 161–72. ICIC. https://doi.org/10.1007/978-3-030-60802-6_15.
18. Chen B, Yang M, Gao L. *et al.* A functional network construction method to interpret the pathological process of colorectal cancer. *Int J Data Min Bioinform* 2020;**23**:251–64. https://doi.org/10.1504/IJDMB.2020.107879.
19. Waddington CH. *The Strategy of the Genes*. London: George Allen & Unwin, 1957. https://doi.org/10.4324/9781315765471.
20. Jin S, MacLean AL, Peng T. *et al.* Scepath: Energy landscape-based inference of transition probabilities and cellular trajectories from single-cell transcriptomic data. *Bioinformatics* 2018;**34**:2077–86. https://doi.org/10.1093/bioinformatics/bty058.
21. Lai Y, Wei X, Lin S. *et al.* Current status and perspectives of patient-derived xenograft models in cancer research. *J Hematol Oncol* 2017;**10**:106. https://doi.org/10.1186/s13045-017-0470-7.
22. Hidalgo M, Amant F, Biankin AV. *et al.* Patient-derived xenograft models: An emerging platform for translational cancer research. *Cancer. Discovery* 2014;**14**:161–74.
23. Sharpless NE, DePinho RA. The mighty mouse: Genetically engineered mouse models in cancer drug development. *Nat Rev Drug Discov* 2006;**5**:741–54.
24. Wright S. Evolution in mendelian populations. *Genetics* 1931;**16**:97–159. https://doi.org/10.1093/genetics/16.2.97.
25. Rogers ZN, McFarland CD, Winters IP. *et al.* Mapping the in vivo fitness landscape of lung adenocarcinoma tumor suppression in mice. *Nat Genet* 2018;**50**:483–6. https://doi.org/10.1038/s41588-018-0083-2.
26. Salehi S, Kabeer F, Ceglia N. *et al.* Imaxt consortium. Clonal fitness inferred from time-series modelling of single-cell cancer genomes. *Nature* 2021;**595**:585–90. https://doi.org/10.1038/s41586-021-03648-3.
27. Bissell MJ, Hines WC. Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. *Nat Med* 2011;**17**:320–9. https://doi.org/10.1038/nm.2328.
28. Sun Y, Yao J, Yang L. *et al.* Computational approach for deriving cancer progression roadmaps from static sample data. *Nucleic Acids Res* 2017;**45**:e69. https://doi.org/10.1093/nar/gkx003.
29. Otano I, Ucero AC, Zugazagoitia J. *et al.* At the crossroads of immunotherapy for oncogene-addicted subsets of nsclc. *Nature reviews. Clin Oncol* 2023;**20**:143–59. https://doi.org/10.1038/s41571-022-00718-x.
30. Gentleman R, Carey VJ, Huber W. *et al.*, (eds.). *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. New York: Springer Science & Business Media, 2005, https://doi.org/10.1007/0-387-29362-0.
31. Szklarczyk D, Franceschini A, Wyder S. *et al.* String v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 2014;**43**:D447–52.
32. Shannon P, Markiel A, Ozier O. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;**13**:2498–504. https://doi.org/10.1101/gr.1239303.
33. Vlasblom J, Wodak SJ. Markov clustering versus affinity propagation for the partitioning of protein interaction graphs. *BMC Bioinformatics* 2009;**10**:99. https://doi.org/10.1186/1471-2105-10-99.
34. Sherman BT, Ming Hao J, Qiu XJ. *et al.* David: A web server for functional enrichment analysis and functional annotation of gene lists. *Nucleic Acids Res* 2022;**50**:w216–21. https://doi.org/10.1093/nar/gkac194.
35. Nacer DF, Liljedahl H, Karlsson A. *et al.* Pan-cancer application of a lung-adenocarcinoma-derived gene-expression-based prognostic predictor. *Brief Bioinform* 2021;**22**:bbab154. https://doi.org/10.1093/bib/bbab154.
36. Liljedahl H, Karlsson A, Oskarsdottir GN. *et al.* A gene expression-based single sample predictor of lung adenocarcinoma molecular subtype and prognosis. *Int J Cancer* 2021;**148**:238–51. https://doi.org/10.1002/ijc.33242.
37. Skoulidis F, Heymach JV. Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy. *Nat Rev Cancer* 2019;**19**:495–509. https://doi.org/10.1038/s41568-019-0179-8.
38. Chan IT, Kutok JL, Williams IR. *et al.* Conditional expression of oncogenic k-ras from its endogenous promoter induces a myeloproliferative disease. *J Clin Invest* 2004;**113**:528–38. https://doi.org/10.1172/JCI20476.
39. Jinesh GG, Sambandam V, Vijayaraghavan S. *et al.* Molecular genetics and cellular events of k-ras-driven tumorigenesis. *Oncogene* 2018;**37**:839–46. https://doi.org/10.1038/onc.2017.377.
40. Inamura K. Clinicopathological characteristics and mutations driving development of early lung adenocarcinoma: Tumor initiation and progression. *Int J Mol Sci* 2018;**19**:1259. https://doi.org/10.3390/ijms19041259.
41. Park M-T, Lee S-J. Cell cycle and cancer. *J Biochem Mol Biol* 2003;**36**:60–5. https://doi.org/10.5483/BMBRep.2003.36.1.060.
42. Kastan MB, Bartek J. Cell-cycle checkpoints and cancer. *Nature* 2004;**432**:316–23. https://doi.org/10.1038/nature03097.

43. Williams GH, Stoeber K. The cell cycle and cancer. *J Pathol* 2012;**226**:352–64. https://doi.org/10.1002/path.3022.
44. Collado M, Blasco MA, Serrano M. Cellular senescence in cancer and aging. *Cell* 2007;**130**:223–33. https://doi.org/10.1016/j.cell.2007.07.003.
45. Campisi J. Aging, cellular senescence, and cancer. *Annu Rev Physiol* 2013;**75**:685–705. https://doi.org/10.1146/annurev-physiol-030212-183653.
46. Suzuki K, Matsubara H. Recent advances in p53 research and cancer treatment. *Biomed Res Int* 2011;**2011**:978312. https://doi.org/10.1155/2011/978312.
47. Achanta G, Sasaki R, Feng L. *et al.* Novel role of p53 in maintaining mitochondrial genetic stability through interaction with dna pol γ. *EMBO J* 2005;**24**:3482–92. https://doi.org/10.1038/sj.emboj.7600819.
48. Liu G, Parant JM, Lang G. *et al.* Chromosome stability, in the absence of apoptosis, is critical for suppression of tumorigenesis in trp53 mutant mice. *Nat Genet* 2004;**36**:63–8. https://doi.org/10.1038/ng1282.
49. Iglesias P, Salas A, Costoya JA. The maintenance of mitochondrial genetic stability is crucial during the oncogenic process. *Commun Integr Biol* 2012;**5**:34–8. https://doi.org/10.4161/cib.18160.
50. Mazieres J, He B, You L. *et al.* Wnt signaling in lung cancer. *Cancer Lett* 2005;**222**:1–10. https://doi.org/10.1016/j.canlet.2004.08.040.
51. Stewart DJ. Wnt signaling pathway in non-small cell lung cancer. *J Natl Cancer Inst* 2014;**106**:djt356. https://doi.org/10.1093/jnci/djt356.
52. Han SW, Roman J. Rosiglitazone suppresses human lung carcinoma cell growth through ppar-dependent and ppar-independent signal pathways. *Mol Cancer Ther* 2006;**5**:430–7. https://doi.org/10.1158/1535-7163.MCT-05-0347.
53. Ammu VVVRK, Garikapati KK, Krishnamurthy PT. *et al.* Possible role of ppar- and cox-2 receptor modulators in the treatment of non-small cell lung carcinoma. *Med Hypotheses* 2019;**124**:98–100. https://doi.org/10.1016/j.mehy.2019.02.024.
54. Qiao M, Sheng S, Pardee AB. Metastasis and akt activation. *Cell Cycle* 2008;**7**:2991–6. https://doi.org/10.4161/cc.7.19.6784.
55. Karar J, Maity A. Pi3k/akt/mtor pathway in angiogenesis. *Front Mol Neurosci* 2011;**4**:1–8. https://doi.org/10.3389/fnmol.2011.00051.
56. Iksen SP, Pongrakhananon V. Targeting the pi3k/akt/mtor signaling pathway in lung cancer: An update regarding potential drugs and natural products. *Molecules* 2021;**26**:4100. https://doi.org/10.3390/molecules26134100.