

LnCeCell: a comprehensive database of predicted lncRNA-associated ceRNA networks at single-cell resolution

Peng Wang^{1,†}, Qiuyan Guo^{2,†}, Yangyang Hao^{1,†}, Qian Liu^{1,†}, Yue Gao¹, Hui Zhi¹, Xin Li¹, Shipeng Shang¹, Shuang Guo¹, Yunpeng Zhang^{1,*}, Shangwei Ning^{1,*} and Xia Li^{1,*}

¹College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China and

²Department of Gynecology, the First Affiliated Hospital of Harbin Medical University, Harbin 150081, China

Received August 14, 2020; Revised October 03, 2020; Editorial Decision October 13, 2020; Accepted November 09, 2020

ABSTRACT

Within the tumour microenvironment, cells exhibit different behaviours driven by fine-tuning of gene regulation. Identification of cellular-specific gene regulatory networks will deepen the understanding of disease pathology at single-cell resolution and contribute to the development of precision medicine. Here, we describe a database, LnCeCell (<http://www.bio-bigdata.net/LnCeCell/> or <http://bio-bigdata.hrbmu.edu.cn/LnCeCell/>), which aims to document cellular-specific long non-coding RNA (lncRNA)-associated competing endogenous RNA (ceRNA) networks for personalised characterisation of diseases based on the ‘One Cell, One World’ theory. LnCeCell is curated with cellular-specific ceRNA regulations from >94 000 cells across 25 types of cancers and provides >9000 experimentally supported lncRNA biomarkers, associated with tumour metastasis, recurrence, prognosis, circulation, drug resistance, etc. For each cell, LnCeCell illustrates a global map of ceRNA sub-cellular locations, which have been manually curated from the literature and related data sources, and portrays a functional state atlas for a single cancer cell. LnCeCell also provides several flexible tools to infer ceRNA functions based on a specific cellular background. LnCeCell serves as an important resource for investigating the gene regulatory networks within a single cell and can help researchers understand the regulatory mechanisms underlying complex microbial ecosystems and individual phenotypes.

INTRODUCTION

Human cancers are complex ecosystems composed of cells with distinct phenotypes, genotypes and epigenetic states (1). This intratumoral heterogeneity is a major obstacle in cancer treatment and a significant confounding factor in bulk-tumour profiling (2). Single-cell RNA sequencing (scRNA-seq) technologies (3,4) are powerful means of exploring tissue heterogeneity and cellular variability (5) and provide an unprecedented opportunity to address increasingly challenging biological questions and understand the atypical functional mechanisms involved in the development of cancers. Currently, many single-cell sequencing databases have been developed based on research requirements. For example, CancerSEA provides functional state activity profiles of single cells for several types of cancers (6). SCPortalen houses human and mouse single-cell transcriptomic datasets and provides access to available single-cell image files (7). However, these databases mainly focus on gene expression, while within a tumour, cells exhibit different behaviours driven by the fine-tuning of gene regulation. Thus, gene–gene associations at the single-cell level in tumours need to be explored.

Accumulating evidence suggests that a subset of long non-coding RNAs (lncRNAs) behave like competing endogenous RNAs (ceRNAs), and compete with microRNAs (miRNAs) to communicate with mRNAs (8). lncRNA-associated ceRNA regulation has been implicated in cell-fate determination and in various human diseases (9). For example, activated by oxidative stress, lncRNAs *H19* and *HULC* compete with *let-7a/let-7b* and *mir-372/mir-373* to target *IL-6* and *CXCR4*, respectively, and promote the tissue invasion and migration of cholangiocarcinoma cells (10). Additionally, the lncRNA *HOTAIR* competes with *mir-206* to regulate *MKLI* expression, promoting the migration and invasion of HeLa cells (11). Several databases

*To whom correspondence should be addressed. Tel: +86 451 86615922; Fax: +86 451 86615922; Email: lixia@hrbmu.edu.cn
Correspondence may also be addressed to Shangwei Ning. Tel: +86 451 86615922; Fax: +86 451 86615922; Email: ningsw@ems.hrbmu.edu.cn
Correspondence may also be addressed to Yunpeng Zhang. Tel: +86 451 86615922; Fax: +86 451 86615922; Email: zhangyp@hrbmu.edu.cn

[†]The authors wish it to be known that, in their opinion, the first four authors should be regarded as joint First Authors.

house ceRNA regulations (12–14); for example, starBase v2.0 has identified several ceRNA regulatory relationships by analysing miRNA binding sites on different types of RNAs (13). LncACTdb 2.0, a recently updated extensive database, provides comprehensive information on ceRNAs in different species and diseases (12). Although these databases are important for the understanding of ceRNA-mediated regulation in tumours, a database dedicated to deciphering ceRNA-mediated regulation within a single cell is still lacking.

Therefore, we developed LnCeCell, a comprehensive database documenting cellular-specific lncRNA-associated ceRNA networks predicted via high-throughput analysis of single-cell genomic data and reliable biomarkers manually curated from published literature. LnCeCell is a curation of cellular-specific ceRNA regulations from thousands of cells across 25 types of cancers, including: (i) >9000 experimentally supported lncRNA biomarkers of tumour metastasis, recurrence, prognosis, circulation, and drug resistance; (ii) cellular-specific ceRNA networks for primary, malignant, and metastatic cancer cells and immune cells; (iii) detailed information of ceRNA sub-cellular locations manually entered from literature and related data sources; (iv) clusters of distinct cellular populations that exhibit diverse behaviours such as angiogenesis, apoptosis, cell cycle, invasion, proliferation and stemness. LnCeCell provides a user-friendly searching and browsing interface. As an important supplement to the database, several flexible tools that facilitate retrieval and analysis of the data are also available. We expect the LnCeCell database to serve as an important resource for the investigation of single-cell ceRNA regulation, aiding in the understanding of regulatory mechanisms behind tiny, yet complex ecosystems.

MATERIALS AND METHODS

Data collection and processing

We collected cancer-related scRNA-seq datasets from Gene Expression Omnibus (GEO) and CancerSEA, ensuring that the number of cancer cells after quality control is >100, and the expression profiles can be divided into profiles of mRNA and lncRNA expression, through annotation using the GENCODE database (release 34, GRCh38). A total of 94 605 single cells derived from 40 single-cell datasets across 25 cancers was used to build the LnCeCell (Supplementary Figure S1). To identify single-cell ceRNA regulations, 108 668 candidate ceRNA regulations were downloaded from starBase (v2.0) (13) and LncACTdb (v2.0) (12). We used a published method, based on probability theory, for cell-specific network construction to identify ceRNA networks in single cells (Supplementary Methods, Supplementary Figure S2) (15). Briefly, the association of a ceRNA pair (lncRNA-mRNA) with a specific cell is estimated by testing statistical independence of their expression values between different cells. We identified 93 307 ceRNA regulations in 94 455 single cells with a false discovery rate (FDR) <0.05. In order to distinguish the functional states of the different cancer cell types, we downloaded the characteristic gene sets of the 14 functional states from CancerSEA (6). Based on the genetic signatures of the 14 functional states, the functional state of single cancer cells in each dataset

were evaluated using the gene set variation analysis (GSVA) package in R (16). The sub-cellular and extracellular vesicle locations of lncRNAs, miRNAs, and mRNAs were collected from related databases (17–21) and published literature. A total of 9306 experimentally supported lncRNA biomarkers were manually curated from the literature and integrated into the LnCeCell database. Detailed information on data collection and processing is described in the Supplementary Methods.

Database construction

LnCeCell is freely available at <http://www.bio-bigdata.net/LnCeCell/> or <http://bio-bigdata.hrbmu.edu.cn/LnCeCell/>. The LnCeCell online web server was developed using Java Server Pages of the Tomcat software (v6). The MySQL (v 5.5) data server was used to document and manage the datasets. Creation of result tables and visualisation of data were performed using jQuery (v1.11.3), Datable (1.10.10) and ECharts (V4.0) plugin software. All statistical analyses were performed using the R framework (V3.6.3). The LnCeCell website is supported on popular web browsers, such as Microsoft Edge, Google Chrome, Firefox and Safari.

RESULTS

Data collection and content of LnCeCell

The current version of LnCeCell contains 40 scRNA-seq expression profiles, covering 94 455 single cells and 25 cancer types. An overview of LnCeCell is shown in Figure 1. The average number of single cells per cancer type was 3778, with Ewing sarcoma having the largest number of single cells ($n = 13\ 170$) and cervical cancer having the smallest number ($n = 126$). For each cell, LnCeCell identified lncRNA-associated ceRNA regulations and provided a cellular-specific ceRNA network. Finally, a total of 93 307 unique ceRNAs were included in the database. Each type of cancer showed a different set of ceRNA regulations, and at the same time, there were many ceRNAs that played a regulatory role in a variety of cancers. A total of 703 pairs of ceRNA, such as *MALAT1-KRAS* and *NEAT1-BACH1*, can be identified in all cancer datasets. On average, each ceRNA was detected in 15 different cancer datasets (Supplementary Figure S3). CeRNA datasets of glioblastoma and cervical cancer showed the highest overlap, with 87 416 pairs of ceRNAs common for both datasets. In addition, several ceRNAs were also common between oesophageal squamous cell carcinoma and glioblastoma. Based on the ceRNA expression profiles, cells of the 40 scRNA-seq datasets were classified into different cell populations and displayed as cluster maps. To explore the diverse cellular behaviours driven by the cellular-specific ceRNA regulations, a comprehensive list of functional annotations has been integrated into LnCeCell, including 14 functional states of cancer cells from CancerSEA, 5917 gene sets from Gene Ontology terms (22), 1329 biological pathways from Molecular Signatures Database (MSigDB) (23) and 10 hallmark cancer processes (24). To further study the molecular function of ceRNAs, sub-cellular localisation of each ceRNA network, including lncRNA, miRNA

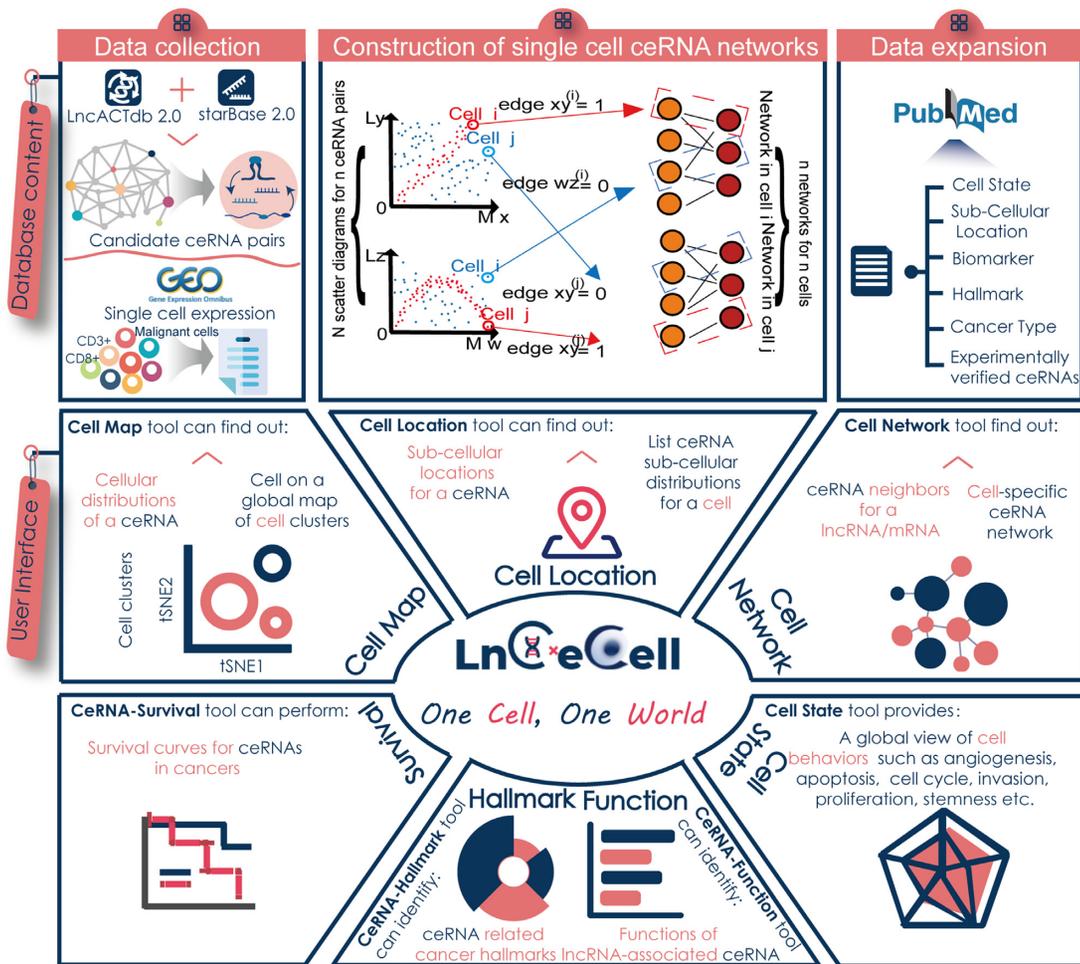


Figure 1. Database content and user interface of LnCeCell. The top area contains the database content, which includes identification of single cell ceRNA networks from high-throughput scRNA-seq datasets and functional annotations in literature. The middle and bottom area contain the user interface of LnCeCell. A panel of tools have been developed to infer ceRNA functions and cell states of a single cell.

and mRNA, were integrated into LnCeCell. For a comprehensive dataset of lncRNA-associated ceRNA, high-confidence lncRNA-ceRNA associations, from published literature in the PubMed database, were integrated in to LnCeCell. lncRNA-ceRNA regulations supported by information from high-confidence experiments, such as PCR, western blot, and luciferase reporter assay, are a part of the database (Supplementary Methods). The database also contains 9,036 experimentally supported lncRNA biomarkers of tumour metastasis, recurrence, prognosis, circulating, drug resistance, etc. For survival analysis of ceRNAs, bulk expression profiles and individual clinical information of 10 141 patients from The Cancer Genome Atlas (TCGA) were also integrated into the LnCeCell. In addition, as an important supplement of the database, several flexible tools that facilitate retrieval and analysis of the data were also made available (Figure 1). These include Cell-Map, which provides a global map of ceRNAs identified in distinct cellular populations, Cell-Location, which provides sub-cellular locations of ceRNAs, Cell-Network, which aids in the visualisation of a dysregulated ceRNA network in a single cell, and Cell-State, which allows users to view the func-

tional state of each cell. CeRNA-Function and CeRNA-Hallmark are used to identify dysregulated functions and cancer hallmarks of ceRNAs, respectively. Furthermore, CeRNA-Survival is used to perform Cox regression analysis and draw survival curves for ceRNAs.

Feature and utility of LnCeCell

LnCeCell allows users to explore lncRNA-associated ceRNA networks in a single cell and to infer cell states and functions within different tumour microenvironments. LnCeCell describes several aspects of single cells from a scRNA-seq dataset, including ceRNA networks, functional states, cell clustering maps, etc., and characterisation of ceRNA regulation consists of the following information: distribution in diverse cell populations, sub-cellular localisations, network of related ceRNA neighbours, functional enrichment, relevant cancer hallmarks, experimentally supported lncRNA biomarkers, and survival analysis for various cancers. The easy-to-use interface allows for searching, browsing, visualising and downloading data.

Global map of distinct cell populations. For each scRNA-seq dataset, LnCeCell classified cells into different cell populations based on the ceRNA expression profile, and provided a global map of cell populations. The global map can be obtained from the Cell Map section of the LnCeCell ‘HOME’ page (Supplementary Figure S4). When ‘ceRNA’ is clicked, the user is required to input the gene name or Ensembl ID of a pair of ceRNA (a mRNA and an lncRNA) and the dataset that the user wishes to explore. LnCeCell will classify cells into different cell populations based on their ceRNA occurrence profiles and provide a global map of the cell populations (Figure 2A). To further illustrate the cells wherein the input ceRNA is present, a global map of ceRNA distribution is also provided by LnCeCell (Figure 2B). In the Cell Map section, users can also determine the location of a cell from different cell populations (Supplementary Figure S5A, B). In addition, LnCeCell can also classify cells based on their scRNA-seq expression profile and provide a global map of cell clusters and ceRNA/cell locations in different cell populations (Supplementary Figure S5C, D).

Sub-cellular locations of ceRNAs. A cell location section was developed as a part of LnCeCell, to illustrate ceRNA sub-cellular locations, by manual curation from literature and related data sources. In this section, users can input the gene name or ID (Ensembl ID for an lncRNA/mRNA, miRBase ID for a miRNA) of a member of the ceRNA network (lncRNA/mRNA/miRNA) of their interest. LnCeCell will identify all possible sub-cellular locations of the input ceRNA, and the identified sub-cellular locations are demarcated with a red frame (Figure 2C). Users can also use this section with a single cell, from a scRNA-seq dataset, as the input. LnCeCell will provide all ceRNA regulations and a global view of all possible sub-cellular locations of these ceRNAs (Supplementary Figure S6A). Detailed sub-cellular location information, including ceRNA names/IDs, possible locations, identified tissues/cell lines, and data sources can be obtained by clicking the ‘Detailed sub-cellular table’ button (Supplementary Figure S6B).

Functional states of cells. To decipher the diverse functional states of cancer cells at a single-cell resolution, LnCeCell provides a global view of cell states/behaviours including stemness, invasion, metastasis, proliferation, epithelial-mesenchymal transition (EMT), angiogenesis, apoptosis, cell cycle, differentiation, DNA damage, DNA repair, hypoxia, inflammation and quiescence. For a single cell, enrichment scores for the 14 functional states, from CancerSEA, are shown as red lines and dots on a radar map (Figure 2D). The average scores of all cells are shown as grey lines and dots.

CeRNA networks of single cells. LnCeCell identifies lncRNA-associated ceRNA regulation and provides a cellular-specific ceRNA network for a single cell. In the Cell Network section, users can input the cell name to obtain a cellular-specific ceRNA network (Figure 2E). In the network, the blue and red nodes represent lncRNAs and mRNAs, respectively, while the edges represent the ceRNA

interactions between lncRNAs and mRNAs. When users move the cursor over a node in the network, all edges and nodes connected to it are highlighted. Users can choose different network layouts, such as the circular layout and force layout. Users can also determine ceRNA neighbours and construct a sub-network by inputting the name of an lncRNA or mRNA of interest (Figure 2F). Clicking a node starts a new search and provides a distribution map of the ceRNA in the human body and cancer biomarker annotations of the node (Figure 2G-H).

A panel of tools for ceRNA analysis. LnCeCell also provides a panel of tools to infer ceRNA functions and their association with cancers, and can be found in the ‘HOME’ page or ‘TOOLS’ section (Supplementary Figure S4). Users can use the CeRNA-Function to identify dysregulated functions of ceRNAs. LnCeCell will perform a ‘guilt-by-association’ strategy for a given lncRNA, (12,25) to infer ceRNA functions based on biological pathways and GO terms (Supplementary Figure S7). The CeRNA-Hallmark tool can be used to identify ceRNA-related cancer hallmarks such as insensitivity to antigrowth signals, tissue invasion, and metastasis. The enrichment scores of the 10 cancer hallmarks are shown as different sectors, and the area of each sector is directly proportional to the $-\log_{10}(P)$ value (Supplementary Figure S8). Further, LnCeCell houses high-throughput expression profiles and clinical follow-up information of thousands of patients across 33 cancer types from TCGA. The CeRNA-Survival tool builds a risk score model based on a linear combination of ceRNA expression values weighted by the Cox regression coefficients (25,26) (Supplementary Figure S9). Using this tool, users can obtain both Cox survival analysis and Kaplan-Meier curves of the ceRNA network members (lncRNA, miRNA, and mRNA) and the ceRNA interaction, respectively.

Flexible ways to access and download the dataset

LnCeCell provides user-friendly searching (Figure 3A) and browsing (Figure 3B) interfaces. In addition, a ‘QUICK SEARCH’ interface is available on the ‘HOME’ page allowing users to directly investigate data or perform analyses (Supplementary Figure S10). All the single cells, lncRNAs, mRNAs, and cancer biomarkers can be viewed comprehensively on the ‘BROWSE’ page (Figure 3B). On clicking a certain dataset listed under the ‘Disease’ catalogue, LnCeCell will provide a graph of all related cells, in which the size and colour of each cell changes with the number of ceRNA regulations identified in the cell. Detailed information can be viewed by clicking each cell in the graph. To further filter the result table, LnCeCell provides a search box on top of the data table, which allows users to obtain more precise results using a keyword (e.g. a certain type of disease or tissue). Users can reorder the data table by clicking on different column headers. The customised results table of ceRNAs and cells can be flexibly downloaded by clicking the ‘Copy’, ‘Excel’, ‘CSV’ and ‘PDF’ buttons. All data in LnCeCell can be freely downloaded on the ‘DOWNLOAD’ page.

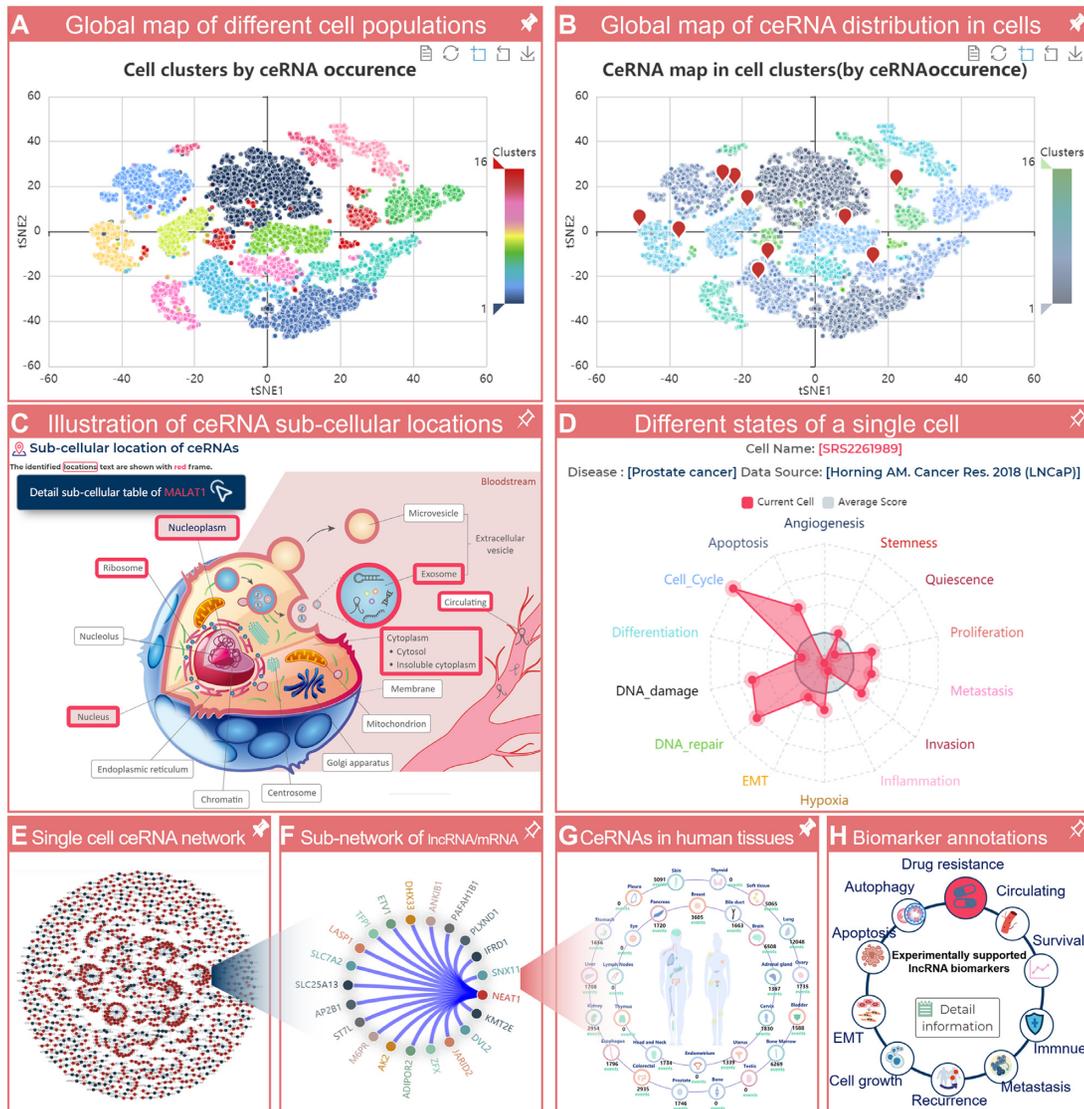


Figure 2. Feature and utility of LnCeCell. (A) For each scRNA-seq dataset, LnCeCell provides a global map of different cell populations. (B) CeRNA distribution in diverse cell populations of the dataset. (C) The possible sub-cellular locations of an input ceRNA. (D) Functional state information of a single cell. (E) The ceRNA regulatory network in a single cell. (F) The sub-network of an lncRNA or mRNA and its neighbours in the single-cell ceRNA network. (G) A ceRNA distribution map in the human body. (H) Cancer biomarker annotations of a lncRNA.

Example application

To demonstrate the usage and potential application of LnCeCell, we used *MALATI-KRAS* as a ceRNA input for the database (Figure 3A). The ceRNA interaction of *MALATI-KRAS* has been confirmed to play important roles in different cancers (27–29). With *MALATI-KRAS* as the search input, LnCeCell returns a result table indicating the number of competing miRNAs, and disease, tissue, and number/percentage of cells in which this ceRNA can be found (Figure 3C). In each line, the first and last columns directly leads users to the detailed information and analysis tools (Figure 3D). In the detail page, we found that the ceRNA of *MALATI-KRAS* was identified in 6.51% of 169 cells from bone marrow tissue of patients with acute myeloid leukaemia. LnCeCell provides a panel of tools for analysing the functions of *MALATI-KRAS* (Figure 3E).

A global map of *MALATI-KRAS* distribution in different cell clusters can be accessed in the Cell Map section of the detail page. In addition, we also found that the possible sub-cellular localisations of this ceRNA are nucleoplasm, ribosome, nucleolus, nucleus, chromatin, and exosomes. The Cell Network section provided a global view of all possible *MALATI*-related ceRNA interactions. Further, we have integrated a Pan-cancer comparison analysis section in our online database. This section provides users with a global view of the ceRNA distribution across different cancer cells (Supplementary Figure S11). Detailed information on the confidence values and experimental annotation can be accessed on this page (Supplementary Figure S12). For a single cancer cell, a functional state activity profile was generated by the Cell State tool. The Hallmarks and Function tools performed functional analyses for *MALATI*

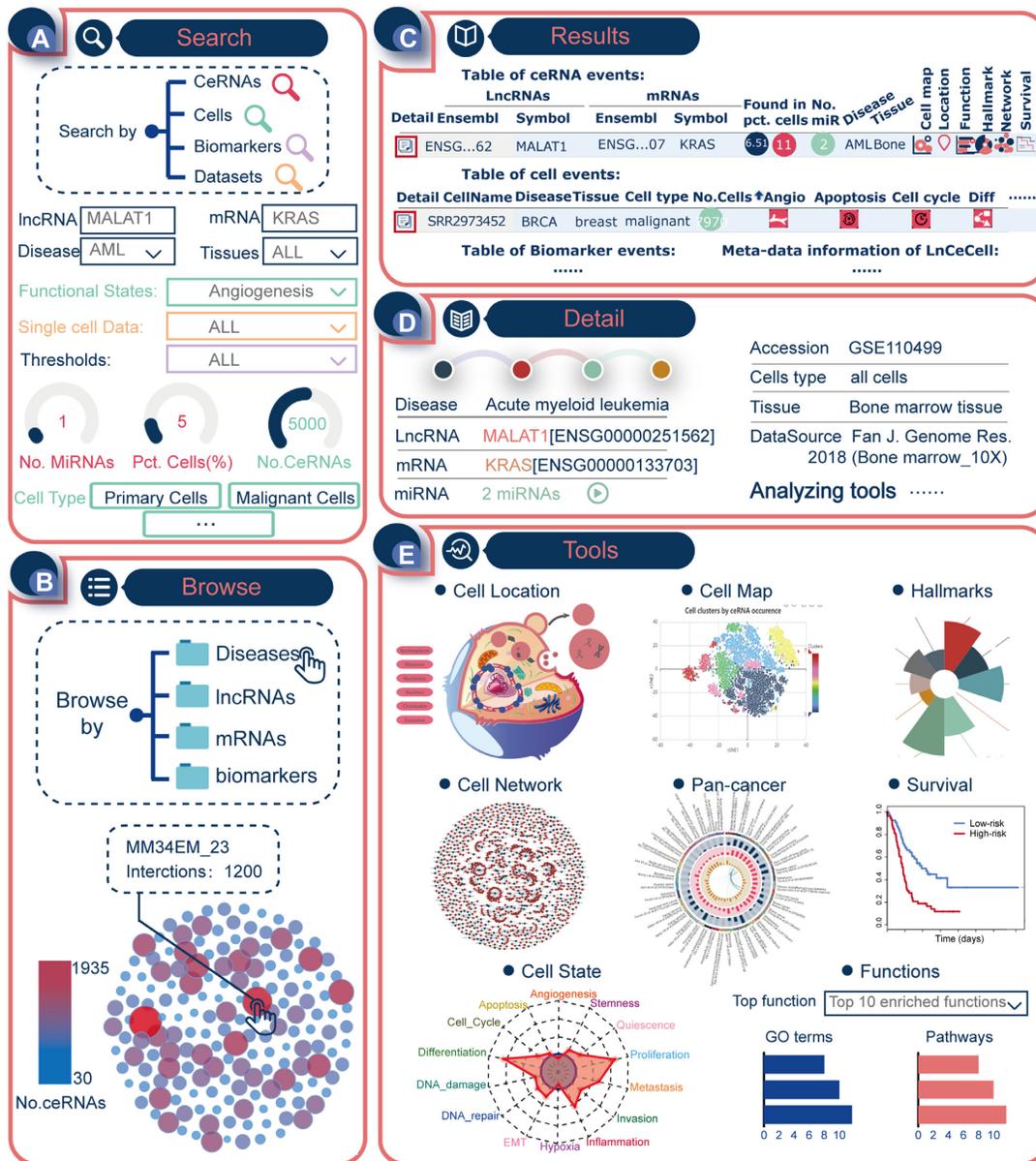


Figure 3. Example of the workflow and application of LnCeCell. (A) The search interface of LnCeCell with *MALAT1-KRAS* as the ceRNA input. (B) The browsing interface for ceRNAs and cells, in LnCeCell. (C) Results, provided by LnCeCell, for ceRNAs, cells, biomarkers and datasets information. (D) The basic information of ceRNA *MALAT1-KRAS*. (E) A panel of tools for analysing the functions of *MALAT1-KRAS* in different cellular backgrounds.

based on molecular pathways and biological processes. The CeRNA-Survival section revealed that ceRNA expression of *MALAT1-KRAS* was associated with survival of acute myeloid leukaemia patients, and the Kaplan-Meier survival curves of the two patient groups were significantly different (Figure 3E, log-rank $P < 0.05$).

We used another example of ceRNA *GAS5-PTEN*, which has been experimentally identified in different cancers (30). It has been shown that downregulation of *GAS5-PTEN* promotes glioma invasion (31). In our study, we found that *GAS5-PTEN* regulation could be identified in 3.9% high-grade glioma cells, wherein *GAS5* and *PTEN* expression was significantly lower than that in other cells (Figure 4A, B). Cells with *GAS5-PTEN* regulation had significantly

higher invasion scores than other cells (Figure 4C). Another study indicated that upregulation of *GAS5-PTEN* could suppress the metastasis of melanoma cells (32). We analysed the melanoma dataset and found *GAS5-PTEN* regulations in 11.9% of tumour cells, wherein *GAS5* and *PTEN* had significantly higher expression levels and lower metastasis scores than that of other cells (Figure 4D–F). The downregulation of lncRNA *GAS5* promotes the invasion and migration of colorectal cancer cells (33). In colorectal adenocarcinoma, we found *GAS5-PTEN* regulation in 2.5% of tumour cells, wherein *GAS5* had significantly lower expression than in other cells (Figure 4G, H). These cells also had significantly higher invasion and metastasis scores, which are consistent with a previous study (Figure 4I, J) (33). These re-

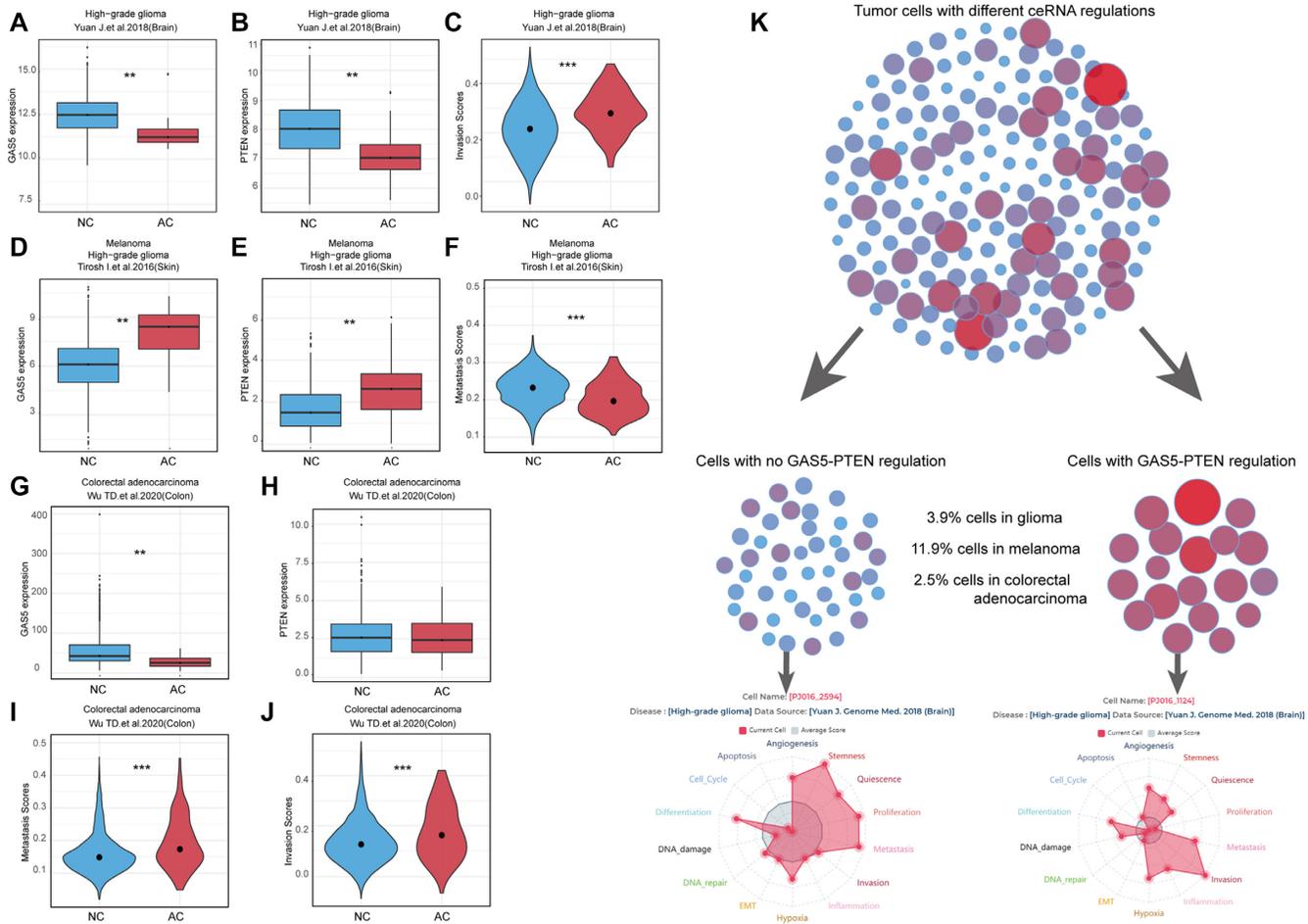


Figure 4. An example of *GASS5-PTEN* expression and regulation in different cancers. (A–C) Cells with a downregulation of *GASS5-PTEN* in glioma are associated with high invasion state. (D–F) Cells with an upregulation of *GASS5-PTEN* have lower metastasis scores in melanoma dataset. (G–J) Cells with a downregulation of *GASS5-PTEN* in colorectal adenocarcinoma are associated with high invasion and metastasis states. (K) Illustration of *GASS5-PTEN* distribution in the tumour microenvironment. Glioma cells with a downregulation of *GASS5-PTEN* have higher invasion state than other cells.

sults reveal that *GASS5-PTEN* distribution within different tumour microenvironments lead to distinct cell states and can be used to study different cellular behaviours driven by the fine-tuning of ceRNA regulation (Figure 4K).

CONCLUSIONS AND FUTURE DEVELOPMENT

The emergence of scRNA-seq technologies have provided researchers with powerful means of exploring intratumoral heterogeneity (34–36), which is a major obstacle to cancer treatment and a significant confounding factor in bulk-tumour profiling (37). A cell-specific gene regulatory network is one of the important factors that lead to different cell behaviours. Understanding and exploring gene-gene associations at the single-cell level will help reveal the mystery of tumours. Thus, LnCeCell, a database of cellular-specific lncRNA-associated ceRNA networks in cancer cells was developed. LnCeCell displays single cell status of cancer cells at multiple levels, and helps to understand the molecular mechanisms of the different behaviours of cancer cells. We believe that LnCeCell will be a useful resource for cancer research.

We identified the cellular-specific lncRNA-associated ceRNA networks and the functional states in >94 000 cells across 25 cancer types, and obtained detailed information on the sub-cellular locations of ceRNAs, from literature and related data sources, using the LnCeCell. The cell clusters based on expression values and regulatory relationships of the ceRNA of interest were also provided. LnCeCell houses over 9000 lncRNA biomarkers of tumour metastasis, recurrence, prognosis, circulation, drug resistance, etc. In addition, LnCeCell also provides several flexible tools for ceRNA functional analyses: CeRNA-Function and CeRNA-Hallmark identify ceRNA-related functional dysregulation and cancer hallmarks, respectively. CeRNA-Survival performs Cox regression analysis and draws survival curves for ceRNAs.

LnCeCell is a comprehensive database that helps to investigate fine-tuning regulatory networks within a single cell, and understand the regulatory mechanisms behind complex microbial ecosystems. We continue to improve the database in the following aspects: (i) continuous collection of newly published datasets of different cancer types, (ii) identification of cell types from cell clusters, (iii) refining of func-

tional states and (iv) collection of newly identified lncRNA biomarkers. We believe that LnCeCell, with continuous improvements, can become an effective tool to analyse the heterogeneity of cancer cells, and contribute to cancer diagnosis and treatment.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Key R&D Program of China [2018YFC2000100]; National Natural Science Foundation of China [32070622, 32070672, 61873075, 81902646]; Heilongjiang Touyan Innovation Team Program; Heilongjiang Provincial Natural Science Foundation [LH2020C057]; Postdoctoral Science Foundation of China [2020M670922]; Postdoctoral Foundation of Heilongjiang Province [LBH-Z19077]. Funding for open access charge: National Key R&D Program of China [2018YFC2000100]; National Natural Science Foundation of China [32070622, 32070672, 61873075, 81902646]; Heilongjiang Touyan Innovation Team Program; Heilongjiang Provincial Natural Science Foundation [LH2020C057]; Postdoctoral Science Foundation of China [2020M670922]; Postdoctoral Foundation of Heilongjiang Province [LBH-Z19077].

Conflict of interest statement. None declared.

REFERENCES

- Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L. *et al.* (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*, **344**, 1396–1401.
- Li, H., Courtois, E.T., Sengupta, D., Tan, Y., Chen, K.H., Goh, J.J.L., Kong, S.L., Chua, C., Hon, L.K., Tan, W.S. *et al.* (2017) Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat. Genet.*, **49**, 708–718.
- Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A. and Kirschner, M.W. (2015) Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, **161**, 1187–1201.
- Pan, X., Durrett, R.E., Zhu, H., Tanaka, Y., Li, Y., Zi, X., Marjani, S.L., Euskirchen, G., Ma, C., Lamotte, R.H. *et al.* (2013) Two methods for full-length RNA sequencing for low quantities of cells and single cells. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 594–599.
- Hedlund, E. and Deng, Q. (2018) Single-cell RNA sequencing: technical advancements and biological applications. *Mol. Aspects Med.*, **59**, 36–46.
- Yuan, H., Yan, M., Zhang, G., Liu, W., Deng, C., Liao, G., Xu, L., Luo, T., Yan, H., Long, Z. *et al.* (2019) CancerSEA: a cancer single-cell state atlas. *Nucleic Acids Res.*, **47**, D900–D908.
- Abugessaisa, I., Noguchi, S., Böttcher, M., Hasegawa, A., Kouno, T., Kato, S., Tada, Y., Ura, H., Abe, K., Shin, J.W. *et al.* (2018) SCPortalen: human and mouse single-cell centric database. *Nucleic Acids Res.*, **46**, D781–D787.
- Militello, G., Weirick, T., John, D., Döring, C., Dimmeler, S. and Uchida, S. (2017) Screening and validation of lncRNAs and circRNAs as miRNA sponges. *Brief. Bioinform.*, **18**, 780–788.
- Long, J., Xiong, J., Bai, Y., Mao, J., Lin, J., Xu, W., Zhang, H., Chen, S. and Zhao, H. (2019) Construction and Investigation of a lncRNA-Associated ceRNA Regulatory Network in Cholangiocarcinoma. *Front. Oncol.*, **9**, 649.
- Wang, W.T., Ye, H., Wei, P.P., Han, B.W., He, B., Chen, Z.H. and Chen, Y.Q. (2016) LncRNAs H19 and HULC, activated by oxidative stress, promote cell migration and invasion in cholangiocarcinoma through a ceRNA manner. *J. Hematol. Oncol.*, **9**, 117.
- Zheng, P., Yin, Z., Wu, Y., Xu, Y., Luo, Y. and Zhang, T.C. (2018) LncRNA HOTAIR promotes cell migration and invasion by regulating MKL1 via inhibition miR206 expression in HeLa cells. *Cell Commun. Signal.: CCS*, **16**, 5.
- Wang, P., Li, X., Gao, Y., Guo, Q., Wang, Y., Fang, Y., Ma, X., Zhi, H., Zhou, D., Shen, W. *et al.* (2019) LncACTdb 2.0: an updated database of experimentally supported ceRNA interactions curated from low- and high-throughput experiments. *Nucleic Acids Res.*, **47**, D121–D127.
- Li, J.H., Liu, S., Zhou, H., Qu, L.H. and Yang, J.H. (2014) starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.*, **42**, D92–D97.
- Wang, P., Li, X., Gao, Y., Guo, Q., Ning, S., Zhang, Y., Shang, S., Wang, J., Wang, Y., Zhi, H. *et al.* (2020) LnCeVar: a comprehensive database of genomic variations that disturb ceRNA network regulation. *Nucleic Acids Res.*, **48**, D111–D117.
- Dai, H., Li, L., Zeng, T. and Chen, L. (2019) Cell-specific network constructed by single-cell RNA sequencing data. *Nucleic Acids Res.*, **47**, e62.
- Hänzelmann, S., Castelo, R. and Guinney, J. (2013) GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*, **14**, 7.
- Li, S., Li, Y., Chen, B., Zhao, J., Yu, S., Tang, Y., Zheng, Q., Li, Y., Wang, P., He, X. *et al.* (2018) exoRBase: a database of circRNA, lncRNA and mRNA in human blood exosomes. *Nucleic Acids Res.*, **46**, D106–D112.
- Liu, T., Zhang, Q., Zhang, J., Li, C., Miao, Y.R., Lei, Q., Li, Q. and Guo, A.Y. (2019) EVmiRNA: a database of miRNA profiling in extracellular vesicles. *Nucleic Acids Res.*, **47**, D89–D93.
- Mas-Ponte, D., Carlevaro-Fita, J., Palumbo, E., Hermoso Pulido, T., Guigo, R. and Johnson, R. (2017) LncATLAS database for subcellular localization of long noncoding RNAs. *RNA*, **23**, 1080–1087.
- Quek, X.C., Thomson, D.W., Maag, J.L., Bartonicek, N., Signal, B., Clark, M.B., Gloss, B.S. and Dinger, M.E. (2015) lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res.*, **43**, D168–D173.
- Xia, S., Feng, J., Chen, K., Ma, Y., Gong, J., Cai, F., Jin, Y., Gao, Y., Xia, L., Chang, H. *et al.* (2018) CSCD: a database for cancer-specific circular RNAs. *Nucleic Acids Res.*, **46**, D925–D929.
- The Gene Ontology, C. (2019) The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.*, **47**, D330–D338.
- Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J.P. and Tamayo, P. (2015) The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.*, **1**, 417–425.
- Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**, 646–674.
- Wang, P., Ning, S., Zhang, Y., Li, R., Ye, J., Zhao, Z., Zhi, H., Wang, T., Guo, Z. and Li, X. (2015) Identification of lncRNA-associated competing triplets reveals global patterns and prognostic markers for cancer. *Nucleic Acids Res.*, **43**, 3478–3489.
- Guo, Q., Wang, J., Gao, Y., Li, X., Hao, Y., Ning, S. and Wang, P. (2020) Dynamic TF-lncRNA regulatory networks revealed prognostic signatures in the development of ovarian cancer. *Front. Bioeng. Biotechnol.*, **8**, 460.
- Chang, J., Xu, W., Du, X. and Hou, J. (2018) MALAT1 silencing suppresses prostate cancer progression by upregulating miR-1 and downregulating KRAS. *Oncotargets Ther.*, **11**, 3461–3473.
- Liu, P., Yang, H., Zhang, J., Peng, X., Lu, Z., Tong, W. and Chen, J. (2017) The lncRNA MALAT1 acts as a competing endogenous RNA to regulate KRAS expression by sponging miR-217 in pancreatic ductal adenocarcinoma. *Sci. Rep.*, **7**, 5186.
- Liu, R., Li, J., Lai, Y., Liao, Y., Liu, R. and Qiu, W. (2015) Hsa-miR-1 suppresses breast cancer development by down-regulating K-ras and long non-coding RNA MALAT1. *Int. J. Biol. Macromol.*, **81**, 491–497.
- Dong, P., Xiong, Y., Yue, J., S.J.B.H., Kobayashi, N., Todo, Y. and Watari, H. (2019) Exploring lncRNA-Mediated regulatory networks in endometrial cancer cells and the tumor microenvironment: Advances and challenges. *Cancers (Basel)*, **11**, 234.
- Ding, Y., Wang, J., Zhang, H. and Li, H. (2020) Long noncoding RNA-GAS5 attenuates progression of glioma by eliminating microRNA-10b and Sirtuin 1 in U251 and A172. *cells*, **46**, 487–496.

32. Bian,D., Shi,W., Shao,Y., Li,P. and Song,G. (2017) Long non-coding RNA GAS5 inhibits tumorigenesis via miR-137 in melanoma. *Am. J. Transl. Res.*, **9**, 1509–1520.
33. Liu,L., Wang,H.J., Meng,T., Lei,C., Yang,X.H., Wang,Q.S., Jin,B. and Zhu,J.F. (2019) lncRNA GAS5 inhibits cell migration and invasion and promotes autophagy by targeting miR-222-3p via the GAS5/PTEN-Signaling pathway in CRC. *Mol. Ther. Nucleic Acids*, **17**, 644–656.
34. Chen,G., Ning,B. and Shi,T. (2019) Single-Cell RNA-Seq technologies and related computational data analysis. *Front. Genet.*, **10**, 317.
35. Hwang,B., Lee,J.H. and Bang,D. (2018) Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.*, **50**, 96.
36. Papalexis,E. and Satija,R. (2018) Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat. Rev. Immunol.*, **18**, 35–45.
37. Roulot,A., Héquet,D., Guinebretière,J.M., Vincent-Salomon,A., Lerebours,F., Dubot,C. and Rouzier,R. (2016) Tumoral heterogeneity of breast cancer. *Ann. Biol. Clin. (Paris)*, **74**, 653–660.