D950–D955 *Nucleic Acids Research, 2018, Vol. 46, Database issue* *Published online 27 October 2017*
doi: 10.1093/nar/gkx1008

HCMDB: the human cancer metastasis database

Guantao Zheng[1,2], Yijie Ma[1], Yang Zou[1], An Yin[2], Wushuang Li[1] and Dong Dong[1,*]

[1]Shanghai Key Laboratory of Regulatory Biology, Institute of Biomedical Sciences, School of Life Sciences, East China Normal University, Shanghai 200241, China and [2]Shanghai Majorbio Bio-pharm Biotechnology Co., Ltd, Shanghai, China

Received August 10, 2017; Revised October 12, 2017; Editorial Decision October 12, 2017; Accepted October 13, 2017

ABSTRACT

Metastasis is the main event leading to death in cancer patients. Over the past decade, high-throughput technologies have provided genome-wide view of transcriptomic changes associated with cancer metastases. Many microarray and RNA sequencing studies have addressed metastases-related expression patterns in various types of cancer, and the number of relevant works continues to increase rapidly. These works have characterized genes that orchestrate the metastatic phenotype of cancer cells. However, these expression data have been deposited in various repositories, and efficiently analyzing these data is still difficult because of the lack of an integrated data mining platform. To facilitate the in-depth analyses of transcriptome data on metastasis, it is quite important to make a comprehensive integration of these metastases-related expression data. Here, we presented a database, HCMDB (the human cancer metastasis database, http://hcmdb.i-sanger.com/index), which is freely accessible to the research community query cross-platform transcriptome data on metastases. HCMDB is developed and maintained as a useful resource for building the systems-biology understanding of metastasis.

INTRODUCTION

Metastasis is the spread of a cancer from one organ to another without being directly connected with it, and it is the principal cause of cancer-related death (1–4). It is of great importance to evaluate the presence of metastasis in cancer patients, when choosing appropriate treatment strategies. A wide variety of approaches have been employed to identify and characterize genes involved in cancer metastasis at the transcriptional level (5–9). It is necessary to characterize the complicated molecular mechanism by integrating different data sources.

Recently, transcriptome data generated by microarray and RNA sequencing technologies have been widely used to explore the molecular nature of metastasis (9–14). Researches on the expression profile of metastases of multiple tumor types have produced a large amount of data at the transcription level. Members of new classes of non-coding RNAs, such as long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) have important roles in tumorigenesis and metastases (5,15–18). High-throughput technologies enable the exploration of transcriptomic changes associated with tumor progression and metastasis at the whole-genome level, which greatly assists our efforts to uncover the underlying molecular mechanism of cancer metastasis (6,19). Furthermore, metastasis contributes to most of the cancer-related death, which emphasizes the importance of metastasis risk prediction. Some works identified gene expression signatures that distinguished primary site from metastatic tumors based on cancer metastasis expression profiles (9,10,20), which would be developed as useful biomarkers for early detection, diagnosis and treatment of cancers.

Most of these high-throughput data have been deposited into NCBI Gene Expression Omnibus (GEO) (21) and the Sequence Read Archive (SRA) (22). The Cancer Genome Atlas (TCGA) dataset also stored expression data of metastasis of various tumor types and matched/unmatched primary tumors. These databases mainly serve as raw data archives, and cannot provide the full utility of transcriptome data for users. It still requires highly developed bioinformatics skills to manipulate metastasis-related data pipelines. These valuable resources provided an opportunity to explore metastasis-associated genes from large amount of samples. Recently, many cancer-related databases have been developed to store cancer-related gene expression and functional information. For example, cBioPortal (23) provided multidimensional cancer genomics data sets, however, only few metastasis-related data were involved. Cancer RNA-seq Nexus (24) only integrated recently published cancer RNA-seq data, and the metastasis-related classes of differential expression analyses were not well defined. dbMEMC (25) is a database of differentially expressed miRNAs in human cancers, and contains few metastasis-related miRNAs. CMGene (26) is a literature-based database for

*To whom correspondence should be addressed. Tel: +86 21 6223 3755; Fax: +86 21 5434 4922; Email: ddong.ecnu@gmail.com

© The Author(s) 2017. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

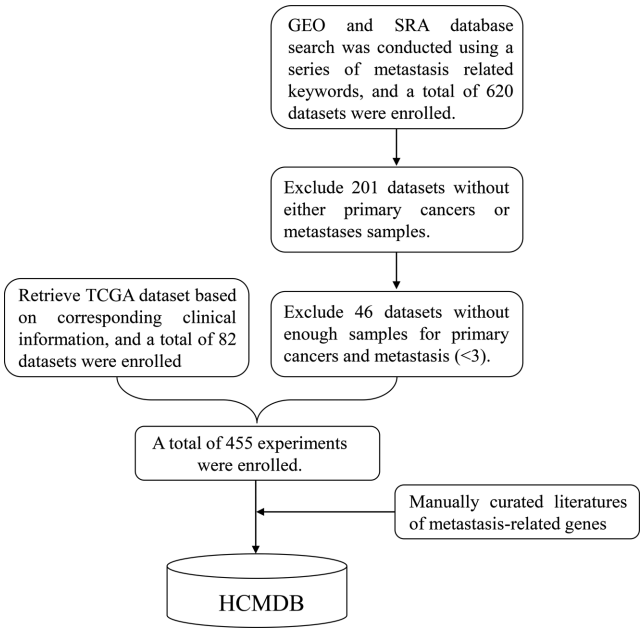

Figure 1. The framework for constructing HCMDB. The transcriptome data were derived from GEO, SRA, TCGA database. A series of filters were used to ensure the data quality of HCMDB. Metastasis-related literatures were manually to annotate the expression dysregulated genes.

cancer metastasis genes, and is not specially for differentially expressed genes of cancer metastasis. These databases provided some information on cancer metastasis, however, our knowledge of metastasis-related gene at the transcriptome level remains limited.

To this end, we presented the human cancer metastasis database (HCMDB), the first public database providing published cancer metastasis expression profiles. It also permits the analysis and visualization of mRNA/lncRNA co-expression networks and miRNA regulatory networks. HCMDB is freely accessible to cancer metastasis research community to query and analyze metastasis-related expression data. We expected this database could facilitate the identification of cancer metastasis associated genes and benefit the examination of their roles in physiological and pathological processes of cancer metastasis. All the data in HCMDB is freely available to the public domain through http://hcmdb.i-sanger.com/download.

DATA COLLECTION AND DATABASE CONTENT

Figure 1 depicts the architecture of the HCMDB database. Systematic data searching was conducted for metastasis-related expression profiles using the following keywords: 'cancer', 'tumor', 'metastasis' and 'epithelial-mesenchymal transition', in combination with 'long non-coding RNA', 'lncRNA' and 'microRNA' from GEO (https://www.ncbi.nlm.nih.gov/geo/) and SRA database (https://www.ncbi.nlm.nih.gov/sra/). The eligible data are limited to human studies published before June 2017. A total of 620 data sets were initially collected. Then, we manually curated these datasets to ensure that the data contains both primary tumors and metastases. Only the datasets have enough samples (at least three) for both primary tumors and metastases were left. Moreover, the TCGA clinical information was retrieved, and corresponding metastasis-related expression data were also enrolled (82 datasets). At last, the gene expression profiles of 29 primary tumor types from 455 experiments remained, including a total of 11 425 samples (Table 1).

The final dataset contains 351 mRNA expression profiles and 160 miRNA expression profiles. The TCGA data were downloaded from https://portal.gdc.cancer.gov/. For SRA data, an in-house bioinformatics pipeline was employed. Briefly, we evaluated RNA-seq quality using FastQC (version 10.10.1), and sequencing reads were aligned to the human genome (hg19) using Bowtie (27) software (version 1.1.1) with default parameters. Then, we re-assembled a transcriptome using Cufflinks (28) (version 2.2.1). The FPKM (fragments per kilobase of transcript per million mapped reads) values were calculated.

To date, lncRNA expression profiles of cancer metastases remain largely unknown. The lncRNA expression of metastases could be detected by mining gene expression microarray data because many lncRNA probes can be identified on the commonly used microarray platforms. So, we used the data mining method to detect lncRNA profiling on the microarray platform. A total of 335 lncRNA expression profiles were involved. For each experiment, the users can compare the expression profiles between different groups. The limma package (29) nested in R (http://www.r-project.org/) was employed to detect differentially expressed genes between different types of samples. Those genes with F.D.R. <0.05 were selected as candidates having significantly different expressions. To explore the biological implications of lncRNA, mRNA-lncRNA co-expressions were calculated using Pearson correlation coefficient. To identify miRNA target genes, PITA (30), miRanda (31) and RNAhybrid (32) software were employed. Those genes that identified by at least two software were regarded as miRNA target genes. The average number of target genes per miRNA is 244.

To further annotate metastasis-related genes, we searched Pubmed database (https://www.ncbi.nlm.nih.gov/pubmed/) using the keywords of 'metastasis' and corresponding gene symbols. All selected literatures were manually curated by at least two researchers. More than 7000 published papers were systematically reviewed, and 2183 metastasis-related genes were manually curated.

DATABASE CONSTRUCTION

We developed a user-friendly web interface to present HCMDB. A MySQL relational database was set up to store the data. The web interface for browsing and searching was implemented by PHP and JavaScript. The data processing was implemented by our in-house R scripts. The web service is based on an Apache Tomcat web server.

DATABASE FEATURES AND UTILITY

Experimental description

For each experiment presented, we carefully reviewed the experiment design of original paper and all the samples profiled. We retrieved the involved samples and classified them

Table 1. Experiment and sample size in the current version of HCMDB

Primary cancers	No. of experiments	No. of samples	Data sources
Bladder cancer	17	423	TCGA
Brain cancer	1	22	GEO
Breast cancer	93	3054	GEO, TCGA
Cervical cancer	11	968	GEO, TCGA
Colorectal cancer	75	2440	GEO, TCGA, SRA
Esophagus cancer	12	171	TCGA
Ewing's sarcoma	1	37	GEO
Eye cancer	3	121	GEO, TCGA
Gastric cancer	3	404	TCGA
Head and neck cancer	2	30	GEO
Kindey cancer	25	353	GEO, TCGA
Laryngeal cancer	2	15	GEO
Liver cancer	30	273	GEO, SRA
Lung cancer	4	46	GEO
Midgut carcinoid tumor	4	39	GEO
Nasopharynx cancer	2	22	GEO
Oral cancer	2	27	GEO
Osteosarcoma	3	30	GEO
Ovarian cancer	1	18	GEO
Pancreatic cancer	27	293	GEO, TCGA
Pancreatic neuroendocrine tumor	18	94	GEO
Penis cancer	3	33	GEO
Prostate cancer	49	863	GEO, TCGA, SRA
Skin cancer	38	598	GEO
Small intestine cancer	11	87	GEO
Synovial sarcoma	1	34	GEO
Testicular cancer	2	142	GEO, TCGA
Thymoma	2	121	TCGA
Thyroid cancer	15	667	GEO, TCGA
Total	455	11 425	

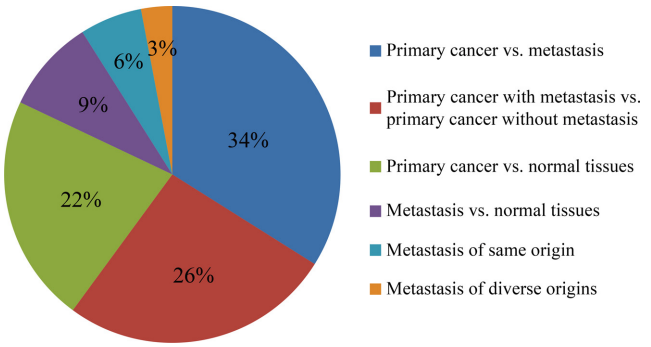

Figure 2. The percentage of categories of differential expression analyses.

as one of the following categories: primary cancer versus metastasis (33,34), primary cancer with metastasis *vs*. primary cancer without metastasis (35,36), primary cancer versus respective normal tissues (37,38), metastasis *vs*. respective normal tissues (39,40), different metastatic tumors of same origin (41), metastatic tumors of diverse origins (42). Most of the groups were defined according to the original experiment design. Primary cancer versus metastasis accounts for the most of the total experiments (Figure 2, 34%). After the samples were assigned to different groups, differential gene expression was assessed with limma package. For each experiment, the detailed information was provided to delineate the experimental design, cancer type and the total number of genes identified.

Database query and search platform

A user-friendly web interface was developed to present the HCMDB (Figure 3). We herein provided several ways to allow database query. First, a search engine was provided in HCMDB using gene names from the 'Search' page. Users can input their interested gene names in the textbox (including mRNAs, lncRNAs and miRNAs), and all the items that contain the query genes in the database can be derived. The searching result page lists the related experiment ID, primary cancer type, metastasis, experiment design, number of samples. A box plot was provided for the comparison of gene expression between different groups. Moreover, manually curated literatures related to the focal genes were also shown in the result page. Second, users can select particular primary cancer type and browse related experiments from the 'Browse' page. We also provided other optional filters to help users focus on the most interested results, such as sample size (<10, 10–50, 50–100, >100) and gene type (mRNA, lncRNA, miRNA).

Detailed experiment page

By clicking the hyperlink of a particular experiment ID, the detailed result of a specific experiment can be provided. This page mainly consists of four different sections: experiment description, expression profile, functional categories and regulatory network. In the 'experiment description' section, a detailed experiment description was listed. In addition, the number of mRNA/lncRNA/miRNA was also provided. In the 'expression profile' section, a list of gene name, differential expression result derived from limma

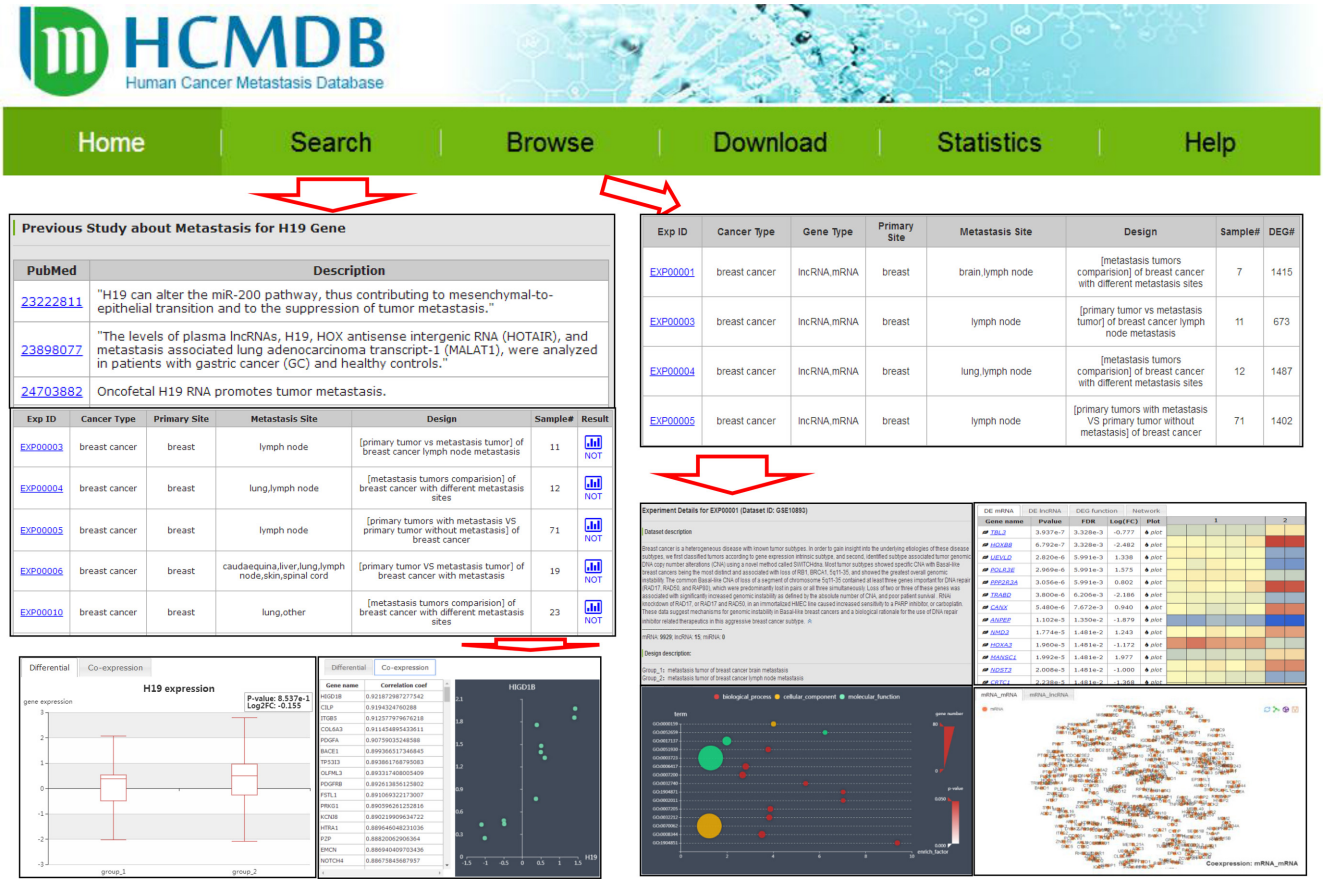

Figure 3. The schematic workflow of HCMDB.

software, such as *P*-value, F.D.R., log Fold Change were displayed. HCMDB creates a heatmap to facilitate users to examine differential expression profiles between different groups. In the 'functional categories' section, HCMDB provides the over-represented GO categories and KEGG pathways of differentially expressed genes. Fisher's exact test was used to identity over-represented GO and KEGG categories. In the 'regulatory network' section, HCMDB presents and visualizes the mRNA-mRNA and mRNA-lncRNA co-expression network, and force-directed and circus layout were employed. If the experiment contains both mRNA and miRNA expression profiles, the miRNA-targets regulatory network was also provided.

DISCUSSION AND CONCLUSION

Integrated analysis of multi-dimensional transcriptomic data is important to our understanding of cancer metastasis (43–49). The decreasing cost of large-scale technologies has led to tremendous amount of transcriptome data from metastasis studies. These data provided us great opportunities to perform gene expression quantification analysis in metastasis. Moreover, these data would help us identify gene expression signature associated with metastasis. Here, we provided the HCMDB to integrate these resources and facilitate the study of gene expression dysregulation in metastasis. HCMDB is freely accessible to the research community to query and analyze transcriptome data on metastasis.

Although some other related works provided some information on human cancer metastasis, such as cBioPortal (23), Cancer RNA-seq Nexus (24), dbMEMC (25) and CMGene (26), there are several advanced features that distinguish HCMDB from other data sources. First, it is a database specific to metastasis-related transcriptome data, and contains a greater number of gene-metastasis associations. Second, the lncRNA expression profiles of metastasis and primary tumors were comprehensively involved. Third, several classes of differential expression analyses were defined when examining the dysregulation pattern. These included primary cancer versus respective normal tissues, metastasis versus respective normal tissues, primary cancer versus metastasis, primary cancer with metastasis versus primary cancer without metastasis, different metastatic tumors of same origin, metastatic tumor of diverse origins. It provided a better way to comprehensively understand the gene expression pattern in metastasis, and the data have implications for our understanding how metastasis arise, and suggest ways in which expression signatures can be improved.

The cancer metastasis might be accomplished by the requirement of multiple genetic or genomics alterations. Recent papers have documented that somatic mutations and copy number alterations may play important roles in driving the development of cancer metastasis (50,51). For example, it has been documented that several recur-

rent copy number alterations might be the driver alterations in metastasis among superficial esophageal squamous cell carcinoma (52). A broader understanding of the genomic changes arising during metastasis will allow us to identify new metastasis-related mechanism and to discover new biomarkers for the diagnostic signatures of metastasis. HCMDB will be continuously updated and provide a unique resource in the following directions: (i) integrate more upcoming metastasis-related transcriptome data. (ii) Collect more comprehensive metastasis genetic and genomic data, including somatic mutation data, copy number alteration data. (iii) Add high-quality drug pharmacological data. Through our efforts, we expect that HCMDB will contribute to research into further understanding of the metastasis transcriptomic regulation mechanisms, and even toward the diagnosis and treatment of cancer metastasis.

FUNDING

National Natural Science Foundation of China to Dong Dong [31200956]. Funding for open access charge: National Natural Science Foundation of China to Dong Dong [31200956].
Conflict of interest statement. None declared.

REFERENCES

1. Fokas,E., Engenhart-Cabillic,R., Daniilidis,K., Rose,F. and An,H.X. (2007) Metastasis: the seed and soil theory gains identity. *Cancer Metast. Rev.*, **26**, 705–715.
2. Fingleton,B. (2007) Molecular targets in metastasis: lessons from genomic approaches. *Cancer Genomics Proteomics*, **4**, 211–221.
3. Hanahan,D. and Weinberg,R.A. (2000) The hallmarks of cancer. *Cell*, **100**, 57–70.
4. Poste,G. and Fidler,I.J. (1980) The pathogenesis of cancer metastasis. *Nature*, **283**, 139–146.
5. Crea,F., Clermont,P.L., Parolia,A., Wang,Y. and Helgason,C.D. (2014) The non-coding transcriptome as a dynamic regulator of cancer metastasis. *Cancer Metast. Rev.*, **33**, 1–16.
6. Crnic,I. and Christofori,G. (2004) Novel technologies and recent advances in metastasis research. *Int. J. Dev. Biol.*, **48**, 573–581.
7. Welch,D.R. (2004) Microarrays bring new insights into understanding of breast cancer metastasis to bone. *Breast Cancer Res.: BCR*, **6**, 61–64.
8. Casimiro,S., Luis,I., Fernandes,A., Pires,R., Pinto,A., Gouveia,A.G., Francisco,A.F., Portela,J., Correia,L. and Costa,L. (2012) Analysis of a bone metastasis gene expression signature in patients with bone metastasis from solid tumors. *Clin. Exp. Metast.*, **29**, 155–164.
9. Ramaswamy,S., Ross,K.N., Lander,E.S. and Golub,T.R. (2003) A molecular signature of metastasis in primary solid tumors. *Nat. Genet.*, **33**, 49–54.
10. Woelfle,U., Cloos,J., Sauter,G., Riethdorf,L., Janicke,F., van Diest,P., Brakenhoff,R. and Pantel,K. (2003) Molecular signature associated with bone marrow micrometastasis in human breast cancer. *Cancer Res.*, **63**, 5679–5684.
11. Mudduluru,G., Abba,M., Batliner,J., Patil,N., Scharp,M., Lunavat,T.R., Leupold,J.H., Oleksiuk,O., Juraeva,D., Thiele,W. *et al.* (2015) A systematic approach to defining the microRNA landscape in metastasis. *Cancer Res.*, **75**, 3010–3019.
12. Yang,Y., Chen,L., Gu,J., Zhang,H., Yuan,J., Lian,Q., Lv,G., Wang,S., Wu,Y., Yang,Y.T. *et al.* (2017) Recurrently deregulated lncRNAs in hepatocellular carcinoma. *Nat. Commun.*, **8**, 14421.
13. Kimbung,S., Johansson,I., Danielsson,A., Veerla,S., Egyhazi Brage,S., Frostvik Stolt,M., Skoog,L., Carlsson,L., Einbeigi,Z., Lidbrink,E. *et al.* (2016) Transcriptional profiling of breast cancer metastases identifies liver metastasis-selective genes associated with adverse outcome in luminal a primary breast cancer. *Clin. Cancer Res.*, **22**, 146–157.
14. Hur,K., Toiyama,Y., Schetter,A.J., Okugawa,Y., Harris,C.C., Boland,C.R. and Goel,A. (2015) Identification of a metastasis-specific MicroRNA signature in human colorectal cancer. *J. Natl. Cancer Inst.e*, **107**, dju492.
15. Lujambio,A. and Esteller,M. (2009) How epigenetics can explain human metastasis: a new role for microRNAs. *Cell Cycle*, **8**, 377–382.
16. Yuan,S., Wang,J., Yang,Y., Zhang,J., Liu,H., Xiao,J., Xu,Q., Huang,X., Xiang,B., Zhu,S. *et al.* (2017) The prediction of clinical outcome in hepatocellular carcinoma based on a six-gene metastasis signature. *Clin. Cancer Res.*, **23**, 289–297.
17. He,A., Hu,R., Chen,Z., Liao,X., Li,J., Wang,D., Lv,Z., Liu,Y., Wang,F. and Mei,H. (2017) Role of long noncoding RNA UCA1 as a common molecular marker for lymph node metastasis and prognosis in various cancers: a meta-analysis. *Oncotarget*, **8**, 1937–1943.
18. Jiang,C., Li,X., Zhao,H. and Liu,H. (2016) Long non-coding RNAs: potential new biomarkers for predicting tumor invasion and metastasis. *Mol. Cancer*, **15**, 62.
19. Liao,J.Y., Wu,J., Wang,Y.J., He,J.H., Deng,W.X., Hu,K., Zhang,Y.C., Zhang,Y., Yan,H., Wang,D.L. *et al.* (2017) Deep sequencing reveals a global reprogramming of lncRNA transcriptome during EMT. *Biochim. Biophys. Acta*, **1864**, 1703–1713.
20. Nadal,C., Maurel,J. and Gascon,P. (2007) Is there a genetic signature for liver metastasis in colorectal cancer? *World journal of gastroenterology*, **13**, 5832–5844.
21. Barrett,T., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M., Holko,M. *et al.* (2013) NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.*, **41**, D991–D995.
22. Wheeler,D.L., Barrett,T., Benson,D.A., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., Dicuccio,M., Edgar,R., Federhen,S. *et al.* (2008) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **36**, D13–D21.
23. Cerami,E., Gao,J., Dogrusoz,U., Gross,B.E., Sumer,S.O., Aksoy,B.A., Jacobsen,A., Byrne,C.J., Heuer,M.L., Larsson,E. *et al.* (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.*, **2**, 401–404.
24. Li,J.R., Sun,C.H., Li,W., Chao,R.F., Huang,C.C., Zhou,X.J. and Liu,C.C. (2016) Cancer RNA-Seq Nexus: a database of phenotype-specific transcriptome profiling in cancer cells. *Nucleic Acids Res.*, **44**, D944–D951.
25. Yang,Z., Ren,F., Liu,C., He,S., Sun,G., Gao,Q., Yao,L., Zhang,Y., Miao,R., Cao,Y. *et al.* (2010) dbDEMC: a database of differentially expressed miRNAs in human cancers. *BMC Genomics*, **11**(Suppl 4), S5.
26. Liu,Y., Li,Z., Lu,J., Zhao,M. and Qu,H. (2017) CMGene: A literature-based database and knowledge resource for cancer metastasis genes. *J. Genet. Genomics*, **44**, 277–279.
27. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
28. Trapnell,C., Williams,B.A., Pertea,G., Mortazavi,A., Kwan,G., van Baren,M.J., Salzberg,S.L., Wold,B.J. and Pachter,L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, **28**, 511–515.
29. Ritchie,M.E., Phipson,B., Wu,D., Hu,Y., Law,C.W., Shi,W. and Smyth,G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47.
30. Kertesz,M., Iovino,N., Unnerstall,U., Gaul,U. and Segal,E. (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.*, **39**, 1278–1284.
31. John,B., Enright,A.J., Aravin,A., Tuschl,T., Sander,C. and Marks,D.S. (2004) Human MicroRNA targets. *PLoS Biol.*, **2**, e363.
32. Kruger,J. and Rehmsmeier,M. (2006) RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res.*, **34**, W451–W454.
33. Weigman,V.J., Chao,H.H., Shabalin,A.A., He,X., Parker,J.S., Nordgard,S.H., Grushko,T., Huo,D., Nwachukwu,C., Nobel,A. *et al.* (2012) Basal-like Breast cancer DNA copy number losses identify genes involved in genomic instability, response to therapy, and patient survival. *Breast Cancer Res. Treatment*, **133**, 865–880.
34. Scotlandi,K., Remondini,D., Castellani,G., Manara,M.C., Nardi,F., Cantiani,L., Francesconi,M., Mercuri,M., Caccuri,A.M., Serra,M. *et al.* (2009) Overcoming resistance to conventional drugs in Ewing

sarcoma and identification of molecular predictors of outcome. *J. Clin. Oncol.*, **27**, 2209–2216.
35. Watanabe,T., Kobunai,T., Yamamoto,Y., Matsuda,K., Ishihara,S., Nozawa,K., Iinuma,H., Konishi,T., Horie,H., Ikeuchi,H. *et al.* (2011) Gene expression signature and response to the use of leucovorin, fluorouracil and oxaliplatin in colorectal cancer patients. *Clin. Transl. Oncol.*, **13**, 419–425.
36. Matsuyama,T., Ishikawa,T., Mogushi,K., Yoshida,T., Iida,S., Uetake,H., Mizushima,H., Tanaka,H. and Sugihara,K. (2010) MUC12 mRNA expression is an independent marker of prognosis in stage II and stage III colorectal cancer. *Int. J. Cancer*, **127**, 2292–2299.
37. Pizzini,S., Bisognin,A., Mandruzzato,S., Biasiolo,M., Facciolli,A., Perilli,L., Rossi,E., Esposito,G., Rugge,M., Pilati,P. *et al.* (2013) Impact of microRNAs on regulatory networks and pathways in human colorectal carcinogenesis and development of metastasis. *BMC Genomics*, **14**, 589.
38. Del Rio,M., Mollevi,C., Vezzio-Vie,N., Bibeau,F., Ychou,M. and Martineau,P. (2013) Specific extracellular matrix remodeling signature of colon hepatic metastases. *PLoS One*, **8**, e74599.
39. Barry,S., Chelala,C., Lines,K., Sunamura,M., Wang,A., Marelli-Berg,F.M., Brennan,C., Lemoine,N.R. and Crnogorac-Jurcevic,T. (2013) S100P is a metastasis-associated gene that facilitates transendothelial migration of pancreatic cancer cells. *Clin. Exp. Metast.*, **30**, 251–264.
40. Wragg,J.W., Finnity,J.P., Anderson,J.A., Ferguson,H.J., Porfiri,E., Bhatt,R.I., Murray,P.G., Heath,V.L. and Bicknell,R. (2016) MCAM and LAMA4 are highly enriched in tumor blood vessels of renal cell carcinoma and predict patient outcome. *Cancer Res.*, **76**, 2314–2326.
41. Kimbung,S., Kovacs,A., Bendahl,P.O., Malmstrom,P., Ferno,M., Hatschek,T. and Hedenfalk,I. (2014) Claudin-2 is an independent negative prognostic factor in breast cancer and specifically predicts early liver recurrences. *Mol. Oncol.*, **8**, 119–128.
42. Meyniel,J.P., Cottu,P.H., Decraene,C., Stern,M.H., Couturier,J., Lebigot,I., Nicolas,A., Weber,N., Fourchotte,V., Alran,S. *et al.* (2010) A genomic and transcriptomic approach for a differential diagnosis between primary and secondary ovarian carcinomas in patients with a previous history of breast cancer. *BMC Cancer*, **10**, 222.
43. Huang,E., Ishida,S., Pittman,J., Dressman,H., Bild,A., Kloos,M., D'Amico,M., Pestell,R.G., West,M. and Nevins,J.R. (2003) Gene expression phenotypic models that predict the activity of oncogenic pathways. *Nat. Genet.*, **34**, 226–230.
44. Pedraza-Farina,L.G. (2006) Mechanisms of oncogenic cooperation in cancer initiation and metastasis. *Yale J. Biol. Med.*, **79**, 95–103.
45. Seiler,R., Lam,L.L., Erho,N., Takhar,M., Mitra,A.P., Buerki,C., Davicioni,E., Skinner,E.C., Daneshmand,S. and Black,P.C. (2016) Prediction of lymph node metastasis in patients with bladder cancer using whole transcriptome gene expression signatures. *J. Urol.*, **196**, 1036–1041.
46. Sun,J., Chen,X., Wang,Z., Guo,M., Shi,H., Wang,X., Cheng,L. and Zhou,M. (2015) A potential prognostic long non-coding RNA signature to predict metastasis-free survival of breast cancer patients. *Scientific Rep.*, **5**, 16553.
47. Lee,J.Y., Park,K., Lim,S.H., Kim,H.S., Yoo,K.H., Jung,K.S., Song,H.N., Hong,M., Do,I.G., Ahn,T. *et al.* (2015) Mutational profiling of brain metastasis from breast cancer: matched pair analysis of targeted sequencing between brain metastasis and primary breast cancer. *Oncotarget*, **6**, 43731–43742.
48. Saiselet,M., Gacquer,D., Spinette,A., Craciun,L., Decaussin-Petrucci,M., Andry,G., Detours,V. and Maenhaut,C. (2015) New global analysis of the microRNA transcriptome of primary tumors and lymph node metastases of papillary thyroid cancer. *BMC Genomics*, **16**, 828.
49. Schell,M.J., Yang,M., Missiaglia,E., Delorenzi,M., Soneson,C., Yue,B., Nebozhyn,M.V., Loboda,A., Bloom,G. and Yeatman,T.J. (2016) A composite gene expression signature optimizes prediction of colorectal cancer metastasis and outcome. *Clin. Cancer Res.*, **22**, 734–745.
50. Xie,T., Cho,Y.B., Wang,K., Huang,D., Hong,H.K., Choi,Y.L., Ko,Y.H., Nam,D.H., Jin,J., Yang,H. *et al.* (2014) Patterns of somatic alterations between matched primary and metastatic colorectal tumors characterized by whole-genome sequencing. *Genomics*, **104**, 234–241.
51. Robinson,D.R., Wu,Y.M., Lonigro,R.J., Vats,P., Cobain,E., Everett,J., Cao,X., Rabban,E., Kumar-Sinha,C., Raymond,V. *et al.* (2017) Integrative clinical genomics of metastatic cancer. *Nature*, **548**, 297–303.
52. Wang,P., Shan,L., Xue,L., Zheng,B., Ying,J. and Lu,N. (2017) Genome wide copy number analyses of superficial esophageal squamous cell carcinoma with and without metastasis. *Oncotarget*, **8**, 5069–5080.