Li et al. BMC Cancer (2019) 19:67
https://doi.org/10.1186/s12885-019-5274-4

BMC Cancer

RESEARCH ARTICLE Open Access

Identifying primary site of lung-limited Cancer of unknown primary based on relative gene expression orderings

Mengyao Li[1†], Hongdong Li[2*†], Guini Hong[1], Zhongjie Tang[1], Guanghao Liu[1], Xiaofang Lin[1], Mingzhang Lin[1], Lishuang Qi[3] and Zheng Guo[1,4*]

Abstract

Background: Precise diagnosis of the tissue origin for metastatic cancer of unknown primary (CUP) is essential for deciding the treatment scheme to improve patients' prognoses, since the treatment for the metastases is the same as their primary counterparts. The purpose of this study is to identify a robust gene signature that can predict the origin for CUPs.

Methods: The within-sample relative gene expression orderings (REOs) of gene pairs within individual samples, which are insensitive to experimental batch effects and data normalizations, were exploited for identifying the prediction signature.

Results: Using gene expression profiles of the lung-limited metastatic colorectal cancer (LmCRC), we firstly showed that the within-sample REOs in lung metastases of colorectal cancer (CRC) samples were concordant with the REOs in primary CRC samples rather than with the REOs in primary lung cancer. Based on this phenomenon, we selected five gene pairs with consistent REOs in 498 primary CRC and reversely consistent REOs in 509 lung cancer samples, which were used as a signature for predicting primary sites of metastatic CRC based on the majority voting rule. Applying the signature to 654 primary CRC and 204 primary lung cancer samples collected from multiple datasets, the prediction accuracy reached 99.36%. This signature was also applied to 24 LmCRC samples collected from three datasets produced by different laboratories and the accuracy reached 100%, suggesting that the within-sample REOs in the primary site could reveal the original tissue of metastatic cancers.

Conclusions: The result demonstrated that the signature based on within-sample REOs of five gene pairs could exactly and robustly identify the primary sites of CUPs.

Keywords: Cancer of unknown primary, Relative gene expression orderings, Metastasis, Lung cancer, Colorectal cancer

Background

Despite the recent advances in pathology investigations and imaging technology, the primary site remains unknown for about 3% of all the malignancies [1–3]. By definition, the cancer of unknown primary (CUP) is metastatic at diagnosis with unknown primary site, which indicates a high malignant degree with poor prognosis [4]. In clinical, the therapeutic strategy for CUPs often needs the recognition of primary sites, as the current clinical guidelines recommend the same or similar treatment scheme for metastases as their primary counterparts [5, 6].

Some investigators have tried to use gene expression profiling to predict the primary tumor sites for CUPs [7, 8]. For example, Greco *et al.* made use of a 92-gene molecular tumor profiling (MTP) assay to predict the primary tissues of CUPs, which showed an accuracy of 75% [8]. However, risk scores of such signatures rely on the

* Correspondence: biomantis_lhd@163.com; guoz@hrbmu.edu.cn
†Mengyao Li and Hongdong Li contributed equally to this work.
[2]Department of Bioinformatics, Gannan Medical University, Ganzhou 341000, China
[1]Key Laboratory of Ministry of Education for Gastrointestinal Cancer, Department of Bioinformatics, Fujian Medical University, Fuzhou 350001, China
Full list of author information is available at the end of the article

absolute expression levels of genes, which could be affected by experimental batch effects [9]. Thus, such signatures often fail in independent samples [10–12]. It has been reported that the within-sample relative gene expression orderings (REOs) of gene pairs within individual samples are insensitive to experimental batch effects [10–12], invariant to monotonic data transformation [13, 14], and robust against partial RNA degradation [15] as well as sampling site uncertainty within a tumor tissue [16]. Based on these unique advantages, some classifiers based on REO signatures, such as TSP [10] and K-TSP [11], were proposed to identify transcriptional signatures for discriminating cancer subtypes [17–19], which obviate the need of data normalization for the discovery and validation datasets and thus can be applied to the individual level [17, 20, 21].

Colorectal cancer (CRC) is the third most frequent cancer worldwide, which accounts for approximately 10% of the global cancer burden [22, 23]. About 25% of the CRC patients present with metastases at diagnosis [24], of which liver and lung are the most frequent metastasis sites [25]. It has been reported that the lung metastases of CRC share high genomic concordance with the primary CRC [23]. Thus, we could hypothesize that the gene expression patterns of lung metastases of CRC would be more similar to the primary tumor on the susceptible primary organ than the metastatic organ. Here, using samples from the lung-limited metastatic colorectal cancer (LmCRC), we validated this hypothesis by comparing the stable REOs in LmCRC with the stable within-sample REOs in primary CRC and primary lung cancer, respectively. Then, we extracted a signature consisting of five gene pairs from primary CRC and lung cancer, and showed that this signature could predict all the LmCRC samples into the CRC-like group.

Methods

Data source and data preprocessing

The gene expression data analyzed in this study were downloaded from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) [26]. Detailed information for each dataset was described in Table 1. Set1~Set3 denoted the gene expression profiles for primary CRC from three datasets respectively. Totally, there were 498 primary CRC samples. Set4~Set6 denoted the gene expression profiles for primary lung cancer from three datasets respectively. Totally, there were 509 primary lung cancer samples. The profiles were used to select the primary CRC and lung cancer characteristic REOs. Set7 denoted the gene expression profiles for LmCRC samples.

The raw data (.CEL files) for each dataset were downloaded from GEO and normalized by the robust multi-array average method (RMA) in the Bioconductor package [27–29] except GSE14095. Because the raw data were not provided, the normalized data provided by the authors were downloaded for GSE14095. The original platform annotation files obtained from GEO for each dataset were used to annotate the CloneIDs to GeneIDs.

Table 1 The datasets analyzed in this study

Label	Dataset	Platform	Sample size	Ref (PMID)
Training sets				
Primary CRC				
Set1	GSE21510	GPL570	123	21270110
Set2	GSE14095	GPL570	189	21680303
Set3	GSE41258	GPL96	186	19359472
Primary lung cancer				
Set4	GSE31210	GPL570	226	22080568
Set5	GSE14814	GPL96	133	20823422
Set6	GSE43580	GPL570	150	23966112
Validation sets				
Primary CRC				
	GSE2138	GPL96	20	16247484
	GSE7208	GPL96	59	17638901
	GSE39582	GPL570	566	23700391
	GSE5364	GPL96	9	18636107
	GSE19249	GPL571	15	20522636
Primary lung cancer				
	GSE19804	GPL570	60	20802022
	GSE33532	GPL570	80	Michael Meister, *et al.*
	GSE18842	GPL570	46	20878980
	GSE5364	GPL96	18	18636107
	GSE19249	GPL571	7	20522636
Lung metastases of CRC				
Set7	GSE41258	GPL96	20	19359472
	GSE5851	GPL571	3	17664471
	GSE28702	GPL570	1	22095227

Detection of stable REO gene pairs

For a dataset, if gene A had a higher expression level than gene B in more than 95% samples, the gene pair (A, B) was defined as a stable REO gene pair. A concordance score was used to evaluate the reproducibility of stable REO gene pairs identified from two inde pendent datasets. If two lists of stable REO gene pairs overlapped k gene pairs, of which s gene pairs had the same REO patterns, the concordance score will be calculated as $s/k \times 100\%$. The cumulative binomial distribution model was used to evaluate the probability of observing a concordance score of $s/k \times 100\%$ by chance as follows:

$$P = 1 - \sum_{i=0}^{s-1} \binom{k}{i} (P_e)^i (1-P_e)^{k-1} \quad (1)$$

where P_e was the probability of a gene pair having the concordant relationship in the two datasets by chance (here, $P_e = 0.5$).

Selection of predictive gene pairs as a candidate signature

If a stable REO gene pair (*A, B*) in class1 showed the reverse REO pattern (*B, A*) in class2, it was defined as a reversed gene pair between these two classes. Theoretically, according to the REOs of the reversed gene pairs, we could classify the samples in these two classes. The procedure for predictive signature selection was as follows:

Firstly, calculated the appearance frequency of each gene in reversed gene pairs.

Secondly, calculated the average rank difference score ΔavgR for each reversed gene pair, as described in formula (2):

$$\Delta avgR_{ij} = \frac{\sum_{n=1}^{N1} \left| R_{n,i} - R_{n,j} \right| + \sum_{m=1}^{N2} \left| R_{m,i} - R_{m,j} \right|}{N1 + N2} \quad (2)$$

Here N1 and N2 represented the number of profiles in class1 and class2, respectively. $R_{n,i}$, $R_{n,j}$, $R_{m,i}$, and $R_{m,j}$ represented the rank of gene *i* or *j* in the *n*-th and *m*-th profile of class1 and class2 respectively.

Thirdly, sorted the genes according to the appearance frequencies in the reverse order and selected one reversed pair with the maximum ΔavgR score for each gene.

At last, the top *n* gene pairs were selected as the candidate predictive signature.

Anti-lung cancer drugs and protein-protein interaction data

The data of antitumor drugs and their target genes were collected from the DrugBank database (http://www.drugbank.ca/) [30], which contains 174 kinds of anticancer drugs approved by the U.S. Food and Drug Administration and 570 corresponding target genes. A total of 10 anti-lung cancer drugs and their 11 corresponding target genes were used in this study (Table 2).

The human protein-protein interaction (PPI) data were constructed as previously described [31]. The PPI data were downloaded from Human Protein Reference Database (HPRD) [32] in November 2016.

Table 2 Anti-lung cancer drugs and their target genes

Drug ID	Drug	FDA	Target genes
DB00317	Gefitinib	approved	EGFR
DB00361	Vinorelbine	approved	TUBB
DB00642	Pemetrexed	approved	TYMS, ATIC, DHFR, GART
DB05390	INS 316	investigational	P2RY2
DB08865	Crizotinib	approved	ALK, MET
DB08916	Afatinib	approved	EGFR, ERBB2, ERBB4
DB09063	Ceritinib	approved	ALK
DB09330	Osimertinib	approved	EGFR
DB09559	Necitumumab	approved	EGFR
DB11363	Alectinib	approved	ALK

Results

High REO concordance between lung metastases of CRC and primary CRC

To evaluate whether the REO patterns of lung metastases of CRC were similar to the primary CRC or lung cancer, a total of 498 primary CRC samples and 509 primary lung cancer samples were used (training set, Table 1).

First, gene pairs with stable REOs in more than 95% samples were identified and referred to as stable REO gene pairs. In the primary CRC Set1, Set2 and Set3, 156,893,727, 120,114,768 and 60,208,179 stable gene pairs were identified, respectively. Each two of the three lists of stable REO gene pairs showed significantly high concordances (Table 3), with a concordance score ranged from 91.3% ($P < 2.20 \times 10^{-16}$) to 99.0% ($P < 2.20 \times 10^{-16}$). There was a total of 35,220,621 stable REO gene pairs overlapped among these three datasets, which was denoted as primary CRC characteristic stable REO gene pairs. In primary lung cancer Set4, Set5 and Set6, 154,434,794, 53,985,252 and 140,533,599 stable REO gene pairs were identified, respectively (Table 3). A total of 31,739,263 stable REO gene pairs overlapped in all of these three datasets were denoted as primary lung cancer characteristic stable gene pairs.

Then, the characteristic stable REO gene pairs for primary CRC and lung cancer were compared to REO gene pairs identified for LmCRC samples. Between primary CRC and lung cancer, 6599 characteristic stable REO gene pairs showed the reverse REO patterns. These 6599 gene pairs (involving 4802 genes) were examined in each of the 20 LmCRC samples in Set7. In these 20 LmCRC samples, the REOs of the 6599 gene pairs were highly concordant with the primary CRC rather than with the primary lung cancer, which varied from the lowest concordance score of 88.36% to the highest concordance score of 99.89% (Fig. 1).

Collectively, these results indicated that, though with some characteristics of lung cancer, the lung metastases of CRC samples were more similar to the primary CRC.

Table 3 Concordance scores of stable REO gene pairs of primary CRC and primary lung cancer datasets

Dataset	Number of stable gene pairs	Number of overlaps	Concordance score	*p*-value
Primary CRC				
GSE21510	156,893,727	108,393,508	99.00%	$< 2.20 \times 10^{-16}$
GSE14095	120,114,768			
GSE21510	156,893,727	43,803,419	91.20%	$< 2.20 \times 10^{-16}$
GSE41258	60,208,179			
GSE14095	120,114,768	38,324,773	94.10%	$< 2.20 \times 10^{-16}$
GSE41258	60,208,179			
Primary lung cancer				
GSE31210	154,434,794	35,809,034	86.00%	$< 2.20 \times 10^{-16}$
GSE14814	53,985,252			
GSE43580	140,533,599	128,549,112	99.60%	$< 2.20 \times 10^{-16}$
GSE31210	154,434,794			
GSE14814	53,985,252	33,941,889	88.10%	$< 2.20 \times 10^{-16}$
GSE43580	140,533,599			

Therefore, the characteristic stable gene pairs of primary tumors could be applied to predict the primary tumor site of CUP.

A robust signature for discriminating lung metastases of CRC from lung cancer

The 6599 gene pairs with reversal REO patterns between the primary CRC and primary lung cancer samples were used in developing the signature for discriminating the primary CRC and lung cancer. The candidate predictive signature was selected based on the appearance frequencies of genes in the reversed gene pairs between primary CRC and lung cancer and the average rank difference score ΔavgR of each gene pair, as described in Methods. Then, sequentially took odd numbered gene pairs (i.e. 1, 3, 5, ... gene pairs) from the candidate gene pair list to classify primary CRC and lung cancer samples by the majority voting rules: if more than half of the REOs of the gene pairs in a sample were consistent with the candidate signature gene pairs, the sample would be predicted into CRC-like group, otherwise, the sample would be predicted into the lung cancer-like group. The classification accuracy was 99.36% when five gene pairs were taken, and kept unchanged at 99.36% from five gene pairs to 43 gene pairs. Interestingly, the gene SLC34A2 was included in all the 5 gene pairs (Table 4). This gene was reported to play an essential role in the tumorigenesis and progression of non-small cell lung cancer [33] and other pneumonosis such as pulmonary alveolar microlithiasis [34].

The prediction capacity of the signature was further tested in an independent dataset comprising 654 primary CRC and 204 primary lung cancer samples collected from seven datasets (Table 1). The result showed that

Fig. 1 REO patterns of the 6599 gene pairs for each lung metastasis of CRC samples. The green bar stood for the proportion of REO pattern the same as that in primary CRC. The yellow bar stood for the proportion of REO pattern the same as that in primary lung cancer

Table 4 Top five genes with highest appearance frequencies and their gene pairs with maximum ΔavgR

Gene Symbol	Appearance Frequency	Gene Pair Symbol[a]	ΔavgR
GUCY2C	1969	GUCY2C, SLC34A2	10,298.34705
CDH17	1322	CDH17, SLC34A2	10,273.16384
FABP1	485	FABP1, SLC34A2	10,001.53088
SLC34A2	474	KRT20, SLC34A2	10,657.31487
USH1C	270	USH1C, SLC34A2	9020.681651

[a]The former genes had higher expression levels than the latter genes in the CRC

99.54% of the 654 primary CRC samples were correctly predicted into the CRC group and 99.51% of the 204 primary lung cancer samples were correctly predicted, rea ching an average prediction accuracy of 99.53%. This result suggested that the signature had a robust discriminating capability to distinguish the primary CRC and lung cancer samples.

Using the signature to predict the 20 LmCRC samples collected from 3 datasets (Table 1), the result showed that all these samples were correctly classified to the CRC group. The accurate prediction indicated that the REO patterns in primary CRC and primary lung cancer could be applied to identify the tissue of origin for pulmonary tumor. There were another three lung metastases of CRC samples in GSE5851 and one in GSE28702, which were also predicted to the CRC group, with the prediction accuracy of 100%.

Lung cancer characteristics of lung metastases of CRC

A total of 2034 differentially expressed genes (DEGs) were distinguished between the primary CRC and lung metastases of CRC in GSE41258 by the Student's *t*-test with false discovery rate (FDR) less than 1%. In the PPI network, 90.91% (10) of the 11 anti-lung cancer drugs target genes had direct PPI links with at least 119 DEGs (Fig. 2). Especially, three anti-lung cancer drugs target genes (DHFR, GART and ALK) also presented in the DEGs. The DEGs centrally had direct interaction with EGFR, ERBB2, ERBB4, MET and TUBB, which suggested that the divergence between the primary CRC and the metastases of CRC was related to lung cancer. Furthermore, the direct interaction with anti-lung cancer drug target genes indicated these target genes could also be regarded as the CRC lung metastases treatment target genes, and their corresponding drugs, including Osimertinib, Necitumumab, Gefitinib, Afatinib, Osimertinib,

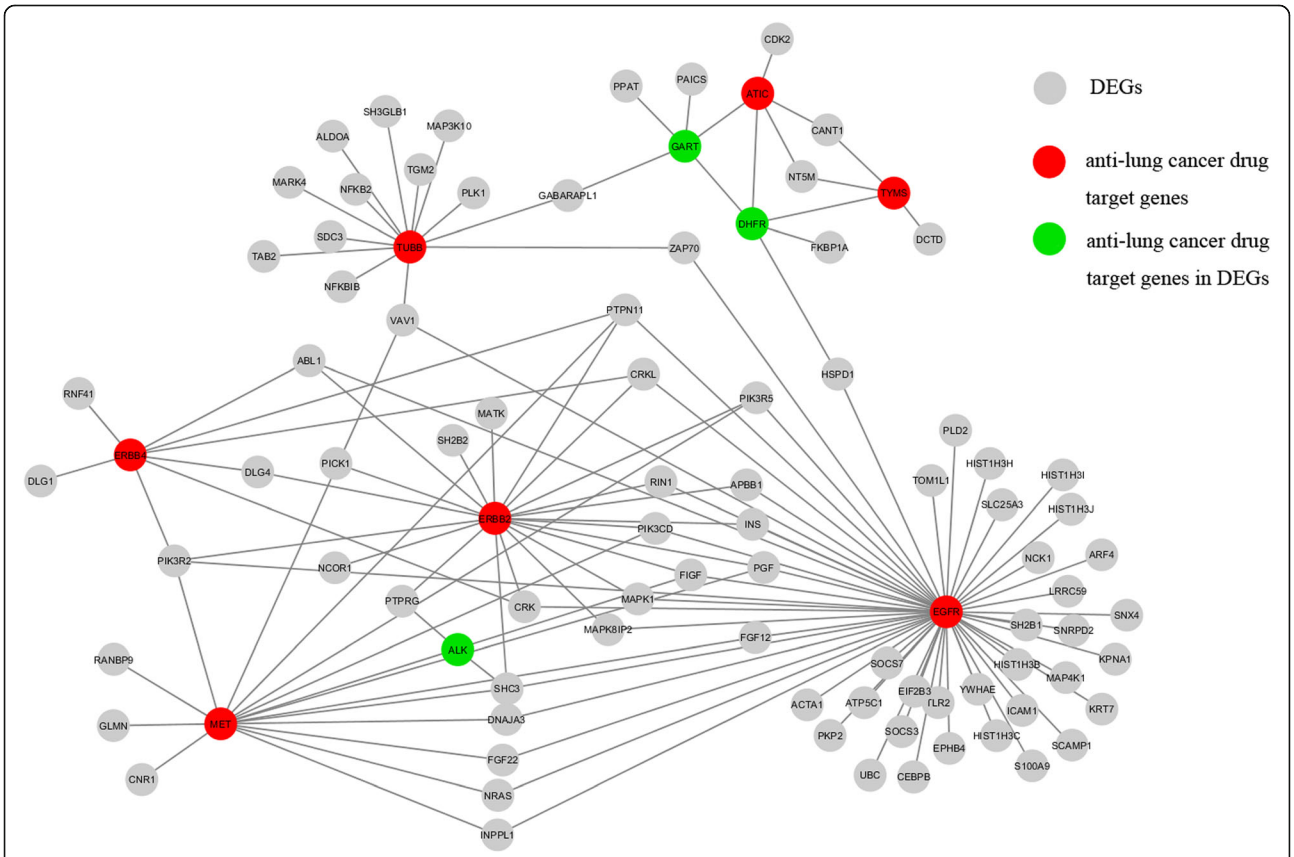

Fig. 2 The PPI links between DEGs and anti-lung cancer drug target genes

Necitumumab, Crizotinib and Vinorelbine could be considered to be included in the LmCRC regimen.

Discussion

Up to now, there are massive data of primary tumors in the public databases, whereas the data for metastases are rare. As the REO-based signature is insensitive to experimental batch effects, we could obviate the need for data normalization for the discovery and validation datasets. Especially, because the REO-based signature obviates the need for data normalization, it can be applied at the individual level [17, 20, 21].

The linear progression model, where the accumulated genetic alterations in the primary tumor could lead to metastases, is generally accepted in researches on cancer progression [23]. In the linear progression model, the metastases acquire most traits of the primary site. Accordingly, the treatment for the metastases is the same or similar to their primary counterparts. Therefore, to precisely identify whether a lesion is the primary tumor or metastasis from other cancerous organs is essential for tailoring the regimen. The liver- and lung-limited metastases are the most frequent migrating targets of CRC [25]. In this study, taking the LmCRC as an example, we showed a high concordance rate between the REOs within lung metastases and their primary CRC tissues other than the primary lung cancer tissues, which provided a direct evidence for the linear progression model.

On the other hand, there might be some differences between the lung metastases and primary CRC. By using the Student's *t*-test with FDR < 0.01, 2034 DEGs were detected between the primary CRC and lung metastases samples in GSE41258, among which 274 genes were also detected as DEGs between the primary CRC and primary lung cancer samples in GSE19249. Then, we made use of the normal colorectal tissue and lung tissue samples to further explore whether these 274 DEGs might exhibit lung tissue-specific characteristics. The result showed that 52 of the 274 DEGs were DEGs between the normal colorectal tissue and lung tissue samples (Student's *t*-test with FDR < 0.01). These results indicated that the lung metastases of CRC might possess some characteristics of the host organ, which needs to be further confirmed by analyzing microdissected samples of lung metastases of CRC to eliminate the possible confounding influence of residual lung tissues in the LmCRC samples. Finally, a PPI network analysis was conducted for DEGs between the primary CRC and their lung counterparts. As shown in Fig. 2, three DEGs (DHFR, GART and ALK) also played roles as anti-lung cancer drugs target genes. As the treatment for lung-limited metastases of CRC was curative resection accompanied with the regimen for CRC [35, 36], the analysis indicated that some lung cancer drugs could be recommended for LmCRC patients, which deserves further study for tailoring the treatment regimen for the LmCRC patients.

Conclusions

The REOs-based signature could identify the primary tissue of LmCRC with an accuracy of 100%. The within-sample REOs in primary sites could be a powerful approach for predicting the origin tissues of CUPs.

Abbreviation

CRC: Colorectal cancer; CUP: Cancer of unknown primary; DEGs: Differentially expressed genes; FDR: False discovery rate; GEO: Gene Expression Omnibus; LmCRC: Lung-limited metastatic colorectal cancer; MTP: Molecular tumor profiling; PPI: Protein-protein interaction; REOs: Relative gene expression orderings; RMA: Robust multi-array average

Acknowledgements

Not applicable.

Funding

This work was supported in part by the National Natural Science Foundation of China (grant numbers 81501215, 81501829, 81372213, 81572935, 61601151 and 21534008) and Natural Science Foundation of Fujian Province, China (grant number 2016 J01706) and Joint Fund for Program of Science and Technology innovation of Fujian Province, China (grant number 2016Y9102, 2016Y9044). The funding bodies had no role in the design of the study, collection, analysis, and interpretation of data and in writing of the manuscript.

Availability of data and materials

All data generated or analysed during this study are included in this published article.

Authors' contributions

MYL, HL and ZG conceived and designed the study; ZT contributed to the data collection and pre-processing; GL, XL and ML contributed to the data collection, literature review, result discussion and code testing; MYL conducted the signature identification and wrote the manuscript; HL, GH, LQ and ZG edited the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Ethics approval and participant consent was not necessary as this study involved the use of a previously-published de-identified database GEO.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details

[1]Key Laboratory of Ministry of Education for Gastrointestinal Cancer, Department of Bioinformatics, Fujian Medical University, Fuzhou 350001, China. [2]Department of Bioinformatics, Gannan Medical University, Ganzhou 341000, China. [3]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150086, China. [4]Fujian Key Laboratory of Tumor Microbiology, Fujian Medical University, Fuzhou 350001, China.

Received: 24 October 2017 Accepted: 3 January 2019
Published online: 14 January 2019

References

1. Pavlidis N, Briasoulis E, Hainsworth J, Greco FA. Diagnostic and therapeutic management of cancer of an unknown primary. Eur J Cancer. 2003;39(14):1990–2005.
2. Schwartz AM, Harpaz N. A primary approach to cancers of unknown primary. J Natl Cancer Inst. 2013;105(11):759–61.
3. Gatalica Z, Millis SZ, Vranic S, Bender R, Basu GD, Voss A, Von Hoff DD. Comprehensive tumor profiling identifies numerous biomarkers of drug response in cancers of unknown primary site: analysis of 1806 cases. Oncotarget. 2014;5(23):12440–7.
4. Golfinopoulos V, Pentheroudakis G, Salanti G, Nearchou AD, Ioannidis JP, Pavlidis N. Comparative survival with diverse chemotherapy regimens for cancer of unknown primary site: multiple-treatments meta-analysis. Cancer Treat Rev. 2009;35(7):570–3.
5. Onaitis MW, Petersen RP, Haney JC, Saltz L, Park B, Flores R, Rizk N, Bains MS, Dycoco J, D'Amico TA, et al. Prognostic factors for recurrence after pulmonary resection of colorectal cancer metastases. Ann Thorac Surg. 2009;87(6):1684–8.
6. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology: Colorectal Cancer Screening. *Version 1,* 2015, *from* http://wwwnccnorg/professionals/physician_gls/pdf/genetics_colonpdf 2015.
7. Hainsworth JD, Rubin MS, Spigel DR, Boccia RV, Raby S, Quinn R, Greco FA. Molecular gene expression profiling to predict the tissue of origin and direct site-specific therapy in patients with carcinoma of unknown primary site: a prospective trial of the Sarah Cannon research institute. J Clin Oncol. 2013;31(2):217–23.
8. Greco FA, Lennington WJ, Spigel DR, Hainsworth JD. Molecular profiling diagnosis in unknown primary cancer: accuracy and ability to complement standard pathology. J Natl Cancer Inst. 2013;105(11):782–90.
9. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, Geman D, Baggerly K, Irizarry RA. Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet. 2010;11(10):733–9.
10. Geman D, d'Avignon C, Naiman DQ, Winslow RL. Classifying gene expression profiles from pairwise mRNA comparisons. Stat Appl Genet Mol Biol. 2004;3(1):1-19.
11. Tan AC, Naiman DQ, Xu L, Winslow RL, Geman D. Simple decision rules for classifying human cancers from gene expression profiles. Bioinformatics. 2005;21(20):3896–904.
12. Heinaniemi M, Nykter M, Kramer R, Wienecke-Baldacchino A, Sinkkonen L, Zhou JX, Kreisberg R, Kauffman SA, Huang S, Shmulevich I. Gene-pair expression signatures reveal lineage control. Nat Methods. 2013;10(6):577–83.
13. Eddy JA, Sung J, Geman D, Price ND. Relative expression analysis for molecular cancer diagnosis and prognosis. Technol Cancer Res Treat. 2010;9(2):149–59.
14. Wang H, Zhang H, Dai Z, Chen MS, Yuan Z. TSG: a new algorithm for binary and multi-class cancer classification and informative genes selection. BMC Med Genet. 2013;6(Suppl 1):S3.
15. Chen R, Guan Q, Cheng J, He J, Liu H, Cai H, Hong G, Zhang J, Li N, Ao L, et al. Robust transcriptional tumor signatures applicable to both formalin-fixed paraffin-embedded and fresh-frozen samples. Oncotarget. 2017;8(4):6652–62.
16. Cheng J, Guo Y, Gao Q, Li H, Yan H, Li M, Cai H, Zheng W, Li X, Jiang W, et al. Circumvent the uncertainty in the applications of transcriptional signatures to tumor tissues sampled from different tumor sites. Oncotarget. 2017;8(18):30265–75.
17. Xu L, Tan AC, Winslow RL, Geman D. Merging microarray data from separate breast cancer studies provides a robust prognostic test. BMC Bioinf. 2008;9:125.
18. Xu L, Tan AC, Naiman DQ, Geman D, Winslow RL. Robust prostate cancer marker genes emerge from direct integration of inter-study microarray data. Bioinformatics. 2005;21(20):3905–11.
19. Xu L, Geman D, Winslow RL. Large-scale integration of cancer microarray data identifies a robust common cancer signature. BMC Bioinf. 2007;8:275.
20. Wang H, Sun Q, Zhao W, Qi L, Gu Y, Li P, Zhang M, Li Y, Liu SL, Guo Z. Individual-level analysis of differential expression of genes and pathways for personalized medicine. Bioinformatics. 2015;31(1):62–8.
21. Zhou X, Li B, Zhang Y, Gu Y, Chen B, Shi T, Ao L, Li P, Li S, Liu C, et al. A relative ordering-based predictor for tamoxifen-treated estrogen receptor-positive breast cancer patients: multi-laboratory cohort validation. Breast Cancer Res Treat. 2013;142(3):505–14.
22. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA Cancer J Clin. 2012;62(1):10–29.
23. Tan IB, Malik S, Ramnarayanan K, McPherson JR, Ho DL, Suzuki Y, Ng SB, Yan S, Lim KH, Koh D, et al. High-depth sequencing of over 750 genes supports linear progression of primary tumors and metastases in most patients with liver-limited metastatic colorectal cancer. Genome Biol. 2015;16:32.
24. Van Cutsem E, Cervantes A, Nordlinger B, Arnold D, Group EGW. Metastatic colorectal cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2014;25(Suppl 3):iii1–9.
25. Pfannschmidt J, Dienemann H, Hoffmann H. Surgical resection of pulmonary metastases from colorectal cancer: a systematic review of published series. Ann Thorac Surg. 2007;84(1):324–38.
26. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res. 2013;41(Database issue):D991–5.
27. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003;4(2):249–64.
28. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res. 2003;31(4):e15.
29. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics. 2003;19(2):185–93.
30. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 2006;34(Database issue):D668–72.
31. Shen X, Li S, Zhang L, Li H, Hong G, Zhou X, Zheng T, Zhang W, Hao C, Shi T, et al. An integrated approach to uncover driver genes in breast cancer methylation genomes. PLoS One. 2013;8(4):e61214.
32. Peri S, Navarro JD, Kristiansen TZ, Amanchy R, Surendranath V, Muthusamy B, Gandhi TK, Chandrika KN, Deshpande N, Suresh S, et al. Human protein reference database as a discovery resource for proteomics. Nucleic Acids Res. 2004;32(Database issue):D497–501.
33. Wang Y, Yang W, Pu Q, Yang Y, Ye S, Ma Q, Ren J, Cao Z, Zhong G, Zhang X, et al. The effects and mechanisms of SLC34A2 in tumorigenesis and progression of human non-small cell lung cancer. J Biomed Sci. 2015;22:52.
34. Yin X, Wang H, Wu D, Zhao G, Shao J, Dai Y. SLC34A2 gene mutation of pulmonary alveolar microlithiasis: report of four cases and review of literatures. Respir Med. 2013;107(2):217–22.
35. Nordlinger B, Sorbye H, Glimelius B, Poston GJ, Schlag PM, Rougier P, Bechstein WO, Primrose JN, Walpole ET, Finch-Jones M, et al. Perioperative chemotherapy with FOLFOX4 and surgery versus surgery alone for resectable liver metastases from colorectal cancer (EORTC intergroup trial 40983): a randomised controlled trial. Lancet. 2008;371(9617):1007–16.
36. Roth AD, Tejpar S, Delorenzi M, Yan P, Fiocca R, Klingbiel D, Dietrich D, Biesmans B, Bodoky G, Barone C, et al. Prognostic role of KRAS and BRAF in stage II and III resected colon cancer: results of the translational study on the PETACC-3, EORTC 40993, SAKK 60-00 trial. J Clin Oncol. 2010;28(3):466–74.