frontiers in Genetics

ORIGINAL RESEARCH
published: 04 April 2019
doi: 10.3389/fgene.2019.00285

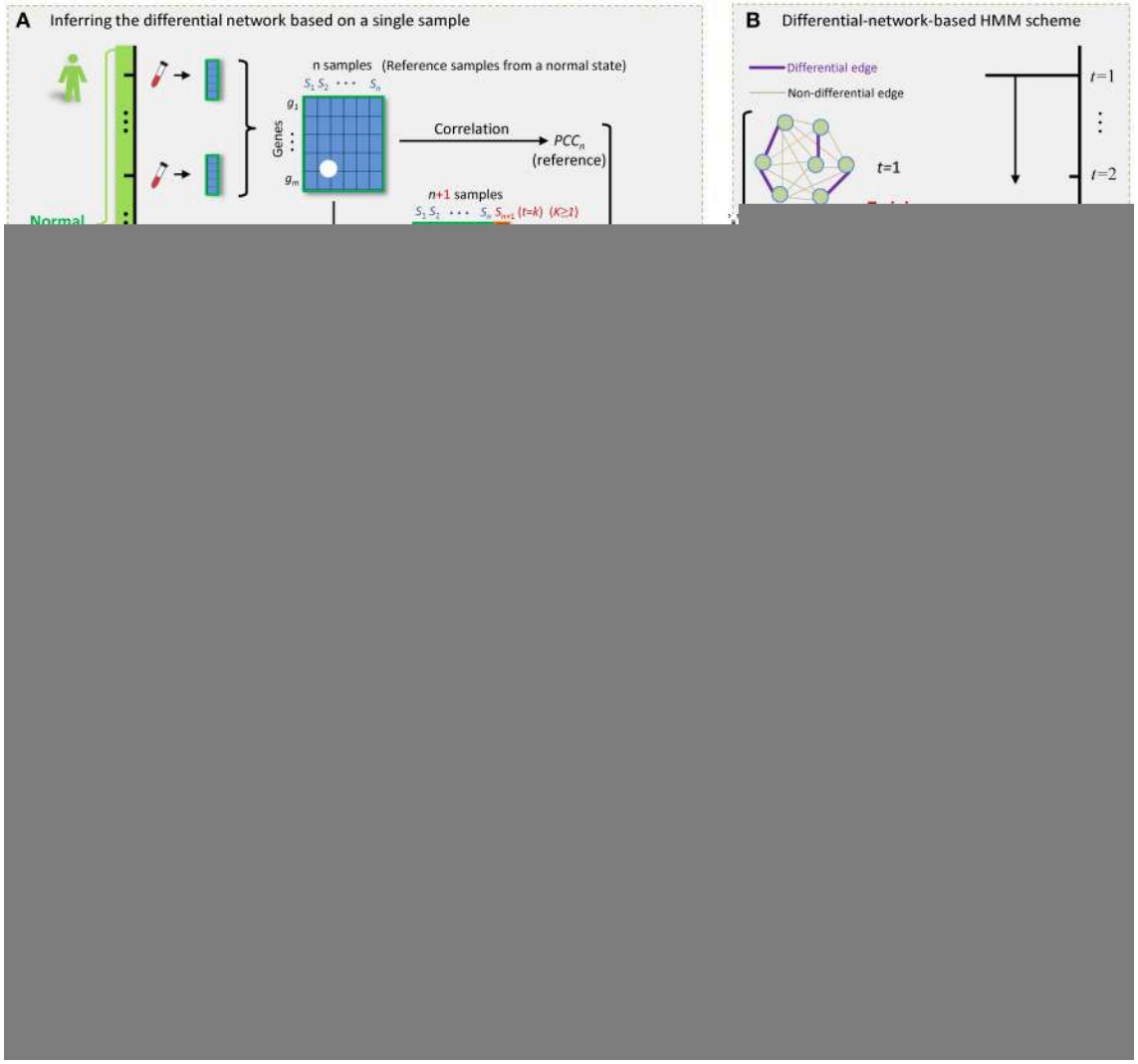

FIGURE 1 | The outline for identifying the SSI score based on HMM. **(A)** The progression of complex diseases is generally modeled as three states, i.e., a normal state, a pre-disease state, and a disease state. The pre-disease state is immediately before the sudden deterioration, which is sensitive to treatment and reversible to the normal state. The disease state is usually0.2047.21(l).283354(s)-277.026(s)0275587(e)0.20647(s)-0.29406(e)-278.537(r)0.112026(s)027557584(e)0.20604(o)-0.22b06(e)0.203l04(o)-0.2

state was viewed as a time-varying Markov process considering its dynamical instability. Taking multiple normal samples as the references or background, a differential network whose edges carried the differential information before and after combing a single sample with references, was obtained specific to the single sample derived at a time point (**Figure 1A**). Then, under the hypothesis that a time point $t = T$ ($T > 2$) is the candidate tipping point, a probabilistic score, namely single-sample-based

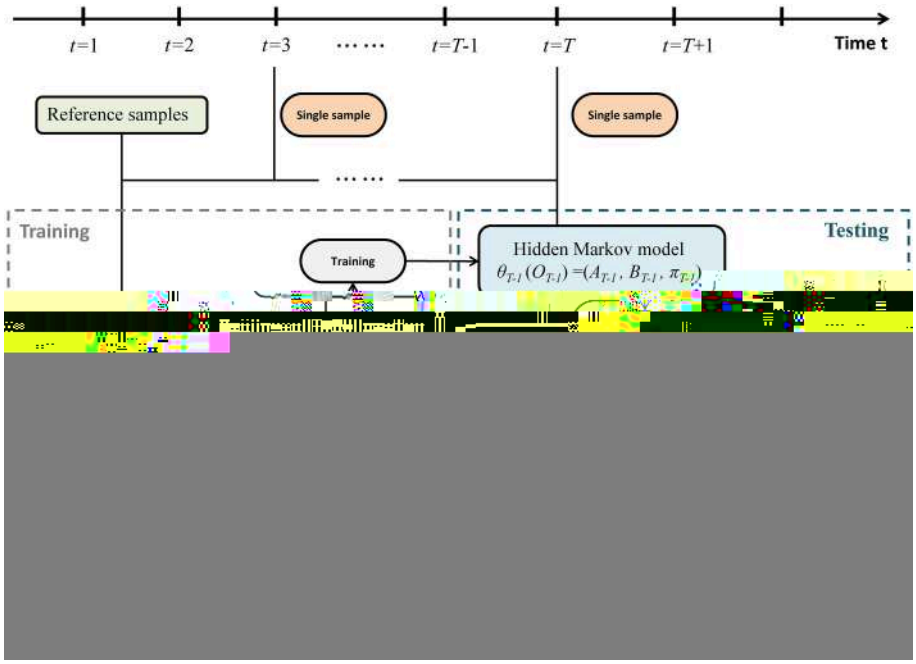

FIGURE 2 | The algorithm of the single-sample-based HMM. The above flowchart shows how the algorithm works based on a series of single case samples. Regarding a point $t = T$ ($T > 2$) as a candidate tipping point, the sample series is divided into training part ranging from $t = 1$ to $t = T-1$, and testing part starting from $t = T$. If a probabilistic score (single-sample-based inconsistency score, SSI score) increases significantly, then the candidate $t = T$ is determined as the identified tipping point, and the algorithm ends. Otherwise, if there is no significant change in SSI score, then $t = T$ is classified as a time point belonging to the normal state, and the algorithm continues with $t = T+1$ being a new candidate tipping point.

transition matrix at time point $T-1$ is

$$A^{T-1} = (a_{ij})_{2\times 2}$$

with

$$a_{ij} = P(s_q = P_i | s_{q-1} = P_j), i, j \in \{0, 1\}.$$

$q - 1 \in \{1, \ldots, T - 2\}$

in the differential network due to the time-varying and fluctuating dynamics of the system. Specifically, the algorithm is guaranteed by the generic properties 2 and 3 listed in section Theoretical Basis.

Data Accessing and Processing for Real Datasets

Two gene expression profiling datasets including the time-course dataset for influenza virus infection process (GSE30550) downloaded from the NCBI GEO database (www.ncbi.nlm.nih.gov/geo) and stage-course dataset for stomach adenocarcinoma (STAD) from TCGA database (http://cancergenome.nih.gov). For omics data (GSE30550), we discarded the probes without corresponding NCBI Entrez gene symbol. After removing any redundancy in dataset GSE30550, we obtained 11,451 molecules through probe mapping. For each gene mapped by multiple probes, the average value was employed as the gene expression.

When applied the algorithm to both two disease datasets, there were two extra steps as follows.

First, the expression profiling information was mapped to the protein-protein interaction networks from STRING (http://stringw-db.org) (Szklarczyk et al., 2014) for *Homo sapiens*. In such a network, the edges were filtered by the confidence level with a threshold of 0.700. All the isolated nodes were discarded. Then we choose the cutoff parameter d so that there are only 10% edges in the first differential network comparing with original STRING network, that is, over 90% edges disappear comparing with the original STRING network due to the generic property that the network structure would remain stable during the normal stage, and thus there are few edges in a differential network based on samples generated from normal stage.

Second, the differential network was partitioned into local networks to reduce computational complexity. Each local network contained a center node and its first-order neighbors. The local SSI score for each local network was calculated through above algorithm. Given k local networks, then a weighted average SSI score was derived as follows,

$$\mathrm{SSI} = \frac{n_1 SSI_1 + n_2 SSI_2 + \ldots + n_k SSI_k}{n_1 + n_2 + \ldots + n_k},$$

Where n_i denotes the number of nodes in the i-th local network ($I = 1, 2, \ldots, k$) and SSI_i stands for the local SSI score of this subnetwork.

The networks were visualized using Cytoscape (www.cytoscape.org) and the functional analysis was based on Ingenuity Pathway Analysis (IPA, http://www.ingenuity.co

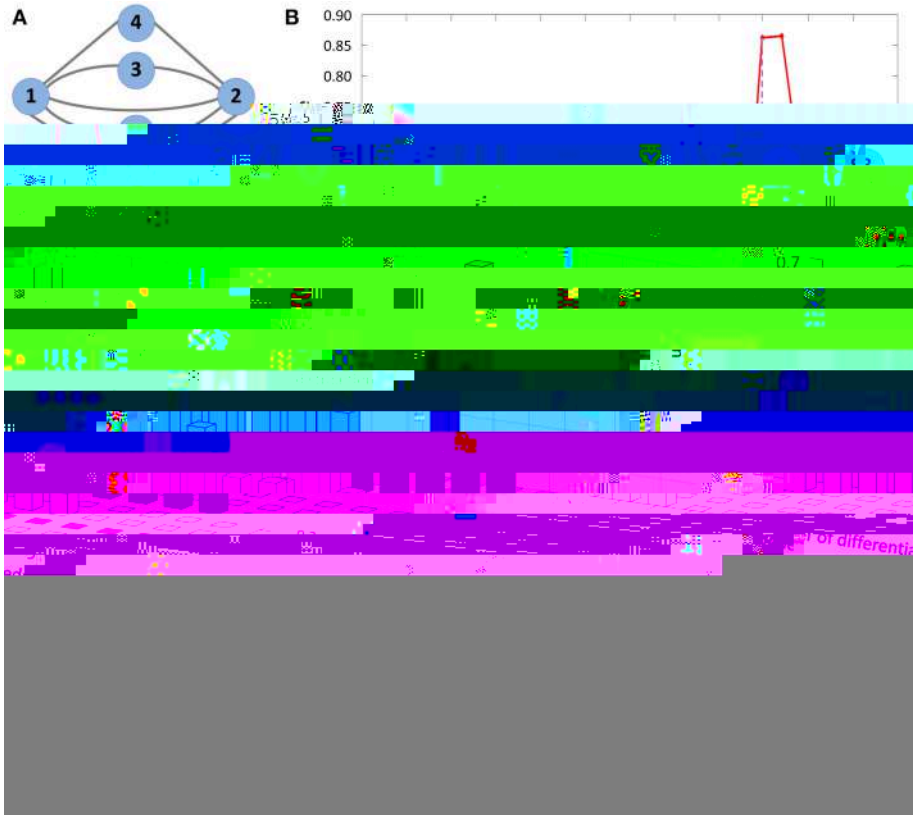

FIGURE 3 | The application of SSI score in numerical simulation. **(A)** The numerical simulation was based on a nine-node regulatory network. **(B)** The abrupt increase of SSI score indicates the tipping point at $P = 0$. **(C)** From the dynamical changes of differential-edge distribut

FIGURE 4 | The application of SSI score in STAD dataset. **(A)** The significant increase of SSI score indicates the tipping point at stage IIIA, before the deterioration into distant metastasis at stage IV. **(B)** The average expression deviation between each single sample and the reference. **(C)** The dynamical evolution of the whole gene regulatory network. The top 1% genes with the largest local SSI scores were arranged at the right-bottom corner.

injection, 0, 5, 12, 21, 29, 36, 45, 53, 60, 69, 77, 84, 93, 101, and 108 h after the injection. At each time point, there was only a single sample for each subject. By employing the proposed method, we obtained the individual-specific SSI score for each subject either in symptomatic or asymptomatic group.

The individual-specific SSI scores in **Figure 5B** demonstrated that there were obvious signals provided by SSI score for all symptomatic subjects (9 red curves), while there were few significant changes in the SSI scores for asymptomatic subjects (8 blue curves). The specific SSI scores for nine symptomatic subjects were shown in **Figure 5C**. Clearly, the SSI score indicated the pre-disease states (the state before the appearance of clinical symptom) for each symptomatic individual, with 100% accuracy. However, there was 25% false positive rate (**Figure 5A**). To demonstrate the evolution of individual-specific differential network, two sets of differential networks, respectively, for two symptomatic subjects, i.e., subject 1 and subject 12, were illustrated in **Figure 6**. Clearly, at the respective tipping point, there were many differential edges arising just before the emergence of clinic symptoms. At the tipping point of each symptomatic subject, the top 1% genes with the largest local SSI scores were regarded as a set of dynamical network biomarker, which were selected for further functional analysis.

Based on IPA analysis, the common SSI-signaling genes were highly related to functions annotation "Quantity of lymphocytes" (P-value = 2.23E-11), "Inflammation" (P-value = 2.47E-10), "Viral Infection" (P-value = 1.06E-09), "Homeostasis of leukocytes" (P-value = 1.14E-08). From KEGG enrichment analysis, the common SSI-signaling genes were enriched in Influenza A, and a variety of cellular pathways including PI3K-Akt signaling pathway, MAPK signaling pathway, NF-kappa B signaling pathway, etc. The functional analysis again validated the effectiveness of SSI-signaling genes. A list of common SSI-signaling genes for influenza infection was provided in **Table S2**.

DISCUSSION

Detecting the early-warning signal before a sudden deterioration into a severe disease state is crucial to patients all over the world. However, it is generally challenging to signal such critical transition through only a single case sample, since the lack of samples disables statistical indices and thus makes conventional methods fail. In this work, we proposed a computational method to identify the pre-disease state on the basis of a single sample. Specifically, given a number of reference samples which can be the normal samples of an individual (**Figure 1A**), the proposed method can distinguish the abnormal single sample by a differential-network-based HMM scheme. The proposed method has been validated by both the numerical simulation (**Figure 3**) and two real datasets (**Figures 4**, **5**).

Comparing with the traditional methods which are mostly based on the differential expression of observed biomolecules, the proposed method aims at exploring the dynamic information of

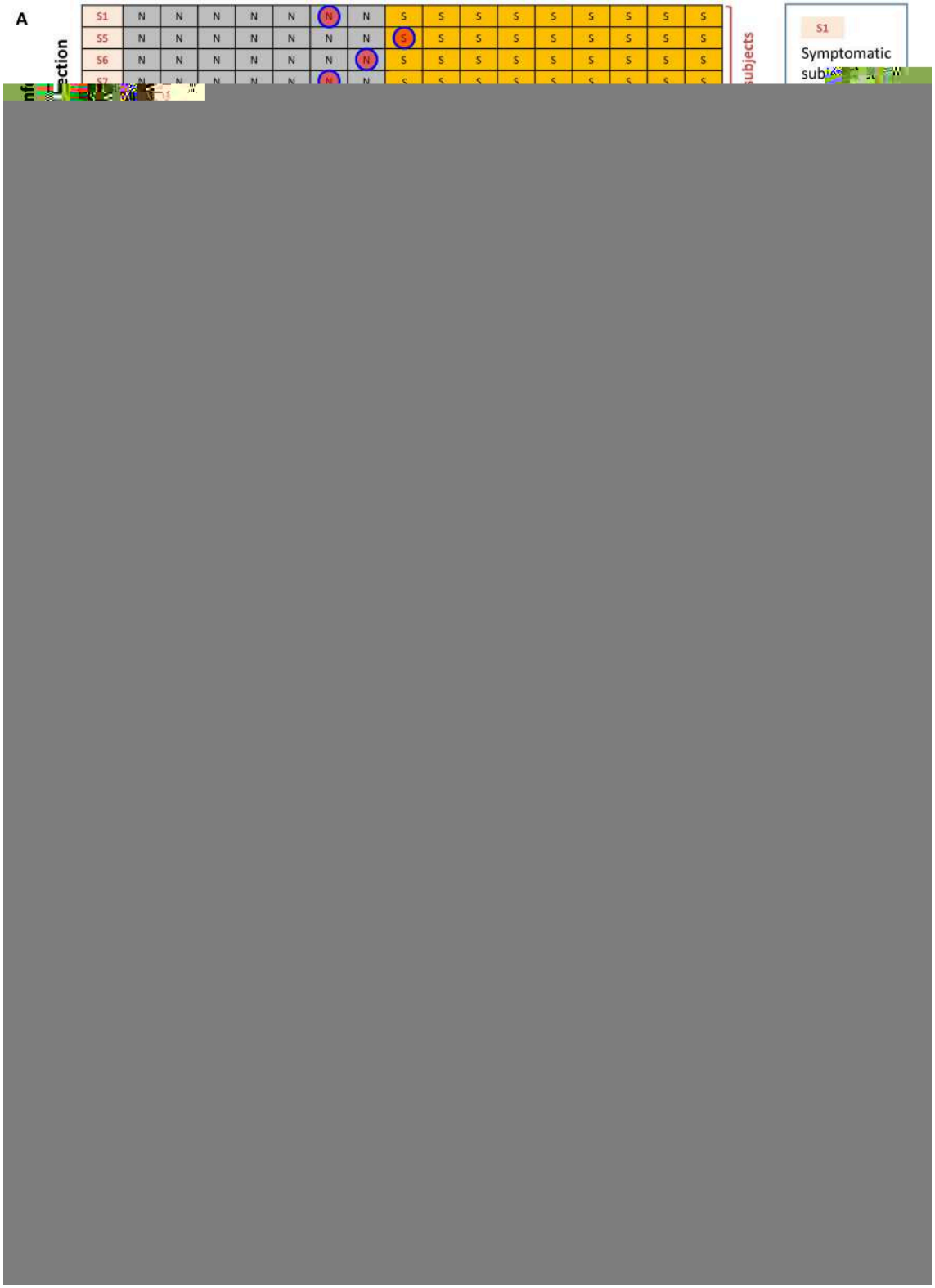

FIGURE 5 | The application of SSI score in influenza infection dataset. **(A)** The overall information of the 17 subjects in the influenza-infection challenge. **(B)** Line chart of SSI score for all 17 subjects. The red curves are for symptomatic subjects, while the blue curves represent asymptomatic subjects. **(C)** The individual-specific SSI scores for 9 symptomatic subjects. For each SSI curve, the star symbol represents the time point when SSI-score signal arises, the diamond symbol represents the time point at which the initial flu symptoms appears.

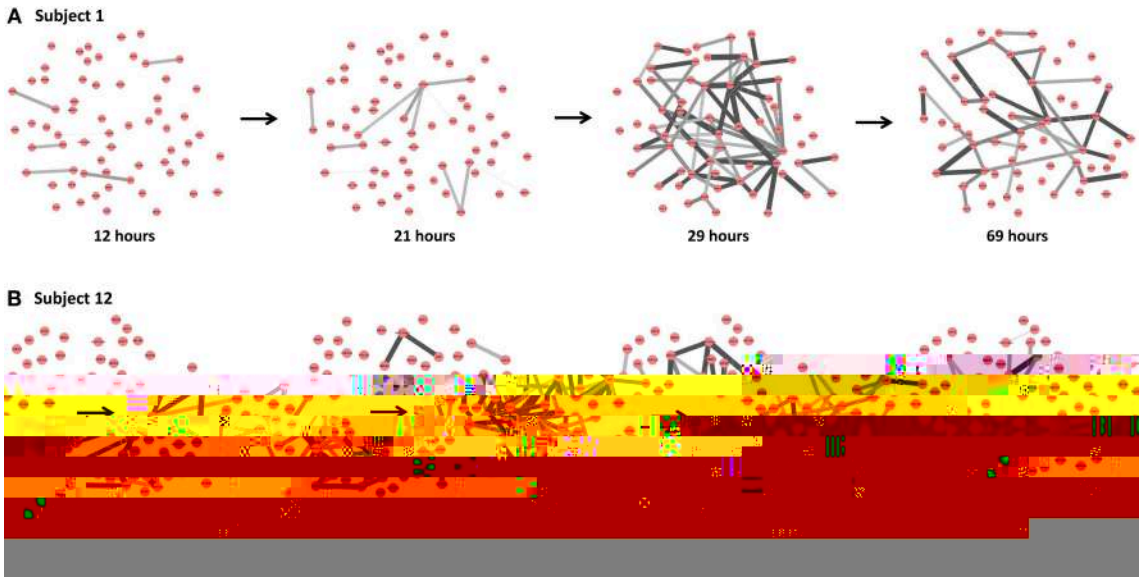

FIGURE 6 | The dynamical evolution of subject-specific networks. To illustrate the dynamical evolution of the differential network, the individual-specific networks of two symptomatic subjects (subjects 1 and 12) were exhibited. **(A)** The individual-specific networks for subject 1. **(B)** The individual-specific networks for subject 12. Clearly, at the identified tipping point of each subject, there were considerably more differential edges than that at other time point.

differential associations among biomolecules when a biological system is in the vicinity of a tipping point. This method thus possesses several obvious advantages. First, it works when only a single case sample is available, which benefits the analysis in personalized medicine. Second, it detects the pre-disease state rather than a disease state, which may help to achieve early diagnosis of some complex diseases. Third, it well-exhibits the critical properties at a network level which may provide new insights into catastrophic deterioration, such as the abnormally arising differential associations.

Although the proposed method is merely a step toward the identification of pre-disease state and the algorithm is expected to be improved in both sensitive and accurate ways, following the idea of personalized medicine, it provides a computational way and achieves individual-specific analysis and prediction by making use of only a single sample.

DATA AVAILABILITY

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30550.

AUTHOR CONTRIBUTIONS

RL and PC conceived the project. PC supervised the project. JZ, XY, and YL performed computational and analysis. All authors wrote the manuscript and read and approved the final manuscript.

FUNDING

This work was supported by National Natural Science Foundation of China (Nos. 11771152, 91530320, 61803360, and 11871456), Pearl River Science and Technology Nova Program of Guangzhou (No. 201610010029), Major Science and Technology Projects in Guangdong Province under Grant No. 2015B010128008.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00285/full#supplementary-material

REFERENCES

Bilmes, J. A. (1998). A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. *Int Comput Sci Inst.* 4:126.

Chen, L., Liu, R., Liu, Z., Li, M., and Aihara, K. (2012). Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. *Sci Rep.* 2, 342–349. doi: 10.1038/srep00342

Chen, L., Wang, R. S., and Zhang, X. S. (2009). *Biomolecular Networks Methods and Applications in Systems Biology* Vol. 10. Hoboken, NJ: John Wiley and Sons.

Chen, P., Chen, E., Chen, L., Zhou, X. J., and Liu, R. (2018). Detecting early-warning signals of influenza outbreak based on dynamic

network marker. *J Cell Mol Med*. 23, 395–404. doi: 10.1111/jcmm.13943

Chen, P., Li, Y., Liu, X., Liu, R., and Chen, L. (2017). Detecting the tipping points in a three-state model of complex diseases by temporal differential networks. *J Transl Med* 15, 217–231. doi: 10.1186/s12967-017-1320-7

Chen, P., Liu, R., Aihara, K., and Chen, L. (2015). Identifying critical differentiation state of MCF-7 cells for breast cancer by dynamical network biomarkers. *Front Genet* 6:252. doi: 10.3389/fgene.2015.00252

Chen, P., Liu, R., Li, Y., and Chen, L. (2016). Detecting critical state before phase transition of complex biological systems by hidden Markov model. *Bioinformatics* 32, 2143–2150. doi: 10.1093/bioinformatics/btw154

Crapoulet, N., O'Brien, P., Ouellette, R. J., and Robichaud, G. A. (2011). Coordinated expression of Pax-5 and FAK1 in metastasis. *Anti Cancer Agents Med Chem*. 11, 643–649. doi: 10.2174/187152011796817637

Guide, D. (2009). *Stomach Cancer Treatment Choices by Type and Stage of Stomach Cancer*. New York, NY: American Cancer Society.

He, D., Liu, Z. P., Honda, M., Kan01eko, S., and Chen, L. (2012). Coexpression network analysis in chronic hepatitis B and C hepatic lesions reveals distinct patterns of disease progression to hepatocellular carcinoma. *J Mol Cell Biol*. 4, 140–152. doi: 10.1093/jmcb/mjs011

Hirata, Y., Bruchovsky, N., and Aihara, K. (2010). Development of a mathematical model that predicts the outcome of hormone therapy for prostate cancer. *J Theor Biol*. 264, 517–527. doi: 10.1016/j.jtbi.2010.02.027

Hwang, S. J., Seol, H. J., Park, Y. M., Kim, K. H., Gorospe, M., Nam, D. H., et al. (2012). MicroRNA-146a suppresses metastatic activity in brain metastasis. *Mol Cells* 34, 329–334. doi: 10.1007/s10059-012-0171-6

Lesterhuis, W. J., Bosco, A., Millward, M. J., Small, M., Nowak, A. K., and Lake, R. A. (2017). Dynamic versus static biomarkers in cancer immune checkpoint blockade: unravelling complexity. *Nat Rev Drug Discov*. 16, 264–272. doi: 10.1038/nrd.2016.233

Litt, B., Esteller, R., Echauz, J., D'Alessandro, M., Shor, R., Henry, T., et al. (2001). Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. *Neuron* 30, 51–64. doi: 10.1016/S0896-6273(01)00262-8

Liu, R., Chen, P., Aihara, K., and Chen, L. (2015). Identifying early-warning signals of critical transitions with strong noise by dynamical network markers. *Sci Rep*. 5, 17501–17513. doi: 10.1038/srep17501

Liu, R., Li, M., Liu, Z. P., Wu, J., Chen, L., and Aihara, K. (2012). Identifying critical transitions and their leading biomolecular networks in complex diseases. *Sci Rep*. 2, 813–821. doi: 10.1038/srep00813

Liu, R., Wang, J., Ukai, M., Sewon, K., Chen, P., Suzuki, Y., et al. (2018). Hunt for the tipping point during endocrine resistance process in breast cancer by dynamic network biomarkers. *J Mol Cell Biol* doi: 10.1093/jmcb/mjy059. [Epub ahead of print].

Liu, R., Wang, X., Aihara, K., and Chen, L. (2014a). Early diagnosis of complex diseases by molecular biomarkers, network biomarkers, and dynamical network biomarkers. *Med Res Rev* 34, 455–478. doi: 10.1002/med.21293

Liu, R., Yu, X., Liu, X., Xu, D., Aihara, K., and Chen, L. (2014b). Identifying critical transitions of complex diseases based on a single sample. *Bioinformatics* 30, 1579–1586. doi: 10.1093/bioinformatics/btu084

Lu, X., Wang, Q., Hu, G., Van Poznak, C., Fleisher, M., Reiss, M., et al. (2009). ADAMTS1 and MMP1 proteolytically engageEGF-like ligands in an osteolytic signaling cascade for bone metastasis. *Genes Dev*. 16, 1882–1894. doi: 10.1101/gad.1824809

McSharry, P. E., Smith, L. A., and Tarassenko, L. (2003). Prediction of epileptic seizures: are nonlinear methods relevant? *Nat Med*. 9, 241–242. doi: 10.1038/nm0303-241

Mojtahedi, M., Skupin, A., Zhou, J., Castano, I. G., Leong-Quong, R. Y., Chang, H., et al. (2016). Cell fate decision as high-dimensional critical state transition. *PLoS Biol* 14:640. doi: 10.1371/journal.pbio.2000640

Murray, J. D. (2002). *Mathematical Biology rd Edn*. New York, NY: Springer. doi: 10.1093/imammb/20.4.377

Park, Y. M., Hwang, S. J., Masuda, K., Choi, K. M., Jeong, M. R., Nam, D. H., et al. (2012). Heterogeneous nuclear ribonucleoprotein C1/C2 (hnRNPC) controls the metastatic potential of glioblastoma by regulating PDCD4. *Mol Cell Biol*. 32, 4237–4244. doi: 10.1128/mcb.00443-12

Richard, A., Boullu, L., Herbach, U., Bonnafoux, A., Morin, V., Vallin, E., et al. (2016). Single-cell-based analysis highlights a surge in cell-to-cell molecular variability preceding irreversible commitment in a differentiation process. *PLoS Biol* 14:e1002585. doi: 10.1371/journal.pbio.1002585

Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., et al. (2009). Early-warning signals for critical transitions. *Nature* 461, 53–59. doi: 10.1038/nature08227

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2014). STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 43, D447–D452. doi: 10.1093/nar/gku1003

Van't Veer, L. J., Dai, H., Van De Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530–536. doi: 10.1038/415530a

Vázquez-Villa, F., García-Ocaña, M., Galván, J. A., García-Martínez, J., García-Pravia, C., Menéndez-Rodríguez, P., et al. (2015). COL11A1/(pro) collagen 11A1 expression is a remarkable biomarker of human invasive carcinoma-associated stromal cells and carcinoma progression. *Tumor Biol*. 36, 2213–2222. doi: 10.1007/s13277-015-3295-4

Venegas, J. G., Winkler, T., Musch, G., Melo, M. F. V., Layfield, D., Tgavalekos, N., et al. (2005). Self-organized patchiness in asthma as a prelude to catastrophic shifts. *Nature* 434, 777–782. doi: 10.1038/nature03490

Wadhwa, R., Taketa, T., Sudo, K., Blum, M. A., and Ajani, J. A. (2013). Modern oncological approaches to gastric adenocarcinoma. *Gastroenterol Clin*. 42, 359–369. doi: 10.1016/j.gtc.2013.01.011

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.