Study of the co-expression gene modules of non-small cell lung cancer metastases

CURRENT STATUS: POSTED

Guanghui Wang
Institute of Oncology

Fenglong Bie
Institute of Oncology

Guangxu Li
Department of Thoracic Surgery

Junping Shi
Medical Department

Yanwu Zeng
OrigiMed

Jiajun Du dujiajun@sdu.edu.cn
Institution of Oncology, Shandong Provincial Hospital Affiliated to Shandong University

Corresponding Author
ORCID: 0000-0003-2406-9435

DOI: 10.21203/rs.2.17505/v1

SUBJECT AREAS Oncology

KEYWORDS non-small cell lung cancer, metastases, gene module, Weighted Correlation Network Analysis, co-expression
Abstract

Background: Metastasis regularly is a marker of the disease development of cancers. Some metastatic sites significantly showed a more serious clinical outcome in non-small cell lung cancer (NSCLC). Whether they are caused by tissue-specific (TS) or non-tissue-specific (NTS) mechanisms is still unclear.

Methods: Weighted Correlation Network Analysis (WGCNA) was used to identify the gene modules among the metastases of NSCLC. The clinical significance of those gene modules was evaluated with the Cox hazard proportional model with another independent dataset. Functions of each gene module were analyzed with gene ontology. Typical genes were further studied.

Results: There were two TS gene modules and two NTS gene modules identified. One TS gene module (green module) and one NTS gene module (purple module) significantly correlated with survival. This NTS gene module (purple module) was significantly enriched in the epithelial-to-mesenchymal transition (EMT) process. Higher expression of the typical genes (CA14, SOX10, TWIST1, and ALX1) from EMT process was significantly associated with a worse survival.

Conclusion: The lethality of NSCLC metastases was caused by TS gene modules and NTS gene modules, among which the EMT-related gene module was critical for a worse clinical outcome.

Background

Lung cancer is the most deadly cause of cancer-related deaths worldwide. Among them, non-small cell lung cancer (NSCLCs) accounts for 80%. Along with disease development, metastasis of lung cancer is one of the most significant markers
(Chadha, Ganti et al. 2005, Oh, Taylor et al. 2009). It is reasonable to hypothesize that there could be some shared mechanism which is responsible for the lethality from the metastases of NSCLC. It has been reported that multiple mechanisms are involved in the process of angiogenesis, hypoxia, circulation, and establishment of a metastatic focus (Popper 2016).

However, tumors are highly heterogeneous. It has long been realized that patients with the same metastatic site showed different survival time (Kochhar, Frytak et al. 1997, Pop, Nadeemy et al. 2009). Recently, the metastatic site was recognized as an important factor that affects clinical outcome (Bauml, Mick et al. 2013, Riihimaki, Hemminki et al. 2014). Riihimaki et al. (Riihimaki, Hemminki et al. 2014) have found that liver and bone metastases caused a worse survival time, and the respiratory and nervous system metastases showed a better survival time. Gibson et al. (Gibson, Li et al. 2018) have found that liver metastasis had worse survival and adrenal metastasis had the best survival. Tamura et al. (Tamura, Kurishima et al. 2015) have found that liver and adrenal gland metastasis had bad survival and brain and bone metastases have better survival. Those studies conflicted with each other for some metastatic sites. Commonly, liver metastasis had bad survival.

As for now, the understanding of the mechanism of the lethality difference within the metastases of lung cancer is still unclear. Only a limited number of studies in genomic variation analysis, especially in some driver genes, were reported. For examples, EGFR mutation could be an important marker for clinical outcome (Hsu, De Caluwe et al. 2017, Li, Zhao et al. 2019). The median survival time for the EGFR+ group and EGFR was 22.4 months and 7.9 months, respectively (Hsu, De Caluwe et al. 2017). However, in the EGFR+ cohort, the brain showed the highest risk ratio,
which was contrary to a previous study (Riihimaki, Hemminki et al. 2014). Besides, there were only 22.3% of the patients tested as $\text{EGFR}^+$ (Hsu, De Caluwe et al. 2017), indicating that there must be other mechanisms involved.

In spite of the success in studying such driver genes, there are many unresolved questions. For examples, which mechanism is most important for the survival? Is there shared mechanism among these metastases? Is there a tissue-specific mechanism among these metastases?”. In order to solve these problems, a global transcriptome study would be necessary.

In this study, the Weighted Correlation Network Analysis (WGCNA) was used to identify the gene modules in NSCLC. The tissue-specificity and function of those gene modules were studied. Further using an independant dataset, the clinical outcomes of these modules were evaluated. Specifically, some typical genes were carefully examined.

Result

Prepare data for WGCNA analysis

By searching in the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) with keywords of “lung cancer AND metastasis” and further checking manually, only one dataset, GSE12630, which contains 12 metastatic samples of lung cancer, was screened (Suppl. Table 1) and then the CEL files for those samples were downloaded. The expression was called with the affy package (Gautier, Cope et al. 2004) from Bioconductor on R platform and normalized with the Robust Multi-array Average (RMA) method (Bolstad, Irizarry et al. 2003). Low variable probes were removed according to the interquartile range (<0.5). In total, 3533 probes were kept for further analysis.
**WGCNA analysis for the gene modules**

In order to get the co-expression gene modules, the WGCNA analysis was conducted with the 12 samples. As denoted by the analysis of the scale-free topology model fit, a soft power threshold at 37 showed a better performance. Unsigned Pearson’s correlation coefficient between genes was used to generate the adjacency matrix. The topological overlap was calculated to construct the co-expression network. Each gene was then hierarchically clustered according to the topological distance in the co-expression network (Figure 2). The hierarchical tree was dynamically cut. Nine gene modules were inferred. The gene modules with high eigengene similarity (>0.75) were merged. Finally, they were merged into four gene modules. For convenience, they were named as black, green, yellow and purple modules, which contained 157, 749, 131, and 47 genes respectively.

**Tissue specificity and functions of the gene modules**

To characterize the gene modules, the eigengene was used as a surrogate measure for each gene module. A Pearson’s correlation coefficient between those eigengenes for each gene modules and the tissue type vector was calculated. As shown in (Figure 3), the green module was highly expressed in the lymph node ($r = 0.67$, $p$-value = 0.02) and expressed lower in the adrenal gland ($r = -0.52$, $p$-value = 0.08); the yellow module was highly expressed in the intestine tissue ($r = 0.87$, $p$-value = 0.0001). The purple and black gene modules showed no significant correlation with any tissue types. Those results showed that there were tissue-specific and non-tissue-specific gene modules in NSCLC metastases.

Next, the function of the four gene modules in the biological process was examined with gene ontology. The yellow module was enriched with “extracellular matrix organization” and “cell adhesion” (Suppl. Figure 1A); the black module was enriched
with “immune response”, “defense response”, “innate immune response” and “regulation of T cell activation” (Suppl. Figure 1B); the green module was enriched with “hormone-mediated signaling pathway”, “cell communication” and “cellular response to endogenous stimulus” (Suppl. Figure 1C); and the purple module was enriched with “neural crest cell migration”, “mesenchyme development” and “stem cell differentiation” (Figure 4).

**EMT process was critical for lethality caused by metastasis**

Except for the biological function of gene modules identified by WGCNA analysis, we also studied their clinical association. Without loss of generality, another dataset, GSE14814, was used to validate the gene modules. That dataset had recorded survival/censor time, which was used as a dependent variable in a Cox hazard proportional model (Table 1). The eigengene expressions of those gene modules was dichotomized by the median. They were used as independent variables in the model. As denoted by the log-likelihood ratio test, the model was significant with p-value = 0.04. Specifically, the green and purple modules were significantly correlated with survival. The p-values were 0.0069 and 0.036, respectively. The purple module had a hazard ratio equal to 2.20 and a 95% confident interval within [1.06–4.59]. The purple module had a hazard ratio equal to 0.36 and a 95% confident interval within [0.17–0.76]. Higher expression of the purple module suggested a lower risk of lethality.

**Clinical significance of the typical genes from the purple module**

Many studies (Nouri, Ratther et al. 2014, Puisieux, Brabletz et al. 2014, Shen, Xu et al. 2019) had shown that EMT was important for the metastasis process, but there were no studies about the clinical association with metastasis. In order to study the clinical significance of the EMT process, the hub genes and EMT related genes from
the purple module were further validated in a wider dataset provided by the web service (Monzon, Lyons-Weiler et al. 2009) (Figure 5). The hub gene was defined as the gene which had the highest correlation with the eigengene of the gene modules. CA14 was the hub gene of the purple module. SOX10, ALX1, and TWIST1 were the genes involved in the EMT in the purple module. Using the median as the cutoff, their Kaplan-Meier curves were plotted as in Figure 5. Higher expression of these genes showed a significantly shorter survival time with the p-values of 0.0023, 0.037, 3.7e–5 and 3.7e–3, respectively. The results further proved the lethality of the purple module and the EMT process for the lung cancer metastasis.

Discussion and Conclusion

Lung metastasis could severely affect the clinical outcome (Riihimaki, Hemminki et al. 2014). Different metastatic sites could cause significantly different survival (Riihimaki, Hemminki et al. 2014, Tamura, Kurishima et al. 2015, Hsu, De Caluwe et al. 2017, Gibson, Li et al. 2018). Those studies have found both common and different results. To resolve those discrepancies, it is urgent to determine the background molecular mechanisms. Compared to the genomic variations, the expression profile can reflect the biological process in those metastasis tissues more directly when considering the difficulties in the evaluation of the effects of genomic variations. Due to the difficult availability of metastatic tissues, there have been only a few studies on the expression profiles of lung cancer metastasis up to now. After searching for the GEO database, only one dataset containing multiple metastatic tissues has been collected. That dataset only contains metastatic sites of the adrenal gland, lymphoid, skin and omentum tissues. There are no expression profiles from the liver, which is the most lethal metastasis (Riihimaki, Hemminki et
al. 2014, Tamura, Kurishima et al. 2015, Gibson, Li et al. 2018). Limited by the data availability, it is hard to know whether the liver has a tissue-specific module or non-tissue-specific module. Although there is a study which contains only a single metastatic site of brain tissue (Luke, Blazquez et al. 2018), to avoid possible error from multiple experiments, it was not included in this study. We look forward to more efforts in this field of research.

In order to infer the gene modules that are tissue-specific and non-tissue-specific, unsupervised clustering is needed. In fact, in considering the high heterogeneity of lung cancer metastasis, Student’s t-test in combination with supervised clustering could always be futile. WGCNA has the advantage to cluster all genes into separate gene modules without the knowledge of the tissue types. In this study, we have successfully identified four gene modules. The function of those gene modules was studied with the tissue-specificity and the gene ontology of biological process. As we have hypothesized, two gene modules (green and yellow) are correlated with the tissue types and two other gene modules (black and purple) were not correlated with any tissue types.

The green module was correlated with lymphoid metastasis tissue. It was enriched with “cell communication”. The typical gene was Wnt10B, which was reported to be associated with lymph node metastasis of gastric cancer (Wu, Bie et al. 2017). However, there is also a report in endometrial cancer that high expression of Wnt10B is prone to not have lymphoid metastasis (Chen, Wang et al. 2013). Those results suggest that the expression of Wnt10B is critical for lymphoid metastasis. The green gene module could be tissue-specific for lymphoid metastasis of pan-cancer.

Similarly, the yellow module was correlated with intestine metastasis. It was
enriched with “extracellular matrix organization” and “cell adhesion”. The typical genes were *MMP2*, *COL1A1*, and *ECM2*. Currently, there are rarely studies in the intestine metastasis of lung cancer, but they have appeared in the metastasis of other cancers, such as, breast cancer (Mendes, Kim *et al.* 2007), ovarian cancer (Kenny, Kaur *et al.* 2008) and lung cancer (Rojiani, Alidina *et al.* 2010).

The black gene module was non-tissue-specific and enriched with “immune response”, “defense response”, “innate immune response” and “regulation of T cell activation”. The black module reflects the function of immune microenvironment. It is a general mechanism in all types of cancers. For example, the inflammatory chemokine C-C motif ligand 4 (*CCL4*) is an “innate immune response”-related gene. *CCL4* can promote the lymphangiogenesis in oral squamous cell carcinoma (OSCC), which is a critical step in tumor metastasis (Lien, Tsai *et al.* 2018). In breast cancer, *CCL4* can contribute to the metastasis to the bone by binding to CC chemokine receptor type 5 (Sasaki, Baba *et al.* 2016). The genes in the “regulation of T cell activation” category, *HLA-DR* or *CD4*, also play a critical role in the metastasis in the metastasis of melanoma (Costantini and Barbieri 2017), breast cancer (DeNardo, Barreto *et al.* 2009), and prostate cancer (Liu, Guo *et al.* 2018).

The most interesting module was the purple module, which was also a non-tissue-specific gene module and had a significant effect on the survival. The purple module was enriched with “neural crest cell migration”, “mesenchyme development” and “stem cell differentiation”. Generally, for cancer metastasis, the nervous system plays a big part of metastatic initiation and clonal expansion, angiogenesis to extravasation and colonization (Kuol, Stojanovska *et al.* 2018). The typical genes from it are *SOX10*, *TWIST1*, and *ALX1*. Those genes also take part in the EMT process. EMT is a process whereby epithelial cells are transformed into
mesenchymal cells (Kalluri and Weinberg 2009). After the EMT process, the transited cells gain migratory/invasive properties (Bonnomet, Syne et al. 2012). This study proves that EMT could be the most important signature for the severity of lung cancer metastasis. In summary, in this article, we have studied the gene co-expression modules in the metastatic tissues of lung cancer. Four gene modules were identified, from which two gene modules were tissue-specific and the other two gene modules were non-tissue-specific. Specifically, one of the non-tissue-specific gene modules was significantly correlated with survival. The function of that gene module demonstrates that the EMT process plays a big role in the metastasis process of lung cancer and survival.

Method

Data collections

Two datasets, GSE12630 (Monzon, Lyons-Weiler et al. 2009) and GSE14814 (Zhu, Ding et al. 2010), were downloaded from the Gene Expression Omnibus (GEO) website (http://www.ncbi.nlm.nih.gov/geo/). For better comparison, only the sample data generated on the Affymetrix U133A microarray platform were chosen as candidates. The samples for primary or metastatic lung cancer were manually curated from dataset GSE12630. The raw CEL files from both datasets were processed by affy package on R platform. The expression was further normalized with the Robust Multi-array Average (RMA) method (Bolstad, Irizarry et al. 2003). The suboptimal probes with the name suffix “s_at” and “x_at” were removed from the analysis. To better fit the normal distribution, the expression was log2-transformed. The annotation package for U133A was downloaded from Bioconductor
(http://bioconductor.org). For dataset GSE12630, the low variant probes were filtered according to the interquartile range (<0.5).

**WGCNA analysis**

WGCNA analysis was conducted on the R platform (Langfelder and Horvath 2008). Briefly, the scale-free topology model fit was used to choose the soft power. An unsigned correlation distance defined as \((1-r)/2\) was calculated, where \(r\) is the Pearson’s correlation coefficient. The adjacency matrix was generated according to the unsigned correlation distance matrix with the soft power. The adjacency correlation was then used to construct the topological network. The genes were hierarchically clustered with the topological distance. Then, the tree was dynamically cut into multiple gene modules with the parameter “minClusterSize = 30, deepSplit = 2, pamRespectsDendro = FALSE”. Those gene modules with high eigengene similarity (>0.75) were merged.

**Gene function analysis**

The biological processes of gene ontology were used to enrich the gene modules with the DAVID web service (Huang da, Sherman et al. 2009). The bar plot is generated on the R platform.

**Survival analysis**

The WGCNA modules were selected by survival analysis of their eigengenes. The eigengenes were dichotomized by the median. The expression over the median will be defined as high expression; the expression below the median will be defined as low expression. The gene modules were regressed against the survival time with Cox proportional hazards model with R package “survival” (Therneau 2015). The log-likelihood test was used to test the significance. The modules with higher survival probability were further enriched with gene ontology. The typical genes
were validated in a large dataset with the web service (Monzon, Lyons-Weiler et al. 2009). The median value was used as a cutoff to categorize the gene expression into high and low. The survival time was plotted against survival probability with the Kaplan-Meier curve. The log-likelihood ratio test was used to compare their different significance.

Abbreviations

NSCLC: non-small cell lung cancer
TS: tissue-specific
NTS: non-tissue-specific
WGCNA: Weighted Correlation Network Analysis
EMT: epithelial-to-mesenchymal transition
GEO: Gene Expression Omnibus
RMA: Robust Multi-array Average
CCL4: C-C motif ligand 4
OSCC: oral squamous cell carcinoma.

Declarations

Ethics approval and consent to participate

Yes.

Consent for publication

Yes.

Availability of data and material

Yes.

Competing interests

Conflict of interest: Guanghui Wang, Fenglong Bie, Guangxu Li, Junping Shi, Yanwu Zeng and Jiajun Du declared no conflicts of interest in this work.

Funding

This work was supported by the National Natural Science Foundation of China, grant nos. 81602009 (funder: Guanghui Wang) and 81672288 (funder: Jiajun Du) and The
Joint Research Funds for Shandong University and Karolinska Institute, grant no. SDU-KI–2019–16 (funder: Jiajun Du).

Authors’ contributions

GHW, FLB and JJD designed this experiment. GHW, FLB, GXL collected and processed data. JPS, YWZ and JJD wrote and polished article. All of the authors reviewed the manuscript. All authors have read and approved the manuscript.

Acknowledgements

Not applicable.

References

Bauml, J., R. Mick, Y. Zhang, C. D. Watt, A. Vachani, C. Aggarwal, T. Evans and C. Langer (2013). “Determinants of survival in advanced non—small-cell lung cancer in the era of targeted therapies.” Clin Lung Cancer 14(5): 581–591.

Bolstad, B. M., R. A. Irizarry, M. Astrand and T. P. Speed (2003). “A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.” Bioinformatics 19(2): 185–193.

Bonnomet, A., L. Syne, A. Brysse, E. Feyereisen, E. W. Thompson, A. Noel, J. M. Foidart, P. Birembaut, M. Polette and C. Gilles (2012). “A dynamic in vivo model of epithelial-to-mesenchymal transitions in circulating tumor cells and metastases of breast cancer.” Oncogene 31(33): 3741–3753.

Chadha, A. S., A. K. Ganti, J. S. Sohi, A. E. Sahmoun and S. A. Mehdi (2005). “Survival in untreated early stage non-small cell lung cancer.” Anticancer Res 25(5): 3517–3520.

Chen, H., Y. Wang and F. Xue (2013). “Expression and the clinical significance of Wnt10a and Wnt10b in endometrial cancer are associated with the Wnt/beta-catenin
Costantini, F. and G. Barbieri (2017). “The HLA-DR mediated signalling increases the migration and invasion of melanoma cells, the expression and lipid raft recruitment of adhesion receptors, PD-L1 and signal transduction proteins.” *Cell Signal* 36: 189–203.

DeNardo, D. G., J. B. Barreto, P. Andreu, L. Vasquez, D. Tawfik, N. Kolhatkar and L. M. Coussens (2009). “CD4(+) T cells regulate pulmonary metastasis of mammary carcinomas by enhancing protumor properties of macrophages.” *Cancer Cell* 16(2): 91–102.

Gautier, L., L. Cope, B. M. Bolstad and R. A. Irizarry (2004). “affy—analysis of Affymetrix GeneChip data at the probe level.” *Bioinformatics* 20(3): 307–315.

Gibson, A. J. W., H. Li, A. D’Silva, R. A. Tudor, A. A. Elegbede, S. M. Otsuka, D. G. Bebb and W. Y. Cheung (2018). “Impact of number versus location of metastases on survival in stage IV M1b non-small cell lung cancer.” *Med Oncol* 35(9): 117.

Hsu, F., A. De Caluwe, D. Anderson, A. Nichol, T. Toriumi and C. Ho (2017). “Patterns of spread and prognostic implications of lung cancer metastasis in an era of driver mutations.” *Curr Oncol* 24(4): 228–233.

Huang da, W., B. T. Sherman and R. A. Lempicki (2009). “Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.” *Nat Protoc* 4(1): 44–57.

Kalluri, R. and R. A. Weinberg (2009). “The basics of epithelial-mesenchymal transition.” *J Clin Invest* 119(6): 1420–1428.

Kenny, H. A., S. Kaur, L. M. Coussens and E. Lengyel (2008). “The initial steps of ovarian cancer cell metastasis are mediated by MMP-2 cleavage of vitronectin and fibronectin.” *J Clin Invest* 118(4): 1367–1379.
Kochhar, R., S. Frytak and E. G. Shaw (1997). “Survival of patients with extensive small-cell lung cancer who have only brain metastases at initial diagnosis.” *Am J Clin Oncol* **20**(2): 125-127.

Kuol, N., L. Stojanovska, V. Apostolopoulos and K. Nurgali (2018). “Role of the nervous system in cancer metastasis.” *J Exp Clin Cancer Res* **37**(1): 5.

Langfelder, P. and S. Horvath (2008). “WGCNA: an R package for weighted correlation network analysis.” *BMC Bioinformatics* **9**: 559.

Li, W. Y., T. T. Zhao, H. M. Xu, Z. N. Wang, Y. Y. Xu, Y. Han, Y. X. Song, J. H. Wu, H. Xu, S. C. Yin, X. Y. Liu and Z. F. Miao (2019). “The role of EGFR mutation as a prognostic factor in survival after diagnosis of brain metastasis in non-small cell lung cancer: a systematic review and meta-analysis.” *BMC Cancer* **19**(1): 145.

Lien, M. Y., H. C. Tsai, A. C. Chang, M. H. Tsai, C. H. Hua, S. W. Wang and C. H. Tang (2018). “Chemokine CCL4 Induces Vascular Endothelial Growth Factor C Expression and Lymphangiogenesis by miR–195–3p in Oral Squamous Cell Carcinoma.” *Front Immunol* **9**: 412.

Liu, L., K. Guo, Z. Liang, F. Li and H. Wang (2018). “Identification of candidate genes that may contribute to the metastasis of prostate cancer by bioinformatics analysis.” *Oncol Lett* **15**(1): 1220-1228.

Luke, F., R. Blazquez, R. F. Yamaci, X. Lu, B. Pregler, S. Hannus, K. Menhart, D. Hellwig, H. J. Wester, S. Kropf, D. Heudobler, J. Grosse, J. Moosbauer, M. Hutterer, P. Hau, M. J. Riemschneider, M. Bayerlova, A. Bleckmann, B. Polzer, T. Beissbarth, C. A. Klein and T. Pukrop (2018). “Isolated metastasis of an EGFR-L858R-mutated NSCLC of the meninges: the potential impact of CXCL12/CXCR4 axis in EGFRmut NSCLC in diagnosis, follow-up and treatment.” *Oncotarget* **9**(27): 18844-18857.

Mendes, O., H. T. Kim, G. Lungu and G. Stoica (2007). “MMP2 role in breast cancer
brain metastasis development and its regulation by TIMP2 and ERK1/2.” *Clin Exp Metastasis* 24(5): 341-351.

Monzon, F. A., M. Lyons-Weiler, L. J. Buturovic, C. T. Rigl, W. D. Henner, C. Sciulli, C. I. Dumur, F. Medeiros and G. G. Anderson (2009). “Multicenter validation of a 1,550-gene expression profile for identification of tumor tissue of origin.” *J Clin Oncol* 27(15): 2503-2508.

Nouri, M., E. Ratther, N. Stylianou, C. C. Nelson, B. G. Hollier and E. D. Williams (2014). “Androgen-targeted therapy-induced epithelial mesenchymal plasticity and neuroendocrine transdifferentiation in prostate cancer: an opportunity for intervention.” *Front Oncol* 4: 370.

Oh, Y., S. Taylor, B. N. Bekele, J. M. Debnam, P. K. Allen, D. Suki, R. Sawaya, R. Komaki, D. J. Stewart and D. D. Karp (2009). “Number of metastatic sites is a strong predictor of survival in patients with nonsmall cell lung cancer with or without brain metastases.” *Cancer* 115(13): 2930-2938.

Pop, D., A. S. Nadeemy, N. Venissac, P. Guiraudet, J. Otto, M. Poudenx and J. Mouroux (2009). “Skeletal muscle metastasis from non-small cell lung cancer.” *J Thorac Oncol* 4(10): 1236-1241.

Popper, H. H. (2016). “Progression and metastasis of lung cancer.” *Cancer Metastasis Rev* 35(1): 75-91.

Puisieux, A., T. Brabletz and J. Caramel (2014). “Oncogenic roles of EMT-inducing transcription factors.” *Nat Cell Biol* 16(6): 488-494.

Riihimaki, M., A. Hemminki, M. Fallah, H. Thomsen, K. Sundquist, J. Sundquist and K. Hemminki (2014). “Metastatic sites and survival in lung cancer.” *Lung Cancer* 86(1): 78-84.

Rojiani, M. V., J. Alidina, N. Esposito and A.M. Rojiani (2010). “Expression of MMP-2
correlates with increased angiogenesis in CNS metastasis of lung carcinoma.” *Int J Clin Exp Pathol* 3(8): 775–781.

Sasaki, S., T. Baba, T. Nishimura, Y. Hayakawa, S. Hashimoto, N. Gotoh and N. Mukaida (2016). “Essential roles of the interaction between cancer cell-derived chemokine, CCL4, and intra-bone CCR5-expressing fibroblasts in breast cancer bone metastasis.” *Cancer Lett* 378(1): 23–32.

Shen, M., Z. Xu, W. Xu, K. Jiang, F. Zhang, Q. Ding, Z. Xu and Y. Chen (2019). “Inhibition of ATM reverses EMT and decreases metastatic potential of cisplatin-resistant lung cancer cells through JAK/STAT3/PD-L1 pathway.” *J Exp Clin Cancer Res* 38(1): 149.

Tamura, T., K. Kurishima, K. Nakazawa, K. Kagohashi, H. Ishikawa, H. Satoh and N. Hizawa (2015). “Specific organ metastases and survival in metastatic non-small-cell lung cancer.” *Mol Clin Oncol* 3(1): 217–221.

Therneau, T. M. (2015). A Package for Survival Analysis in S.

Wu, X. D., Q. L. Bie, B. Zhang, Z. H. Yan and Z. J. Han (2017). “Wnt10B is critical for the progression of gastric cancer.” *Oncol Lett* 13(6): 4231–4237.

Zhu, C. Q., K. Ding, D. Strumpf, B. A. Weir, M. Meyerson, N. Pennell, R. K. Thomas, K. Naoki, C. Ladd-Acosta, N. Liu, M. Pintilie, S. Der, L. Seymour, I. Jurisica, F. A. Shepherd and M. S. Tsao (2010). “Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer.” *J Clin Oncol* 28(29): 4417–4424.

Tables

**Table 1. Study of module eigengenes by survival analysis**
Table 1. Study of module eigengenes by survival analysis

|        | Coef | Exp(coef) | Se(coef) | z    | P (>|z|) |
|--------|------|-----------|----------|------|----------|
| MEgreen| 0.7889 | 2.201     | 0.3754   | 2.102 | 0.03559  |
| MEPurple| -1.0137 | 0.3629    | 0.375    | -2.703| 0.00686  |
| MEblack| -0.5291 | 0.5892    | 0.2973   | -1.779| 0.07516  |
| MEyellow| 0.1942  | 1.2144    | 0.2989   | 0.65  | 0.51579  |

The study was conducted in a dataset consisting of 133 samples. The log-likelihood ratio test showed a significant correlation between the gene modules and the survival time. The green and purple gene modules were the two significant gene modules with hazard ratios, 0.45 and 2.76, respectively.

Suppl. Table 1 The data used in this study

| GSE  | GSM          | GPL    | primary | metastasis      | shortDesc |
|------|--------------|--------|---------|-----------------|-----------|
| GSE12630 | GSM319679 | GPL96  | lung    | Adrenal gland   | ad        |
| GSE12630 | GSM319681 | GPL96  | lung    | Adrenal gland   | ad        |
| GSE12630 | GSM319674 | GPL96  | lung    | Adrenal gland   | ad        |
| GSE12630 | GSM319739 | GPL96  | lung    | Adrenal gland   | ad        |
| GSE12630 | GSM319669 | GPL96  | Lung    | Lung            | lu        |
| GSE12630 | GSM319684 | GPL96  | Lung    | Lung            | lu        |
| GSE12630 | GSM319724 | GPL96  | lung    | Lung            | lu        |
| GSE12630 | GSM319699 | GPL96  | lung    | Lymph node      | ly        |
| GSE12630 | GSM319671 | GPL96  | lung    | Lymph node      | ly        |
| GSE12630 | GSM319785 | GPL96  | lung    | Lymph node      | ly        |
| GSE12630 | GSM319786 | GPL96  | lung    | Lymph node      | ly        |
| GSE12630 | GSM319823 | GPL96  | lung    | Lymph node      | ly        |
| GSE12630 | GSM319790 | GPL96  | lung    | Omentum         | om        |
| GSE12630 | GSM319700 | GPL96  | lung    | Skin            | sk        |
| GSE14814 | GSM1267450 | GPL96  | lung    | Unknown         | lu        |
| GSE14814 | GSM1267451 | GPL96  | lung    | Unknown         | lu        |
| GSE14814 | GSM1267452 | GPL96  | lung    | Unknown         | lu        |
| GSE14814 | GSM1267453 | GPL96  | lung    | Unknown         | lu        |
| GSE14814 | GSM1267454 | GPL96  | lung    | Unknown         | lu        |
| GSE14814 | GSM1267455 | GPL96  | lung    | Unknown         | lu        |
| GSE14814 | GSM1267456 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267457 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267458 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267459 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267460 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267461 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267462 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267463 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267464 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267465 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267466 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267467 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267468 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267469 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267470 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267471 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267472 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267473 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267474 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267475 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267476 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267477 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267478 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267479 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267480 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267481 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267482 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267483 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267484 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267485 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267486 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267487 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267488 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267489 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267490 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267491 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM1267492 | GPL96 | lung | Unknown | lu |
| Run Set | Sample ID | Platform | Tissue | Experiment Type | Description |
|---------|-----------|----------|--------|-----------------|-------------|
| GSE14814 | GSM370913 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370914 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370915 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370916 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370917 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370918 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370919 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370920 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370921 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370922 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370923 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370924 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370925 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370926 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370927 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370928 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370929 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370930 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370931 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370932 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370933 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370934 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370935 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370936 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370937 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370938 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370939 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370940 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370941 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370942 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370943 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370944 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370945 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370946 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370947 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370948 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370949 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370950 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370951 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370952 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370953 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370954 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370955 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370956 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370957 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370958 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370959 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370960 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370961 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370962 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370963 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370964 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370965 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370966 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370967 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370968 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370969 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370970 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370971 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370972 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370973 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370974 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370975 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370976 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370977 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370978 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370979 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370980 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370981 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370982 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370983 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370984 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370985 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370986 | GPL96 | lung | Unknown | lu |
| GSE14814 | GSM370987 | GPL96 | lung | Unknown | lu |
GSE14814  GSM370988  GPL96  lung  Unknown  lu
GSE14814  GSM370989  GPL96  lung  Unknown  lu
GSE14814  GSM370990  GPL96  lung  Unknown  lu
GSE14814  GSM370991  GPL96  lung  Unknown  lu
GSE14814  GSM370992  GPL96  lung  Unknown  lu
GSE14814  GSM370993  GPL96  lung  Unknown  lu
GSE14814  GSM370994  GPL96  lung  Unknown  lu
GSE14814  GSM370995  GPL96  lung  Unknown  lu
GSE14814  GSM370996  GPL96  lung  Unknown  lu
GSE14814  GSM370997  GPL96  lung  Unknown  lu
GSE14814  GSM370998  GPL96  lung  Unknown  lu
GSE14814  GSM370999  GPL96  lung  Unknown  lu
GSE14814  GSM371000  GPL96  lung  Unknown  lu
GSE14814  GSM371001  GPL96  lung  Unknown  lu
GSE14814  GSM371002  GPL96  lung  Unknown  lu

Figures

Figure 1

The workflow of this study
Figure 2

WGCNA analysis of the metastasis tissues
Figure 3

The relationship between modules and tissues

Figure 4

Biological processes enriched in purple module
Figure 5
Validation of typical genes

Supplementary Files
This is a list of supplementary files associated with the primary manuscript. Click to download.
S Figure 1.docx