RNA-seq of 272 gliomas revealed a novel, recurrent *PTPRZ1-MET* fusion transcript in secondary glioblastomas

Zhao-Shi Bao, Hui-Min Chen, Ming-Yu Yang, et al.

Genome Res. published online August 18, 2014
Access the most recent version at doi:10.1101/gr.165126.113

RNA-seq of 272 gliomas revealed a novel, recurrent *PTPRZ1-MET* fusion transcript in secondary glioblastomas

Zhao-Shi Bao[1,2,10*], Hui-Min Chen[5*], Ming-Yu Yang[5*], Chuan-Bao Zhang[1,2,10], Kai Yu[5], Wan-Lu Ye[5], Bo-Qiang Hu[5], Wei Yan[8], Wei Zhang[2,10], Johnny Akers[11], Valya Ramakrishnan[11], Jie Li[11], Bob Carter[11] ,Yan-Wei Liu[1, 2,10],Hui-Min Hu[1], Zheng Wang[1, 2,10], Ming-Yang Li[1, 2,10], Kun Yao[6,10], Xiao-Guang Qiu[10,12], Chun-Sheng Kang[7,10], Yong-Ping You[8,10], Xiao-Long Fan[9], Wei Sonya Song [1,3], Rui-Qiang Li[5] , Xiao-Dong Su[5**],Clark C. Chen[11**], Tao Jiang[1, 2, 3, 4, 10**]

Affiliation:
[1] Beijing Neurosurgical Institute, Beijing 100050, China
[2] Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, Beijing 100050, China
[3]Center of Brain Tumor, Beijing Institute for Brain Disorders
[4]China National Clinical Research Center for Neurological Diseases
[5]Biodynamic Optical Imaging Center (BIOPIC), School of Life Sciences, Peking University, Beijing 100871, China
[6]Department of Pathology, Beijing Sanbo Brain Hospital, Capital Medical University, Beijing 100093, China
[7]Department of Neurosurgery, Tianjin Medical University General Hospital, Key Laboratory of Post-trauma Neuro-repair and Regeneration in Central Nervous System, Ministry of Education, Tianjin, China
[8]Department of Neurosurgery, the First Affiliated Hospital of Nanjing Medical University, Nanjing, China
[9]Laboratory of Neuroscience and Brain Development, Beijing Key Laboratory of Gene Resources and Molecular Development, Beijing Normal University, Beijing, China
[10]Chinese Glioma Cooperative Group (CGCG)
[11]Center for Theoretical and Applied Neuro-Oncology (CTAN), Division of Neurosurgery, University of California, San Diego
[12]Department of Radio-therapy, Beijing Tiantan Hospital, Capital Medical University, Beijing 100050, China

* These authors contributed equally to this study.
** Corresponding authors.

Corresponding author: Tao Jiang, M.D., Beijing Neurosurgical Institute, Beijing 100050, China, Department of Neurosurgery, Beijing Tiantan Hospital, No.6 Tiantan Xili, Dongcheng District, Beijing 100050, China; e-mail: taojiang1964@163.com; Clark C. Chen, M.D, Ph.D., University of California, San Diego, clarkchen@ucsd.edu. Xiao-Dong Su, Ph.D., Biodynamic Optical Imaging Center (BIOPIC), School of Life Sciences, Peking University, Beijing 100871, China, e-mail: xdsu@pku.edu.cn;

Funding: This work was supported by grants from National High Technology Research and Development Program (No.2012AA02A508), International Science and Technology Cooperation

Program (No. 2012DFA30470), and National Natural Science Foundation of China (No. 91229121). CCC is supported by the Doris Duke Charitable Foundation, Sontag Foundation, Burroughs Wellcome Fund, Forbeck Foundation, and Kimmel Foundation.

ABSTRACT

Studies of gene rearrangements and the consequent oncogenic fusion proteins have laid the foundation for targeted cancer therapy. To identify oncogenic fusions associated with glioma progression, we catalogued fusion transcripts by RNA-seq of 272 WHO grade II, III, and IV gliomas. Fusion transcripts were more frequently found in high grade gliomas, in the classical subtype of gliomas, and in gliomas treated with radiation / temozolomide. 67 in-frame fusion transcripts were identified, including three recurrent fusion transcripts: *FGFR3-TACC3*, *RNF213-SLC26A11*, and *PTPRZ1-MET* (*ZM*). Interestingly, the *ZM* fusion was found only in grade III astrocytomas (1/13; 7.7%) or secondary GBMs (sGBMs, 3/20; 15.0%). In an independent cohort of sGBMs, the ZM fusion was found in 3 of 20 (15%) specimens. The fusion was additionally detected in 3 of 19 (16%) glioblastoma cell lines, confirming the recurrent nature of this transcript. Genomic analysis revealed that the fusion arose from translocation events involving introns 3 or 8 of *PTPRZ* and intron 1 of *MET*. *ZM* fusion transcripts were found in GBMs irrespective of isocitrate dehydrogenase 1 (*IDH1*) mutation status. sGBMs harboring *ZM* fusion showed higher expression of genes required for *PIK3CA* signaling and lowered expression of genes that suppressed *RB1* or *TP53* function. Expression of the *ZM* fusion was mutually exclusive with *EGFR* over-expression in sGBMs. Exogenous expression of the *ZM* fusion in the U87MG glioblastoma line enhanced cell-migration and invasion. Clinically, patients afflicted with *ZM* fusion harboring glioblastomas survived poorly relative to those afflicted with non-*ZM* harboring sGBMs ($p<0.001$). Our study profiles the shifting RNA landscape of gliomas during progression and reveled *ZM* as a novel, recurrent fusion transcript in sGBMs.

INTRODUCTION

The paradigm of oncogene addiction is predicated on the premise that some oncogenes perform essential and irreplaceable functions required for the survival of cancer cells (Weinstein 2002). The aggregate of studies spanning the past two decades, however, reveal that very few oncogenes actually fulfill this criteria (Torti and Trusolino 2011). Most oncogenic pathways appear dynamic, with highly redundant circuitry (Stommel et al. 2007; Nitta et al. 2010). One of the notable exception to these observations involve fusion proteins (Ren 2005). These fusion proteins typically resulted from chromosomal translocations (Nambiar et al. 2008) and executed novel functions that cannot be reconstituted by the expression of either parental protein (Ren 2005; Singh et al. 2012). Importantly, these novel functions re-programmed the cellular circuitry to a state of exquisite addiction (Ren 2005; Sasaki et al. 2010). Some of the most promising clinical results have arisen from targeted inhibition of these fusion proteins, including the *BCR-ABL1* fusion (Ren 2005) and the *EML4–ALK* fusion (Sasaki et al. 2010).

As a first step toward the identification of novel fusion proteins, we performed RNA sequencing of 272 WHO grade II, III, and IV gliomas. Gliomas are the most common form of brain cancer and can be classified grade I to grade IV based on histologic features (Louis et al. 2007; Wang and Jiang 2013). Grade II gliomas are also known as low grade glioma, whereas grade III and IV tumor are frequently termed high grade glioma (Wen and Kesari 2008). The term glioblastoma (GBM) is synonymous with grade IV glioma (Wen and Kesari 2008). GBM is one of the deadliest of human cancers, with median survival of 14 months after maximal surgical resection, chemotherapy, and

radiation therapy (Stupp et al. 2005). Based on clinical history, GBMs can generally be classified into two subtypes (Ohgaki and Kleihues 2009). Primary GBM (pGBM) refers to the vast majority of GBMs that are thought to form *de novo* in the elderly. On the other hand, secondary GBMs (sGBMs) typically progress from lower grade tumor and affect younger patient populations. While pGBMs and sGBMs are indistinguishable histologically (Louis et al. 2007), emerging genomic profiling revealed distinct genetic landscape between these two tumor types (Ohgaki and Kleihues 2009). For instance, mutation in the metabolic enzyme isocitrate dehydrogenase (*IDH*) are found almost exclusively in the sGBMs (Yan et al. 2009)

Our efforts unveiled a novel, recurrent fusion transcript involving the *PTPRZ1* and the *MET* gene (*ZM*) that was found in 15% of sGBMs. sGBMs harboring this fusion exhibit aggressive clinical behavior and are associated with poor prognosis to afflicted patients.

RESULTS

RNA-seq of 272 gliomas. 272 freshly frozen glioma samples were collected for the initial exploratory analysis using RNA-seq. The demographics of this patient population can be found in **Table 1 and Supplemental Table 1**. Central pathology reviews of these specimens were performed by independent board certified neuro-pathologists and graded based on the 2007 WHO classification (Louis et al. 2007). All samples were collected based on The Cancer Genome Atlas criteria (The Cancer Genome Atlas 2008), such that collected specimens contained at least 80% viable GBM tissue, frozen within 5 minutes of resection, and subjected to RNA-seq analysis. In total, 1386 Gbs of 101bp pair-end reads were generated, with an average of 50 million reads per sample. The

sequencing data were mapped to the human reference gene set and reference genome RefSeq (hg19) using BWA (Li and Durbin 2009).

TopHat-Fusion (Kim and Salzberg 2011b) and deFuse (McPherson et al. 2011a) algorithms were used for fusion detection. Only fusions that scored positive on both algorithms were subsequently analyzed (See **Methods**). In total, 214 fusion transcripts were detected in 38% of the samples (n=104, **Supplemental Table 2**). With the exception of the fusion transcripts that fell in GC-rich regions (4.7%), validation of all fusion transcripts were performed using conventional PCR amplification followed by Sanger sequencing.

In general, fusion transcripts were more frequent in high grade gliomas. Only 18.0 % of grade II gliomas harbored fusion transcripts. In contrast, nearly half of the high grade gliomas (42.5 % of grade III glioma and 55.6 % grade IV glioma) harbored fusion transcripts (**Figure 1A**). These results were generally consistent with the progressive increase in genomic instability during advancing tumor grade (Negrini et al. 2010). The highest numbers of fusion transcripts were found in gliomas that recurred after radiation and/or temozolomide treatment (**Figure 1A**), suggesting that DNA damage accumulation contributes to fusion transcript formation. The number of fusion transcripts detected did not significantly differ between pGMB and sGBM. However, the classical subtype of gliomas (Verhaak et al. 2010) was more likely to harbor fusion transcripts relative to the other The Cancer Genome Atlas defined transcriptional subtypes ($p<0.03$, **Figure 1B**)

The most common form of fusion transcripts (75.7%) arose from joining of sequences from the same

chromosome (**Figure 1C**). These fusion transcripts most commonly mapped to chromosomes 12 and 19, suggesting these chromosomes may represent "hot-spots" for deletional instability. Fusion transcripts involving sequences from distinct chromosomes constituted 24.3 % of all fusion transcripts. These events most commonly involved exchanges between chromosomes 4 and 16, suggesting that these chromosomes may be located in physical proximity in a three dimensional chromatin structure (Sajan and Hawkins 2012) **(Figure 1D)**. The ratio of intra-chromosomal and inter-chromosomal translocations was consistent with those reported in breast cancer (Kangaspeska et al. 2012).

Functional annotation of the in-frame fusion transcripts. Of the 214 fusion transcripts, 147 were out-of-frame and 67 were in-frame (sequences found on **Supplemental Table 2**). The distribution of these fusion transcripts as a function of age, sex, and glioma grade are shown in **Supplemental Figure 1A**. Of these in-frame fusions, 55 arose from fusion of sequences located on the same chromosome and 12 from joining of sequences derived from different chromosomes.

Functional annotation of the fusion transcript revealed as the following. We identified 14 fusion transcripts containing sequences of genes involved in the canonical GBM signaling pathways (The Cancer Genome Atlas 2008), including *RTK/PIK3CA*, *RB1*, and *TP53* signaling pathways, including *CBL-FBXO2*, *FGFR3-TACC3*, *PTPRZ1-MET (ZM)*, *VHL-BRK1*, *EGFR-VSTM2A*, *JAK1-HIVEP3*, *PTEN-COL17A1*, *CDK4-TSFM*, *IL1RAP-FGF12*, *NFATC3-CPNE2*, *PLA2G6-CRYBB1*, *APEH-SHISA5*, *TCF7L1-KIF1B*, *TGFB1-SAE1* (**Supplemental Figure 1B**). Additionally, we identified 11 fusion transcripts containing sequences of genes with metabolic function: *CCM2-OGDH*, *FRMD4A-PFKP*, *TPT1-AADAT*, *AHCYL2-TMEM178B*, *PTN-DGKI*,

ZMIZ1-MAT1A, MTAP-C9orf92, CD81-SPAG6, CDK17-KCNC2, BCR-LZTR1, and *AP2A2-SBF2*. We found five fusion transcripts containing sequences encoding kinase domains: *FGFR3-TACC3*, *PTPRZ1-MET* (*ZM*), *ST7-CTTNBP2*, *PLAGL2-HCK*, and *NEK6-RXRA*. Finally, we identified two fusion transcripts containing sequences from genes implicated in chromatin remodeling: *IFT80-MLH1* and *MLL3-CHGB*.

We identified 3 recurrent fusion transcripts: *FGFR3-TACC3, RNF213-SLC26A11,* and the *ZM* fusion. The *FGFR3-TACC3* fusion was previously reported by Singh et al. (Singh et al. 2012). This fusion transcript was found in 3/59 pGBMs **(**5.1%**, Supplemental Figure 1C)**. The *RNF213-SLC26A11* fusion transcript was previously identified in a chronic myeloid leukemia specimen (Zhou et al. 2013). This fusion transcript was detected in 1 pGBM (1/59 or 1.7%) and a recurrent grade III glioma (1/13 or 7.7%, **Supplemental Figure 1D**). The *ZM* fusion has not been previously reported and was detected in 1 grade III glioma and 3 sGBMs (**Supplemental Figure 1E**). The four fusion transcripts involved four different breakpoints within the *PTPRZ1* coding sequence. In contrast, the breakpoints in the *MET* gene were located at the same junction.

Validation of the *PTPRZ1-MET* as a recurrent fusion transcript. Three fusion transcripts contained sequences encoding the carbonic anhydrase (CA) domain and fibronectin type III (Fib III) domain of PTPRZ1 (Mohebiany et al. 2013) fused to the dimerization domain, immunoglobulin-like domains, transmembrane domain, and the tyrosine kinase domain of MET (Organ and Tsao 2011) (**Figure 2A**). To confirm the recurrent nature of the *ZM* fusion, we collected an additional 192 glioma samples (**Supplemental Table 3**) and screened for the presence of the fusion transcript in

these samples using fusion-specific PCR primers (**Supplemental Table 4)**. Of the samples tested, the *ZM* fusion transcript was detected in five additional specimens **(Figure 2B)**. These fusion breakpoints were confirmed by direct Sanger sequencing. Importantly, a variety of fusion junctions were detected, rendering cross-contamination as the reason for fusion detection unlikely. Interestingly, 6 out of 40 (15%) specimens in total were derived from patients afflicted with sGBM.

To further confirm the recurrent nature of the *ZM* fusion, we screened 19 GBM cell lines using fusion-specific PCR primers and detected the fusion sequence in three cell lines (two long-term passaged lines (U118 and LN18) and one primary neurosphere line (CMK3), **Figure 2C**). In U118 and LN18, a T->C transition mutation was found in the second nucleotide of the *MET* sequence, changing Methionine into Threonine. In CMK3, additional mutations were found in addition to the T->C transition. A stretch of nucleotide changes between nucleotide 73 to 123 of *PTPRZ1* coding sequence changed the stretch of amino acids from Tyr-Leu-Lys-Arg-Phe-Leu-Ala-Cys-Ile-Gln-Leu-Leu-Leu-Cys-Val-Cys-Arg-Leu toTyr-Tyr-Arg-Gln-Gln-Arg-Lys-Leu-Val-Glu-Glu-Ile-Gly-Try-Ser-Tyr-Thr. These results together confirmed the recurrent nature of ZM fusion in GBMs.

Genomic Translocation Events that give rise to *ZM* fusion transcripts. We wished to characterize the genomic translocation that gave rise to the *ZM* fusion transcript. To this end, we extracted DNA from two *ZM*-harboring glioblastoma specimens: CGGA_D64 and CGGA_1068. CGGA_D64 harbored a *ZM* fusion that fused exon 8 of *PTPRZ* to exon 2 of *MET* (**Supplemental Figure 2A**). PCR amplification and Sanger sequencing of the genomic breakpoint in CGGA_D64 revealed a translocation fusing DNA sequences from intron 8 of *PTPRZ1* and intron 1 of *MET* (**Supplemental Figure 2B**). Splicing of the fuse intron is expected to give rise to the fusion

transcript observed in CGGA_D64. CGGA_1068 harbored a RNA transcript fusing exon 2 of *PTPRZ* to exon 2 of *MET* (**Supplemental Figure 2C**). Sanger sequencing of the genomic breakpoint in CGGA_1068 revealed a translocation fusing DNA sequences from intron 1 of *PTPRZ1* and intron 1 of *MET* (**Supplemental Figure 2D**). Splicing of the fuse intron is expected to give rise to the fusion transcript observed in CGGA_1068. Of note, intact *PTPRZ* transcripts can be detected in both specimens (**Supplemental Figure 2E**), suggesting tandem duplication of the region of *PTPRZ* involved in *ZM* fusion.

Genetic landscape of *ZM* fusion containing sGBMs. We wished to assess whether the genetic landscape of the *ZM* fusion containing sGBM differed from those of non-fusion containing sGBMs. Given the emerging importance of *IDH1* mutations in sGBM (Yan et al. 2009), we first tested whether the presence of *ZM* fusion transcripts coincided with the presence of *IDH1* mutations. The *ZM* fusion transcripts were found in one *IDH1* mutated GBM and two *IDH* wild-type GBMs (**Figure 3A**). We next tested whether *ZM*-fusion transcripts were enriched in any particular The Cancer Genome Atlas transcriptional subtypes (Verhaak et al. 2010). The fusion transcript was detected in a proneural subtype and two classical subtype GBMs. These results suggest that the *ZM* fusion was not tightly coupled to *IDH1* mutation status or transcriptional subtypes.

We next determined whether the expression of nodal genes involved in canonical GBM signaling pathways (*PIK3CA*, *RB1*, and *TP53*) (The Cancer Genome Atlas 2008) differed between sGBMs with or without *ZM* fusion transcripts. Of the key genes that mediate *PIK3CA* signaling, there was significant over-expression of *MET*, *PIK3CA*, and *AKT1* transcripts in the *ZM* fusion bearing GBMs, suggesting hyper-activation of the *PIK3CA* pathway (Ng et al. 2012). Of the key genes mediating *RB1* function, the expression of *CDKN2A*, a suppressor in this pathway (Foulkes et

al. 1997), appeared significantly suppressed while *CDK6*, an activator of this pathway (Deshpande et al. 2005), was over-expressed in the *ZM* fusion harboring sGBMs. Of the key genes mediating *TP53* function, *MDM2* and *MDM4* (Wade et al. 2013) appeared highly over-expressed in the fusion bearing GBMs, implicating suppression of *TP53* mediated DNA damage response (Bartkova et al. 2005; Bartek et al. 2007; Bartkova et al. 2010) as a key step in the pathogenesis of these tumors **(Figure 3B)**.

These results suggest that the *ZM* fusion harboring tumors may exhibit a more aggressive phenotype. Clinically, this aggressive phenotype would manifest in the form of decreased overall survival (OS). Indeed, we found that the patients afflicted with *ZM* fusion harboring sGBMs fared particularly poorly, with significantly compromised overall survival relative to those afflicted with sGBMs without the *ZM* fusion **(**median OS with *ZM* fusion Vs. without *ZM* fusion: 127 days Vs. 248 days, $p<0.001$, log-rank test, **Figure 3C)**.

Mutual exclusivity of Epidermal Growth Factor Receptor (*EGFR*) expression and *ZM* fusion in sGBMs. *EGFR* is a receptor tyrosine kinase that is frequently amplified in GBMs (Wen and Kesari 2008) and signals through the *RTK/PIK3CA* cascade. Our analysis suggested that this pathway is hyperactive in *ZM* fusion harboring sGBMs. Since genetic events that are functionally redundant (Ciriello et al. 2012) frequently demonstrate mutually exclusive patterns in genomic analysis, we tested whether the expression of *ZM* fusion and *EGFR* over-expression were mutually exclusive in sGBMs. Consistent with our hypothesis, over-expression of *EGFR* and the presence of the *ZM* fusion transcript appeared mutually exclusive (**Figure 3D**).

TERT promoter mutations are often associated with *EGFR* amplification or over-expression (Nonoguchi et al. 2013; Appin and Brat 2014). We thus, hypothesize that *TERT* mutations will co-segregate with *EGFR* over-expression and will be mutually exclusive to the presence of *ZM* fusion transcript. To test this hypothesis, we performed Sanger sequencing of the *TERT* promoter region in 20 sGBMs (3 samples harboring *ZM* fusion transcripts and 17 *ZM* negative samples). *TERT* mutation was found in approximately a third of the *ZM*-negative sGBMs but none of the *ZM* harboring sGBMs (**Figure 3D**), further confirming our hypothesis.

Importantly, *EGFR* over-expression was not associated with changes in the overall survival of sGBM patients (median OS with high *EGFR* expression vs. low *EGFR* expression: 233 days vs. 240 days, p=0.730, log-rank test, **Figure 3E**), suggesting that *ZM* fusion protein likelymediate activities not attributed to *EGFR*.

Expression of the *ZM* fusion protein.

We characterized *MET* expression at the protein level in two *ZM*-harboring sGBM specimens with sufficient quantity for immunoblotting analysis (CGGA_1475 and CGGA_D64) and two *ZM*-negative samples (CGGA_822 and CGGA_1285). CGGA_D64 harbored a *ZM* fusion that fused exon 8 of *PTPRZ* to exon 2 of *MET*. The expected molecular weight of this fusion protein is approximately 190 kD. This prediction is born out when the D64 ZM fusion protein was exogenously expressed in 293 T (Supplemental Figure 2). Indeed, when protein extract from CGGA-D64 was probed with an anti-Met antibody, a 190 kD band was observed in addition to the 145 kD band. These results suggest that the *ZM* fusion transcript was translated into protein (**Figure 4A**).

CGGA_1475 harbored a *ZM* fusion that fused exon 2 of *PTPRZ* to exon 2 of *MET*. The

anticipated molecular weight of exon 1 and 2 of *PTPRZ* are 2.3 kD and 2.7 kD, respectively. The molecular weight of Met is approximately 145 kD. Exon 1 of *MET* encodes the 5'untranslated sequence (394 bp). The anticipated molecular weight of the *ZM* fusion in CGGA_1475, where exons 1 and 2 of *PTPRZ* are fused to exon 2 of MET therefore approximates that of the native *MET* (145 kD). The molecular weight of this ZM fusion approximates that of the wild type *MET* (~145 kD). As such, these two species cannot be discriminated based on SDS-PAGE. This prediction was born out when the CGGA_1475 *ZM* fusion was over-expressed in 293T cells (**Supplemental Figure 3**). Given this ambiguity, it is difficult to interpret whether the strong 145 kD band in the CGGA_1475 extract represents MET or ZM fusion protein expression.

To explore the oncologic function of the *ZM* protein, we cloned a His tagged version of the CGGA_1475 *ZM* fusion into an adenovirus vector and stably expressed this protein in the U87MG glioblasotma line. Stable expression of this *ZM* protein can be detected as evidenced by a 145 kD band when probed with an anti-His tag antibody or an anti-MET antibody. Importantly, the MET endogenously expressed in U87MG is not phosphorylated at residue 1234/5(Cooper et al. 1984; Eder et al. 2009). This phosphorylation event occurs upon dimerization and activation of MET (Wickramasinghe and Kong-Beltran 2005). In contrast, exogenously expressed MET or ZM fusion harbor this phosphorylation (**Figure 4B**).

As *MET* signaling is known to enhance glioblastoma migration and invasion (The Cancer Genome Atlas 2008), we tested whether *ZM* fusion modulated these properties in U87MG. Consistent with prior reports(Brockmann et al. 2003), MET expression caused a 7.55 fold increase in the migratory activity of U87MG in a Matrigel-coated transwell assay relative to a vector control. Expression of the ZM fusion caused a 9.9 fold increase in the migratory activity of U87MG cells

relative to vector control. This represents an approximate 30% increase in cellular migratory activity in ZM expressing U87MG relative to cells expressing wild type MET (**Figure 4C**). In sum, these results suggest that the recurrent nature of the ZM fusion is unlikely a statistical artifact and that the ZM fusion contribute to the oncologic function of glioblasotmas.

Discussion

Our study is the first study to profile the shifting RNA landscape of gliomas as a function of tumor grade and the identification of *ZM* as a recurrent fusion gene in sGBMs. In total, 214 fusion transcripts were detected. We observed a notable increase in the proportion of gliomas harboring fusion transcripts during the grade II-III transition. However, this proportion remained somewhat constant between grade III or IV gliomas. These results suggest chromosomal instability as a phenotype associated with the transition from low grade glioma (grade II) to high grade glioma (grade III and IV). Highest number of fusion transcripts was found in GBMs that recurred after radiation/temozolomide treatment, suggesting that DNA damage accumulation contributed to fusion transcript formation. Supporting this hypothesis, the classical subtypes of glioblastoma was more likely to harbor fusion transcripts relative to other subtypes. A hallmark of the classical subtype involved aberrant *EGFR* signaling (Verhaak et al. 2010), and this aberrant signaling has been shown to induce increased DNA damage accumulation and chromosomal instability (Nitta et al. 2010).

Two of the three recurrent fusion transcripts have been previously reported. The *RNF213-SLC26A11* fusion transcript was found in a chronic myeloid leukemia patient (Zhou et al. 2013). The *FGFR3-TACC3* fusion transcript was initially reported by Singh et al (Singh et al. 2012) as a recurrent fusion transcript in glioblastomas. The finding was subsequently confirmed by an

independent group (Parker et al. 2013). Singh et. al. reported transcripts fusing exon 16 of *FGFR3* to exon 8 of *TACC3*. Soon after this initial report, Parker et al. reported fusion transcripts involving exon 18 or 19 of *FGFR3* and exon 4, 10 or 11 of *TACC3*. Our analysis showed fusion involving exon 17 of *FGFR3* and exon 8, 10 or 11 of *TACC3*. We found *FGFR3-TACC3* fusion transcripts in 5% of primary glioblastoma. This prevalence is largely consistent with those reported by Singh at al and Parker et al(Singh et al. 2012; Parker et al. 2013)..

In addition the two previously reported fusion transcripts, we identified a novel recurrent fusion transcript that we termed the *ZM* fusion transcript. *ZM* is detected in approximately 15% of sGBMs in independent cohorts, rendering it the most frequently recurring transcripts for sGBMs. This transcript arose as a result of translocation events between the introns of *PTPRZ* and the *MET* proto-oncogene, resulting in a fusion containing variable numbers of *PTPRZ1* extra-cellular domains (including the carbonic anhydrase domain and the fibronectin domain) and the entire intracellular domain of *MET* (Figure 2A). *MET* encodes a well-studied proto-oncogene whose activation is triggered by dimerization of the intracellular domain upon binding of hepatocyte growth Factor (HGF) (Stommel et al. 2007; Li et al. 2011). *PTPRZ1* encodes a membrane associated tyrosine phosphatase that is highly expressed in the central nervous system (Muller et al. 2003) and signals through beta -catenin mediated functions (Diamantopoulou et al. 2012). The recurrent nature of the *ZM* fusion suggests that it plays an active role in glioblastoma biology. Consistent with this hypothesis 1) stable expression of *ZM* fusion enhanced glioblastoma migration and invasion; 2) *ZM* fusion harboring sGBM harbored over-expression of genes involved in *PIK3CA* signaling; 3) The expression of the *ZM* fusion transcript was associated with worsened overall survival in sGBM patients (127 days Vs. 248 days). It remains to be determined whether *ZM* expressing glioblastomas

are sensitive to MET inhibitors.

The heterogeneity of most fusion transcripts observed in this study and their non-recurrent nature are highly reminiscent of RNA-seq efforts in other solid tumor types, including melanoma (Berger et al. 2010) and lung cancer (Seo et al. 2012). In our profiling of 272 gliomas, 67 in-frame fusion transcripts were identified. While many of these fusion proteins involve genes that participate in canonical GBM signaling pathways, including *RTK/PIK3CA*, *RB1*, and *TP53* (The Cancer Genome Atlas 2008), most of the fusion transcripts identified here involved gene sequences that have not been well-studied in glioblastoma. Characterization of these fusion sequences may unveil novel biologic insights.

In sum, we reported a comprehensive RNA-seq analysis that characterized the RNA fusion landscape during glioma progression. The study provided a catalogue of novel fusion transcripts as potential GBM therapeutic targets and revealed *ZM* as a novel, recurrent fusion transcript in sGBMs.

Methods

Clinical specimen collection

All research performed were approved by Institutional review board (IRB) boards at Tiantan Hospital and were in accordance with the principles expressed at the declaration at Helsinki. Each patient was consented by a dedicated clinical research specialist prior to collection. Written consent was obtained for each patient. The specimens were collected under IRB KY2013-017-01. For each patient, the following clinical information was collected: diagnosis, gender, age, WHO grades and

overall survival (OS).

RNA-seq and Quality control (QC)

The libraries were sequenced on the Illumina HiSeq 2000 platform using the 101-bp pair-end sequencing strategy. The original image data generated by the sequencing machine were converted into sequence data via base calling (Illumina pipeline CASAVA v1.8.2) and then subjected to standard QC criteria to remove all of the reads that fit any of the following parameters:

1) The reads that aligned to adaptors or primers with no more than two mismatches.

2) The reads with more than 10% unknown bases (N bases).

3) The reads with more than 50% of low quality bases (quality value ≤ 5) in one read.

Finally, 1308.3 Gb (94.4%) filtered reads were left for further analysis after QC.

Read mapping

Hg 19 RefSeq (RNA sequences, GRCh37) was downloaded from the UCSC Genome Browser (http://genome.ucsc.edu).

Candidate gene fusion identification

We used two algorithms, deFuse(McPherson et al. 2011b) (deFuse-0.6.1) and TopHat-Fusion (Kim and Salzberg 2011a)(TopHatFusion-0.1.0), to detect gene fusion based on the pair-end reads in different samples. The candidates simultaneously detected by both deFuse and TopHat-Fusion were regarded as reliable candidate gene fusions, which was carried forward for further analysis.

Expression analysis of RefSeq genes

The gene expression was calculated using the RPKM method (Audic and Claverie 1997; Mortazavi et al. 2008) (Reads Per Kilobase transcriptome per million reads). The RPKM method is

able to remove the influence of different gene lengths and sequencing discrepancies from the calculation of gene expression. Therefore, the calculated gene expression can be directly used to comparing the differences in gene expression among samples.

RNA extraction and PCR validation

First strand cDNA was synthesized from 500-1000 ng total RNA with random hexamer primers (Promega) using SuperScript III reverse transcriptase (Invitrogen). Reverse transcription was performed at 55°C for 60 min followed by 70°C 15 min to inactivate the reaction. E.coli RNase H (New England Biolabs) was added to remove RNA complementary to cDNA. Primers (Supplemental Table 3) were designed flanking fusion point. PCR products were purified using QIAquick PCR purification kit (Qiagen) and cloned into pGEM-T easy vector (Promega), then sequenced by ABI Prism 3730×l DNA Sequencer (Applied Biosystems). sixty-four out of 67 (95.5%) fusion genes were confirmed by sequencing.

Statistical analysis

Overall survival time (OS) was calculated from the date of diagnosis until death or the last follow-up. The survival curve was calculated with the Kaplan-Meier method and the difference was analyzed using the two-sided log-rank test. A p-value < 0.05 was considered statistically significant. All the data analysis was performed in GraphPad Prism and R.

Data Access

The raw sequencing data for 272 gliomas have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE48865.

Reference

Appin CL, Brat DJ. 2014. Molecular genetics of gliomas. *Cancer J* **20**(1): 66-72.

Audic Sp, Claverie J-M. 1997. The Significance of Digital Gene Expression Profiles. *Genome Research* **7**(10): 986-995.

Bartek J, Bartkova J, Lukas J. 2007. DNA damage signalling guards against activated oncogenes and tumour progression. *Oncogene* **26**(56): 7773-7779.

Bartkova J, Hamerlik P, Stockhausen MT, Ehrmann J, Hlobilkova A, Laursen H, Kalita O, Kolar Z, Poulsen HS, Broholm H et al. 2010. Replication stress and oxidative damage contribute to aberrant constitutive activation of DNA damage signalling in human gliomas. *Oncogene* **29**(36): 5095-5102.

Bartkova J, Horejsi Z, Koed K, Kramer A, Tort F, Zieger K, Guldberg P, Sehested M, Nesland JM, Lukas C et al. 2005. DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* **434**(7035): 864-870.

Berger MF, Levin JZ, Vijayendran K, Sivachenko A, Adiconis X, Maguire J, Johnson LA, Robinson J, Verhaak RG, Sougnez C et al. 2010. Integrative analysis of the melanoma transcriptome. *Genome Res* **20**(4): 413-427.

Brockmann MA, Ulbricht U, Gruner K, Fillbrandt R, Westphal M, Lamszus K. 2003. Glioblastoma and cerebral microvascular endothelial cell migration in response to tumor-associated growth factors. *Neurosurgery* **52**(6): 1391-1399; discussion 1399.

Ciriello G, Cerami E, Sander C, Schultz N. 2012. Mutual exclusivity analysis identifies oncogenic network modules. *Genome Res* **22**(2): 398-406.

Cooper CS, Park M, Blair DG, Tainsky MA, Huebner K, Croce CM, Vande Woude GF. 1984. Molecular cloning of a new transforming gene from a chemically transformed human cell line. *Nature* **311**(5981): 29-33.

Deshpande A, Sicinski P, Hinds PW. 2005. Cyclins and cdks in development and cancer: a perspective. *Oncogene* **24**(17): 2909-2915.

Diamantopoulou Z, Kitsou P, Menashi S, Courty J, Katsoris P. 2012. Loss of receptor protein tyrosine phosphatase beta/zeta (RPTPbeta/zeta) promotes prostate cancer metastasis. *J Biol Chem* **287**(48): 40339-40349.

Eder JP, Vande Woude GF, Boerner SA, LoRusso PM. 2009. Novel therapeutic inhibitors of the c-Met signaling pathway in cancer. *Clin Cancer Res* **15**(7): 2207-2214.

Foulkes WD, Flanders TY, Pollock PM, Hayward NK. 1997. The CDKN2A (p16) gene and human cancer. *Mol Med* **3**(1): 5-20.

Kangaspeska S, Hultsch S, Edgren H, Nicorici D, Murumagi A, Kallioniemi O. 2012. Reanalysis of RNA-sequencing data reveals several additional fusion genes with multiple isoforms. *PLoS One* **7**(10): e48745.

Kim D, Salzberg S. 2011a. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biology* **12**(8): 1-15.

Kim D, Salzberg SL. 2011b. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol* **12**(8): R72.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(14): 1754-1760.

Li Y, Li A, Glas M, Lal B, Ying M, Sang Y, Xia S, Trageser D, Guerrero-Cazares H, Eberhart CG et al. 2011. c-Met signaling induces a reprogramming network and supports the glioblastoma stem-like phenotype. *Proc Natl Acad Sci U S A* **108**(24): 9951-9956.

Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A, Scheithauer BW, Kleihues P. 2007. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol* **114**(2): 97-109.

McPherson A, Hormozdiari F, Zayed A, Giuliany R, Ha G, Sun MG, Griffith M, Heravi Moussavi A, Senz J, Melnyk N et al. 2011a. deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. *PLoS Comput Biol* **7**(5): e1001138.

McPherson A, Hormozdiari F, Zayed A, Giuliany R, Ha G, Sun MGF, Griffith M, Heravi Moussavi A, Senz J, Melnyk N et al.

2011b. deFuse: An Algorithm for Gene Fusion Discovery in Tumor RNA-Seq Data. *PLoS Comput Biol* **7**(5): e1001138.

Mohebiany AN, Nikolaienko RM, Bouyain S, Harroch S. 2013. Receptor-type tyrosine phosphatase ligands: looking for the needle in the haystack. *FEBS J* **280**(2): 388-400.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Meth* **5**(7): 621-628.

Muller S, Kunkel P, Lamszus K, Ulbricht U, Lorente GA, Nelson AM, von Schack D, Chin DJ, Lohr SC, Westphal M et al. 2003. A role for receptor tyrosine phosphatase zeta in glioma cell migration. *Oncogene* **22**(43): 6661-6668.

Nambiar M, Kari V, Raghavan SC. 2008. Chromosomal translocations in cancer. *Biochim Biophys Acta* **1786**(2): 139-152.

Negrini S, Gorgoulis VG, Halazonetis TD. 2010. Genomic instability--an evolving hallmark of cancer. *Nat Rev Mol Cell Biol* **11**(3): 220-228.

Ng K, Kim R, Kesari S, Carter B, Chen CC. 2012. Genomic profiling of glioblastoma: convergence of fundamental biologic tenets and novel insights. *J Neurooncol* **107**(1): 1-12.

Nitta M, Kozono D, Kennedy R, Stommel J, Ng K, Zinn PO, Kushwaha D, Kesari S, Furnari F, Hoadley KA et al. 2010. Targeting EGFR induced oxidative stress by PARP1 inhibition in glioblastoma therapy. *PLoS One* **5**(5): e10767.

Nonoguchi N, Ohta T, Oh JE, Kim YH, Kleihues P, Ohgaki H. 2013. TERT promoter mutations in primary and secondary glioblastomas. *Acta Neuropathol* **126**(6): 931-937.

Ohgaki H, Kleihues P. 2009. Genetic alterations and signaling pathways in the evolution of gliomas. *Cancer Sci* **100**(12): 2235-2241.

Organ SL, Tsao MS. 2011. An overview of the c-MET signaling pathway. *Ther Adv Med Oncol* **3**(1 Suppl): S7-S19.

Parker BC, Annala MJ, Cogdell DE, Granberg KJ, Sun Y, Ji P, Li X, Gumin J, Zheng H, Hu L et al. 2013. The tumorigenic FGFR3-TACC3 gene fusion escapes miR-99a regulation in glioblastoma. *J Clin Invest* **123**(2): 855-865.

Ren R. 2005. Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia. *Nat Rev Cancer* **5**(3): 172-183.

Sajan SA, Hawkins RD. 2012. Methods for identifying higher-order chromatin structure. *Annu Rev Genomics Hum Genet* **13**: 59-82.

Sasaki T, Rodig SJ, Chirieac LR, Janne PA. 2010. The biology and treatment of EML4-ALK non-small cell lung cancer. *Eur J Cancer* **46**(10): 1773-1780.

Seo JS, Ju YS, Lee WC, Shin JY, Lee JK, Bleazard T, Lee J, Jung YJ, Kim JO, Yu SB et al. 2012. The transcriptional landscape and mutational profile of lung adenocarcinoma. *Genome Res* **22**(11): 2109-2119.

Singh D, Chan JM, Zoppoli P, Niola F, Sullivan R, Castano A, Liu EM, Reichel J, Porrati P, Pellegatta S et al. 2012. Transforming fusions of FGFR and TACC genes in human glioblastoma. *Science* **337**(6099): 1231-1235.

Stommel JM, Kimmelman AC, Ying H, Nabioullin R, Ponugoti AH, Wiedemeyer R, Stegh AH, Bradner JE, Ligon KL, Brennan C et al. 2007. Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies. *Science* **318**(5848): 287-290.

Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJ, Belanger K, Brandes AA, Marosi C, Bogdahn U et al. 2005. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med* **352**(10): 987-996.

The Cancer Genome Atlas. 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**(7216): 1061-1068.

Torti D, Trusolino L. 2011. Oncogene addiction as a foundational rationale for targeted anti-cancer therapy: promises and perils. *EMBO Mol Med* **3**(11): 623-636.

Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP et al. 2010. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by

abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**(1): 98-110.

Wade M, Li YC, Wahl GM. 2013. MDM2, MDMX and p53 in oncogenesis and cancer therapy. *Nat Rev Cancer* **13**(2): 83-96.

Wang Y, Jiang T. 2013. Understanding high grade glioma: molecular mechanism, therapy and comprehensive management. *Cancer Lett* **331**(2): 139-146.

Weinstein IB. 2002. Cancer. Addiction to oncogenes--the Achilles heal of cancer. *Science* **297**(5578): 63-64.

Wen PY, Kesari S. 2008. Malignant gliomas in adults. *N Engl J Med* **359**(5): 492-507.

Wickramasinghe D, Kong-Beltran M. 2005. Met activation and receptor dimerization in cancer: a role for the Sema domain. *Cell Cycle* **4**(5): 683-685.

Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, Kos I, Batinic-Haberle I, Jones S, Riggins GJ et al. 2009. IDH1 and IDH2 mutations in gliomas. *N Engl J Med* **360**(8): 765-773.

Zhou JB, Zhang T, Wang BF, Gao HZ, Xu X. 2013. Identification of a novel gene fusion RNF213SLC26A11 in chronic myeloid leukemia by RNA-Seq. *Mol Med Rep* **7**(2): 591-597.

Figure Legends

Figure 1. Fusion distribution depending on WHO classification, The Cancer Genome Atlas subtypes or chromosomes. A, Circos plot of genomic distribution of fusion genes in grade II, grade III, primary GBM, secondary GBM and recurrent gliomas. B, fusion distribution in the four The Cancer Genome Atlas subtypes. There was a distinct higher proportion of patients with fusion in Classical subtype (p=0.012). C, genomic distribution of fusion genes, indicating that, chromosome 12 was the hot spot for intra-chromosome fusion. D, for the inter-chromosome fusion detection, Chromosome 4,6 and chromosome 2, 12 were the fusion pairs with the most inter-chromosome fusion frequency.

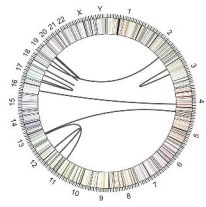
Figure 2. *ZM* fusion in training and validation sets. A, schematic of *PTPRZ1*, *MET* and the resulting *PTPRZ1-MET* fusion proteins. B, PCR and sanger sequencing validation of the positive fusion samples in training and validation sets. C, *ZM* fusion screening in 19 glioma cell lines. Three cell lines (U118, LN18 and CMK3) showed *ZM* fusion positive.

Figure 3. *PTPRZ1-MET* likely confers unique function. A, fusion distribution in The Cancer Genome Atlas subtypes in sGBM samples. The three *ZM* fused sGBM samples consisted of one proneural and two classical subtypes. Due to the small sample size, *ZM* fusion did not showed much

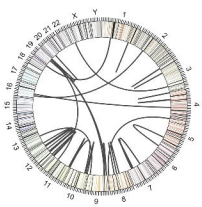
association with *IDH1* mutation. B, expression alteration in *ZM* fused sGBMs compared with those without it. The former group showed higher expression of *MET-PIK3CA-AKT1* axis as well as *MDM2* and *MDM4*. C, *ZM* fused sGBM samples had a significantly shorter overall survival. D, *ZM* fused sGBM samples showed a higher proportion of *MET* overexpression and *EGFR* underexpression. E, *EGFR* showed no prognostic value in the sGBM samples.

Figure 4. Immunoblot analysis and invasion assay used for oncogenic alteration detection of ZM fusion in vitro. A, Immunoblot analysis of ZM-negative samples (822, 1285) and ZM-positive samples (1475, D64). Other than the band of WT MET at 145kDa, D64 showed a distinct band at 190kDa, in accordance with ZM fusion protein (ZM is 942bp larger than WT MET at nucleotide level), which also hybridized with antibody to MET. STAT3 and MAPK1/3 pathway were intensively activated in ZM-positive samples. B, His tagged version of the CGGA_1475 ZM fusion was cloned into an adenovirus vector and stably expressed this protein in the U87MG glioblasotma line. The MET endogenously expressed in U87MG is not phosphorylated at residue 1234/5. In contrast, exogenously expressed MET or ZM fusion harbor this phosphorylation . C, Contrasted with scrambled, notably more U87 cells infected by ZM and MET adenovirus penetrated the Matrigel-coated transwell at 24h after cells seeded. Meanwhile, the ZM group showed more intensive invasion than cells over expressing MET. (left panel:1×; right panel: 10×; scale bar in right panel: 200μm). D, The fold induction in migration relative to SC group. (SC, Scramble Control)

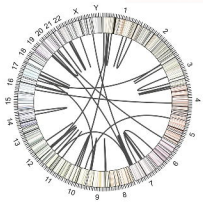
A



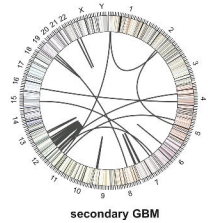
WHO grade II



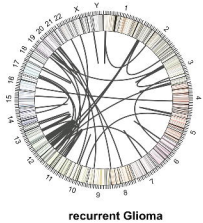
WHO grade III



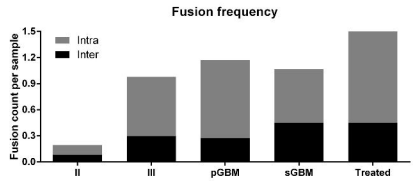
primary GBM



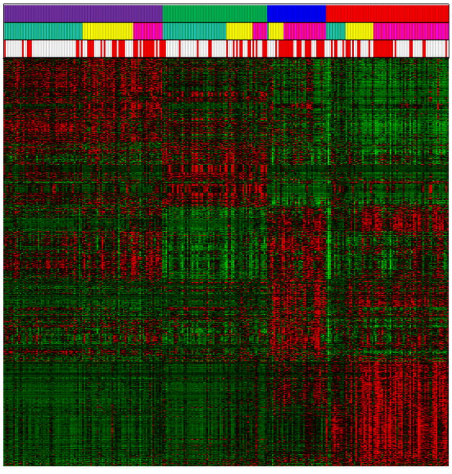
secondary GBM

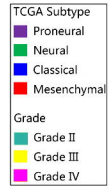


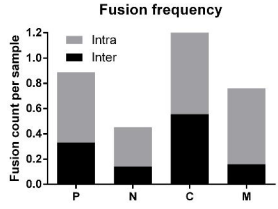
recurrent Glioma



B

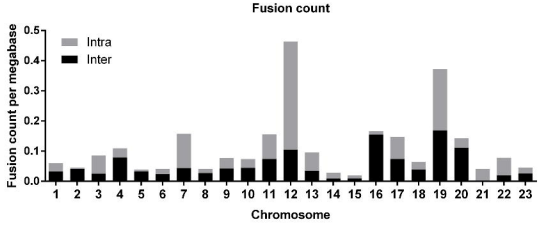






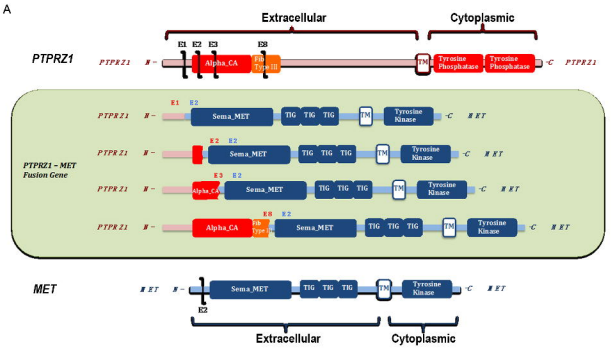
	with	without	p
P	35	62	0.012
N	18	46	
C	22	14	
M	29	46	

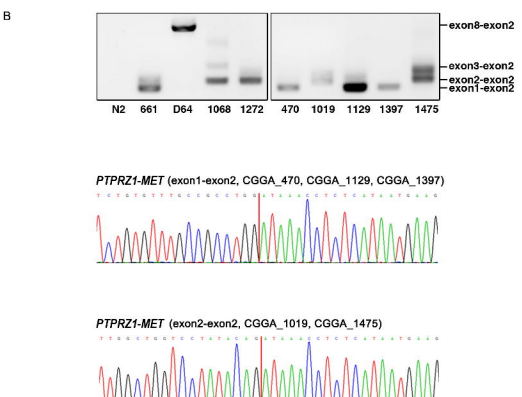
C

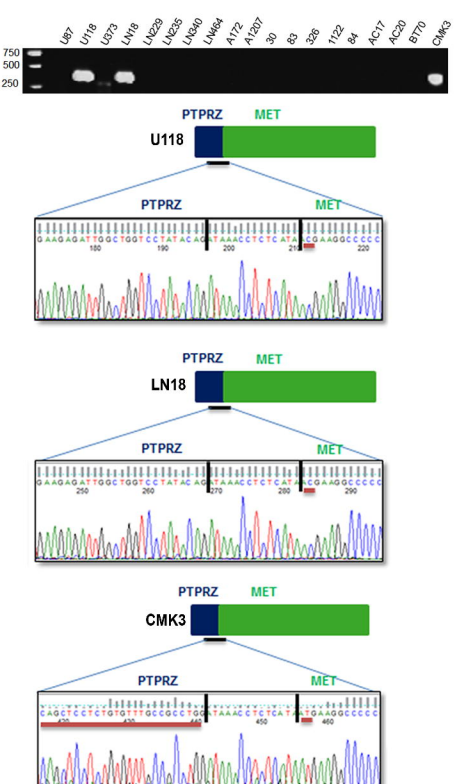


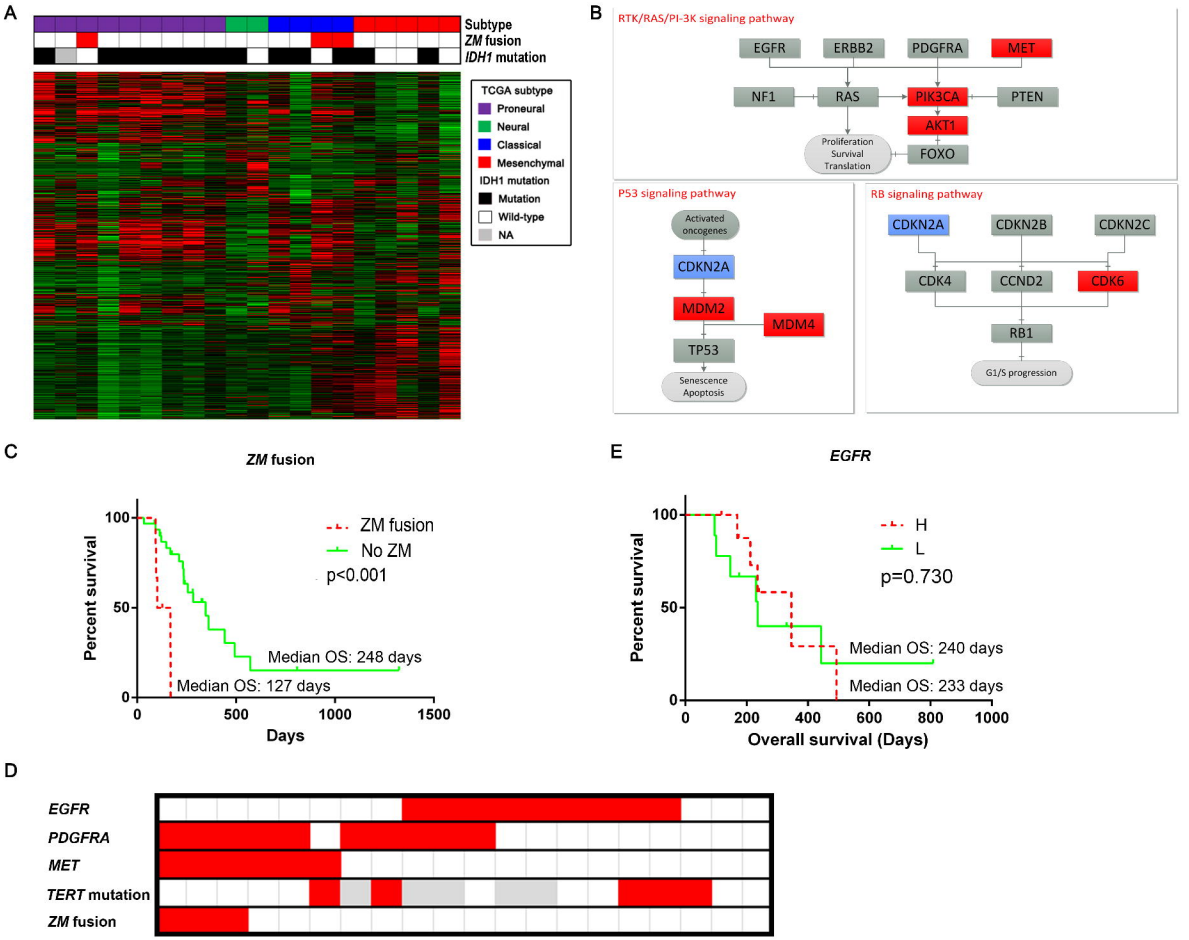
D

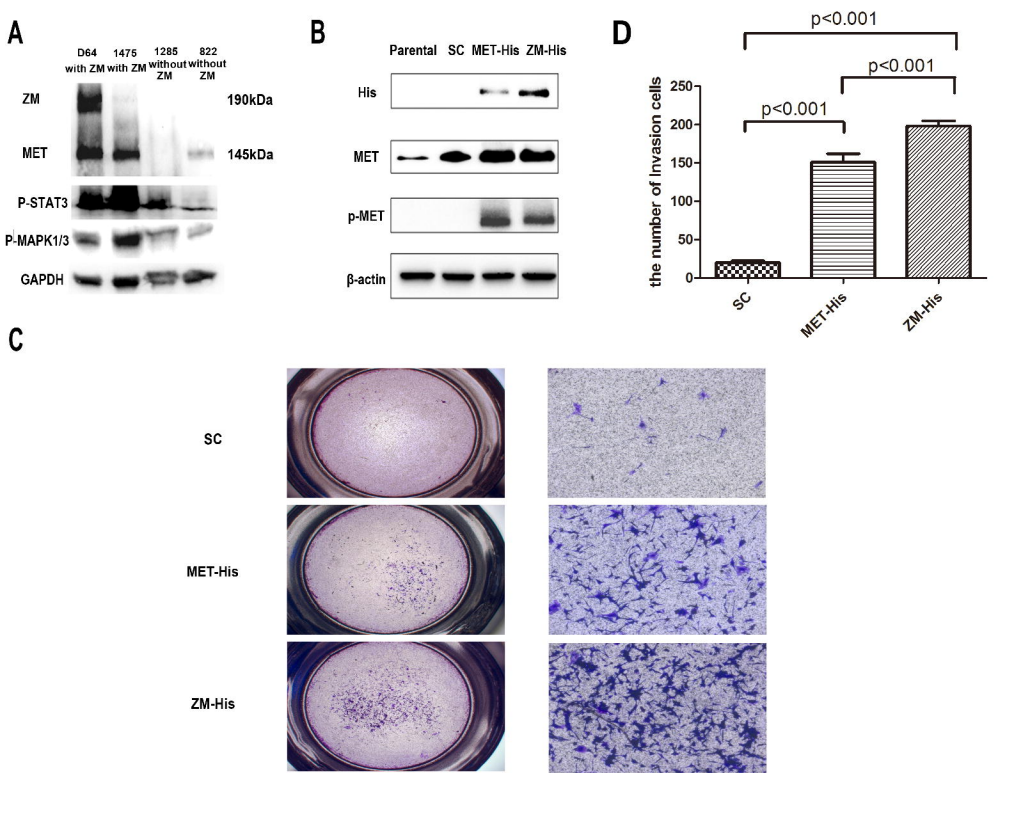
Fusion pairs (chr)	Fusion genes
4, 16	RRN3P2-ANK2 (13)
2, 12	C2orf48-PRMT8 (1)
	VSNL1-AVIL (1)
	A2M-NOL10 (1)
	AVIL-VSNL1 (1)
	PRMT8-LAPTM4A (1)


A
Extracellular
Cytoplasmic
PTPRZ1
E1 E2 E3
Alpha_CA
E8
Fibronectin type III
TM
Tyrosine Phosphatase
Tyrosine Phosphatase
-C
PTPRZ1
PTPRZ1 - MET Fusion Gene
E1 E2
Sema_MET
TIG TIG TIG
TM
Tyrosine Kinase
-C
MET
E2 E2
E3 E2
E8 E2
MET
Extracellular
Cytoplasmic


B
exon8-exon2
exon3-exon2
exon2-exon2
exon1-exon2
N2 661 D64 1068 1272 470 1019 1129 1397 1475
PTPRZ1-MET (exon1-exon2, CGGA_470, CGGA_1129, CGGA_1397)
PTPRZ1-MET (exon2-exon2, CGGA_1019, CGGA_1475)


C
U87 U118 U373 LN18 LN229 LN235 LN340 LN464 A172 A1207 30 83 326 1122 84 AC17 AC20 BT70 CMK3
750
500
250
PTPRZ MET
U118
PTPRZ MET
LN18
PTPRZ MET
CMK3
PTPRZ MET


A
Subtype
ZM fusion
IDH1 mutation
TCGA subtype
Proneural
Neural
Classical
Mesenchymal
IDH1 mutation
Mutation
Wild-type
NA
B
RTK/RAS/PI-3K signaling pathway
EGFR
ERBB2
PDGFRA
MET
NF1
RAS
PIK3CA
PTEN
AKT1
Proliferation
Survival
Translation
FOXO
P53 signaling pathway
Activated oncogenes
CDKN2A
MDM2
MDM4
TP53
Senescence
Apoptosis
RB signaling pathway
CDKN2A
CDKN2B
CDKN2C
CDK4
CCND2
CDK6
RB1
G1/S progression
C
ZM fusion
Percent survival
ZM fusion
No ZM
p<0.001
Median OS: 248 days
Median OS: 127 days
Days
E
EGFR
Percent survival
H
L
p=0.730
Median OS: 240 days
Median OS: 233 days
Overall survival (Days)
D
EGFR
PDGFRA
MET
TERT mutation
ZM fusion


A
D64 with ZM
1475 with ZM
1285 without ZM
822 without ZM
ZM
MET
190kDa
145kDa
P-STAT3
P-MAPK1/3
GAPDH
B
Parental
SC
MET-His
ZM-His
His
MET
p-MET
β-actin
D
p<0.001
p<0.001
p<0.001
the number of Invasion cells
250
200
150
100
50
0
SC
MET-His
ZM-His
C
SC
MET-His
ZM-His