BMC Bioinformatics

METHODOLOGY

Open Access

Detect the early-warning signals of diseases based on signaling pathway perturbations on a single sample

Yanhao Huo[1], Geng Zhao[2], Luoshan Ruan[5], Peng Xu[1,3], Gang Fang[1], Fengyue Zhang[4], Zhenshen Bao[1]* and Xin Li[5]*

*Correspondence:
bzsbao@163.com; xin.li@whu.edu.cn
[1] Institute of Computational Science and Technology, Guangzhou University, Guangzhou 510006, Guangdong, China
[5] Department of Gynecology, Renmin Hospital of Wuhan University, Wuhan 430000, Hubei, China
Full list of author information is available at the end of the article

Abstract

Background: During the pathogenesisof complex diseases, a sudden health deterioration will occur as results of the cumulative effect of various internal or external factors. The prediction of an early warning signal (pre-disease state) before such deterioration is very important in clinical practice, especially for a single sample. The single-sample landscape entropy (SLE) was proposed to tackle this issue. However, the PPI used in SLE was lack of definite biological meanings. Besides, the calculation of multiple correlations based on limited reference samples in SLE is time-consuming and suspect.

Results: Abnormal signals generally exert their effect through the static definite biological functions in signaling pathways across the development of diseases. Thus, it is a natural way to study the propagation of the early-warning signals based on the signaling pathways in the KEGG database. In this paper, we propose a signaling perturbation method named SSP, to study the early-warning signal in signaling pathways for single dynamic time-series data. Results in three real datasets including the influenza virus infection, lung adenocarcinoma, and acute lung injury show that the proposed SSP outperformed the SLE. Moreover, the early-warning signal can be detected by one important signaling pathway PI3K-Akt.

Conclusions: These results all indicate that the static model in pathways could simplify the detection of the early-warning signals.

Keywords: Pre-disease state, Early warning signals, Single sample, Signal perturbation, Signaling pathway

Huo *et al. BMC Bioinformatics* 2022, **22**(Suppl 12):367

Page 2 of 15

Background

Theoretical considerations and computational studies suggest that many types of complex dynamical systems may have a critical points between an ordered and a disordered dynamical regime [1, 2]. This regime provides complex systems to have an optimal balance between robustness and adaptability. Specifically, they can execute normal functions in a variable environment while being responsive to specific changes in the environment. Studies of human brain oscillations [3], computer network traffic and the Internet [4, 5], financial markets [6], forest fires [7], ecosystems [8, 9], climate systems [10, 11], economics and global finance [12], neuronal networks [13], and biological macroevolution have all revealed such critical dynamics [14].

Mathematically, the dynamic behavior of complex systems can be characterized by state space. One state can be represented as the behavior of the composite elements in the system. In a relatively long time, some initial states of the system will eventually settle down to one of a limited set of stable states. Bifurcation theory demonstrates that complex systems may undergo a sudden state transition under some critical continuous perturbations of various internal or externals. Such a change often occurs at a critical threshold, or the so-called "tipping point", at which the system shifts abruptly from one state to another.

The time evolution of complex diseases may follow the bifurcation theory, that is, a sudden health deterioration will occur during these diseases' gradual progression at a tripping point time [15–17]. For example, it takes at least a decade or even decades for cancers to develop before getting worse [18, 19]. According to these concepts, the disease progression can be divided into three states: the normal state, the disease state, and the pre-disease state between them [20]. The normal state represents a relatively healthy stage during which the disease is under control, in an incubation period or a chronic inflammation period. The disease stage represents a continuous health deterioration that is hard to reverse. The pre-disease stage is a relatively unstable and reversible state which may transition to the normal state if appropriate treatment is applied [17, 21–25]. From the perspective of disease prevention and a better understanding of disease progression, detecting the early-warning signals for the pre-disease state is an important issue in clinical practice [17].

The execution of various physiological processes in cells is carried out by complex biomolecular systems. The massively parallel dynamics of complex molecular networks furnish the cell with the ability to process information from its environment and mount appropriate responses [26, 27]. Therefore, one direct way is determining the pre-disease state based on the state transition based on a complex network model. However, this is limited by the curse of a high dimension of genomic data and the lack of data for a single sample. The model-free direction is to identify a group of biomarkers and define a measurement based on the single sample expression data. Due to the rapid development of high-throughput technologies, innovative biomarkers are identified as the unprecedented rich information of genotypes and phenotypes of diseases. Molecular biomarkers, for example, genes, RNAs, proteins, and metabolites, are widely adopted traditional model-driven method and still play an important role in analyzing data, because of their simplicity of measurement and implementation [28–33]. Because the progression process of diseases is affected by the complications and variations of genetic, epigenetic,

Huo *et al. BMC Bioinformatics* 2022, **22**(Suppl 12):367

Page 3 of 15

and environmental factors, the molecular biomarkers have a high false-positive rate and low coverage. To surmount these shortages, network biomarkers are proposed [34, 35]. Network biomarkers are considered to be more robust because a complex disease is generally caused not by the malfunction of individual molecules but by a network that contains the interplay of a group of correlated molecules [36]. However, molecular biomarkers and network biomarkers are mainly used to distinguish the disease state from the normal state by the great distinction between them [22]. Thus, they are not suitable for detecting the early-warning signals. To solve this problem, Chen and Liu et al. proposed a dynamic network biomarker (DNB), which could be applied to different application scenarios. The DNBs-based method and its subsequent modifications have been used to identify the pre-disease states of several diseases [37–41]. These methods have also been applied to detect the tipping points of cell fate decision and differentiation [42, 43], and the immune checkpoint blockade [44]. A DNB must appear and satisfy three statistic conditions which require multiple samples from an individual [17]: i.e. correlations between the variables among this group rapidly increase, correlations between this group and other variables rapidly decrease and standard deviations of the variables among this group drastically increase. The calculation of correlations and standard deviations limits the application of DNBs-based methods not fit to single case samples in clinical practice but appropriate for multiple samples. Recently, Liu et.al. proposed a new conception named the single-sample landscape entropy (SLE) based on DNB theory to detect the early-warning signals of diseases (detailed information can be seen section materials and method) [15]. The limitation of SLE lies in the following two aspects. First, the lack of definite biological meanings in PPI makes SLE indirectly depict the deviation of an individual at a time point from the health state based on the Pearson correlation change. Second, the calculation of various correlations between genes using reference samples from an individual is time consuming and suspect.

Signaling pathways in the KEGG database include a series of validated enzymatic reactions from biological experiments. And they can transmit the extracellular molecular signals into cells to exert effects through specific biological functions. The development of diseases results from the continuous perturbation of the various abnormal signals in some of these pathways. Therefore, Tarca et al. proposed a signaling pathway impact analysis (SPIA) method to determine significant disease-related signaling pathways by normal and disease samples [45]. From the perspective of system biology, signaling pathways are a kind of static biological model of the corresponding functions. In this paper, we adapt the signal perturbation to study the propagation of the early-warning signals based on the signaling pathways in the KEGG database. We name it as single-sample signal perturbation (SSP). We applied SSP to three datasets (an individual-sample dataset of influenza virus infection, the TCGA dataset of lung adenocarcinoma, and a dataset of acute lung injury). our results show that SSP outperforms SLE in predicting the early-warning signals. Also, SSP is relatively simple to avoid calculating the correlation between genes in the large PPI network. Furthermore, SSP can be further simplified by using some crucial pathways, i.e., the PI3K-Akt signaling pathway.

Huo *et al. BMC Bioinformatics* 2022, **22**(Suppl 12):367

Page 4 of 15

Results

Complex diseases arise is due to the accumulation of differential expressions and signal perturbations in a subgroup of genes allowing uncontrolled biological functions. The accumulation of differential expressions and signal perturbations in signaling pathways is an invertible dynamic process before the biological functions become uncontrolled. Thus, when the accumulation changes in signaling pathways, it will signal a piece of pre-disease information. We use three single-sample datasets to illustrate how SSP works, including influenza virus infection (GSE30550), lung adenocarcinoma (LUAD) from the TCGA database, and acute lung injury (GSE2565). In this paper, we download 178 signaling pathways of human and 174 mmu signaling pathways from the KEGG PATHWAY dataset (https://www.kegg.jp/kegg/pathway.html). The pathways in KEGG dataset are all hand-painted and validated by scientific pieces of literature. In this section, we compare the early-warning signals detection of the proposed SSP with SLE using all these signaling pathways.
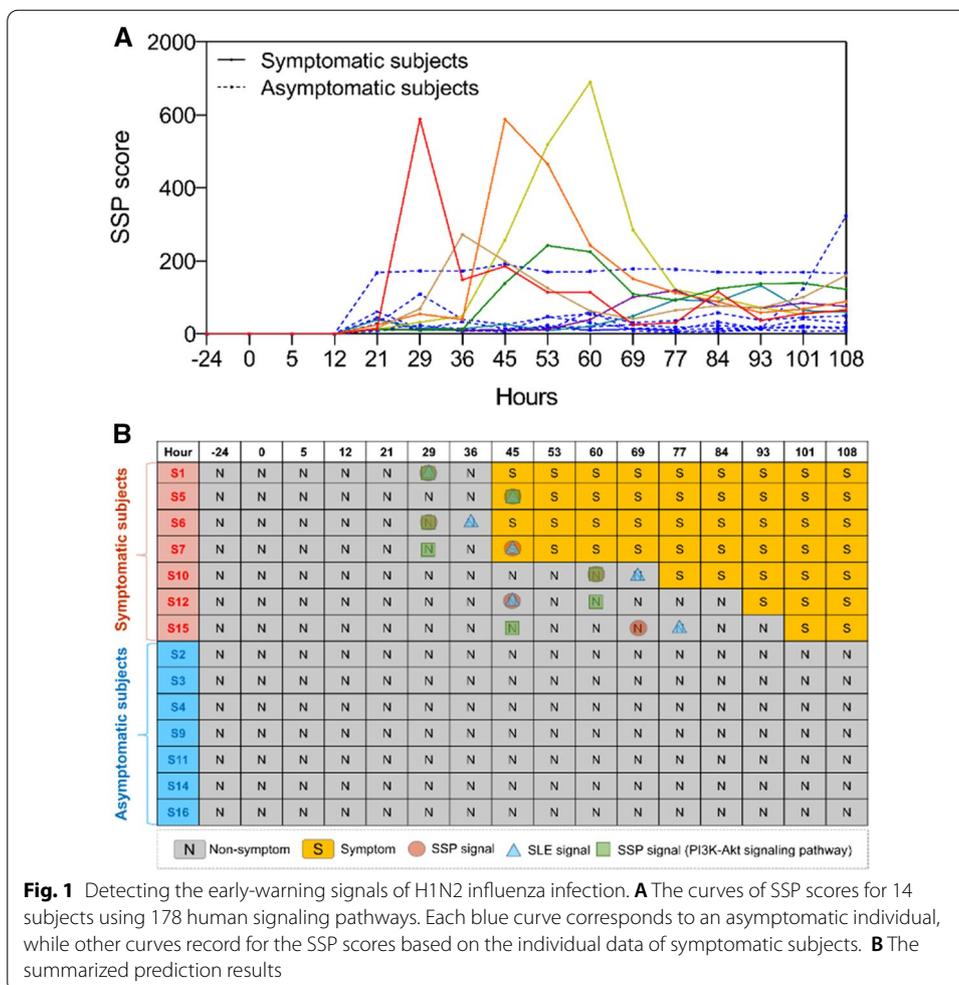
The PI3K-Akt is an important signaling pathway associated with many complex diseases, such as the later stages of influenza virus infection [46], acute lung injury [47–49], non-small cell lung cancer [50]. To demonstrate the simplicity of the proposed SSP method, we also present the predicted results on the three datasets by only using this pathway instead of all the signaling pathways.

Early-warning signals of individual influenza infection

For each subject in the individual influenza infection dataset (GSE30550), the gene expression profiles of the first four time points, i.e., Baseline, 0, 5, and 12 h are regarded as reference samples. Figure 1A shows the global SSP scores of the 14 subjects by all signaling pathways. First, the SSP scores of the symptomatic subjects are relatively higher and more unstable than that of the asymptomatic subjects. Second, there is a drastic increase of SSP scores of the symptomatic subjects at some middle time points, which provides an early-warning signal for these subjects. Figure 1A shows the predicted warning signals for the 14 subjects of SLE, SSP with all signaling pathways, and SSP with only PI3K-Akt pathway.

Figure 1B presents the predicted warning signals by the SLE method, SSP with all signaling pathways, and only the PI3K-Akt pathway. For the seven symptomatic subjects, three predicted warning signals by SSP with all signaling pathways are earlier than that by the SLE method. The other four warning signals are predicted at the same time point which including two signals are overlapped with the appearance of the symptom. For the SSP with only PI3K-Akt pathway, four predicted warning signals are earlier than that by the SLE method, one later and two equals. These observations indicate that the proposed SSP method has a better performance than the SLE method.

Figure 2 shows the curves of the SSP scores and SLE scores of the seven symptomatic subjects with times. The two score curves present a similar trend across all time points. Figure 3 shows the curves of the SSP scores by only PI3K-Akt pathway with times which also has a similar trend with the previous two. Compared with the SLE method, the proposed SSP methods capture the significant change of the development of diseases based on the signal perturbation in signaling pathways while the former detect it based on the

Huo *et al. BMC Bioinformatics* 2022, **22**(Suppl 12):367

Page 5 of 15

Fig. 1 Detecting the early-warning signals of H1N2 influenza infection. **A** The curves of SSP scores for 14 subjects using 178 human signaling pathways. Each blue curve corresponds to an asymptomatic individual, while other curves record for the SSP scores based on the individual data of symptomatic subjects. **B** The summarized prediction results
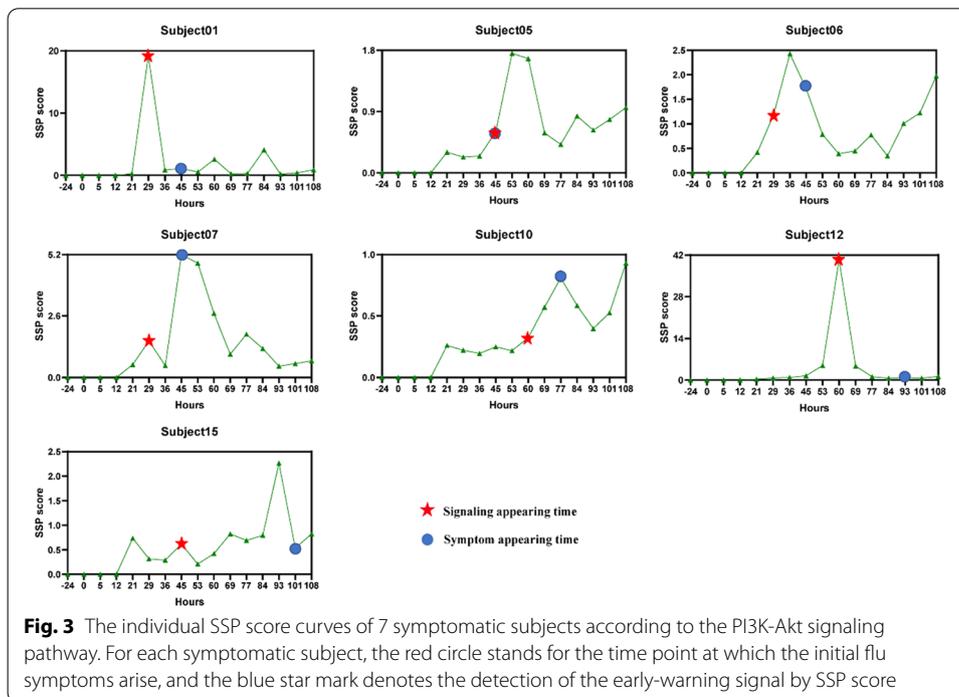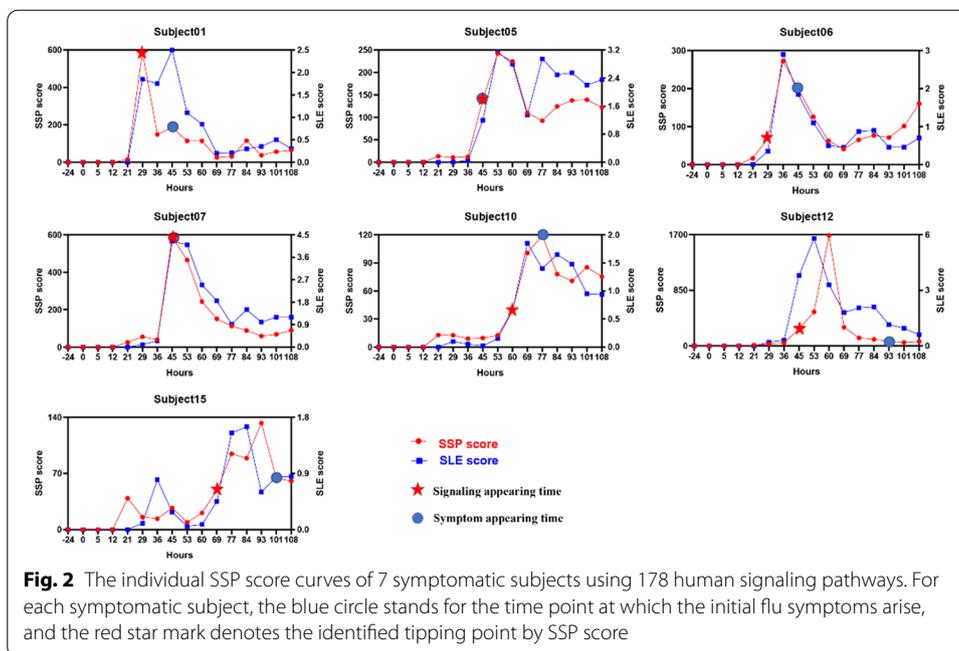
correlation change between genes. Despite their difference, the results indicate that the two kinds of scores could detect the underlying molecular interaction changes in the development of diseases. However, the SSP score is easier to understand as it directly integrates the underlying molecular interactions models. Therefore, this score is more believable and easier to understand than the SLE scores.

As the perturbations of upstream genes usually contribute abnormal signals in a signaling pathway, we present the fold change of the 30 upstream genes in the PI3K-Akt signaling pathway in Fig. 4 by Subject 01. We can see that some genes are significantly highly expressed at time 29 h, which is consistent with the detected warning signal for Subject01 in Fig. 3. Therefore, the pathway-based method may further help to uncover the underlying mechanisms for the pre-disease state and to predict a more precise warning signal in the future.

Early-warning signals of lung adenocarcinoma (LUAD)

For dataset LUAD, the 58 tumor adjacent (TA) samples are considered as reference samples. Because of no individual-based samples across all time points, the expressions of genes are obtained by their average value in each time point. Metastasis is the

Huo *et al. BMC Bioinformatics* 2022, **22**(Suppl 12):367

Page 6 of 15

Fig. 2 The individual SSP score curves of 7 symptomatic subjects using 178 human signaling pathways. For each symptomatic subject, the blue circle stands for the time point at which the initial flu symptoms arise, and the red star mark denotes the identified tipping point by SSP score

Fig. 3 The individual SSP score curves of 7 symptomatic subjects according to the PI3K-Akt signaling pathway. For each symptomatic subject, the red circle stands for the time point at which the initial flu symptoms arise, and the blue star mark denotes the detection of the early-warning signal by SSP score

culprit behind most cancer-related deaths and the ultimate challenge in our effort to fight cancer as a life-threatening disease [51]. Stage II means cancer may have spread from the lung to the nearby lymph nodes and stage IV means the tumor cells have invaded into distant tissues of other organs [52].
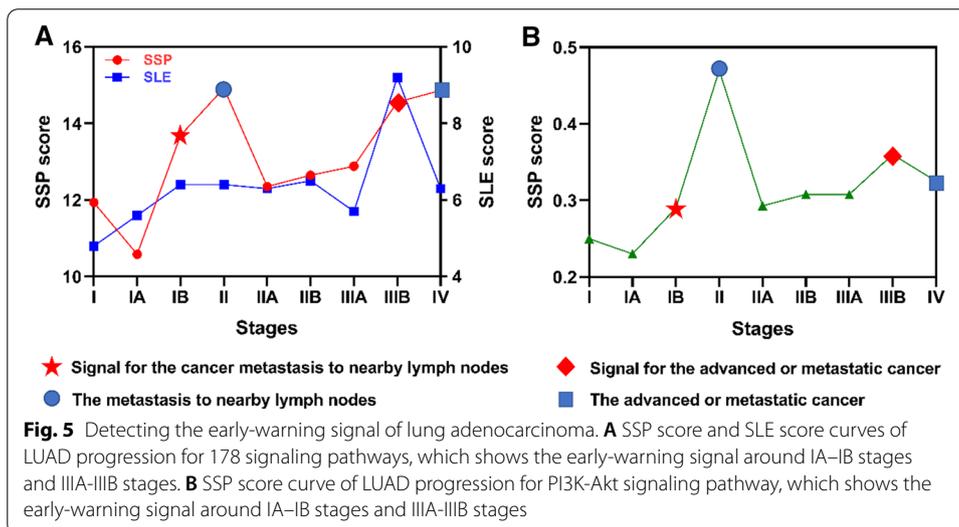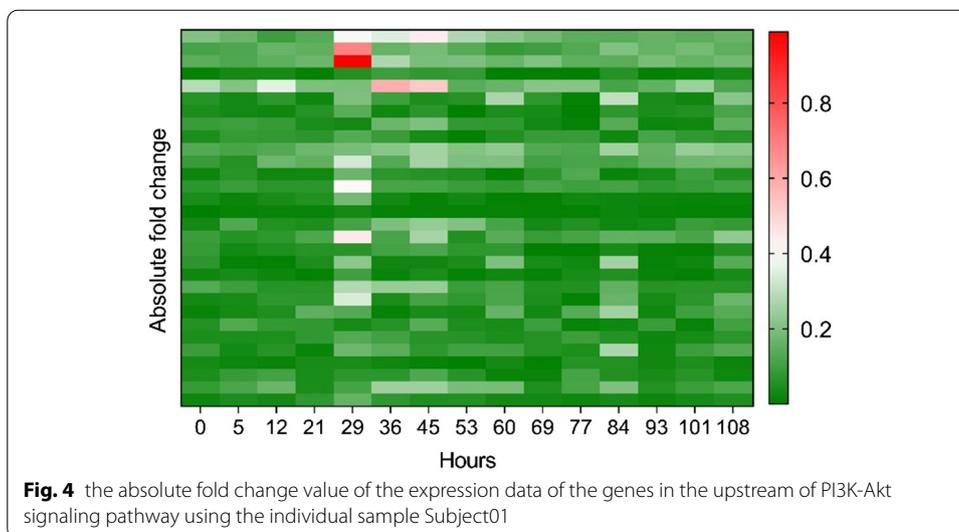
Huo *et al. BMC Bioinformatics* 2022, **22**(Suppl 12):367

Page 7 of 15

Fig. 4 the absolute fold change value of the expression data of the genes in the upstream of PI3K-Akt signaling pathway using the individual sample Subject01

Fig. 5 Detecting the early-warning signal of lung adenocarcinoma. **A** SSP score and SLE score curves of LUAD progression for 178 signaling pathways, which shows the early-warning signal around IA–IB stages and IIIA-IIIB stages. **B** SSP score curve of LUAD progression for PI3K-Akt signaling pathway, which shows the early-warning signal around IA–IB stages and IIIA-IIIB stages

Figure 5A shows the predicted warning signals by the SLE and SPP with all signaling pathways. the former only correctly give the warning signals before stage IV, while the latter can correctly give the warning signals for both stage II and IV. Figure 5B shows that the same results can even be obtained by SSP with only PI3K-Akt pathway. Therefore, the proposed SSP method gives better prediction performance than the SLE method. In practice, the precise warning signals will help to take chemotherapy and radiotherapy timely to prevent serious deterioration or slow down cancer progression [53].

Early-warning signals of acute lung injury

For the acute lung injury dataset GSE2565, the samples collected from the air- or phosgene-exposed mice at 0 h are considered as reference samples. Because of no individual-based samples across all time points, the expressions of genes are obtained by their average value in each time point.
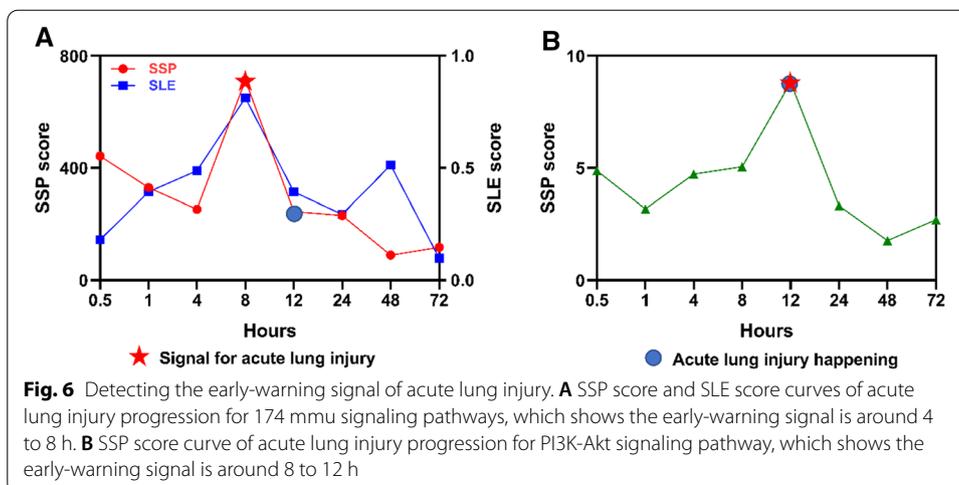
Huo *et al. BMC Bioinformatics* 2022, **22**(Suppl 12):367

Page 8 of 15

Fig. 6 Detecting the early-warning signal of acute lung injury. **A** SSP score and SLE score curves of acute lung injury progression for 174 mmu signaling pathways, which shows the early-warning signal is around 4 to 8 h. **B** SSP score curve of acute lung injury progression for PI3K-Akt signaling pathway, which shows the early-warning signal is around 8 to 12 h

Figure 6A shows that both the SLE and SPP with all signaling pathways predict the warning signals at 8 h after exposure which is 4 h before the happening of acute lung injury. Figure 6B shows the predicted warning signal just at the observed happening time by SSP with only the PI3K-Akt pathway. The original experiment found that the most severe phosgene-induced acute lung injury ranged from 4 to 12 h after exposure. Especially, the main physiological effects occurred within the first 8 h after exposure, resulting in common observations of enhanced bronchi alveolar lavage fluid (BALF) protein levels, increased pulmonary edema, and ultimately decreased survival rates [54]. Therefore, the predicted signals are consistent with the actual disease development. In addition, in Fig. 6A, b, the SSP score from time point 0.5 to 8 h has a down behavior. The expressions of genes are obtained by their average value of all samples at the time point in each time point. Thus, the noise may be introduced in the SSP score and will result in this case. But this case does not influence the predicted signals.

Discussion

In the progression of complex diseases, a sudden health deterioration (pre-disease state) usually occurs due to the gradual accumulating effect of various internal or external factors. An early warning signal before this deterioration is essentially important for clinical practice. Some effective treatments could be applied to prevent the happening of the irreversible process of diseases. In recent years, Chen et al. developed a new framework to tackle this issue. Especially the SLE method can detect early warning signals based on the time point expression data of an individual through the correlation changes between genes.

In this paper, we apply the signal perturbation in pathways to this framework to predict the early warning signals of complex diseases. Results from three datasets demonstrate that the proposed method SSP outperforms the SLE method in both the influenza virus infection dataset and LUAD dataset, and equals in the acute lung injury dataset. The consideration of the static model of pathways provides the proposed SSP method a relatively concrete biological mechanism to detect the essential changes. In addition, it reduces the computation complexity of the proposed SSP compared with the SLE

Huo *et al. BMC Bioinformatics* 2022, **22**(Suppl 12):367

Page 9 of 15

Table 1 The number of tumor samples within each stage in the LUAD dataset from TCGA

Stage	I	IA	IB	II	IIA	IIB	IIIA	IIIB	IV	TA
Sample	5	134	147	1	50	71	73	11	26	58

method. Finally, based on the result by only the PI3K-Akt signaling pathway, the proposed method SSP can be further simplified by using some crucial pathways if enough data is available.

However, the SSP method still has its limitations. SSP mainly relies on the signaling pathways from KEGG datasets. The biological information included in these pathways is incomplete. Thus, the SSP score may have a fluctuation under some circumstances. The noise in initial gene expression data is also the reason for the fluctuation of the SSP score. In addition, the strength of the interactions between genes in pathways is not considered in SSP. This may result in the loss of some real biological information.

Conclusions

In this paper, a method called SSP to identify the pre-disease state of complex diseases is presented. SSP adapted a static biological model (signaling pathways with specific biological functions) to a dynamic system to character the dynamic change of system for an individual sample, which has a better performace than previous method SLE. Besides, benefiting from the use of static signaling pathways, the calculation of SSP method takes less time than previous method SLE. Therefore, SSP can identify the pre-disease state in less time that previous methods cannot.

Methods

Datasets

We download three time-course or stage-course datasets including the the microarray data of influenza virus infection process (GSE30550) and acute lung injury induced by phosgene inhalation(GSE2565) from the NCBI GEO database (www.ncbi.nlm.nih.gov/geo), and data of the lung adenocarcinoma (LUAD) from the TCGA database. For all these omics genomic data, we discard the probes without the corresponding NCBI Entrez gene symbol. For each gene mapped by multiple probes, the average value is employed as the gene expression.

The dataset GSE30550 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30550) records influenza virus infection of 16 human adult subjects inoculated with live H3N2/ Wisconsin influenza virus [55]. Each subject includes 16 time point gene expression samples (Baseline, 0, 5, 12, 21, 29, 36, 45, 53, 60, 69, 77, 84, 93, 101, and 108 h). The gene expression profiles are measured on whole peripheral blood at an interval of 8 h post-inoculation (hpi) through 108 hpi. We only analyze 14 subjects because the other 2 subjects lack data at some time points. The 14 subjects include 7 subjects with clinical symptoms of influenza infection and 7 subjects without any clinical symptom at all-time points.

The lung adenocarcinoma (LUAD) dataset includes 518 tumor samples and 58 tumor-adjacent samples. The tumor samples are divided into different stages based on clinical (stage) information from TCGA. Based on clinical information, the samples are grouped into ten stages, i.e., stage I, IA, IB, II, IIA, IIB, IIIA, IIIB, and IV of lung cancer (Table 1).

Huo *et al. BMC Bioinformatics* 2022, **22**(Suppl 12):367

Page 10 of 15

Table 2 The number of samples within each hour in the dataset GSE2565

Hour	0	0.5	1	12	24	4	48	72	8
Sample	6	12	12	12	12	12	14	12	12

The dataset GSE2565 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE2565) comprises expression profiles of the mouse with acute lung injury induced by phosgene inhalation. These lung tissues are collected from air- or phosgene-exposed mice at 0, 0.5, 1, 4, 8, 12, 24, 48, and 72 h after exposure (Table 2) [54]. RNA is extracted from the lung and used as starting material for the probing of oligonucleotide microarrays to determine changes in gene expression following phosgene exposure.

Algorithm to detect the pre-disease state based on SSP

Because it is hard to characterize the health degree of an individual using one time point expression data, Chen et al. developed a novel framework that indirectly depicts the deviation of an individual at time point t from the health state based on the Pearson correlation changes between two groups of samples. Specifically, the first is the reference group composed of n samples with a normal/healthy state, the other is the mixed group composed of the n reference samples and one sample at a time point t. Given gene g_i and its k th neighbor gene g_{ik} (total M neighbor genes) in a PPI network, the local entropy of the gene g_i at time point t is defined as

$$H^n(g_i, t) = -\frac{1}{\log M} \sum_{k=1}^{M} p_k^n(t) \log p_k^n(t) \tag{1}$$

with

$$p_k^n(t) = \frac{\left| PCC^n(g_i^n(t), g_{ik}^n(t)) \right|}{\sum\limits_{j=1}^{M} \left| PCC^n(g_i^n(t), g_{ij}^n(t)) \right|} \tag{2}$$

where $PCC^n(g_i(t), g_{ik}(t))$ represents the Pearson Correlation Coefficient between the gene g_i and the k th neighbor gene g_{ik} based on n reference samples. $g_i^n(t)$ and $g_{ik}^n(t)$ respectively denote the expressions of genes g_i and g_{ik} at time point t based on n reference samples. Then for a single sample at the time point t, SLE mixes it with the n reference samples. The local entropy $H^{n+1}(g_i, t)$ of the gene g_i at the time point t of the mixed $n + 1$ samples is calculated in a similar way to that in Eq. (1) and Eq. (2), but is based on the mixed $n + 1$ samples instead of n reference samples.

They also measured the expression perturbation of gene g_i at time point t by the differential standard deviation

$$\Delta SD(g_i, t) = \left| SD^{n+1}(g_i, t) - SD^n(g_i, t) \right| \tag{3}$$

where $SD^{n+1}(g_i, t)$ and $SD^n(g_i, t)$ are the standard deviations of the gene g_i respectively based on the reference samples and the mixed samples.

Huo *et al. BMC Bioinformatics* 2022, **22**(Suppl 12):367

Page 11 of 15

Then, the absolute differential entropy of the gene g_i at the time point t between the local entropies $H^{n+1}(g_i, t)$ and $H^n(g_i, t)$ is weighted by $\Delta SD(g_i, t)$ as

$$\Delta H(g_i, t) = \Delta SD(g_i, t) \left| H^{n+1}(g_i, t) - H^n(g_i, t) \right| \tag{4}$$

Finally, the abnormal score of an individual at the time t is the summation of $\Delta H(g_i, t)$ for all Q genes as

$$\Delta H(t) = \frac{1}{Q} \sum \Delta H(g_i, t) \tag{5}$$

As the development of disease results from the abnormal molecular perturbation through some important signaling pathways, Tarca et al. proposed a signaling pathway impact analysis (SPIA) method to determine the significant disease-related signaling pathways by normal and disease samples [45]. They defined the signal perturbation $PF(g_i)$ of a gene g_i in a specific signaling pathway as

$$PF(g_i) = FC(g_i) + \sum_k \beta_{ik} \frac{PF(g_k)}{N_k} \tag{6}$$

where gene g_k is the direct upstream gene of gene g_i in the specific signaling pathway. $PF(g_k)$ is the signal perturbation of the gene g_k. N_k is the number of the direct downstream genes of gene g_k. β_{ik} is the strength of the interaction between gene g_k and g_i. β_{ik} is 1 when the interaction between the two genes is activated. β_{ik} is -1 when the interaction between the two genes is inhibited. $FC(g_i)$ is the fold change of the expression of the gene g_i by normal and disease samples.

Signaling pathways in the KEGG database can transmit the extracellular molecular signals into cellcells to exert effects through specific biological functions. Complex diseases always associate with the abnormal biological functions that direct impacted by the most downstream genes. The accumulation of differential expressions and signal perturbations in signaling pathways is an invertible dynamic process before the biological functions become uncontrolled. The upstream genes in signaling may be very differential expression in the pre-disease state. As demonstrated by SPIA, the amount of signal perturbations in signaling pathways are directly impact by the expression value of upstream genes. Therefore, following SPIA, we adapt the dynamic time-series expression data to static signaling pathwaysin the same wayinstead of the correlations calculated in SLE.

$$PF(g_i, t) = FC(g_i, t) + \sum_k \beta_{ik} \frac{PF(g_k, t)}{N_k} \tag{7}$$

where gene g_k is the direct upstream gene of gene g_i in the specific signaling pathway. $PF(g_i, t)$ is the signal perturbation of the gene g_k at a time point t. N_k is the number of the direct downstream genes of gene g_k in the specific signaling pathway. β_{ik} is the strength of the interaction between gene g_i and g_k. β_{ik} is 1 when the interaction between the two genes is activated. β_{ik} is -1 when the interaction between the two genes is inhibited. $FC(g_i, t)$ is the fold change of the gene g_i at a time point t.

Huo *et al. BMC Bioinformatics* 2022, **22**(Suppl 12):367
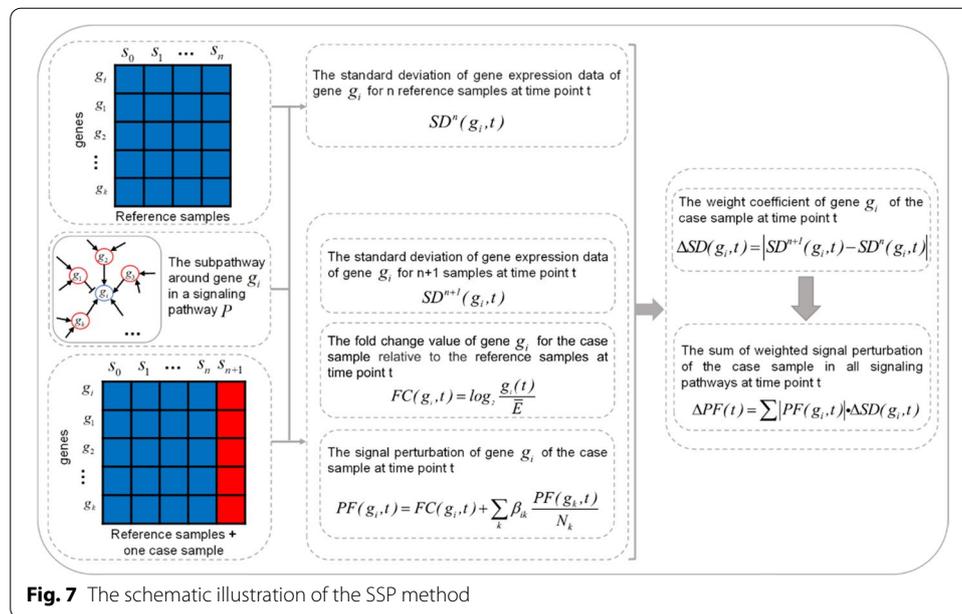
Page 12 of 15

Fig. 7 The schematic illustration of the SSP method

$$FC(g_i, t) = \log_2 \frac{g_i(t)}{\overline{E}} \tag{8}$$

$g_i(t)$ is the expression of the gene g_i at the time point t. $\overline{E}$ is the average expression of all n reference samples.

Then, the weighted signal perturbation (SSP score) for all genes in all signaling pathways at the time point t is calculated, i.e.

$$\Delta PF(t) = \sum \left| PF(g_i, t) \right| \cdot \Delta SD(g_i, t) \tag{9}$$

If the SSP score has a sharp increase from the previous time point to a time point t, the time point t will be considered as an early-warning signal for the disease. Figure 7 shows the schematic diagram of the proposed SSP method. The main difference between SSP and SLE based method is that SSP directly describes the deviation of an individual at a time point from a normal state by adapting the signal perturbation transmitted in static signaling pathways instead of the indirect way in SLE.

Huo *et al. BMC Bioinformatics* 2022, **22**(Suppl 12):367

Page 13 of 15

Availability of data and materials
SSP is available and open source at https://github.com/ZhenshenBao/SSP. The datasets used and/or analyzed during the current study are available and are listed in the manuscript.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Institute of Computational Science and Technology, Guangzhou University, Guangzhou 510006, Guangdong, China. [2]Netease Youdao Information Technology (Hangzhou) Co., Ltd., Hangzhou 310000, Zhejiang, China. [3]School of Computer Science of Information Technology, Qiannan Normal University for Nationalities, Duyun 558000, Guizhou, China. [4]Department of Biomedical Engineering, School of Life Science and Technology, Beijing Institute of Technology, Beijing 100081, China. [5]Department of Gynecology, Renmin Hospital of Wuhan University, Wuhan 430000, Hubei, China.

References

1. Langton CGJPD. Computation at the edge of chaos: Phase transitions and emergent computation. Phys D Nonlinear Phen. 1990;42(1):12–37.
2. Krawitz P, Shmulevich I. Basin entropy in Boolean network ensembles. Phys Rev Lett. 2007;98(15):158701.
3. Linkenkaer-Hansen K, Nikouline VV, Palva JM, Ilmoniemi RJ. Long-range temporal correlations and scaling behavior in human brain oscillations. J Neurosci. 2001;21(4):1370–7.
4. Valverde S. Mechanics RVSJPAS, Applications I: Self-organized critical traffic in parallel computer networks. Phys A. 2002;312(3–4):636–48.
5. Fukuda K, Takayasu H. Takayasu MJPASM, Applications I: Origin of critical behavior in Ethernet traffic. Phys A. 2000;287(1):289–301.
6. Lux T, Marchesi MJN. Scaling and criticality in a stochastic multi-agent model of a financial market. Nature. 1999;397(6719):74–5.
7. Malamud BD, Morein G, Turcotte DL. Forest fires: An example of self-organized critical behavior. Science (New York, NY). 1998;281(5384):1840–2.
8. Drake JM, Griffen BD. Early warning signals of extinction in deteriorating environments. Nature. 2010;467(7314):456–9.
9. Scheffer M, Carpenter S, Foley JA, Folke C, Walker B. Catastrophic shifts in ecosystems. Nature. 2001;413(6856):591–6.
10. Dakos V, Scheffer M, van Nes EH, Brovkin V, Petoukhov V, Held H. Slowing down as an early warning signal for abrupt climate change. Proc Natl Acad Sci USA. 2008;105(38):14308–12.
11. Lenton TM, Held H, Kriegler E, Hall JW, Lucht W, Rahmstorf S, Schellnhuber HJ. Tipping elements in the Earth's climate system. Proc Natl Acad Sci USA. 2008;105(6):1786–93.
12. May RM, Levin SA, Sugihara G. Ecology for bankers. Nature. 2008;451(7181):893–4.
13. Kinouchi O, Copelli M. Optimal dynamical range of excitable networks at criticality. Nat Phys. 2006;2(5):348–51.
14. Sneppen K, Bak P, Flyvbjerg H, Jensen MH. Evolution as a self-organized critical phenomenon. Proc Natl Acad Sci USA. 1995;92(11):5209–13.
15. Liu R, Chen P, Chen L. Single-sample landscape entropy reveals the imminent phase transition during disease progression. Bioinformatics. 2020;36(8):1522–32.
16. Liu R, Li M, Liu ZP, Wu J, Chen L, Aihara K. Identifying critical transitions and their leading biomolecular networks in complex diseases. Sentific Reports. 2012;2:813.
17. Chen L, Liu R, Liu ZP, Li M, Aihara K. Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. Sci Rep. 2012;2:342.
18. Danning H, Zhi-Ping L, Masao H, Shuichi K, Luonan C. Coexpression network analysis in chronic hepatitis B and C hepatic lesions reveals distinct patterns of disease progression to hepatocellular carcinoma. J Mol Cell Biol. 2012;3:140.
19. Scheffer M, Bascompte J, Brock WA, Brovkin V, Carpenter SR, Dakos V, Held H, Nes EHV, Rietkerk M, Sugihara G. Early-warning signals for critical transitions. Nature. 2009;461(7260):53–9.

20. Xiaoping L, Xiao C, Rui L, Xiangtian Y, Luonan C, Kazuyuki A, Papin JA. Quantifying critical states of complex diseases using single-sample dynamic network biomarkers. Plos Comput Biol. 2017;13(7):e1005633.

21. Achiron A, Grotto I, Balicer R, Magalashvili D, Feldman A, Gurevich M. Microarray analysis identifies altered regulation of nuclear receptor family members in the pre-disease state of multiple sclerosis. Neurobiol Dis. 2010;38(2):201–9.

22. Liu R, Wang X, Aihara K, Chen L. Early diagnosis of complex diseases by molecular biomarkers, network biomarkers, and dynamical network biomarkers. Med Res Rev. 2014;34(3):455–78.

23. Litt B, Esteller R, Echauz J, D'Alessandro M, Shor R, Henry T, Pennell P, Epstein C, Bakay R, Dichter M, et al. Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. Neuron. 2001;30(1):51–64.

24. McSharry PE, Smith LA, Tarassenko L: Prediction of epileptic seizures: are nonlinear methods relevant? *Nature Med.* 2003, 9(3):241–242; author reply 242.

25. Paek SH, Chung HT, Jeong SS, Park CK, Kim CY, Kim JE, Kim DG, Jung HW. Hearing preservation after gamma knife stereotactic radiosurgery of vestibular schwannoma. Cancer. 2005;104(3):580–90.

26. Liu W, Lähdesmäki H, Dougherty ER. Shmulevich I (2008) Inference of Boolean networks using sensitivity regularization. EURASIP J Bioinformatics Syst Biol. 2008;1:780541.

27. Xu P, Wu Q, Lu D, Yu J, Rao Y, Kou Z, Fang G, Liu W, Han H. A systematic study of critical miRNAs on cells proliferation and apoptosis by the shortest path. BMC Bioinformatics. 2020;21(1):396.

28. Rhodes DR, Sanda MG, Otte AP, Chinnaiyan AM, Rubin MA. Multiplex biomarker approach for determining risk of prostate-specific antigen-defined recurrence of prostate cancer. J Natl Cancer Inst. 2003;95(9):661–8.

29. Mikolajczyk SD, Song Y, Wong JR, Matson RS, Rittenhouse HG. Are multiple markers the future of prostate cancer diagnostics? Clin Biochem. 2004;37(7):519–28.

30. Huang Y, Isharwal S, Haese A, Chun FK, Makarov DV, Feng Z, Han M, Humphreys E, Epstein JI, Partin AW, et al. Prediction of patient-specific risk and percentile cohort risk of pathological stage outcome using continuous prostate-specific antigen measurement, clinical stage and biopsy Gleason score. BJU Int. 2011;107(10):1562–9.

31. Berchuck A. Biomarkers in the ovary. J Cell Biochem Suppl. 1995;23:223–6.

32. Soussi T, Wiman KG. Shaping genetic alterations in human cancer: the p53 mutation paradigm. Cancer Cell. 2007;12(4):303–12.

33. Han H, Liu W. The coming era of artificial intelligence in biological data science. BMC Bioinformatics. 2019;20(Suppl 22):712–712.

34. Jin G, Zhou X, Wang H, Zhao H, Cui K, Zhang XS, Chen L, Hazen SL, Li K, Wong ST. The knowledge-integrated network biomarkers discovery for major adverse cardiac events. J Proteome Res. 2008;7(9):4013–21.

35. Ideker T, Sharan R. Protein networks in disease. Genome Res. 2008;18(4):644–52.

36. Schadt EE. Molecular networks as sensors and drivers of common human diseases. Nature. 2009;461(7261):218–23.

37. Chen P, Liu R, Li Y, Chen L. Detecting critical state before phase transition of complex biological systems by hidden Markov model. Bioinformatics. 2016;32(14):2143–50.

38. Chen P, Li Y, Liu X, Liu R, Chen L. Detecting the tipping points in a three-state model of complex diseases by temporal differential networks. J Transl Med. 2017;15(1):217.

39. Chen P, Chen E, Chen L, Zhou XJ, Liu R. Detecting early-warning signals of influenza outbreak based on dynamic network marker. J Cell Mol Med. 2019;23(1):395–404.

40. Liu R, Yu X, Liu X, Xu D, Aihara K, Chen L. Identifying critical transitions of complex diseases based on a single sample. Bioinformatics. 2014;30(11):1579–86.

41. Liu R, Chen P, Aihara K, Chen L. Identifying early-warning signals of critical transitions with strong noise by dynamical network markers. Sci Rep. 2015;5:17501.

42. Mitra M, Alexander S, Joseph Z, Castaño IG, Leong-Quong RYY, Hannah C, Kalliopi T, Alessandro G, Sui H. Cell fate decision as high-dimensional critical state transition. PLOS Biol. 2016;14(12):2000640.

43. Richard A, Boullu L, Herbach U, Bonnafoux A, Gandrillon O. Single-cell-based analysis highlights a surge in cell-to-cell molecular variability preceding irreversible commitment in a differentiation process. PLOS Biol. 2016;14(12):e1002585.

44. Lesterhuis WJ, Bosco A, Millward MJ, Small M, Nowak AK, Lake RA. Dynamic versus static biomarkers in cancer immune checkpoint blockade: unravelling complexity. Nat Rev Drug Discovery. 2017;16(4):264–72.

45. Tarca AL, Draghici S, Khatri P, Hassan SS, Mittal P, Kim JS, Kim CJ, Kusanovic JP, Romero R. A novel signaling pathway impact analysis. Bioinformatics. 2009;25(1):75–82.

46. Ehrhardt C, Ludwig S. A new player in a deadly game: influenza viruses and the PI3K/Akt signalling pathway. Cell Microbiol. 2010;11(6):863–71.

47. Lee JP, Kuan YH, Yi-Ching LI, Chen HY, Lin RH, Huang SS, Chen HL, Kuan PC, Liao MF, Chen CJ. Protective effects of luteolin against lipopolysaccharide-induced acute lung injury involves inhibition of MEK/ERK and PI3K/Akt pathways in neutrophils. Acta Pharmacol Sin. 2010;7:831–8.

48. Deng W, Li CY, Tong J, Zhang W, Wang DX. Regulation of ENaC-mediated alveolar fluid clearance by insulin via PI3K/Akt pathway in LPS-induced acute lung injury. Respir Res. 2012;13(1):29–29.

49. Yang Y, Cheng Y, Lian QQ, Yang L. Fang GS (2013) Contribution of CFTR to alveolar fluid clearance by lipoxin A4 via PI3K/Akt pathway in LPS-induced acute lung injury. Med Inflam. 2013;6:862628.

50. Phillips RJ, Mestas J, Gharaee-Kermani M, Burdick MD, Sica A, Belperio JA, Keane MP, Strieter RM. Epidermal growth factor and hypoxia-induced expression of CXC chemokine receptor 4 on non-small cell lung cancer cells is regulated by the phosphatidylinositol 3-kinase/PTEN/AKT/mammalian target of rapamycin signaling pathway and activation of hypoxia inducible factor-1alpha. J Biol Chem. 2005;280(23):22473–81.

51. Wan L, Pantel K, Kang Y. Tumor metastasis: moving new biological insights into the clinic. Nat Med. 2013;19(11):1450–64.

52. Chiang AC. J M: Molecular basis of metastasis. N Engl J Med. 2008;359(26):2814–23.

53. Hansen HH. Treatment of advanced non-small cell lung cancer. BMJ (International Edition). 2002;325(7362):452–453.

54. Sciuto AM, Phillips CS, Orzolek LD, Hege AI, Moran TS, Dillman JF. Genomic analysis of murine pulmonary tissue following carbonyl chloride inhalation. Chem Res Toxicol. 2005;18(11):1654–60.

Huo *et al. BMC Bioinformatics* 2022, **22**(Suppl 12):367

Page 15 of 15

55. Huang Y, Zaas AK, Rao A, Dobigeon N, Woolf PJ, Veldman T, Øien NC, McClain MT, Varkey JB, Nicholson B, et al. Temporal dynamics of host molecular responses differentiate symptomatic and asymptomatic influenza a infection. PLoS Genet. 2011;7(8):e1002234.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.