

Whole-Exome Sequencing Identifies Somatic Mutations Associated With Lung Cancer Metastasis to the Brain

Zhenghao Liu

Department of Neurosurgery, Sun Yat-sen Memorial Hospital, Sun Yat-sen University

Meiguang Zheng

Department of Neurosurgery, Sun Yat-sen Memorial Hospital, Sun Yat-sen University

Bingxi Lei

Department of Neurosurgery, Sun Yat-sen Memorial Hospital, Sun Yat-sen University

Zhiwei Zhou

Department of Neurosurgery, Sun Yat-sen Memorial Hospital, Sun Yat-sen University

Yutao Huang

Department of Neurosurgery, Sun Yat-sen Memorial Hospital, Sun Yat-sen University

Wenpeng Li

Department of Neurosurgery, Sun Yat-sen Memorial Hospital, Sun Yat-sen University

Qinbiao Chen

Department of Neurosurgery, Sun Yat-sen Memorial Hospital, Sun Yat-sen University

Pengcheng Li (✉ lipc1981@163.com)

Department of Thoracic Oncology, Wuhan Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei 430022, China

Yuefei Deng

Sun Yat-sen University, Department of Neurosurgery

Research

Keywords: whole-exome sequencing, lung cancer, somatic mutation, brain metastasis

Posted Date: September 21st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-74758/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Lung cancer is the most aggressive cancer which representing one-quarter of all cancer-related deaths, and metastatic spread accounts for >70% of these deaths, especially brain metastasis. Metastasis associated mutations are important biomarkers for metastasis prediction and outcome improvement.

Methods: In this study, we applied whole-exome sequencing to identify potential metastasis related mutation in 12 paired lung cancer and brain metastasis samples.

Results: We identified 1,702 SNVs and 6,131 mutation events in 1,220 genes. Furthermore, we identified several lung cancer metastases associated genes (*KMT2C*, *AHNAK2*). A mean of 3.1 driver gene mutation events per tumor with the dN/dS of 2.13 indicating a significant enrichment for cancer driver gene mutations. Mutation spectrum analysis found lung-brain metastasis samples have more similar Ti/Tv(transition/transversion) profile with brain cancer in which C>T transitions are more frequently while lung cancer has more C>A transversion. We also found the most important tumor onset and metastasis pathways such as chronic myeloid leukemia, ErbB signaling pathway and glioma pathway. Finally, we identified a significant survival associated mutation gene *ERF* in both TCGA (P=0.01) and our dataset (P=0.012). **Conclusion:** In summary, we conducted a pairwise lung-brain metastasis based exome-wide sequencing and identified some novel metastasis related mutations which provided potential biomarkers for prognosis and targeted therapeutics.

Background

Lung cancer(LC) is the leading cause of cancer death in both men and women, accounting for one-quarter of all cancer deaths(1). The five-year survival rate has failed to improve significantly over the last 30 years and remains a mere 19%, due to recurrence and metastasis. Metastasis accounts for about 90% of cancer-related deaths and is the inevitable outcome of most human tumors. The most common site of lung cancer metastasis the brain, and about 50% of all lung cancers develop into brain metastasis (BM) during the process of the disease (2, 3). It has been reported that the rate of brain metastasis from lung cancer has recently increased, placing a great burden on public health services.

'Seed-and-soil' hypothesis, which is the most widely accepted hypothesis for the formation of metastasis, denotes that the growth of metastatic cancer cells depends on the intrinsic abilities of the cancer cells themselves ('seeds') and the target organ microenvironment ('soil')(4). The cancer cell population has multiple, genetically heterogeneous subpopulations(5). Metastasis is a Darwinian natural selection process in which cancer cells(seeds) with distinct metastatic traits that enable them to obtain metastatic advantage are selected from a genetically- and epigenetically-heterogeneous tumor cell subpopulations (6, 7). The advantageous lung cancer cells('seeds') proliferate in brain('soil') that provide a congenial ground and form metastatic brain tumors, whose genetic landscape is reshaped(8).

A large number of studies has attempted to predict high rate for BM in lung cancer, and it has been found that the factors include young age (<60 years) (9, 10), non-squamous cell carcinoma (10, 11) and the presence of clinical bulky mediastinal lymph nodes (≥ 2 cm)(9) are associated with a high BM rate. However, other studies reported conflicting results (12, 13). Many candidate metastasis genes also have been found to be involved in metastasis through changes in gene expression levels(14). The expression levels of E-cadherin, N-cadherin, KIFC1, and FALZ may be used to identify patients at high risk of lung-brain metastasis (15, 16). However, the molecular basis of metastatic gene expression remains largely unknown, and the genetic profiles of brain metastases from Lung cancer might give us a closely insight into tumour initial, dissemination and local progression (14).

To reveal the molecular mechanisms and the genetic alterations involved in metastasis from lung tumors to the brain, we carried out whole-exome sequencing (WES) of the primary tumors and the corresponding brain metastases from 12 patients with metastatic non-small-cell lung carcinoma. Our study can be instrumental for the identification of new genetic targets which may provide new therapeutic strategies for the design of drug intervention to improve the severity of the disease.

Materials And Methods

Patients and specimens

The pairwise lung-brain tumor samples and adjacent histologically normal tissue samples from 12 patients were collected in Sun Yat-sen memorial hospital from 2010 to 2015. The Ethics Committee at Sun Yat-sen memorial hospital approved the utilization of samples, and all patients signed the informed consent form. 12 paired samples were subjected to HE staining, then isolated the normal cells from the tumor cells through histopathological examination. The normal cells were regard as matched normal controls. The total amount of DNA extracted from the archived formalin-fixed paraffin-embedded (FFPE) samples of tumor tissue was up to standard and qualified.

Next-generation Sequencing, Variant Calling And Annotation

The genomic DNAs were exacted and sonicated to an average size of 200 bp. The targeted DNA fragments were captured pulldown and exon-wide libraries were created using the Roche SeqCap EZ Exome V3 and TruePrep DNA Library Prep Kit V2 for Illumina (#TD501, Vazyme, Nanjing, China), and paired-end sequence data were generated using Illumina HiSeq machines. The sequence data were aligned to the human reference genome (NCBI build 37) using BWA and sorted and removed PCR duplication using GATK 4.1.2.0 (17). Somatic mutation calling was performed using Mutect1, Mutect2 (17) and VarDict (18). Somatic mutations existing in at least two of the results of the three software were selected as high confident mutations and be involved in the further bioinformatics and bio-statistical analysis. Copy number variants (CNVs) from whole-exome sequencing data were detected by CNVKIT(19). The dN/dS ratios of each domain were calculated via The DiversiTools according to Xia

described(20). The GISTIC2.0 was used to identify regions of the genome that are significantly amplified or deleted across a set of samples(21). Somatic variants were annotated by Ensembl Variant Effect Predictor (22). Transition (Ti) and transversion (Tv) ratio were applied to measure the selection in cancer genomes and to show mutation characteristics between different cancer types.

Spectrum, Signatures Of Somatic Mutations And Subclonal Architecture Inference

Non-negative matrix factorization and model-selection were applied to delineate mutational processes underlying genome-wide SNVs and to identify the major mutational signatures(23). In the cohort, one sample was considered as a strong association with one mutational signature if the proportion of the contribution > 20% with MutationalPatterns (version 1.10) (24) and deconstructSigs (25). The sub-clonal architecture of tumors was inferred by sciClone (26) and clonevol(27).

Survival Analysis, Enrichment Analysis And Other Related Statistics

Pathway analysis was based on DAVID bioinformatics(28), and webgestalt (29) with significant Benjamini adjusted p -value ($p < 0.05$). The co-mutation profile was prepared with R package ComplexHeatmap (30). Cox regression was applied for survival analysis between mutation and overall survival time and the K-M plot were used to show the difference between the survival time among different groups. TCGA mutation and survival data were downloaded from the GDC database (<https://portal.gdc.cancer.gov/exploration>). In the validation study to ERF, we download the expression and survival data of ERF mutation and expression from TCGA project, Cox-regression was conducted to binary gene expression data dichotomized by median expression level for TCGA dataset. Since our research subjects in a small sample size, all the clinical related statistical analysis is considered to be significant when $p < 0.05$ without multiple correction test.

Results

Landscape of somatic mutations in primary lung cancer and brain metastasis

We collected and quantified DNA from the original 12 non-small cell lung cancer (NSCLC) patient FFPE samples and the matched brain metastasis samples. The average coverage depth for the tumor cells and normal cells were $194\times$ and $120\times$. Detailed clinicopathological information is summarized in Supplementary Table 1. We identified 1,702 SNVs and 6,131 mutation events in 1,220 genes from 12 paired lung cancer (LC) and brain metastasis (BM), including LC most frequently driver gene mutations such as *TP53*, *EGFR*, *BRCA1*, *BRCA2*, and *BRAF*. We identified a mean of 3.1 driver gene mutation events

per tumor with the dN/dS of 2.13 which is slightly higher than non-metastasis lung cancer samples in The Cancer Genome Atlas (TCGA), indicating a significant enrichment for the cancer driver gene mutations. We did not find any difference of the dN/dS ratio between primary tumor (dN/dS = 2.20), brain metastasis tumor (dN/dS = 2.06) and shared mutations between lung cancer and brain (dN/dS = 2.25).

We found that more somatic mutations in BM lesions (median 71, range 23–180) than in LC lesions (median 48.5, range 13–187), while the difference was not statistically significant ($p = 0.069$, Student's t test) (Supplementary Fig. 1B). High correlation between TMB(tumor mutation burden) of LC and TMB of BM were confirmed by Pearson coefficient 0.65 ($p = 0.02$) (Supplementary Fig. 1C,1D), indicating the TMB of BM can be estimated by that of primary lung cancer when brain tissue is not available, so as to screen patients who will most likely benefit from PD-L1 immunotherapy. 18.2% (0.5–35.9%) of all SNVs were shared between LC and BM, clearly suggesting a common ancestral truncal clone with 30.0% (9.3–60.8%) LC-specific and 51.8% (19.9–79.7%) BM-specific, respectively (Fig. 1B). Although metastases had more private SNVs than the primary tumor, they were not enriched for the pan-cancer driver genes (31) (Fig. 1C). It suggested that few additional private genomic driver genes were required for metastasis when the primary cancer is already advanced.

We identified several lung cancer metastases associated genes (*KMT2C*, *AHNAK2*, *PDE4DIP*, *ANKRD36C*, and *BAGE2*), and the mutations of these genes showed distribution diversity among the LC and BM samples (Fig. 1A and 1D). *KMT2C* mutations were found in 25% samples in LC, however, the mutation frequency in BM was up to 50%, indicating the positive selection of *KMT2C* mutations during metastasis. *AHNAK2* have significant enrichment in LC according to TCGA dataset with mutation ratio of 18.8% in lung cancer while 9.98% in Pan-cancer ($p = 7.2 \times 10^{-9}$, Chi-square test). However, mutation frequency of *AHNAK2* in our dataset is as high as 26.9% which is 1.43 times of LC population in TCGA dataset ($p = 0.02$, Chi-square test). We also observed all *EGFR* mutations were shared between LC and BM, suggesting that *EGFR* mutations are drivers and likely to be an early event before BM.

In order to provide more landscape for the mutations identified in our study, we conducted a pathway analysis to the most frequently mutation genes (mutation frequency > 5%) (Supplementary table2). We found the most important tumor pathways were chronic myeloid leukemia ($p = 0.002$), ErbB signaling pathway ($p = 0.0014$) and glioma pathway($p = 0.05$). Keyword enrichment indicates important metabolic abnormal for the lung-metastasis cancers including EGF-like domain and tyrosine-specific phosphatase (Supplementary table3).

Copy Number Variations

To further explore BM-related molecular events, genomic copy number variations (CNV) were analyzed: 8q21.2, 6p22.1, 12p13.33, 5q35.3 were the most common chromosome deleted regions in both LC and BM, and 8q24.13 were the most commonly regions with gain copy numbers in both LC and BM (Supplementary Fig. 2A). Loss of 6p22.1, which harbors *HLA-A*, *HLA-G* and *HLA-H*, was most frequent in both LC and BM of patient 9 and in BM of patient 11 (Supplementary Fig. 2B). Interestingly, these

samples also had high TMB (Supplementary Fig. 1C). This may be due to the loss of *HLA* function associated with higher overall mutation burden and a larger fraction of HLA-binding neoantigens(32). The recurrent deletion of *HLA* was detected as the early events, indicating the important role of the immune system in LCBM, and these patients may benefit from immunotherapy.

The differences include gains of chromosomes 7q35 and loss of 7q22.1, 7q36.3, which were more frequent in metastasis samples, and gains of chromosomes 11q13.2 and losses of 7q11.23, 2q13 which were less frequent. Most recurrent CNV regions were shared in LC and BM samples, indicating that CNVs are early molecular events in tumorigenesis and metastasis.

Spectrum And Signatures

To determine the relationships between the mutational spectra and tumor organ sites, we analyzed the spectra of LC, BM from our study and primary lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), low grade glioma (LGG), glioblastoma multiforme (GBM) from TCGA dataset. C > T was the most common base substitutions in our LC and BM samples, which was much closer to primary brain cancer (LGG and GBM) but significantly different from primary lung cancer (LUAD and LUSC) which has higher C > A transversion (Fig. 2A). These evidences consistent with our hypothesis that the mutations identified in our study have higher probability to be associated with brain metastasis. The mutational spectrum of LC and BM samples from the same individuals are more similar to each other than that from different patients, implying that different mutational processes were involved during the development of metastasis between the different patients (Fig. 2B).

We further analyzed mutational signatures in BM and LC and signatures 1, 3 and 4, which have been linked to aging, BRCA1/2 mutations and smoking, respectively, were identified as dominant in either BM or LC samples (33) (Fig. 2C). There was no significant difference between the signature levels of LC and BM tissues(Wilcoxon rank sum test, $p \geq 0.05$), indicating the change of mutational signatures happened before metastasis and may not lead to their great difference between the two groups.

Clonal Evolution During The Development Of Lcbm

Phylogenetic trees give clear overviews of the order of mutation events, allowing the track of emergence and movement of clones from LC to BM (Fig. 3). Phylogenetic trees of the 9 patients showed that mutations on the trunk were probably earlier genetic alteration events, followed by those on the branch mutations occurred later during tumorigenesis and BM development. Clonal evolution analyses revealed that LC tumors and BM tumors had the same evolutionary process in three patients (P02, P07 and P10), LC tumors harbor a cluster of LC-private clones in other three patients (P05 and P08) and BM tumors harbor clones that are nonexistent in matched LC tumors in other three patients (P06, P13 and P15), indicating the mutations on BM-private clones may contribute to metastatic progression.

Overall Survival By Genotype

In order to identify independent prediction factor for outcomes, we conducted survival analysis to several potential factors. Patients with the aberrations of 18 genes in LC (Supplementary Fig. 3) and 15 genes in BM (Supplementary Fig. 4) had significantly worse OS than those without these aberrations ($p < 0.05$). Of these genes, we identified a significant survival associated mutation gene *ERF* which was confirmed by both TCGA ($p = 0.01$) (Fig. 4A) and our dataset ($p = 0.012$) (Fig. 4C). What's more, in order to show the prognostic roles of *ERF*, we also found high expression of *ERF* genes in TCGA is a significant risk factors for the overall survival time (HR = 1.46, $p < 1.2 \times 10^{-22}$, Fig. 4B). Taken together, our findings reveal an important role for *ERF* in prognostic prediction of lung cancer.

Multivariate analysis demonstrated gender ($p = 2.02 \times 10^{-119}$), smoking status ($p = 1.21 \times 10^{-269}$), metastatic tumor size ($p = 0$), and the ratio of shared mutations in lung and brain cancers ($p = 0.019$) were significant associate with overall survival time while no significant association were found in drinking status ($p = 0.996$), the number of metastatic tumors ($p = 0.746$), the mutation numbers of primary tumor ($p = 0.840$) or metastatic tumor ($p = 0.248$) (Fig. 4D).

Discussion

The metastatic cascade involves multiple steps, including invasion, entry into the circulation from the primary tumor, systemic dissemination, arrest and extravasation in secondary organs, settlement into latency, reactivation, outgrowth, and potential seeding of tertiary metastasis (34). Genetic and epigenetic changes accumulating among primary tumor cells and metastases may contribute to these multiple steps of metastatic cascade (35, 36). Hence, it is necessary to collect a well-defined cohort of matched primary tumors and BM to perform comparative deep sequencing analyses to acquire some biomarkers of metastasis. However, most patients with brain metastasis of lung cancer (LCBM), which were in the late stage at diagnosis, typically are treated with palliative approaches such as chemotherapy, targeted therapy and whole-brain radiotherapy instead of neurosurgical resection. Thus rarely have researchers have the opportunity to investigate matched primary-metastatic tumors in the mutation status of analyzed genes between tumor sites on a large-scale(37).

In this study, we collected 12 paired lung cancer and brain metastasis samples fortunately and identified some genes associated with LCBM. *KMT2C* (lysine-specific methyltransferase 2C, also known as MLL3), which belongs to mixed-lineage leukemia (MLL) family of histone methyltransferases, was the most commonly mutated gene in our samples. Recent studies have revealed frequent mutations of *KMT2C* in several epithelia and myeloid cells, and it was identified as a putative tumor suppressor (38, 39). Metastatic spread has been proposed that circulating tumor cell (CTC) populations in the blood of carcinoma patients contain cells with the clonal capacity to initiate metastatic growth in distant organs, thus having similarity with the hematologic tumor. *KMT2C* originally identified as oncogenic fusions in leukemia (40), and the most enriched pathway is chronic myeloid leukemia in our study. Hashim *et al.* reported that the *Keap1-Nrf2*-ARE pathway is mutated in NSCLC patients that metastasised to the brain

and in CTC according to WGS(41). Recent reports demonstrated that metastatic brain tumors of NSCLC and lung adenocarcinoma patients frequently carried EGFR mutations than those who without(42, 43). Our results further emphasized the association between the brain metastasis of lung cancer and leukemia. *AHNAK2*(AHNAK nucleoprotein 2) is a prognostic marker and an oncogenic protein for clear cell renal cell carcinoma and hypoxic upregulation of *AHNAK2* support EMT (epithelial-mesenchymal transition) and cancer cell stemness (44). Cancer cells acquire characteristics of self-renewal, motility, and invasiveness, traits that facilitate metastatic dissemination during EMT(45). That is, the driver gene mutation of *AHNAK2* may promote metastatic colonization of the lung to brain by supporting EMT.

Our finding depicted that chronic myeloid leukemia and ErbB signaling pathway were mutated in the majority of LCBM patients supports our hypothesis that mutations in these pathways may indeed provide a survival advantage to these cells and help them reach distant sites. Of note, Glioma pathways have also been identified. Ti/Tv profile showed our mutation profile is much closer to brain cancer mutation profile since brain cancer has high C > T transitions are more frequently while lung cancer has a higher frequency of C > A transversion. These evidences highly indicate that brain tumor-related events are involved in the process of LCBM.

Multivariate analysis demonstrated the high ratio of shared mutations were associated with better prognosis(46, 47). This implies the patients might have a preferable prognosis when the tendency of metastatic cancer mutation was more inclined to primary cancer, that is the BM sample might not have evolved from the primary cancer but rather they had a shared antecedent. Limited inter-tumor heterogeneity between LC and BM within the same patients results in effective postoperative chemotherapy and radiotherapy in this situation.

Overall, we revealed the genomic difference between metastatic and primary tissues, and identified several genes associated with LCBM. While further molecular biology studies to validate the role of identified candidates and a larger-scale LCBM samples would be required to confirm our findings, this study provides a preliminary evidence of the genetic evolution of LC metastases and further gives us an advantage to reveal the therapeutic vulnerabilities of LC metastatic tumor.

Abbreviations

WES
whole- exome sequencing; LC:lung cancer; BM:brain metastasis; LCBM:Brain metastasis of lung cancer; Ti/Tv:transition/transversion; FFPE:formalin-fixed paraffin-embedded; HE staining:hematoxylin-eosin staining ; NSCLC:non-small cell lung cancer; CNV:copy number variations; LUAD:lung adenocarcinoma; LUSC:lung squamous cell carcinoma; LGG:low grade glioma; GBM:glioblastoma multiforme; CTC:circulating tumor cell; EMT:epithelial-mesenchymal transition;

Declarations

Acknowledgements

This work was sponsored by Shanghai Tongshu Biotechnology Co., Ltd

Authors' contributions

Yuefei Deng and Pengcheng Li designed the study and devised the experiments. Zhiwei Zhou and Yutao Huang provided tumor samples and clinical information. Zhenghao Liu, Meiguang Zheng and Bingxi Lei dealt with the data analysis and prepared the main manuscript. Zhenghao Liu, Wenpeng Li and Qinbiao Chen edited figures and searched literature. All authors contributed to the discussions and manuscript preparation.

Funding

None

Availability of data and materials

All the sequencing data was deposited in the NCBI database under the BioProject accession code SRP182103 and PRJNA515561

Ethics approval and consent to participate

This study protocol was approved by the Ethics Committee of Sun Yat-sen memorial hospital and Tongji Medical college.

Consent for publication

Not applicable.

Competing interests

The authors declare no conflict of interest.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* 2019;69(1):7-34.
2. Mujoomdar A, Austin JH, Malhotra R, Powell CA, Pearson GD, Shiau MC, et al. Clinical predictors of metastatic disease to the brain from non-small cell lung carcinoma: primary tumor size, cell type, and lymph node metastases. *Radiology.* 2007;242(3):882-8.
3. Guo S, Yan F, Xu J, Bao Y, Zhu J, Wang X, et al. Identification and validation of the methylation biomarkers of non-small cell lung cancer (NSCLC). *Clin Epigenetics.* 2015;7:3.
4. Langley RR, Fidler IJ. The seed and soil hypothesis revisited—the role of tumor-stroma interactions in metastasis to different organs. *Int J Cancer.* 2011;128(11):2527-35.

5. Landau DA, Carter SL, Getz G, Wu CJ. Clonal evolution in hematological malignancies and therapeutic implications. *Leukemia*. 2014;28(1):34-43.
6. Greaves M, Maley CC. Clonal evolution in cancer. *Nature*. 2012;481(7381):306-13.
7. Naxerova K, Jain RK. Using tumour phylogenetics to identify the roots of metastasis in humans. *Nat Rev Clin Oncol*. 2015;12(5):258-72.
8. Yousefi M, Bahrami T, Salmaninejad A, Nosrati R, Ghaffari P, Ghaffari SH. Lung cancer-associated brain metastasis: Molecular mechanisms and therapeutic options. *Cell Oncol (Dordr)*. 2017;40(5):419-41.
9. Ceresoli GL, Reni M, Chiesa G, Carretta A, Schipani S, Passoni P, et al. Brain metastases in locally advanced nonsmall cell lung carcinoma after multimodality treatment: risk factors analysis. *Cancer*. 2002;95(3):605-12.
10. Ji Z, Bi N, Wang J, Hui Z, Xiao Z, Feng Q, et al. Risk factors for brain metastases in locally advanced non-small cell lung cancer with definitive chest radiation. *Int J Radiat Oncol Biol Phys*. 2014;89(2):330-7.
11. Wang SY, Ye X, Ou W, Lin YB, Zhang BB, Yang H. Risk of cerebral metastases for postoperative locally advanced non-small-cell lung cancer. *Lung Cancer*. 2009;64(2):238-43.
12. Cox JD, Scott CB, Byhardt RW, Emami B, Russell AH, Fu KK, et al. Addition of chemotherapy to radiation therapy alters failure patterns by cell type within non-small cell carcinoma of lung (NSCCL): analysis of radiation therapy oncology group (RTOG) trials. *Int J Radiat Oncol Biol Phys*. 1999;43(3):505-9.
13. Robnett TJ, Machtay M, Stevenson JP, Algazy KM, Hahn SM. Factors affecting the risk of brain metastases after definitive chemoradiation for locally advanced non-small-cell lung carcinoma. *J Clin Oncol*. 2001;19(5):1344-9.
14. Nguyen DX, Massague J. Genetic determinants of cancer metastasis. *Nat Rev Genet*. 2007;8(5):341-52.
15. JY Y, SH Y, JE L, DG C, HK K, SH K, et al. E-cadherin as a predictive marker of brain metastasis in non-small-cell lung cancer, and its regulation by pioglitazone in a preclinical model. *Journal of neuro-oncology*. 2012;109(2):219-27.
16. H G-R, E O, M P, J S, P Y, M H, et al. The expression of three genes in primary non-small cell lung cancer is associated with metastatic spread to the brain. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2009;15(5):1755-61.
17. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-303.
18. Lai Z, Markovets A, Ahdesmaki M, Chapman B, Hofmann O, McEwen R, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res*. 2016;44(11):e108.

19. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol.* 2016;12(4):e1004873.
20. Xia H, Beck AS, Gargili A, Forrester N, Barrett AD, Bente DA. Transstadial Transmission and Long-term Association of Crimean-Congo Hemorrhagic Fever Virus in Ticks Shapes Genome Plasticity. *Sci Rep.* 2016;6:35819.
21. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhir R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 2011;12(4):R41.
22. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 2016;17(1):122.
23. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* 2013;3(1):246-59.
24. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med.* 2018;10(1):33.
25. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* 2016;17:31.
26. Miller CA, White BS, Dees ND, Griffith M, Welch JS, Griffith OL, et al. SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. *PLoS Comput Biol.* 2014;10(8):e1003665.
27. Dang HX, White BS, Foltz SM, Miller CA, Luo J, Fields RC, et al. ClonEvol: clonal ordering and visualization in cancer sequencing. *Ann Oncol.* 2017;28(12):3076-82.
28. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4(1):44-57.
29. Wang J, Vasaike S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* 2017;45(W1):W130-W7.
30. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* 2016;32(18):2847-9.
31. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell.* 2018;174(4):1034-5.
32. Castro A, Ozturk K, Pyke RM, Xian S, Zanetti M, Carter H. Elevated neoantigen levels in tumors with somatic mutations in the HLA-A, HLA-B, HLA-C and B2M genes. *BMC Med Genomics.* 2019;12(Suppl 6):107.
33. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature.* 2013;500(7463):415-21.

34. Paterlini-Brechot P. Organ-specific markers in circulating tumor cell screening: an early indicator of metastasis-capable malignancy. *Future Oncol.* 2011;7(7):849-71.
35. Popper HH. Progression and metastasis of lung cancer. *Cancer Metastasis Rev.* 2016;35(1):75-91.
36. Turajlic S, Swanton C. Metastasis as an evolutionary process. *Science.* 2016;352(6282):169-75.
37. M P, AS B, R K, CC Z, JA H, S L-R, et al. Spectrum of gene mutations detected by next generation exome sequencing in brain metastases of lung adenocarcinoma. *European journal of cancer (Oxford, England : 1990).* 2015;51(13):1803-11.
38. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, et al. Mutational landscape and significance across 12 major cancer types. *Nature.* 2013;502(7471):333-9.
39. Chen C, Liu Y, Rappaport AR, Kitzing T, Schultz N, Zhao Z, et al. MLL3 is a haploinsufficient 7q tumor suppressor in acute myeloid leukemia. *Cancer Cell.* 2014;25(5):652-65.
40. Ziemer-van der Poel S, McCabe NR, Gill HJ, Espinosa R, 3rd, Patel Y, Harden A, et al. Identification of a gene, MLL, that spans the breakpoint in 11q23 translocations associated with human leukemias. *Proc Natl Acad Sci U S A.* 1991;88(23):10735-9.
41. Aljohani HM, Aittaleb M, Furgason JM, Amaya P, Deeb A, Chalmers JJ, et al. Genetic mutations associated with lung cancer metastasis to the brain. *Mutagenesis.* 2018;33(2):137-45.
42. Matsumoto S, Takahashi K, Iwakawa R, Matsuno Y, Nakanishi Y, Kohno T, et al. Frequent EGFR mutations in brain metastases of lung adenocarcinoma. *Int J Cancer.* 2006;119(6):1491-4.
43. Liao L, Ji X, Ge M, Zhan Q, Huang R, Liang X, et al. Characterization of genetic alterations in brain metastases from non-small cell lung cancer. *FEBS Open Bio.* 2018;8(9):1544-52.
44. M W, X L, J Z, Q Y, W C, W J, et al. AHNAK2 is a Novel Prognostic Marker and Oncogenic Protein for Clear Cell Renal Cell Carcinoma. *Theranostics.* 2017;7(5):1100-13.
45. L W, K P, Y K. Tumor metastasis: moving new biological insights into the clinic. *Nature medicine.* 2013;19(11):1450-64.
46. Gaidzik VI, Teleanu V, Papaemmanuil E, Weber D, Paschka P, Hahn J, et al. RUNX1 mutations in acute myeloid leukemia are associated with distinct clinico-pathologic and genetic features. *Leukemia.* 2016;30(11):2160-8.
47. Zhong Q, Peng HL, Zhao X, Zhang L, Hwang WT. Effects of BRCA1- and BRCA2-related mutations on ovarian and breast cancer survival: a meta-analysis. *Clin Cancer Res.* 2015;21(1):211-20.

Figures

Figure 1

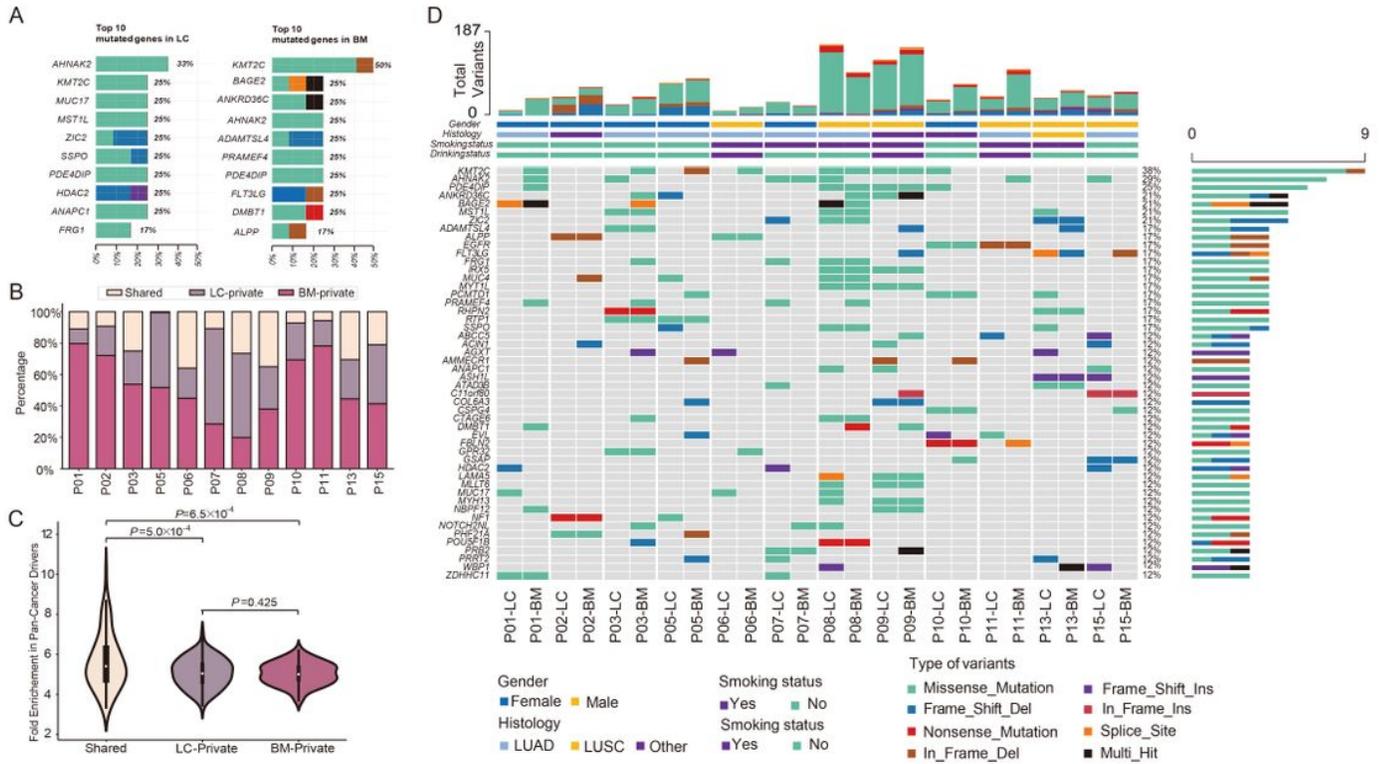


Figure 1

The mutational landscape and patterns of genetic divergence in paired primary lung cancer and brain metastases. (A) Top mutated genes in lung cancer and brain cancer samples. (B) Tumor specific mutations and lung-brain shared mutations. (C) Drive mutation genes in tumor specific mutations and shared mutations. (D) OncoPrint of lung and brain mutations in all the samples.

Figure 2

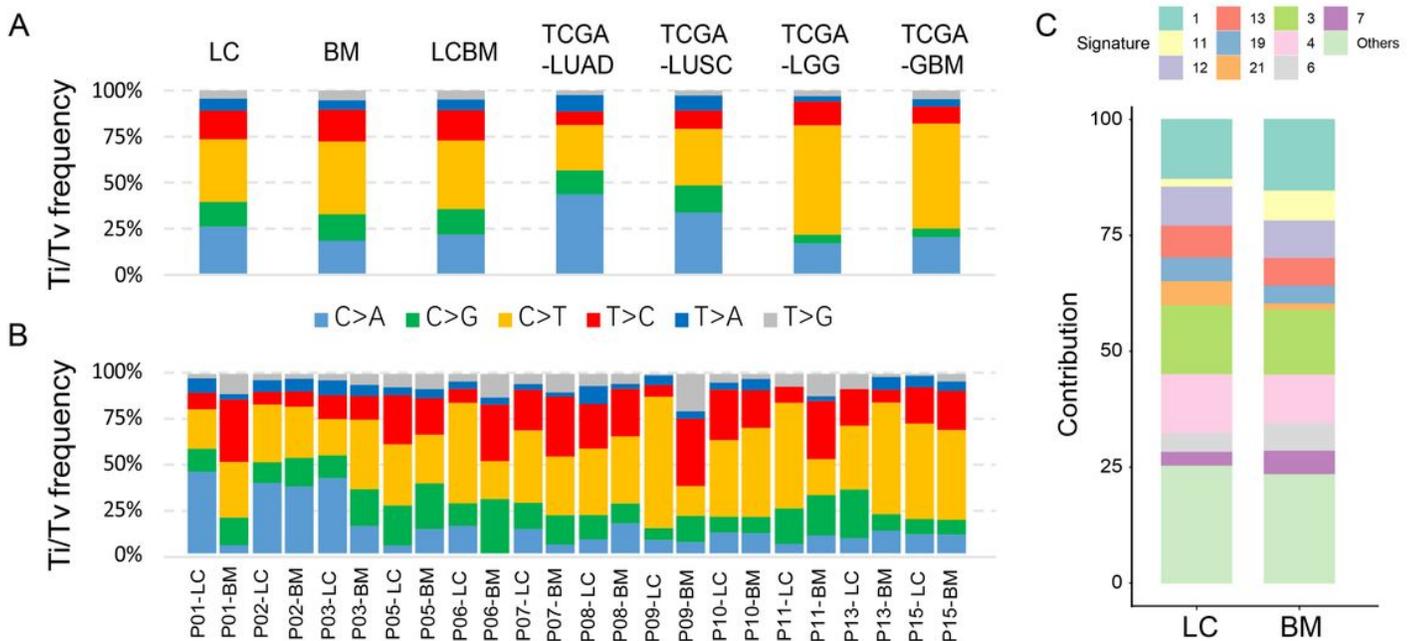


Figure 2

Spectrum and signatures. (A) Mutation spectrum of lung tumors specific mutations, brain tumors specific mutations, shared mutations and TCGA lung and brain mutations. (B) Mutation spectrum of all the samples in our cohort. (C) Mutation signatures of mutations for lung tumor samples and brain tumor samples.

Figure 3

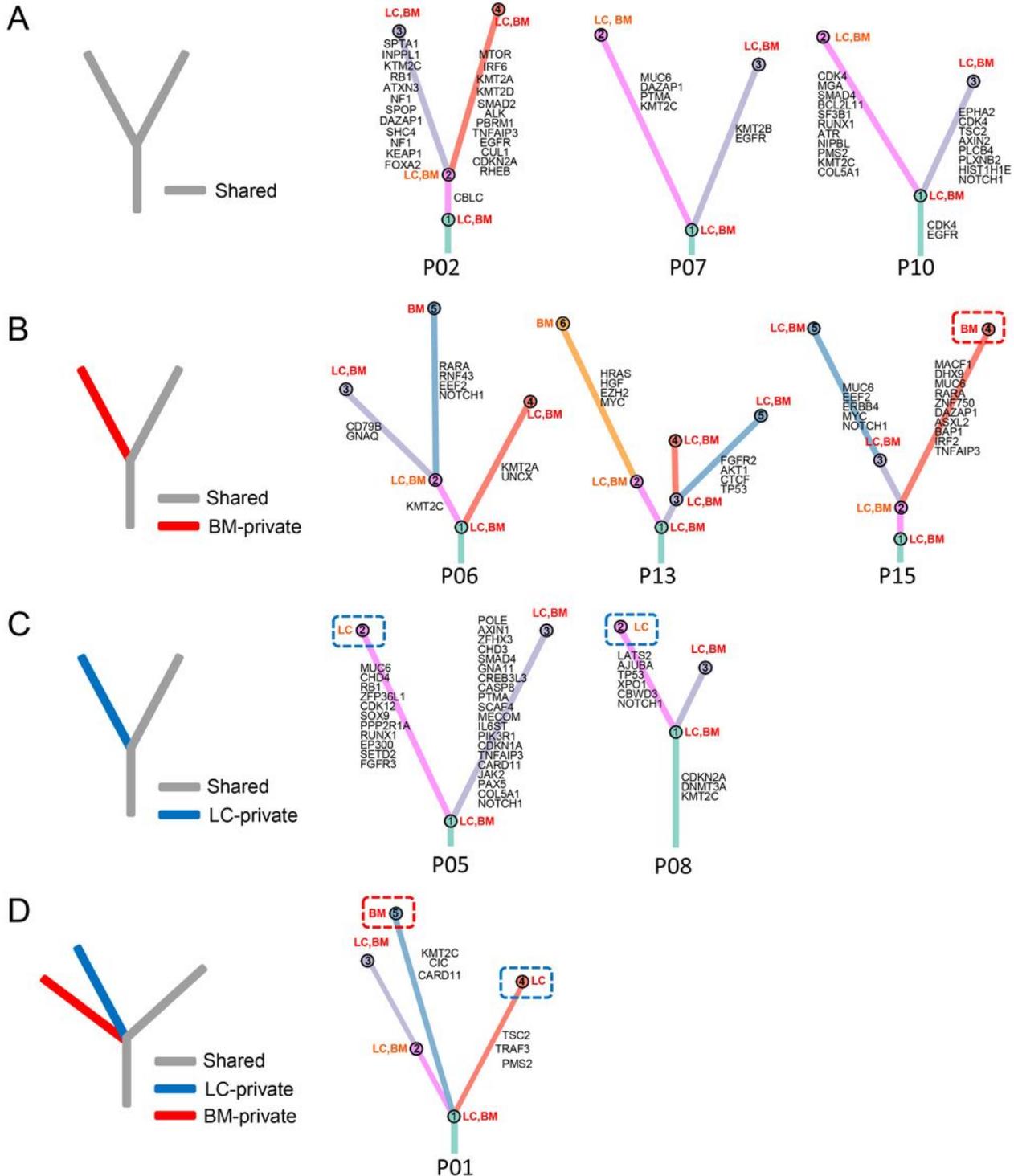


Figure 3

Phylogenetic trees indicating four types of evolution (A) LC tumors and BM tumors had the same evolutionary process. (B) BM tumors harbor a cluster of BM-private clones. (C) LC tumors harbor a cluster of LC-private clones. (D) LC and BM tumors harbor a cluster of LC-private clones and a cluster of BM-private clones respectively.

Figure 4

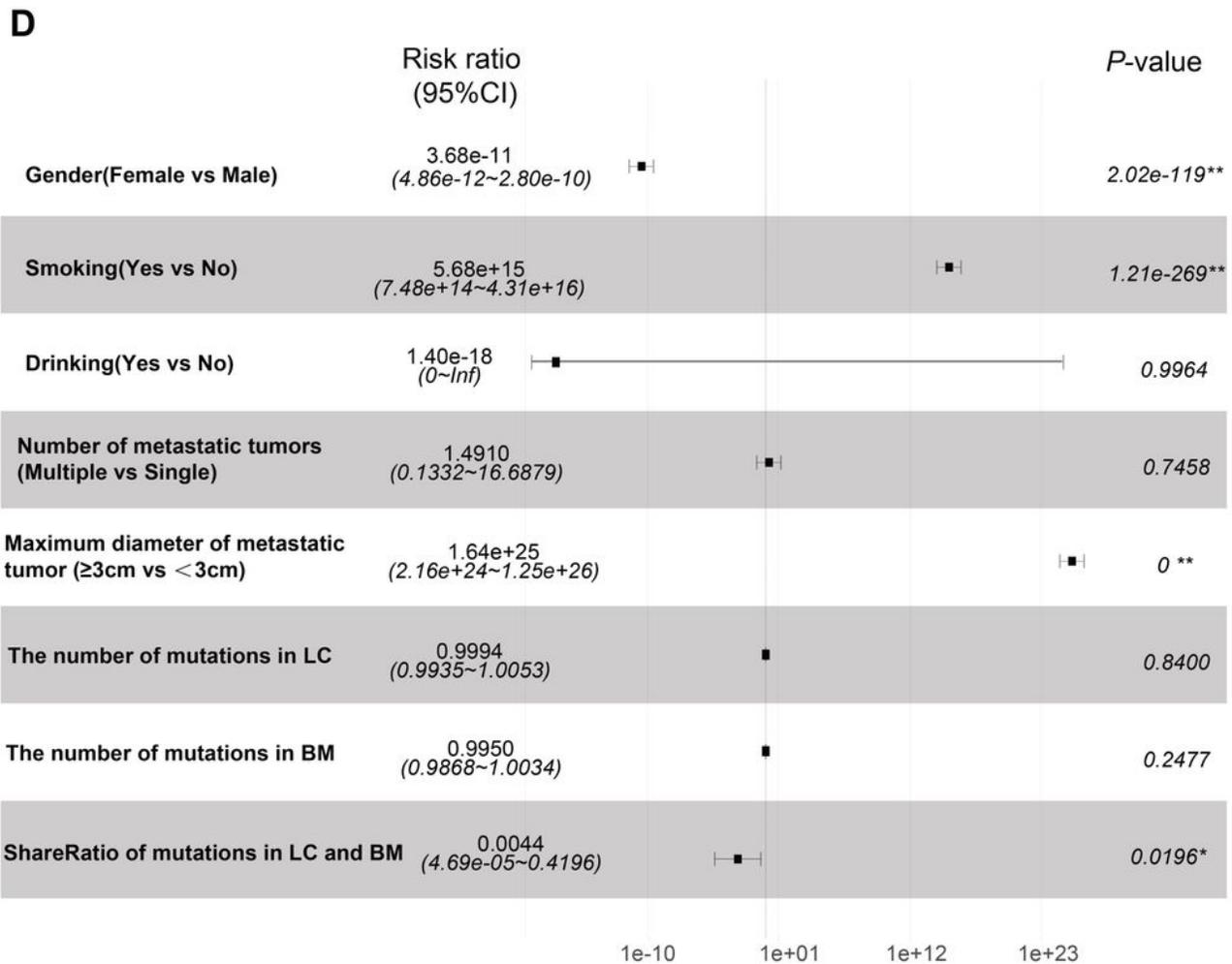
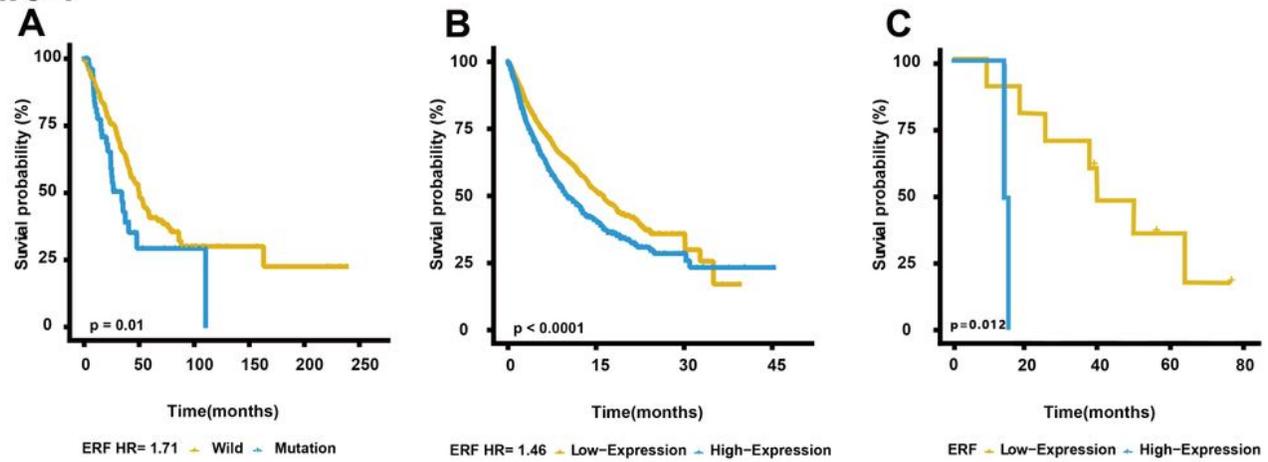


Figure 4

Overall survival analysis (A) Survival analysis to several potential factors (B) Survival analysis between ERF mutation and overall survival times in TCGA. (C) Survival analysis between ERF expression and overall survival times in TCGA. (D) Survival analysis between ERF mutation and overall survival times in our cohort.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementarytable1.xlsx](#)
- [Supplementarytable2.xlsx](#)
- [Supplementarytable3.xlsx](#)
- [supplementaryfigure1.tif](#)
- [supplementaryfigure2.tif](#)
- [supplementaryfigure3.tif](#)
- [supplementaryfigure4.tif](#)