**Whole-exome sequencing identifies somatic mutations associated with lung cancer metastasis to the brain**

**Abstract**

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 lung cancer metastasis prediction and outcome improvement. In this study, we applied whole-exome sequencing to identify potential metastasis related mutation in paired lung cancer and brain metastasis samples. We identified 6,318 SNVs and 56,686 mutation events in 3863 genes 13 paired lung cancer and brain metastasis samples including 46 lung cancer most frequently driven gene mutations. Furthermore, we identified several lung cancer metastases associated genes (BAGE2 and AHNAK2) and epigenetic factors (miR-4436A, miR-6077). We identified a mean of 3.1 driver mutation events per tumor with the dN/dS of 2.06 (95%CI: 1.73-2.4) indicating a significant enrichment for the cancer driven mutations. Mutation spectrum analysis found lung-brain metastasis samples have more similar transition (Ti) and transversion (Tv) profile with brain cancers in which C->T transitions is more frequently while lung cancer has higher frequency of C->A transversion. We also found the most important tumor onset and metastasis pathways such as focal adhesion, PI3K-Akt signaling pathway, MAPK signaling pathway. What’s more, Glioma pathway were also identified which highly indicating the solid finding of the study. Finally, we also identified a significant survival associated mutation gene ERF which was confirmed by both TCGA (P=0.02) and our dataset (P=0.007). In summary, we conducted a pairwise lung-brain metastasis based exome-wide sequencing and identified some novel cancer and metastasis related mutation which provided potential biomarkers for prognosis and targeted therapeutics.

**Key words:** whole-exome sequencing, lung cancer, somatic mutation, brain metastasis

**Introduction**

Lung cancer 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%, which is much lower than that of other common cancers, due to recurrence and metastasis[1]. Metastasis accounts for about 90% of cancer-related deaths and is the inevitable outcome of most human tumors. Different primary tumors tend to metastasize to distinct organs. The most common site of lung cancer metastasis is brain, and about 50% of all lung cancers develop into brain metastasis (BM) during the process of the disease [2, 3]. Conversely, more than half of BMs originate from lung cancer [4]. Moreover, it has been reported that the rate of brain metastasis from lung cancer has recently increased [5], 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’)[6]. The cancer cell population has multiple, genetically heterogeneous subpopulations[7]. 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[8, 9]. 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[10].

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)[11, 12], non-squamous cell carcinoma[12, 13] and the presence of clinical bulky mediastinal lymph nodes (≥2 cm)[11] are associated with a high BM rate. However, other studies reported conflicting results[14, 15]. Many candidate metastasis genes also have been found to be involved in metastasis through changes in gene expression levels[16]. The expression levels of E-cadherin, N- cadherin, KIFC1, and FALZ may be used to identify patients at high risk of lung-brain metastasis [17, 18]. Although the molecular basis of metastatic gene expression is largely unknown, it can sometimes be traced back to tumor-initiating mutations[16].While the process of metastasis is the least well understood on the alterations of genetic landscape. Whereas the role of driver genes ,such as EGFR, ALK, KRAS, BRAF and HER-2 has been identified between progressive tumor stages [19-21], their genetic alterations that would specifically mediate BM remains controversial [22-29].

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 13 patients with metastatic non-small-cell lung carcinoma. Our study may 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.

**Methods**

**Patients and specimens**

The pairwise lung-brain tumor samples and adjacent histologically normal tissue samples from 13 patients were collected in\*\*\*\*hospital from 2010 to 2015. The Ethics Committee at \*\*\*\* approved the utilization of samples, and all patients signed the informed consent form. 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 200bp (range 100-500bp). 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 was generated using Illumina HiSeq machines. The sequence data, aligned to the human reference genome (NCBI build 37) using BWA [30], and sorted and removed PCR duplication using GATK 4.0 [31]. Somatic mutation calling was performed using Mutect1, Mutect2 [31] and VarDict [32]. 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. Transition (Ti) and transversion (Tv) ratio was applied to measure the selection in cancer genomes and to show mutation characteristics between different cancer types.

**Enriched pathway of frequently mutated genes**

Pathway analysis were based on DAVID bioinformatics[33] and webgestalt [34] with significant Benjamini adjusted *P*-value (P<0.05). All the statistical analysis was based on R-3.5.2. Co-Mutation profile was prepared with R package ComplexHeatmap[35]. Cox regression were applied for survival analysis between mutation and overall survival time and 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. Since our aim are validation study design and small sample size, all the clinical related statistical analysis is considered to be significant when P<0.05 without multiple correction test.

**Results**

We collected and quantified genetic material from the original 13 non-small cell lung cancer (NSCLC) patient FFPE samples. Detailed clinicopathological information of the 13 NSCLC patients with brain metastasis (8 adenocarcinomas, 1 squamous cell carcinomas, 2 adeno-squamous carcinoma, 1 large cell carcinoma and 1 clear cell carcinoma + tubular adenocarcinoma) are summarized in Supplementary Table 1. The median age of the NSCLC patients at the initial diagnosis of the primary cancers was 59 years old (range from 42 to 79), with a male percentage of 53.8%. Ten (76.9%) of these patients died of cancer-related causes during the course of the study. Nine pairs (69.2%) were obtained from cases with synchronous disease at their initial diagnosis, and 4 pairs (30.8%) were obtained from patients with metachronous disease who had BM 3 or more months after the initial diagnosis. The median lag (range) till BM for synchronous tumors was 0 (0-1 months) while for metachronous tumors the lag was 16 months. In order to identify independent prediction factor for outcomes, we conducted survival analysis to several potential factors. We found tumor size (P=0.007) was the most significant association factors with overall survival time while no significant association were found in gender (P=0.259), smoking patterns (P=0.134), drinking patterns (P=0.224), the mutation numbers of primary tumor (P=0.14) or the metastatic tumor (P=0.39) (more details see Supplementary Figure 1).

We identified 6,318 SNVs and 56,686 mutation events in 3,863 genes from 13 paired lung cancer and brain metastasis samples including 46 lung cancer most frequently driven gene mutations such as TP53, EGFR, BRCA1, BRCA2 and BRAF (Supplementary Table 2 and Supplementary Table 3). Furthermore, we identified several lung cancer metastases associated genes (BAGE2 and AHNAK2) and epigenetic factors (miR-4436A, miR-6077). BAGE2 mutation were found in 53.8% samples of our cohort, however, the mutation frequency in TCGA are 10.14% (lung cancer dataset) and 3.2% (Pan-cancer dataset). The significant over-representation (fold-change=5.3, P<5.2×10-4, Chi-square test) suggest BAGE2 might play roles in brain metastasis for lung cancers. AHNAK2 have significant enrichment in lung cancer according to TCGA dataset with mutation ratio of 14.16% in lung cancer while 9.98% in Pan-cancer (P=9.8×10-14, Chi-square test). However, mutation frequency of AHNAK2 in our dataset is as high as 53.8% which is 2.45 times of lung cancer population in TCGA dataset (P=6.5×10-3, Chi-square test).

We identified a mean of 3.1 driver mutation events per tumor with the dN/dS of 2.06 (95%CI: 1.73-2.4) which is slightly higher than non-metastasis lung cancer samples, indicating a significant enrichment for the cancer driven mutations. We did not find any difference of the dN/dS ratio between primary tumor (dN/dS=2.14, 95% CI: 1.65-2.64), brain metastasis tumor (dN/dS=1.99, 95% CI: 1.42-2.56) and shared mutations between lung cancer and brain (dN/dS=2.2, 95%CI: 1.39-3.0). Transition (Ti) and transversion (Tv) profile showed our mutation profile is much closer to brain cancer mutation profile since brain cancer have high C->T transitions is more frequently while lung cancer have higher frequency of C->A transversion(Figure 1A) [36]. These evidences consistent with our hypothesis that the mutation identified in our study have higher probability to be associated with brain metastasis.

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%). We found the most important tumor onset and metastasis pathways such as focal adhesion (*P*=0.0018), PI3K-Akt signaling pathway (*P*=0.0176), MAPK signaling pathway (*P*=0.0518). What’s more, Glioma pathway were also identified which highly indicating the solid finding of the study (*P*=0.0572) (Figure 1B). Although *P* values of the two latter are slightly higher than 0.05, they still attract our attention. Keyword enrichment indicates important metabolic abnormal for the lung-metastasis cancers including Alternative splicing, Methylation and EGF-like domain (Figure 1C). The most frequent mutation genes and epigenetic factors of co-mutation profiles are KMT2C, BAGE2, AHNAK2, BAIAP3(Figure 2B), which means these molecular markers might be helpful to define a lung cancer patient at risk of BM. Finally, we also identified a significant survival associated mutation gene ERF which was confirmed by both TCGA (*P*=0.01) and our dataset (*P*=7.0×10-3) (Figure 2C, Figure 2E). What’s more, in order to show the prognostic roles of ERF, we also found high expression of ERF genes is a significant risk factors for the overall survival time (HR=1.46, *P*<1.2×10-22, Figure 2D). Taken together, our findings reveal an important role for ERF in prognostic prediction of lung cancer. What’s more, very interestingly, we found the ratio of shared mutations in lung and brain cancers was associated with better prognosis with marginal significance (*P*=0.095). This implies the patients might live longer 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 in this situation.

**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 [37, 38]. 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 BM 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. So some researchers did not have the opportunity to investigate matched primary-metastatic tumors in the mutation status of analyzed genes between tumor sites[39].

In this study, we collected 13 paired lung cancer and brain metastasis samples fortunately and identified some novel genes associated with LCBM. For example, BAGE (B melanoma antigen) gene family, which is consists of at least 15 loci with up to 99% nucleotide identity among its members[40], was generated by juxtacentromeric reshuffling of the MLL3 gene that occurred during the evolution of hominoids[41]. 14.6% of ovarian cancer tissues and 14.3% in metastatic lesions of ovarian cancer[42], 35% of melanoma cell lines[43], 63% of ovarian carcinomas[44], 6% of rhabdomyosarcomas[45], 21% of hepatocellular carcinomas[46], 8% of breast cancers[47], and 23% of gastric carcinomas [48]express at least one BAGE gene. The sequences at or near the promoters of BAGE genes were hypermethylated in normal tissues and hypomethylated in 98% of analyzed human cancers[40]. Colon cancer could be diagnosed by DNA methylation analysis of the BAGE loci with 94% specificity, 83% sensitivity, and 89% accuracy[49]. The high mutation frequencies of BAGE gene family were found in our LCBM samples. Therefore, in addition to the methylation mechanism, the mutations of BAGE may also lead to its over-expression, leading to the occurrence and metastasis of tumors. 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[50]. During EMT cancer cells acquire characteristics of self-renewal, motility, and invasiveness, traits that facilitate metastatic dissemination[51, 52].That is, the driver mutation of AHNAK2 may promote metastatic colonization of the lung to brain by supporting EMT.

Aside the altered genes, further BM from lung cancer comes from the deregulation of non-coding RNAs (ncRNAs). The role of miRNAs in the development of brain metastases has only recently been established [53].Many mutations were found in miR-4436a and miR-6077 on our lung-brain metastasis samples. CNR1 (cannabinoid receptor 1) and PDX1 were inferred as potential targets of miR-4436a and miR-6077, respectively, via the algorithms of TargetScan 7.2 (http://targetscan.org/). CNR1 is closely associated with a variety of different cell signaling pathways including G-proteins, adenylyl cyclase, PI3K and mitogen-activated protein kinase (MAPK) as a G-protein–coupled receptor [54]. Treatment with the CNR1 antagonist, or loss of the CNR1 gene reduced lung metastasis formation in an alveolar rhabdomyosarcoma metastasis mouse model [55]. PDX-1 (pancreatic and duodenal homeobox 1) induced increased cell proliferation, invasion, and colony formation in vitro, and resulted in markedly increased tumor formation of pancreatic ductal adenocarcinoma cell line (MIA PaCa2) and insulinoma cell lines(βTC6 cells) in xenograft severe combined immunodeficiency mice[56, 57]. PDX1 is also aberrantly over-expressed in other types of cancers, including gastric carcinoma[58], cholangiocarcinoma[59], small-cell neuroendocrine prostate cancer[60].Therefore, the mutation of miR-4436a and miR-6077 might reduce the inhibition of CNR1 and PDX1, respectively, and the increased expression of CNR1 and PDX1 promoted tumorigenesis and metastasis.

Our finding that focal adhesion, pi3k-akt signaling pathway and MAPK signaling pathway related of tumorigenesis and metastasis 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, which highly indicates that brain tumor-related pathway is involved in the process of LCBM.

After all, we identified several genes and epigenetic factors associated with LCBM. While further biological studies to validate the role of identified candidates and testing the predictive power of our mutation signature in larger LCBM samples would be required to confirm our findings, it provides some preliminary evidence that inhibiting these targets may prevent tumor metastases.

**Disclosure of Conflicts of Interest**

The authors declare no conflict of interest.

**Abbreviations**

LCBM: Brain metastasis of lung cancer

**Data Access**

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

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin 2019; 69: 7-34.

2. Gavrilovic IT, Posner JB. Brain metastases: epidemiology and pathophysiology. J Neurooncol 2005; 75: 5-14.

3. Mujoomdar A, Austin JH, Malhotra R 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: 882-888.

4. Budczies J, von Winterfeld M, Klauschen F et al. The landscape of metastatic progression patterns across major human cancers. Oncotarget 2015; 6: 570-583.

5. Fox BD, Cheung VJ, Patel AJ et al. Epidemiology of metastatic brain tumors. Neurosurg Clin N Am 2011; 22: 1-6, v.

6. 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: 2527-2535.

7. Landau DA, Carter SL, Getz G, Wu CJ. Clonal evolution in hematological malignancies and therapeutic implications. Leukemia 2014; 28: 34-43.

8. Greaves M, Maley CC. Clonal evolution in cancer. Nature 2012; 481: 306-313.

9. Naxerova K, Jain RK. Using tumour phylogenetics to identify the roots of metastasis in humans. Nat Rev Clin Oncol 2015; 12: 258-272.

10. Yousefi M, Bahrami T, Salmaninejad A et al. Lung cancer-associated brain metastasis: Molecular mechanisms and therapeutic options. Cell Oncol (Dordr) 2017; 40: 419-441.

11. Ceresoli GL, Reni M, Chiesa G et al. Brain metastases in locally advanced nonsmall cell lung carcinoma after multimodality treatment: risk factors analysis. Cancer 2002; 95: 605-612.

12. Ji Z, Bi N, Wang J 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: 330-337.

13. Wang SY, Ye X, Ou W et al. Risk of cerebral metastases for postoperative locally advanced non-small-cell lung cancer. Lung Cancer 2009; 64: 238-243.

14. Cox JD, Scott CB, Byhardt RW 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: 505-509.

15. Robnett TJ, Machtay M, Stevenson JP et al. Factors affecting the risk of brain metastases after definitive chemoradiation for locally advanced non-small-cell lung carcinoma. J Clin Oncol 2001; 19: 1344-1349.

16. Nguyen DX, Massague J. Genetic determinants of cancer metastasis. Nat Rev Genet 2007; 8: 341-352.

17. JY Y, SH Y, JE L 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: 219-227.

18. H G-R, E O, M P 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: 1755-1761.

19. Ding L, Getz G, Wheeler DA et al. Somatic mutations affect key pathways in lung adenocarcinoma. Nature 2008; 455: 1069-1075.

20. Gautschi O, Milia J, Cabarrou B et al. Targeted Therapy for Patients with BRAF-Mutant Lung Cancer: Results from the European EURAF Cohort. J Thorac Oncol 2015; 10: 1451-1457.

21. Arcila ME, Chaft JE, Nafa K et al. Prevalence, clinicopathologic associations, and molecular spectrum of ERBB2 (HER2) tyrosine kinase mutations in lung adenocarcinomas. Clin Cancer Res 2012; 18: 4910-4918.

22. Doebele RC, Lu X, Sumey C et al. Oncogene status predicts patterns of metastatic spread in treatment-naive nonsmall cell lung cancer. Cancer 2012; 118: 4502-4511.

23. Villalva C, Duranton-Tanneur V, Guilloteau K et al. EGFR, KRAS, BRAF, and HER-2 molecular status in brain metastases from 77 NSCLC patients. Cancer Med 2013; 2: 296-304.

24. Cortot AB, Italiano A, Burel-Vandenbos F et al. KRAS mutation status in primary nonsmall cell lung cancer and matched metastases. Cancer 2010; 116: 2682-2687.

25. Monaco SE, Nikiforova MN, Cieply K et al. A comparison of EGFR and KRAS status in primary lung carcinoma and matched metastases. Hum Pathol 2010; 41: 94-102.

26. Arrieta O, Saavedra-Perez D, Kuri R et al. Brain metastasis development and poor survival associated with carcinoembryonic antigen (CEA) level in advanced non-small cell lung cancer: a prospective analysis. BMC Cancer 2009; 9: 119.

27. Ding X, Zhang Z, Jiang T et al. Clinicopathologic characteristics and outcomes of Chinese patients with non-small-cell lung cancer and BRAF mutation. Cancer Med 2017; 6: 555-562.

28. Couraud S, Vaca-Paniagua F, Villar S et al. Noninvasive diagnosis of actionable mutations by deep sequencing of circulating free DNA in lung cancer from never-smokers: a proof-of-concept study from BioCAST/IFCT-1002. Clin Cancer Res 2014; 20: 4613-4624.

29. Quere G, Descourt R, Robinet G et al. Mutational status of synchronous and metachronous tumor samples in patients with metastatic non-small-cell lung cancer. BMC Cancer 2016; 16: 210.

30. Slater PM, Grivell R, Cyna AM. Labour management of a woman with carnitine palmitoyl transferase type 2 deficiency. Anaesth Intensive Care 2009; 37: 305-308.

31. McKenna A, Hanna M, Banks E et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010; 20: 1297-1303.

32. Lai Z, Markovets A, Ahdesmaki M et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. Nucleic Acids Res 2016; 44: e108.

33. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009; 4: 44-57.

34. Wang J, Vasaikar S, Shi Z et al. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. Nucleic Acids Res 2017; 45: W130-W137.

35. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016; 32: 2847-2849.

36. C K, MD M, F V et al. Mutational landscape and significance across 12 major cancer types. Nature 2013; 502: 333-339.

37. Nguyen DX, Bos PD, Massague J. Metastasis: from dissemination to organ-specific colonization. Nat Rev Cancer 2009; 9: 274-284.

38. Paterlini-Brechot P. Organ-specific markers in circulating tumor cell screening: an early indicator of metastasis-capable malignancy. Future Oncol 2011; 7: 849-871.

39. M P, AS B, R K 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: 1803-1811.

40. C G, C S, M E et al. Frequent DNA hypomethylation of human juxtacentromeric BAGE loci in cancer. Genes, chromosomes & cancer 2005; 43: 11-24.

41. M R, M V, N G et al. BAGE genes generated by juxtacentromeric reshuffling in the Hominidae lineage are under selective pressure. Genomics 2003; 81: 391-399.

42. S Z, X Z, H Y, Y Y. Expression of tumor-specific antigen MAGE, GAGE and BAGE in ovarian cancer tissues and cell lines. BMC cancer 2010; 10: 163.

43. M R, P vdB, ME B et al. New BAGE (B melanoma antigen) genes mapping to the juxtacentromeric regions of human chromosomes 13 and 21 have a cancer/testis expression profile. European journal of human genetics : EJHG 2002; 10: 833-840.

44. M H, I R. mRNA detection of tumor-rejection genes BAGE, GAGE, and MAGE in peritoneal fluid from patients with ovarian carcinoma as a potential diagnostic tool. Cancer 2002; 96: 187-193.

45. P D, E F, B M et al. MAGE, BAGE and GAGE gene expression in human rhabdomyosarcomas. International journal of cancer 2001; 93: 85-90.

46. Y K, T H, K N et al. Expression of MAGE, GAGE and BAGE genes in human liver diseases: utility as molecular markers for hepatocellular carcinoma. Journal of hepatology 2000; 32: 612-617.

47. T F, M M, H U et al. Expression of MAGE and BAGE genes in Japanese breast cancers. Annals of oncology : official journal of the European Society for Medical Oncology 1997; 8: 369-372.

48. J L, Y Y, T F et al. Expression of BAGE, GAGE, and MAGE genes in human gastric carcinoma. Clinical cancer research : an official journal of the American Association for Cancer Research 1996; 2: 1619-1625.

49. C G, ME B, I R et al. BAGE hypomethylation, a new epigenetic biomarker for colon cancer detection. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2008; 17: 1374-1379.

50. M W, X L, J Z et al. AHNAK2 is a Novel Prognostic Marker and Oncogenic Protein for Clear Cell Renal Cell Carcinoma. Theranostics 2017; 7: 1100-1113.

51. L W, K P, Y K. Tumor metastasis: moving new biological insights into the clinic. Nature medicine 2013; 19: 1450-1464.

52. Huang RY. Epithelial-mesenchymal transitions in development and disease. Cell 2009; 139: 871-890.

53. Alsidawi S, Malek E, Driscoll JJ. MicroRNAs in Brain Metastases: Potential Role as Diagnostics and Therapeutics. International Journal of Molecular Sciences 2014; 15: 10508.

54. Kim CG, Lee H, Gupta N et al. Role of Forkhead Box Class O proteins in cancer progression and metastasis. Semin Cancer Biol 2018; 50: 142-151.

55. Marshall AD, Lagutina I, Grosveld GC. PAX3-FOXO1 induces cannabinoid receptor 1 to enhance cell invasion and metastasis. Cancer Res 2011; 71: 7471-7480.

56. Liu SH, Patel S, Gingras MC et al. PDX-1: demonstration of oncogenic properties in pancreatic cancer. Cancer 2011; 117: 723-733.

57. SH L, DD R, J N et al. PDX-1 is a therapeutic target for pancreatic cancer, insulinoma and islet neoplasia using a novel RNA interference platform. PloS one 2012; 7: e40452.

58. F OP, N C, F O et al. The relationship among PDX1, CDX2, and mucin profiles in gastric carcinomas; correlations with clinicopathologic parameters. Journal of cancer research and clinical oncology 2011; 137: 1749-1762.

59. S I, T M, K H et al. Bile duct expression of pancreatic and duodenal homeobox 1 in perihilar cholangiocarcinogenesis. Histopathology 2012; 61: 266-276.

60. R A, J H, JJ A et al. Clinical and Genomic Characterization of Treatment-Emergent Small-Cell Neuroendocrine Prostate Cancer: A Multi-institutional Prospective Study. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2018; 36: 2492-2503.

**Legends of figures**

**Figure 1 Mutation spectra, Enrichment of pathway and keywords in LCBM.** (A) Mutation spectrum of six transition (Ti) and transversion (Tv) categories. (B) Enriched pathway of frequently mutated genes. (C) Keywords Enriched analysis of frequently mutated genes.



**Figure 2 Functional annotation, co-Mutation profiles and survival analysis.** (A) Functional annotation to 6318 somatic variants (left) and variants in coding consequences (right). (B) Co-Mutation profiles for the most frequent mutation genes in 26 samples. (C) Survival analysis between ERF mutation and overall survival times in TCGA. (D) Survival analysis between ERF expression and overall survival times in TCGA. (E) Survival analysis between ERF mutation and overall survival times in our cohort

