



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

Pharmaco-proteogenomic characterization of glioblastoma

Sejin Oh

Department of Medical Science

The Graduate School, Yonsei University

Pharmaco-proteogenomic characterization of glioblastoma

Sejin Oh

Department of Medical Science

The Graduate School, Yonsei University

Pharmaco-proteogenomic characterization of glioblastoma

Directed by Professor Hyun Seok Kim

Doctoral Dissertation
submitted to the Department of Medical Science and
the Graduate School of Yonsei University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy of Medical Science

Sejin Oh

June 2021

This certifies that the Doctoral
Dissertation of Sejin Oh is approved.



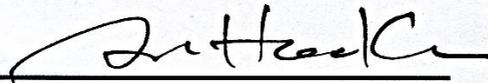
Thesis Supervisor : Hyun Seok Kim



Thesis Committee Member#1 : Cheolju Lee



Thesis Committee Member#2 : Seok-Gu Kang



Thesis Committee Member#3 : Se Hoon Kim



Thesis Committee Member#4 : Min Goo Lee

The Graduate School
Yonsei University

June 2021

ACKNOWLEDGEMENTS

나, 오세진은 박사 학위를 취득한다. 가까운 친구, 동료들과 다소 멧쩍은 이들 모두에게 애정을 담아 감사를 전한다. 여백의 한계로 모두의 귀중한 이름을 적으며 나의 감사를 표하는 것은 불가능하다.

하지만, 김현석 교수님은 꼭 언급하고 싶다. 당신 같은 거인의 어깨에 올라 학문의 세계를 본 것은 나의 25세부터 시작된 인생의 행운이며 특별한 행복이었다고 전하고 싶다. 나의 잘못을 인내하고 용서해주며 끝까지 지도해주신 은사께 나의 마음을 글로 전달하는 것은 불가능한 점을 이해하시길 바란다. 모든 방면에서 번창하시길 빈다.

나의 major revision을 성공시켜준 김주화 선생님께는 무한한 감사를 전한다. 당신은 내 인생에서 각별한 동료이다. 이 시점에서 진행 중인 대장암 논문의 성공적 마무리와 꽃길을 걷길 바란다.

나를 오랫동안 도와주던 김학현에게 따뜻한 감사를 전한다. 논문에서 pathway 그림을 잘 그려줘서 고맙다. 또한, 오명준과 이환호에게도 고난한 revision과정에서 많은 도움을 받았다. 영감을 주는 오랜 동료였던 그대들이 있어서 행운이었다. 각자의 연구가 성공하길 빈다.

김효실 박사님, 이주영 박사님, 윤현주 선생님, 한장희 형, 권성진 형, 성여진 그리고 이윤지와 김용건에게 나는 여러가지 점을 배웠고 도움을 받았다. 인생이 항상 장미빛이길 빈다.

박사과정 말에 GBM 스크리닝을 이끌어준 김상범 박사님과 도와주었던 김민서 선생님께도 감사인사를 드린다. 늘 행복하길 바란다.

학위논문에서 단백질 데이터 사용을 허락해준 염정훈, 권유미, 주신영, 최승혁 선생님들께 감사인사를 드린다. 각자의 영역에서 최고가 되길 빈다. 아울러, 백은옥 교수님과 남도현 교수님 등 GBM연구과정에 함께했던 모든 분들께 감사드린다. 함께한 경험은 내겐 대단히 귀중한 자산이 될 것이다.

최근 들어왔던 김민지와 인턴으로 만났던 김예지 등을 비롯한 후배들의 건투를 빈다. 언젠가 성큼 성장해 있을 모습을 볼 수 있길 기대한다.

박사과정 동안 나의 벗이었던 윤선진 선생님께도 감사를 전한다. 좋은 연구 많이 하고 가족이 오랫동안 건강하길 빈다.

이 논문을 심사를 해주신 이철주 박사님, 강석구 교수님, 김세훈 교수님, 그리고 이민구 교수님께 대단히 감사드린다. 역시, 하시는 일 모두 잘 되시길 빈다.

뜻 깊은 대학생활을 내게 주었던 동대신문사와 일원들인 유권준, 정동훈, 정웅재, 이활란, 이신혜, 이정민, 김윤수, 이영호 형들과 백선아, 이지연, 최지나, 이슬비에게 미안함과 감사함을 이곳에 담는다.
동대신문사여 영광이 있어라!

마지막 감사 인사는 나의 가족에게 보낸다. 나를 홀로 끝까지 키운 어머니에게 어머니의 수고로움을 내가 갚을 길이 마땅히 없다는게 죄송하다. 항상 사랑하고 감사합니다. 역시, 내 동생 세학이에게도 고맙다.

나의 끝내 런슈오, 당신에게 지극한 사랑을 담아 감사를 전하오. 대학원 입학 때, 당신의 조언이 없었다면 나는 좋은 지도교수를 못 만났고 내 박사 과정이 없었을 것이오. 앞으로 당신의 행복을 위해 좋은 남편이 되겠소.
我真的很感谢你的奉献和爱。

끝내와의 결혼을 허락해준 장인 런용성과 장모 평슈명께도 말 못 할 감사를 전합니다. 기다려주신 세월을 꼭 보답하겠습니다.
我将一生和爱心奉献给您的女儿。

자! 이제, 연구실의 키보드에서 손을 뗄 때가 된 거 같다.
나의 박사과정 때의 후회, 걱정과 회노애락들도 손에서 내려 놓겠다.
그리고, 오세진 박사라는 새로운 인생을 살아가겠다.
그 첫날, 내 영광의 박사학위기를 하늘 높이 올리고 흔들며 가슴을 당당히 펴고 세상에 우렁차게 소리치겠다.

야! 기분 좋다!

본 연구는 2020년도 대한암연구재단 암연구 박사학위논문 저술지원사업 지원에 의하여 수행 되었음.

본 연구에서 사용된 GBM환자 샘플의 WES, RNAseq, 약물반응 및 환자 정보 등은 삼성병원 남도현 교수팀으로부터 제공 되었으며 alignment는 조희진 박사와 이인희 박사가 수행함.

본 연구에서 사용된 단백질체 데이터들은 KIST의 이철주 박사팀으로부터 제공 되었으며 염정훈 박사, 주신영 박사, 권유미 박사들이 정량을 함. Proteomic-variant calling은 한양대의 백은옥 교수팀의 최승혁과 이원엽 대학원생들이 수행함.

본 연구의 Westernblot, invasion assay과 PHGDH-activity assay는 김주화 대학원생이 수행함.

TABLE OF CONTENTS

ABSTRACT.....	1
I. INTRODUCTION	2
II. MATERIALS AND METHODS	5
1. Sample acquisition	5
2. Whole-exome sequencing (WES) data analysis.....	5
3. RNA-sequencing data analysis.....	6
4. Materials for quantitative proteomic analysis.....	7
5. Protein extraction, digestion, and TMT peptide labeling·	7
6. Basic pH reversed-phase liquid chromatography (bRPLC) fractionation of peptides and preparation of global proteome and phosphor-proteome samples.....	9
7. Liquid chromatography and tandem mass spectrometry·	10
8. Peptide and protein identification and quantification.....	11
9. Peptide and protein intensity normalization.....	11
10. Variant peptide identification.....	13
11. Protein isoform analysis.....	14
12. Calculation of RNA to protein correlation.....	15
13. Identification of proteomic subtypes.....	15
14. Gene set enrichment analyses (GSEA) and network analysis	16
15. Processing of public data.....	16
16. Surrogate GBM proteomic cluster (sGPC) subtyping.....	17

17. Survival analysis	18
18. Multiplex fluorescent immunohistochemistry	18
19. Cell lines	19
20. Western blot and antibodies	19
21. cDNA transfection	20
22. 3D spheroid cell invasion assay	20
23. Phosphoglycerate dehydrogenase activity assay	21
24. Analysis of patient-derived cell (PDC) based high throughput drug screening dataset	21
25. Murine-GBM cell model	22
26. Chemical compound library	22
27. Pharmacological characterization	22
28. Statistical analysis	23

III. RESULTS

1. Proteomic data portray glioma disease state	24
2. Two proteomic subgroups of <i>IDH</i> -wildtype GBM	33
3. OXPHOS-related proteins separate proteomic subtypes	37
4. GPC-subtype-dependent expression of cell-of-origin markers	42
5. PHGDH predicts favorable prognosis in <i>IDH</i> -wildtype GBM	45
6. Spatial and cell-type characteristics of GPC subtypes	52
7. Subtype-dependent sensitivities to targeted therapies	57
8. Protein markers inform sensitivities to targeted therapies	61

9. Screening for pharmacological agents suppressing growth and invasion of GBM cell of origin.....	66
IV. DISCUSSION.....	71
V. CONCLUSION.....	75
REFERENCES.....	76
ABSTRACT(IN KOREAN)	81
PUBLICATION LIST.....	82

LIST OF FIGURES

Figure 1. Overall characteristics of samples and consistency between genome and proteome data.....	26
Figure 2. Proteome data reliably captures genomic characteristics of samples.....	27
Figure 3. Proteome data captures distinct molecular features of GBM.....	28
Figure 4. Gene set analysis of proteins elevated in glioma tissue	29
Figure 5. Activated STAT1 signaling in <i>IDH</i> -wildtype tumors	30
Figure 6. <i>IDH</i> -wildtype GBM tumors are separated into two proteomic subtypes.....	34
Figure 7. Proteomic signatures and clinical characteristics of GPC subtypes.....	35
Figure 8. Genomic characteristics of GPC subtypes.....	36
Figure 9. Proteins involved in oxidative-phosphorylation (OXPHOS) are major determinants of <i>IDH</i> -wildtype GBM proteomic subtypes.....	39
Figure 10. Protein expressions in central carbon metabolism.....	40
Figure 11. Expression levels of proteins involved in glycolysis and lactate level.....	41

Figure 12. Protein expression of cellular markers and Arp2/3 complex.....43

Figure 13. Subtype-specific expression of *CD274* (PD-L1) and *PDCD1LG2* (PD-L2) in the five datasets.....44

Figure 14. Prognostic biomarker proteins in *IDH*-wildtype GBM46

Figure 15. Kaplan-Meier plots for three indicated prognostic markers in the three independent cohorts.....47

Figure 16. Expression level of PHGDH is associated with patient survival.....48

Figure 17. PHGDH suppresses invasion of GBM.....49

Figure 18. Correlation between PHGDH expression and 2-hydroxyglutarate (2-HG) levels in 878 *IDH*-wildtype cancer cell lines..... 51

Figure 19. Classification of GBM single cells and cellular marker expressions.....54

Figure 20. Spatial, cellular, and immuno characteristics of GPC subtypes.....55

Figure 21. GPC subtypes are determined by dominant population of cell subtypes56

Figure 22. Correlation tests between the drug-response (AUC and ED50) and the abundance of global proteins58

Figure 23. Protein expressions are the most correlative markers for targeted anti-cancer drugs.....59

Figure 24. Subtype-specific anti-cancer drugs and activated pathways.....	60
Figure 25. Correlation tests between the drug–response (AUC and ED50) and the phosphorylation level of proteins.....	63
Figure 26. An integrative analysis with pharmacological profiles and phosphoproteome data.....	64
Figure 27. PHGDH and FKBP9 levels are associated with sensitivity to AZD2014.....	65
Figure 28. Classification of the 6,869 compounds.....	67
Figure 29. Primary screening results.....	68
Figure 30. Representative images for primary screening.....	69
Figure 31. Candidate pathways that significantly suppressed the invasiveness of SVZ-1.....	70

LIST OF TABLES

Table 1. Clinical and molecular annotation of the SMC1 cohort.....	31
--	----

ABSTRACT

Pharmaco-proteogenomic characterization of glioblastoma

Sejin Oh

*Department of Medical Science
The Graduate School, Yonsei University*

(Directed by Professor Hyun Seok Kim)

Glioblastoma (GBM) is the most malignant brain cancer with extremely limited clinical option. Here, I adopted phamarco-proteogenomic approach to identify core oncogenic pathways and to locate vulnerable points in human GBM with 39 *IDH*-wildtype GBMs as well as *IDH*-mutant and low-grade glioma controls. I found two proteomic GBM subtypes, which are largely separated by oxidative-phosphorylation proteins. Specifically, GBM proteomic cluster 1 (GPC1) tumors exhibit Warburg-like features, poor prognostic signature, and neural stem-cell markers. To locate important pathways related to GPC1 tumor, I conducted high-throughput pharmacological characterization with murine-GBM model derived from glioma-stem-cell (GSC) origin. My research will suggest novel therapeutic strategies for GBM.

Key words : glioblastoma, pharmaco-proteogenomics, tumor evolution

Pharmaco-proteogenomic characterization of glioblastoma

Sejin Oh

*Department of Medical Science
The Graduate School, Yonsei University*

(Directed by Professor Hyun Seok Kim)

I. INTRODUCTION

Glioblastoma (GBM) is the most deadly brain tumor characterized by a high recurrence rate (~90%) and one of the shortest median survival time (15~16 months) across all cancers¹. Currently available therapeutic procedures—namely, gross total resection, followed by a combination of radiotherapy and chemotherapy with DNA-alkylating temozolomide^{2,3}—are largely ineffective. Indeed, 90% of patients treated according to the standard clinical procedures experience tumor recurrence within 6–9 months after initial treatment⁴. Much research has been conducted using high-dimensional molecular data in an attempt to locate major oncogenic events and therapeutically actionable intervention points and to classify GBM patients into diagnostic and prognostic subgroups⁵⁻⁷. The vast majority of these efforts have been confined to the genomic, transcriptomic, and epigenetic level. For example, a notable study performed by The Cancer Genome Atlas (TCGA) group involved sequencing several hundred GBM specimens, from which three major oncogenic signaling pathways were identified: receptor tyrosine kinase/RAS/PI3K, p53 and RB⁷. Based on the expression of 840 genes, the researchers classified GBM into four distinct subtypes: classical (*EGFR* amplification and *CDKN2A* deletion), mesenchymal (*NFI* deletion and expression of mesenchymal markers), proneural (*PDGFRA* amplification, *IDH1* mutation and expression of proneural

development genes), and neural (expression of neuronal markers)⁸. Despite these efforts, these mutation-based and transcriptome-based approaches have found limited clinical application, and only a few biomarkers, including *IDH* mutation (favorable prognoses, secondary GBM)⁹, MGMT promoter methylation (benefit from temozolomide)¹⁰, and 1p/19q co-deletion (chemosensitivity)¹¹ are being used in clinic. Meanwhile, *IDH*-wildtype GBM, which is found in ~90% of all GBM cases, represents the most aggressive glioma subtype¹². Establishing predictive biomarkers or patient stratification strategies for use in developing targeted therapies and identifying determinants of long-term survival of *IDH*-wildtype GBM remain challenges.

In this regard, recent proteogenomic studies conducted by Clinical Proteomic Tumor Analysis Consortium (CPTAC) demonstrated that DNA and RNA alterations are insufficient to predict proteomic activities, which may explain disappointing failures of genomic-driven clinical trials¹³⁻¹⁵. Therefore, proteome-centered multi-omic approach may provide more direct and informative insight to combat GBM.

Extreme heterogeneity is another major challenge in GBM treatment. Recent advances on single-cell sequencing enable researchers to explore GBM heterogeneity at single-cell resolution¹⁶. Using this technology, Patel et al. demonstrated that a wide spectrum of GBM subtypes could present in a single tumor mass¹⁶. Indeed, a longitudinal study has shown that GBM with the higher cellular complexity were more likely to change their subtypes when they relapsed¹⁷. Moreover, spatially distant GBM tumor masses can possess different characteristics in not only subtypes but also oncogenomic architectures, resulting in different efficacies of anti-cancer drugs¹⁷. Therefore, it is imperative to identify the core oncogenic pathway and its vulnerable points shared by multiple clones. Recently, it has been reported that *IDH*-wildtype GBM, the most aggressive type of GBM, originates from neural stem cells in

subventricular zone (SVZ), suggesting that migration process from SVZ to cortex may be a pivotal step in glioblastomagenesis¹⁸.

Here, I adopted pharmaco-proteogenomic approach to identify core oncogenic pathways and to locate vulnerable points in human GBM. I collected genome, transcriptome, proteome and pharmacological profile data of 51 anti-cancer drugs for 50 human glioma samples and conducted multi-omic analysis. I found two proteomic GBM subtypes (GPC1 and 2). Particularly, GPC1 showed poor prognostic signatures including high stem cell marker and proteins related to invadopodia, which is likely originated from neural stem cell in SVZ. To locate important pathways and therapeutic targets related to its migration, I screened 6,859 chemical compounds murine GBM model in SVZ. My research will suggest novel therapeutic strategies for GBM.

II. MATERIALS AND METHODS

1. Sample acquisition

Tumor specimens and their corresponding clinical records were kindly provided by Do-hyun Nam. The materials were obtained from patients with informed consent at Samsung Medical Center (SMC). SMC cohort 1 (N = 50) was used for the proteomic analysis. SMC cohort 2 (N = 106) was used as an independent dataset for validation using RNA-based surrogate proteome signatures, and SMC-TMA cohort (N = 120) was used as an independent dataset for validation by multiplex fluorescent immunohistochemistry. Detailed information about the specimens is provided in table 1. This study was approved by the Institutional Review Board (201004004 and 200504001).

2. Whole-exome sequencing (WES) data analysis

The paired-end reads from FASTQ files were aligned to the UCSC human reference genome assembly (hg19) using Burrows-Wheeler Aligner (BWA version 0.6.2)¹⁹. The duplicated reads were removed using Picard (version 1.73), and local realignment was then performed around known insertions and deletions (indels) using SAMtools (version 0.1.18)²⁰ and the Genome Analysis ToolKit (GATK version 2.5-2)²¹. dbSNP (version 135) was subsequently used for the realignment and recalibration process. The resulting bam files were used for mutation calling and copy number analysis. SNVs were identified using MuTect (version 1.1.4)²² with the following criteria: a Phred score > 15 and coverage > 20 in both the tumor tissue and the matched normal tissue. Indels were identified by SomaticIndelDetector (GATK version 2.2) based on a mapping quality score > 15 and a coverage > 10 in both the tumor tissue and the matched normal tissue. The variants were annotated using dbSNP (version 135), 1000 Genomes Project (Phase I) and Exome Sequencing Project (ESP6500SI-V2). The resulting variants were annotated using the Variant Effect Predictor (VEP version 37.75)²³. After

removing synonymous variants, the nonsynonymous and splicing variants were taken forward for further study. Hotspot mutations were annotated using databases downloaded from www.cancerhotspots.org (Hotspot Results V1) and www.3dhotspots.org (3D Hotspot Results)^{24,25}. For the copy number analyses, BAM files from WES of 50 glioma (39 *IDH*-wildtype GBM) and 24 matched normal blood samples were used to generate gene-based read-count matrices using Bedtools, according to the Ensembl gene table (version 37.75). Subsequently, 1 was added to the read counts to prevent negative infinity values for the \log_2 transformation. Genes with mean read counts < 20 were filtered out. The read counts were normalized to counts per million (CPMs) using the edgeR package (version 3.20.9)²⁶. The normalized values of the tumors were divided by that of the matched normal samples to calculate the \log_2 -ratio. A circular binary segmentation algorithm was implemented with the \log_2 -ratio values using the DNACopy package (version 1.52.0). Genomic identification of significant targets in cancer (GISTIC) 2.0 using the default parameters was applied to the segmented data to identify regions that were frequently altered in DNA copy number²⁷.

3. RNA-sequencing data analysis

GSNAP (version 2012-12-20) was used to align the reads of SMC cohort 1 to the reference genome GRCh37, and STAR (version 2.5.4b) aligner was applied to SMC cohort 2. The normalized gene expression values in fragments per kilobase of exon per million fragments mapped (FPKMs) were quantified by Cufflinks (version 2.0.2). Finally, the (FPKM + 1) values were transformed to the \log_2 scale. To determine the RNA subtype, single-sample gene set enrichment analysis (ssGSEA)²⁸ was applied to the Z-score normalized expression data to calculate enrichment scores (ESs) for RNA subtypes as defined by Verhaak et al. The subtype with the highest ES was used as the representative subtype for each sample.

4. Materials for quantitative proteomic analysis

A sequencing-grade modified trypsin/LysC mix was purchased from Promega (Madison, WI, USA), and Tris (2-carboxyethyl) phosphine hydrochloride (TCEP-HCl), tandem mass tag (TMT) isobaric reagents and a Pierce graphite spin column were purchased from Thermo Fisher Scientific (Waltham, MA, USA). Water and organic solvents were obtained from J.T. Baker (Center Valley, PA, USA). Titansphere™ Phos-TiO (10- μ m bulk) was purchased from GL Science (Tokyo, Japan).

5. Protein extraction, digestion, and TMT peptide labeling

Tumor and adjacent normal tissue samples were carefully washed in PBS buffer on ice to remove the blood and then individually cryo-pulverized using a cryoPREP device (CP02, Covaris, Woburn, MA, USA). Each tissue specimen (32-243 mg in total wet tissue weight) was placed in a cryovial (430487, Covaris, Woburn, MA, USA) on dry ice, transferred to a Covaris tissue bag (TT1, Covaris, Woburn, MA, USA), placed into liquid nitrogen for 30 s and then pulverized at impact level 3. The tissue powder from each tissue was then placed in a sonication tube (002109, Covaris, Woburn, MA, USA) and mixed with lysis buffer [8 M urea, 0.1 M NH_5CO_3 , 1 mM PMSF and 1X phosphatase inhibitor cocktail (Thermo Scientific, Pittsburgh, PA, USA)]. The lysis buffer volume varied depending on the total tissue weight (ca. 1 mL for 20 mg). Tissue lysis was performed by sonication using a focused ultrasonicator (S220, Covaris, Woburn, MA, USA) at a setting of 2 W (intensity of 5) for 5 s followed by 36 W (intensity of 10) for 20 s and 0 W (intensity of 0) for 10 s. The sonication cycle was repeated 20 times at 16°C. The homogenate was centrifuged at 16,000 g and 20°C for 10 min (5810 R, Eppendorf), and the supernatant was transferred to a new tube. The protein concentration was determined by BCA protein assay (Thermo Fisher Scientific, Waltham, MA, USA).

The proteins (700 μ g) were subjected to disulfide reduction with 5 mM Tris (2-

carboxyethyl) phosphine (TCEP) at room temperature for 2 hrs and alkylation with 15 mM iodoacetamide at room temperature for 1 hr in the dark. Subsequently, the samples were diluted 10-fold with 0.1 M NH_5CO_3 to reduce the concentration of urea to 0.8 M. The protein sample was digested overnight at 37°C using a Trypsin/LysC protease mixture at a 1:25 enzyme-substrate ratio. The digested samples were allowed to cool at room temperature, and the digestion was quenched by acidification with trifluoroacetic acid (TFA) at a final concentration of 0.5%. The sample was subsequently purified/desalted through HLB solid-phase extraction (SPE) (Sep-Pak, Waters, Milford, MA, USA), dried *in vacuo* and stored at -20°C until further use.

The dried peptides were resuspended in 0.1M Triethylammonium bicarbonate buffer (18597, Merck, Munich, Germany), and the peptide concentration was determined using a Nanodrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) at 280 nm wavelength. An aliquot equivalent to 500 μg of each sample was immediately labeled with 4 mg of each TMT channel, except for TMT 126, which was prepared according to the manufacturer's instructions. For the first batch (sets 1-6), samples from sets 1 through 6 were combined to obtain a global internal standard 1 (GIS1); this standard was labeled with TMT channel 126. GIS2 was used for the second batch (sets 7-11), which was obtained by pooling sets 7 to 11. Following incubation at room temperature for 1 hr, the reaction was quenched with hydroxylamine at a final concentration of 0.3% (v/v). The TMT-labeled samples were pooled at a 1:1:1:1:1 ratio. The combined sample was subsequently purified/desalted using HLB-SPE, dried *in vacuo*, and stored at -20°C until further use. The combined sample comprising the first batch was dried *in vacuo* and subsequently desalted using HLB-SPE. By contrast, the combined sample for the second batch was directly desalted and subsequently dried *in vacuo* because drying samples in the presence of hydroxylamine is detrimental to phospho-peptides²⁹.

6. Basic pH reversed-phase liquid chromatography (bRPLC) fractionation of peptides and preparation of global proteome and phospho-proteome samples

The TMT-labeled peptides were fractionated by bRPLC using an Agilent 1290 Infinity LC System (Agilent Technologies, Santa Clara, CA, USA). Chromatography was performed with an XBridge BEH130 C18 column (4.6- μm i.d. \times 250-mm length; pore size of 130 Å and particle size of 3.5 μm ; Waters Corporation, Milford, MA, USA) at a flow rate of 0.5 mL/min. The mobile phases were 10 mM NH_4HCO_2 (pH 10) as phase A and 10 mM NH_4HCO_2 (pH 10) in 90% ACN (pH 10) as phase B. The peptides were dissolved in 110 μL mobile phase A and then injected into a 100- μL sample loop. The gradient was 2-5% B for 10 min, 5-40% B for 40 min, 40-70% B for 15 min, 70% B for 10 min, and 70-5% B for 15 min. Fractionation was performed by collecting 84 tubes (0.8 min/tube) throughout the chromatographic run. Eighty-four fractions were pooled to obtain 12 concatenated fractions based on the following rule: a set of an arithmetic sequence with a common difference of 12 was pooled into one concatenated fraction; for instance, fractions with numbers 1, 13, 25, 37, 49, 61 and 73 were pooled to generate concatenated fraction 1. A total of 5% of the volume of each fraction was allocated to global proteome analysis and dried. The remaining 95% of the concatenated fractions were further combined into 12 fractions, and the flow-through fractions from bRPLC were also combined into one fraction for phospho-peptide enrichment and dried. For all experiments, the phospho-peptides were subjected to metal oxide affinity chromatography (MOAC) using titanium dioxide beads (10 μm , Titansphere Phos-TiO Bulk)^{30,31}. The dried peptide and TiO₂ beads were preincubated separately in a solution of 3.45 M lactic acid (302 mg/ml), 60% ACN and 0.3% TFA (one fraction of peptide in 100 μL of the solution; 2 mg of beads in 10 μL of the solution). The two preincubated mixtures were combined and further incubated for 30 min at 25°C with agitation. After incubation, the beads enriched with phospho-peptides were collected by

centrifugation, and the unbound supernatant from the three fractions was pooled into one fraction for double TiO₂ enrichment. The beads were washed with 1% TFA in 30% ACN and loaded onto a C8-plugged tip (Diatech Korea, Seoul, Korea). The bound phospho-peptides were eluted with 1.5% NH₄OH and then with 5% pyrrolidine in a single tube. The eluates were directly acidified with 1% TFA and desalted using graphite spin columns (88302, Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions. The phospho-peptides were dried and resuspended in 0.4% acetic acid.

7. Liquid chromatography and tandem mass spectrometry

The dried peptide samples were reconstituted in 0.4% acetic acid, and an aliquot containing ~1 µg of the sample was injected from a cooled (10°C) autosampler into a reversed-phase Magic C18aq (Michrom BioResources, Auburn, CA, USA) column (20 cm × 75 µm, packed in-house) on an Eksigent nanoLC-ultra 1D plus system at a flow rate of 300 nL/min. Before use, the column was equilibrated with 90% buffer A (0.1% formic acid in water) and 10% buffer B (0.1% formic acid in acetonitrile). The peptides were eluted with a linear gradient from 5% to 40% buffer B over 100 min and 40% to 80% buffer B over 5 min and then subjected to an organic wash and aqueous re-equilibration at a flow rate of 300 nL/min with a total run time of 130 min. The HPLC system was coupled to a Q-Exactive mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) operated in data-dependent acquisition (DDA) mode. Survey full-scan MS spectra (m/z 400–2000) were acquired at a resolution of 70,000. The source ionization parameters were as follows: spray voltage, 2.5 kV; capillary temperature, 300°C; and s-lens level, 44.0. The MS/MS spectra of the 12 most intense ions from the MS1 scan with a charge state of 1-5 were acquired with a fixed first m/z of 120 along with the following options: resolution, 17,500; automatic gain control (AGC) target, 1E5; isolation width, 2.0 m/z; normalized collision energy, 27%; dynamic exclusion duration, 90 s; and ion selection threshold, 4.00E+03 counts.

8. Peptide and protein identification and quantification

Peptide and protein identification and quantification were performed using MaxQuant (version 1.5.6.0)³². The mass spectrometry raw files were searched against the Swiss-Prot human database (released in March 2014; <http://www.uniprot.org>) using the Andromeda search engine included in MaxQuant. The following MaxQuant search parameters were used: semispecific trypsin was selected as the enzyme; the carbamidomethylation of cysteine was set as a fixed modification; N-terminal protein acetylation and oxidation (M) were set as variable modifications; and phosphorylation (STY) was set as a variable modification for phosphorylation-enriched samples. The reporter ion was set as six-plex TMT for quantification. Peptide matches were filtered by a minimum length of eight amino acids and no miscleavages were allowed. The false discovery rate (FDR) was set to 0.01 at both the protein and peptide spectrum match (PSM) levels. Proteins identified by at least two unique peptides were used. For protein quantification, the minimum ratio count was set to two, and the peptide for protein quantification was set as unique. Other settings were kept at their default values. In total, 9,367 protein groups, 179,234 stripped peptides, and 2,750,407 peptide spectral matches (PSMs) were identified from the global proteome. In the case of the phosphoproteome, 8,019 phosphorylation sites, 16,377 phosphorylated peptides, and 276,153 PSMs were identified. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository with the dataset identifier PXD015545.

9. Peptide and protein intensity normalization

The protein intensity of a sample was divided by the corresponding global internal standard (GIS) and converted to the \log_2 scale. The normalized \log_2 intensities were then sample-wise median centered across all proteins before protein-wise median centering across samples belonging to a GIS batch. Finally,

the normalized abundance values obtained from two GIS batches (30 samples from batch 1, 24 samples from batch 2) were combined. Samples with no intensity values were arbitrarily given the second smallest value of all samples to avoid negative infinity. A total of 8,034 proteins quantified in all GISs were used for analysis. To convert the abundances of the phospho-peptides to that of phospho-sites, the mass spectral intensity values from all the phospho-peptides containing a particular phospho-site (s) were averaged at the levels of mono-, di-, or tri- or more phosphorylated peptide signal intensities, respectively. These three levels of intensity values (I_{sn}) for a single phospho-site were subjected to the following equation:

$$I_s = \sum_{n=1}^3 \frac{I_{sn}^2}{\sum_{n=1}^3 I_{sn}}$$

, where I_s represents a weighted squared sum of the intensity value for a phospho-site (s). This weighting was performed to give more weight to the phosphorylation class providing higher intensities, and the I_s values were then further \log_2 transformed. For the samples with no intensity value for a phospho-site, the second minimum value of all the samples was arbitrarily set to avoid negative infinity. Due to the relatively high sparsity of phosphorylation data, phospho-peptides (N = 11,346) quantified in at least three GISs (approximately 10 samples or more) were considered for the quantification of the phospho-site level.

Based on the observation that the global and phosphoprotein abundance data of the six subgroups (normal, tumor, GPC1, GPC2, *IDH*-wildtype, *IDH*-mutant, low-grade glioma (LGG), and GBM) largely follow a Gaussian distribution, as determined by the Shapiro-Wilk test, Student's t-test was used to identify differentially expressed proteins (DEPs) and differentially expressed phosphoproteins (DEPPs) from the four comparisons (tumor vs. normal, GPC1 vs. GPC2, GBM vs. LGG, and *IDH*-wildtype vs. mutant). The DEPs and DEPPs whose *P*-values were < 0.05 with their corresponding FDR scores were selected

for further analyses. The DEPPs were further filtered when these were found as DEPs in the same direction in the comparison.

10. Variant peptide identification

To determine the overlaps between the variants called from the WES data and those identified from the mass spectrometry data, variant peptides were identified using a multistage search approach³³ and a unified protein database. A unified protein database consists of both a reference and a sample-specific protein database. The Swiss-Prot human protein database (version 2014/03) was used as the reference protein database. A sample-specific protein database was constructed using the following four types of information: sample-specific protein expression, sample-specific genomic variations, fusion gene prediction, and common contaminants. For this database, the transcript models in Ensembl 75 (released in February 2014) whose FPKM values were > 1 were used. The sample-specific variant peptide database was built by applying SNVs and indels called from the WES data and observed in the RNA-seq data to the transcript models. The resulting RNA sequences were then converted into amino acid sequences allowing up to three missed cleavages on both sides, in accordance with the methodology used in our previous study³⁴. Sometimes, SNVs and indels can result in stop gain or stop loss. The variants resulting in stop gains were ignored because it was impossible to distinguish the variant peptides whose translations were stopped by missed cleavage parameters or a novel stop codon. Conversely, if a stop loss occurred, the sequences were translated if 1) up to 20 amino acids whose read depths at the translated positions were > 3 or 2) until a new stop codon was found. Due to the constraint on the read depth, short peptides could be generated due to early termination. To prevent short peptide generation, up to five missed cleavages in the N-terminus direction were allowed during translation. For the sample-specific fusion gene database, fusion genes were predicted from the RNA-seq data and translated from the fusion junction in both directions based on the

frame of the upstream gene.

To identify variant peptides, the MS/MS spectra were searched using a multistage approach³³, which consisted of the sequential application of MS-GF+³⁵ with a unified database and the in-house software MODplus with the unified database over the set of spectra that were not identified during the first stage. The parameters used for the MS-GF+ search were as follows: precursor error tolerance, 10 ppm; isotope error range, -1 to 2; fragmentation method, HCD; and instrument, Q-Exactive; variable posttranslational modification (PTM), M-oxidation; and two fixed modifications, C-carbamidomethyl and K/N-term-TMT. The second-stage MODplus search focused on the identification of a variety of PTMs. The MODplus search parameters were the following: precursor error tolerance, 10 ppm; fragment ion tolerance, 0.025 Da; isotope error range -1 to 2; instrument, QTOF (equivalent to Q-Exactive with HCD); 46 variable PTMs (this parameter allows identification of modified peptides with multiple modifications within a range of -480 to 470 Da and two fixed modifications, C-carbamidomethyl and K/N-term-TMT. Furthermore, MODplus identified modified peptides with multiple modifications within a given mass range, which is a user-specific parameter. All search steps were executed with the trypsin enzyme, TMT protocol, and semi-tryptic search. In each step, peptides < 8 amino acids and an estimated 1% FDR were discarded using a target-decoy approach. The MODplus search considered De-TMT modifications that discriminated TMT-free PSMs. The TMT-free PSMs were discarded because it was unable to determine in which sample the peptide was expressed. The overlapping somatic variants with WES data were identified based on their genomic coordinates. Finally, the Jaccard coefficient was calculated.

11. Protein isoform analysis

In total, 18 protein isoforms (sharing a gene symbol) were quantified across the GISs and had different quantification values; these corresponded to nine genes

(CAPZB, EPB41L3, IKBIP, MAP2, MAP4, PFN2, PKM, RTN1 and SNX32). Among the 18 isoform groups, only PKM isoforms exhibited mutually exclusive expression patterns in two GPC subtypes: PKM1 was significantly elevated in GPC2, whereas PKM2 was in GPC1. For the comparison of PKM isoforms, we selected peptides belong either to PKM1-specific exon 9 or PKM2-specific exon 10. The peptide intensities were normalized based on the corresponding GISs and transformed to the \log_2 scale. The average peptide expression values were then calculated according to the specific sequences for either PKM1 or PKM2. Five PKM1 isoforms (*ENST00000319622*, *ENST00000389093*, *ENST00000565154*, *ENST00000565184*, and *ENST00000568459*) and one PKM2 isoform (*ENST00000335181*) were used. The average expression of the isoforms was used for PKM1.

12. Calculation of RNA to protein correlation

The Ensembl gene IDs and UniProt IDs were converted into an official gene symbol using the BioMart (version 2.34.2), org.Hs.eg.db (version 3.5), EnsDb.Hsapiens.v75 (version 2.99), UniProt.ws (version 2.18) and rentrez (version 1.2.1) packages. The Spearman correlation coefficients and *P*-values for 4,071 genes were calculated using the R function `cor.test`.

13. Identification of proteomic subtypes

Multiple protein IDs with identical numeric values across all samples, which likely indicate isoforms, were grouped into one ID to prevent undesirable effects from redundant protein IDs. Subsequently, a consensus clustering algorithm was applied to the protein expression matrix with default parameters and 1,000 iterations. The optimal number of clusters was determined by assuming that the *K* value represents the minimal PAC score³⁶.

14. Gene set enrichment analyses (GSEA) and network analysis

Gene set analysis (GSA) was conducted with official symbols from the top 10% PC1 high loadings and DEPs/DEPPs. A total of 9,404 unique pre-annotated gene sets obtained from the CORUM (version 3.0), C2, and C5 MSigDB were used after removing redundant, too-small (< 5), and too-large (> 150) gene sets. Gene sets with at least four hits were then selected from the given symbol lists. The gene symbols corresponding to all global proteins or all global and phosphoproteins were used as the background for PC1-GSA and DEPs/DEPPs-GSA, respectively. The statistical significance of the gene set enrichment was evaluated based on the hypergeometric test, followed by multiple testing correction with the false discovery rate (FDR). Gene sets with FDR < 1% were selected for normal/tumor, whereas those with FDR < 10% were selected for PC1-loadings to examine the relationships between gene sets using the EnrichmentMap algorithm based on an overlap coefficient cutoff of 0.5³⁷. For the DEPs and DEPPs, the signaling network, enzyme-substrate interactions, and transcription factor-target interactions were further analyzed using the literature-curated OmniPath database³⁸. BRCAness score was estimated at the mRNA level using the methods described by Konstantinopoulos et al.³⁹

15. Processing of public data

Gene expression and metadata of TCGA were downloaded using the TCGAbiolinks package (version 2.6.12)⁴⁰. The normalized single-cell gene expression file (GBM_normalized_gene_counts.csv) was downloaded from <http://gbmseq.org>. After 1 was added to each gene expression value from the TCGA data, the values were transformed to the log₂ scale. Mutation (CCLE_DepMap_18Q1_maf_20180207.txt), gene expression (CCLE_RNAseq_genes_rpkm_20180929.gct.gz), and metabolome data (CCLE_metabolomics_20190502.csv) of CCLE were downloaded from <https://portals.broadinstitute.org/ccle/data>. Microarray data for the Yonsei (GSE131837)⁴¹ and ANOCEF (E-TABM-898)⁴² cohorts were downloaded from GEO

(<https://www.ncbi.nlm.nih.gov/geo/>) and EBI (<http://www.ebi.ac.uk/microarray-as/ae>), respectively.

16. Surrogate GBM proteomic cluster (sGPC) subtyping

To extend my analysis to larger cohorts lacking proteome data, I constructed a random forest model using the differentially expressed genes (DEGs) between GPC1 and GPC2 samples and predicted surrogate GPC subtypes (sGPCs) of *IDH*-wildtype GBM tumor samples in a separate independent cohort (N = 106, SMC cohort 2) and TCGA cohort (N = 149). DEGs between two GPC subtypes in *IDH*-wildtype GBMs of SMC cohort 1 were obtained by Student's t-test, and the top 100 DEGs were selected based on the *P*-values. The individual RNA expression data of the other cohort were then merged with the corresponding SMC cohort 1 data, and the merged matrix was subjected to quantile normalization to reduce the batch effects. The random forest model of the normalized SMC cohort 1 was trained using the randomForest package (version 4.6-14), and the optimal parameters were selected using the caret package (version 6.0-8) with 1,000 iterations of five-fold cross-validation. The model was applied to the normalized SMC cohort 2 and samples with a GPC subtyping probability of at least 60% were used for downstream analyses.

To address the sparsity issues in the Darmanis single-cell data⁴³, DEGs were identified from the 1,000 genes with the highest expression in the SMC 1 cohort and those that were expressed in at least 50% of all single cells. To predict the sGPC subtype of each single cell, an equal number of DEGs (FDR < 10%) were selected from both sides. Likewise, an equal number of DEGs (100 GPC1-high and 100 GPC2-high) from the CCLE, Yonsei, and ANOCEF cohort data were selected according to Student's t-test *P*-values. The selected gene expression values were converted into Z scores. ΔZ score was calculated using the following equation:

ΔZ score = mean(Z score of GPC1 high) - mean(Z score of GPC2 high)

The *P*-value for the ΔZ score was estimated by 1,000X permutations of the gene labels. The surrogate GPC subtype was determined if a single cell had permutation *P*-values < 0.05.

17. Survival analysis

Multi-samples with different GPC subtypes were excluded to prevent the effects of multiple samples on the survival analysis. If samples shared a GPC subtype, they were considered a single sample. The survival analysis was performed using the survival package in R (version 2.42-6). 271 previously reported gene expression-based prognostic markers⁴⁴ were validated in the SMC1 cohort by univariate Cox regression analysis of survival rate under the proportional hazards assumption using the *coxph* function in ‘survival’ package (version 2.43-3) of R. If there are multiple samples per patient, a mean expression value was used for a protein.

18. Multiplex fluorescent immunohistochemistry

The tissue microarray (SMC-TMA) consisted of 120 tissue samples that were formalin fixed, paraffin embedded (FFPE), and sectioned (2-mm thickness): 6 normal samples, 35 low grade gliomas, 1 *IDH*-mutant GBM, 3 *IDH*-status unknown GBMs, and 75 *IDH*-wildtype GBMs, including 14 SMC2 tumors. The FFPE tissues on slides were deparaffinized and rehydrated for multiplex immunohistochemistry staining. Epitope retrieval was performed using BOND Epitope Retrieval Solution 2 kits (AR9640, Leica Biosystems, Wetzlar, Germany). Immunofluorescent signals were visualized using the OPAL 7-Color automation IHC kit (NEL82100KT, Akoya, Marlborough, MA, USA), TSA dyes 570 (PHGDH; 1/1000, AB_1855299, Atlas Antibodies, Bromma, Sweden), 690 (Nestin; 1/700, AB_1854381, Atlas Antibodies, Bromma, Sweden), and spectral

DAPI. The stained slides were coverslipped using HIGHDEF® IHC fluoromount (ADI-950-260-0025, Enzo, New York, NY, USA) and scanned using a Vectra® 3.0 Automated Quantitative Pathology Imaging System (PerkinElmer, Waltham, MA, USA). Color separation, cell segmentation, and cell phenotyping were performed on inForm Advanced Image Analysis software (version 2.2, PerkinElmer, Waltham, MA, USA) to extract image data. PHGDH and Nestin positive cells were determined by thresholds of 0.6 (PHGDH) and 1.25 (Nestin), respectively. The experiment was conducted by prismCDX in Korea.

19. Cell lines

All GBM cancer cell lines, except for KNS81 and two U87MG isogenic pairs, were purchased from the Korea Cell Line Bank (KCLB, Seoul, Korea). KNS81 cells were obtained from the JCRB cell bank. U87MG and U87MG-IDH1-R132H cells were purchased from ATCC (Manassas, VA, USA). Cells were grown in RPMI-1640 medium (11875-093, Gibco, Waltham, MA, USA) supplemented with 10% fetal bovine serum (16000-044, Thermo Fisher Scientific, Waltham, MA, USA) and 1% penicillin–streptomycin (15140122, Thermo Fisher Scientific, Waltham, MA, USA). The absence of mycoplasma contamination was confirmed in all cell lines by e-Myco VALiD Mycoplasma PCR detection kits (25299, iNtRON, Seongnam, Korea).

20. Western blot and antibodies

Cells were lysed in RIPA buffer (R0278, Merck, Munich, Germany) with protease inhibitor cocktail (P3100, Genedepot, Katy, TX, USA) and phosphatase inhibitor cocktail (P3200, Genedepot, Katy, TX, USA). Protein concentrations were determined by the Bradford protein assay (500-0006, BIORAD, Hercules, CA, USA), and equal amounts of protein were loaded and separated in sodium dodecyl sulfate-polyacrylamide gels (SDS-PAGE). Proteins were then transferred to nitrocellulose membranes (1620177, BIORAD, Hercules, CA, USA). After

blocking with 5% skim milk, the membranes were probed with primary antibodies. Antibodies used in this study were as follows: STAT1 (1/1000, AB_2799965, Cell Signaling Technology, Danvers, MA, USA); pSTAT1 Serine-727 (1/1000, AB_2773718, Cell Signaling Technology, Danvers, MA, USA); PHGDH (1/1000, AB_2737030, Cell Signaling Technology, Danvers, MA, USA); Nestin (1/1000, AB_10859398, Abcam, Cambridge, UK); FKBP9 (1/1000, AB_11005959, Novus Biologicals, Centennial, CO, USA); β -actin (1/1000, AB_2223172, Cell Signaling Technology, Danvers, MA, USA); IDH1-R132H (1/1000, AB_2335716, DIANOVA, Barcelona, Spain); and IDH1 (1/1000, AB_10950504, Cell Signaling Technology, Danvers, MA, USA). After washing three times, the membranes were incubated with secondary antibodies. Band signals were developed with ECL buffer (32106, Pierce, Waltham, MA, USA).

21. cDNA Transfection

The pCMV6-GFP-PHGDH (RG203949) was purchased from OriGene (Rockville, MD, USA). A172 and SNU201 cells were seeded at 300,000 cells per well in 6-well plates. After overnight incubation at 37°C, cells were transiently transfected with 2 μ g of PHGDH cDNA or an empty vector using Lipofectamine 2000 (1668019, Invitrogen, Waltham, MA, USA) according to the manufacturer's instructions. After 24 hrs, PHGDH expressing cells were collected by trypsinization and resuspended in spheroid forming ECM solution (3500-096-K, R&D Systems, Minneapolis, MN, USA).

22. 3D spheroid cell invasion assay

3D invasion assay was performed using Cultrex 3D Spheroid Cell Invasion Assay (3500-096-K, R&D Systems, Minneapolis, MN, USA) according to the manufacturer's instructions. Briefly, resuspended cells in 1X spheroid forming ECM solution were seeded at 4,000 cells per well in a 96-well plates and incubated at 37°C for 48 hrs for spheroid formation. Once the spheroids were formed, they

were embedded in an invasion matrix and supplemented with culture medium containing NCT-502 or DMSO. 3D spheroid invasion assay plates were incubated for 72 hrs. Invading spheroids were photographed using an inverted phase-contrast microscope (IX73, Olympus, Tokyo, Japan) with CellSence standard 1.15 software (Olympus, Tokyo, Japan) at 4x magnification. For quantification of spheroid invasion, distances of invading cells were measured by ImageJ software. For this, the three longest protrusions or migrated single cells from each of the quadrants were combined, and the median value of the 12 distances per condition were used to compare the invasiveness of the spheroids.

23. Phosphoglycerate dehydrogenase activity assay

Basal PHGDH activities of the nine GBM cell lines were measured by PHGDH activity assay kits (K569, BioVision, Milpitas, CA, USA) according to the manufacturer's instructions. Briefly, 10⁶ cells were lysed and centrifuged at 10,000g for 5 min at 4°C. The supernatant was mixed with saturated 4.32M ammonium sulfate (7096, BioVision, Milpitas, CA, USA) to remove interferences. The mixture was then placed on ice for 30 min and centrifuged at 10,000g at 4°C for 10 min. The pellet was resuspended in the assay buffer. PHGDH induced changes in the probe signals were measured by EnVision 2105 (PerkinElmer, Waltham, MA, USA) at OD 450nm. HS683 and SNU1105 cells were treated with NCT-502 (HY-117240, MedChemExpress, Monmouth, NJ, USA) or DMSO for 2 hrs before measuring enzyme activity.

24. Analysis of patient-derived cell (PDC) based high throughput drug screening dataset

Drug screening data were kindly provided by Do-hyun Nam. The normalized data per drug per concentration were smoothed by removing data points with an abnormally high viability value (> 1.5 and $> 3 \times$ interquartile range from the third quartile viability value of all PDCs) based on the assumption that the anti-cancer

drugs in my panel are not expected to increase the proliferation of PDCs significantly. If a PDC showed no related data for any drug concentration, the pair was not used for further studies. Area under the curve (AUC) values were calculated using the trapezoidal method: a low AUC indicated high cell line sensitivity to the drug. ED50 values were calculated using the drc package (version 3.0-1) with a four-parameter log-logistic fit. If the imputed ED50 values were higher than the highest tested dose (20 μ M), 20 μ M was assigned as the ED50 of the drug. Statistically significant associations between phosphoproteins and drug responses across 50 PDCs were identified by subjecting all the phosphoproteins to a Spearman's correlation test with the ED50 and AUC drug response data.

25. Murine-GBM cell model

The murine GBM cell (*Trp53-/Pten-/EGFRvIII*) and normal control cell models were kindly provided by Seok-Gu Kang. The cells were generated as previously described¹⁸. Cells were cultured in DMEM/F12 (10-090-CVR, Corning, Corning, NY, USA) supplemented with 100 μ g/ml of FGF (C046, Novoprotein, Shanghai, China), 100 μ g/ml of EGF (C029, Novoprotein, Shanghai, China), 20% of B27 (17504044, Thermo Fisher Scientific, Waltham, MA, USA) and 1% penicillin-streptomycin (15140122, Thermo Fisher Scientific, Waltham, MA, USA).

26. Chemical compound library

Bioactive compounds (#L1700) and investigational anti-cancer compounds (#L2000) were purchased from Selleckchem (Houston, TX, USA) and supplied as 10 mM solutions in DMSO. In addition to these library, I added 10 compounds; W54011, Gossypol, Streptonigrin, AJ-D, NBQX, D-AP5, Picrotoxin, IM156, KN612, and Gabazine.

27. Pharmacological characterization

Sub-cultured cells were seeded onto M-shaped ultra-low attachment 96-well

plates (MS-9096MZ, S-Bio, Hudson, NH, USA) at 5,000 cells per well. 24hrs after seeding, spheroids were embedded in 2.5 μ g/ μ l of growth factor reduced matrigel (356231, Corning, Corning, NY, USA) and centrifuged at 4°C and 1,200 rpm for 3 min. Next, 1mM (primary/counter screen) of the pharmacological compound prepared in DMSO were robotically added to cells by using a BioMek FXP liquid handler (Beckman Coulter, Brea, CA, USA), thus yielding final drug concentrations 10 μ M and a DMSO control (0.5%). The spheroids were further incubated at 37°C in a 5% CO₂ incubator for 72 hrs and their images were obtained by Operetta CLS (PerkinElmer, Waltham, MA, USA).

28. Statistical analysis

All statistical analyses were conducted in the R language environment.

III. RESULTS

1. Proteomic data portray glioma disease state

Previously genomic, transcriptomic, and pharmacological profiling have been conducted on 39 *IDH*-wildtype GBM samples as well as two *IDH*-mutant GBM and nine low-grade glioma (LGG) by the Samsung Medical Center (SMC). These samples displayed genetic portrait of major driver mutations⁵, including *EGFR*, *EGFRvIII* (deletion in exon 2–7), *TP53*, *RB1*, *PTEN*, and *PIK3CA* (Figure 1a), and copy number alterations (CNAs) in *CDKN2A/TP53* (deletion) and *EGFR/PDGFR*A (amplification) (Figure 1b), indicating that these samples represented the GBM mutational spectrum. The samples also spanned all four RNA subtypes⁸ (Figure 1a). 20 out of 50 samples were obtained redundantly from multiple regions or at different time points and had different properties regarding mutation, RNA subtype, 5-aminolevulinic acid (5-ALA) positivity, location (locally adjacent or core and margin of tumors), or primary/relapse status (Table 1). Unsupervised clustering showed that samples from the same patient showed a high degree of DNA-level similarity (Figure 1a).

To gain insight into GBM at the proteomic level, their proteome data was generated by KIST using liquid-tandem mass spectrometry (LC-MS/MS) with six-plex tandem mass tag (TMT) reagents. I first measured the global proteome and phosphoproteome levels relative to pooled global internal standards (GIS). I obtained measurements of 9,367 protein groups and 8,020 phosphorylated amino acid sites for all 50 tumor samples and four normal brain tissues from 11 TMT sets where samples were randomly assigned (Figure 2a). Each TMT set possessed an average of 6,294 protein groups and 2,796 phosphorylation sites. I selected for further analysis 3,909 protein groups that were quantified in all GIS and localized to all cellular compartments as well as 4,489 phospho-sites quantified in three or more GIS. Significant overlap between the single amino acid variants (SAVs) detected in this study by LC-MS/MS and the previously annotated single

nucleotide variants (SNVs; Figure 2b) indicated that our proteomic assay could successfully detect mutant proteins. In addition, the protein expression levels were generally positively correlated with the RNA levels, except for genes involved in certain housekeeping functions, including genes associated with ribosomes and oxidative phosphorylation (OXPHOS—the mitochondrial process through which ATP is synthesized via the electron transport chain coupled to substrate oxidation; Figure 3a), which is concordant with the findings of previous studies on other tumor types¹³⁻¹⁵.

Next, I conducted unsupervised hierarchical clustering with global proteome data to evaluate the degree of similarity between samples. Interestingly, samples were clustered primarily by RNA subtypes ($P < 0.001$; permutation test) or clinical phenotypes, such as 5-ALA positivity, tumor grade and primary/relapse status ($P < 0.05$; permutation tests) (Figure 3b), while DNA-level similarity merely clustered samples according to their origins. This result suggests that proteome might better represent the disease state and underlying biology than genome. Given the high quality of the proteome data, I identified differentially expressed proteins by comparing glioma samples with adjacent normal tissue samples. As expected, compared to normal tissues, gliomas showed elevated levels of proteins involved in cell proliferation and immune responses (Figure 4). I also compared *IDH*-mutant ($N = 6$; 2 grade IV and 4 low grade) and *IDH*-wildtype ($N = 44$; 39 grade IV and 5 low grade) gliomas, because the latter are associated with a significantly worse prognosis than the former. Compared with *IDH*-mutant gliomas, *IDH*-wildtype gliomas displayed significantly elevated phosphorylation level of signal transducer and activator of transcription 1 (STAT1) at serine-727, which is a marker of STAT1 activation⁴⁵, correlated with elevated target protein levels in *IDH*-wildtype tumors (Figure 5a). As shown in an *IDH*-wildtype and mutant U-87MG cell line pair, *IDH*-mutation status directly influences pSTAT1-S727 levels (Figure 5b). Collectively, my results suggest that quantitative proteomic data can accurately reflect clinical phenotype and provide additional insight into the activated pathway.

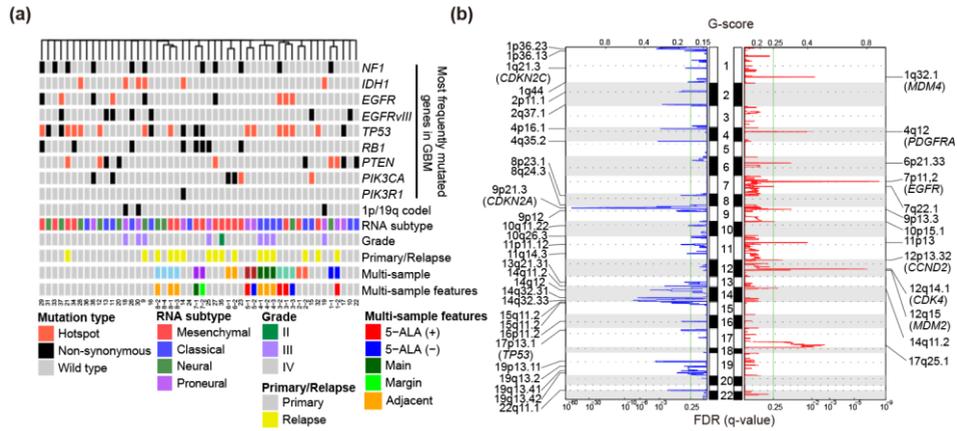


Figure 1. Overall characteristics of samples and consistency between genome and proteome data. (a) Characteristics of *IDH*-wildtype GBM (N = 39), *IDH*-mutant GBM (N= 2) and low-grade glioma (N = 9) tissue samples. Unsupervised hierarchical clustering with complete linkage was used to cluster samples based on the 1 – Jaccard coefficient as the distance metric. The type of mutations in the 8 most frequently mutated GBM genes are color-coded according to the legend. The multi-sample row displays multiple tumor samples obtained from the same patient as the same color; no color indicates unique samples. 5-ALA (within multi-sample features) indicates the intensity of the 5-aminolevulinic acid-induced fluorescence. (b) Copy number variations in 50 glioma samples. Red and blue colors indicate amplification and deletion, respectively.

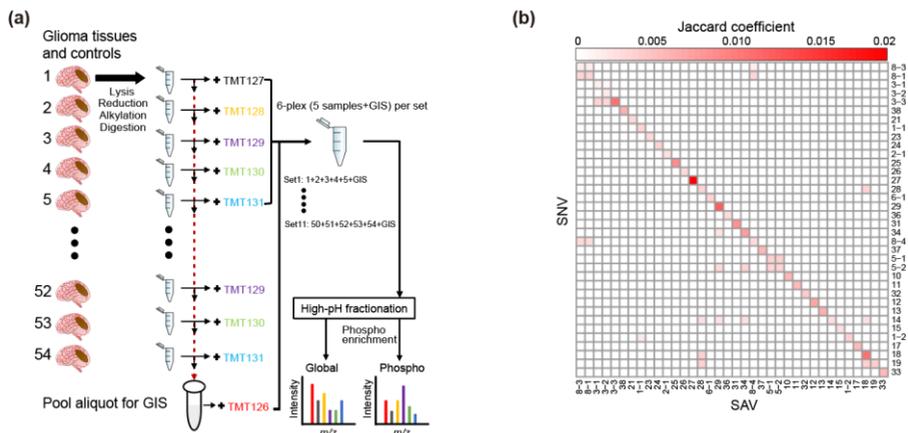


Figure 2. Proteome data reliably captures genomic characteristics of samples. (a) Overview of the multiplexed quantitative proteomic assay of glioma tissues. Trypsin-digested glioma (N = 50) and control normal tissues (N = 4) were tagged with a six-plex tandem mass tag (TMT): TMT127-131 for samples and TMT126 for the global internal standard (GIS) control. A total of 11 sets for 54 samples were prepared. High-pH fractionated peptides were subjected to liquid chromatography-tandem mass spectrometry to identify and quantify phosphopeptides and global proteins. See “Methods” for further details. (b) Coherence map of single-nucleotide variants (SNV) and single amino acid variants (SAVs).

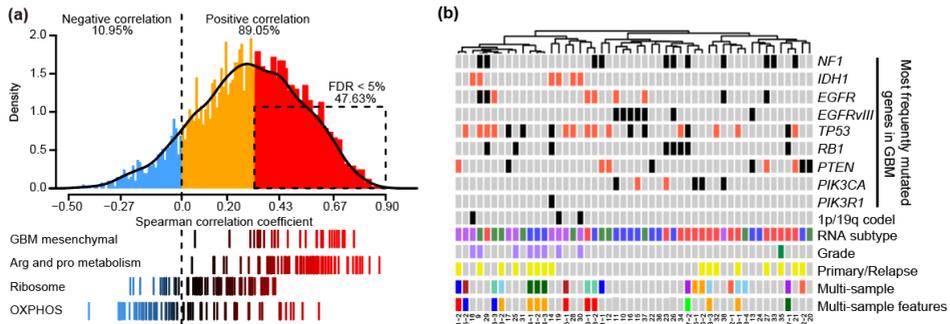


Figure 3. Proteome data captures distinct molecular features of GBM. (a) Correlations between mRNA and protein levels in glioma tissue samples. (Top) Density plot of Spearman's correlation coefficients between mRNA and protein abundance using the 8,034 proteins detected in all GIS samples ($N = 4,071$ at the gene level). Statistically significant positive correlations with a false discovery rate (FDR) $< 5\%$ are indicated by the dashed-line box. (Bottom) Distribution of correlation coefficients for gene sets of interest. (b) Unsupervised hierarchical clustering of the 50 samples with global-proteomic data. Complete linkage and the distance metric $1 - \text{Pearson's correlation coefficient}$ was used for clustering.

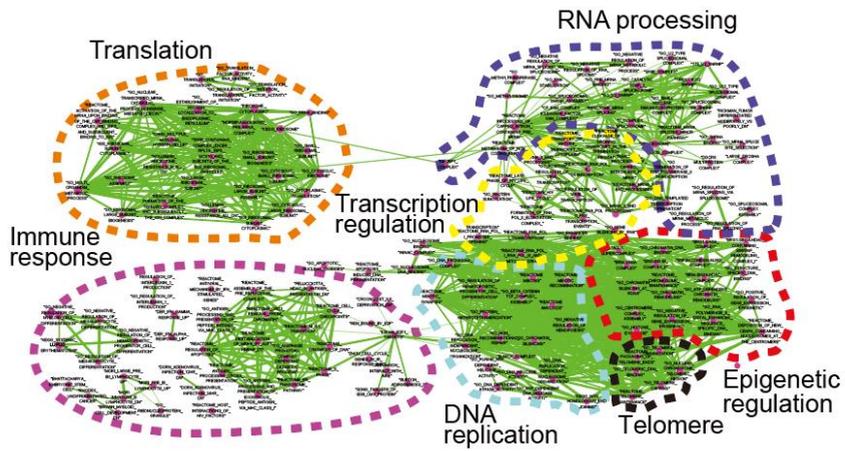


Figure 4. Gene set analysis of proteins elevated in glioma tissues. The over-represented gene sets (FDR < 1%) were visualized using EnrichmentMap with default parameters. The connected gene sets were manually annotated as indicated, based on the corresponding function.

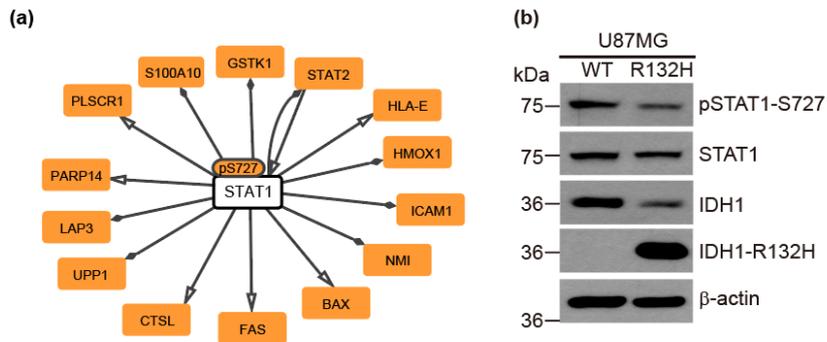


Figure 5. Activated STAT1 signaling in *IDH*-wildtype tumors. (a) Genetic regulatory network activated in *IDH*-wildtype tumors. The transcription factor–target gene regulatory network was formed by the significantly upregulated phosphoproteins and global proteins in *IDH*-wildtype tumors using the OmniPath database in Cytoscape. (b) Comparison of STAT1 (pS727) levels in U87MG isogenic cell lines. Molecular weight markers (kDa) are shown on the left. WT: wildtype.

Table 1. Clinical and molecular annotation of the SMC1 cohort

Set	TMT	Sample	GPC	RNA_subtype4	Expired	OS	PFS	Longitudinal pair	Relapsed sample	Grade	Pathology	Features
2	131	1-1	1	NEURAL	1	650	212	58	I	IV	Glioblastoma	
6	131	1-2	2	PRONEURAL	1	650	62	58	R	IV	Glioblastoma	5-ALA(+)
6	127	2-1	1	MESENCHYMAL	1	324	144	59	I	IV	Glioblastoma	
4	130	2-2	2	CLASSICAL	1	324	95	59	R	IV	Glioblastoma	
8	129	3-1	2	MESENCHYMAL	0	405	381		I	IV	Glioblastoma	5-ALA(+)
10	128	3-2	2	CLASSICAL	0	405	381		I	IV	Glioblastoma	5-ALA(++)
11	131	3-3	2	NEURAL	0	405	381		I	IV	Glioblastoma	5-ALA(-)
1	129	4-1	2	CLASSICAL	1	758	67		R	III	Gliomatosis_cerebri	locally adjacent
6	130	4-2	2	CLASSICAL	1	758	67		R	III	Gliomatosis_cerebri	locally adjacent
5	128	4-3	2	CLASSICAL	1	758	67		R	III	Gliomatosis_cerebri	locally adjacent
5	127	5-1	2	PRONEURAL	1	885	386		I	IV	Glioblastoma	5-ALA(+)
6	129	5-2	2	PRONEURAL	1	885	386		I	IV	Glioblastoma	5-ALA(-)
9	127	6-1	1	MESENCHYMAL	1	333	68	61	I	IV	Glioblastoma	
2	127	6-2	1	MESENCHYMAL	1	333	148	61	R	IV	Glioblastoma	
9	130	7-1	1	MESENCHYMAL	1	228	53		I	IV	Glioblastoma	main
7	129	7-2	1	MESENCHYMAL	1	228	53		I	IV	Glioblastoma	margin
9	128	8-1	1	MESENCHYMAL	1	578	123	65	R	IV	Glioblastoma, recurrent chemoradiation change	locally adjacent
11	130	8-2	2	NEURAL	1	578	123	65	R	IV	Glioblastoma, recurrent chemoradiation change	locally adjacent
10	129	8-3	2	MESENCHYMAL	1	578	123	65	R	IV	Glioblastoma, recurrent chemoradiation change	locally adjacent
3	130	8-4	1	NEURAL	1	578	376	65	I	IV	Glioblastoma	
10	130	9	2	NEURAL	0	2802	1136		R	III	Anaplastic_oligodendroglioma	
4	127	10	1	CLASSICAL	1	175	169		I	IV	Glioblastoma	
4	131	11	1	CLASSICAL	1	463	258		I	IV	Glioblastoma	
1	128	12	1	NEURAL	1	275	253		I	IV	Glioblastoma	
3	127	13	1	CLASSICAL	1	390	92		I	IV	Glioblastoma	
5	129	14	2	PRONEURAL	1	2247	28	55	R	IV	Glioblastoma	
2	130	15	1	CLASSICAL	1	474	275		I	IV	Glioblastoma	
4	128	16	1	CLASSICAL	1	368	153		I	IV	Glioblastoma	
3	129	17	2	PRONEURAL	1	551	502		I	IV	Glioblastoma	
1	131	18	2	PRONEURAL	0	1028	697		I	III	Anaplastic_oligodendroglioma	
4	129	19	2	PRONEURAL	0	978	871		I	III	Anaplastic_oligodendroglioma	
1	130	20	1	NEURAL	1	652	643		I	IV	Glioblastoma	
9	131	21	1	MESENCHYMAL	1	429	28	57	R	IV	Glioblastoma	
10	131	22	1	CLASSICAL	1	155	44		I	IV	Glioblastoma	
8	128	23	1	MESENCHYMAL	1	519	497		I	IV	Glioblastoma	
11	128	24	1	CLASSICAL	1	600	208		I	IV	Glioblastoma	
7	128	25	2	PRONEURAL	1	1758	83		R	III	Anaplastic_oligoastrocytoma	
8	130	26	1	CLASSICAL	1	379	88		I	IV	Glioblastoma	

Set	TMT	Sample	GPC	RNA_subtype4	Expired	OS	PFS	Longitudinal pair	Relapsed sample	Grade	Pathology	Features
11	127	27	1	MESENCHYMAL	0	2254	1138		Secondary	IV	Glioblastoma	
7	127	28	2	NEURAL	0	1086	1084		I	IV	Glioblastoma	
10	127	29	2	MESENCHYMAL	1	166	142		I	IV	Glioblastoma	
2	128	30	2	PRONEURAL	0	1749	1097		I	III	Anaplastic_oligodendroglioma	
9	129	31	2	NEURAL	1	843	390		I	IV	Glioblastoma	
8	127	32	2	MESENCHYMAL	1	3584	2		R	IV	Glioblastoma	
11	129	33	1	MESENCHYMAL	1	345	329		I	IV	Glioblastoma	
8	131	34	1	MESENCHYMAL	1	720	88		I	IV	Glioblastoma	
7	130	35	1	MESENCHYMAL	1	829	215		R	II	Chordoid_glioma_of_the_third_ventricle	
6	128	36	1	CLASSICAL	0	1238	1168		I	IV	Glioblastoma	
3	131	37	1	NEURAL	0	400	343		I	IV	Glioblastoma	
5	131	38	1	PRONEURAL	1	883	284		I	IV	Glioblastoma	
3	128	586N									Normal	
1	127	655N									Normal	
2	129	753N									Normal	
5	130	608N									Normal	

2. Two proteomic subgroups of *IDH*-wildtype GBM

I next classified the tumors based on the proteome data with the 3,909 protein groups detected in all the tumor samples. A consensus clustering algorithm analysis identified two stable proteomic subtypes for *IDH*-wildtype GBM: glioblastoma proteome cluster 1 (GPC1, N = 26) and GPC2 (N = 13) (Figure 6a). Inclusion of *IDH*-mutant GBMs and LGGs did not alter the classification of *IDH*-wildtype GBM tumors (Figure 6b): the two *IDH*-mutant GBMs were classified with GPC2 tumors, and the nine LGGs were significantly associated with GPC2 tumors (χ^2 test $P = 0.0029$). Removing non-unique samples (i.e., any sample that shared a patient of origin) or taking all the GBMs regardless of *IDH* genotype did not affect the binary nature of the classification (Figure 6b). The stability of this classification supports the robustness of the subtypes and implies that binary classification is applicable to GBM in general.

The GPC1 global proteome expression pattern was distinct from that of normal brain tissues, whereas the GPC2 global proteome expression displayed a normal brain tissue-like pattern (Figure 7a, top panel). However, the purities of GPC2 tumor were similar to those of GPC1 based on comparable variant allele frequency (VAF) distribution (Figure 7a, bottom panel), which suggests that the abundance of normal cells in tumors does not contribute to these proteomic differences. A subtype switch, likely driven by therapeutic treatment, was observed in recurrent tumors, with two of the three samples switching from GPC1 into GPC2 (Figure 7b). This finding demonstrates protein-level subtype plasticity, which is in accordance with previous findings from a longitudinal transcriptome analysis¹⁷. The two GPC subtypes were largely independent of the four RNA subtypes⁸ (χ^2 test $P = 0.122$; Figure 8a). Notably, *EGFRvIII* and *PIK3CA* mutations were exclusively found in GPC1 tumors, whereas other GBM driver mutations in *TP53*, *NF1*, *PTEN*, *RBI*, and *EGFR (non-vIII)* were relatively evenly distributed between the two subtypes (Figure 8b and 8c). These results suggest that *IDH*-wildtype GBM can be classified into two stable protein subtypes that are distinct from RNA

subtypes by different molecular determinants.

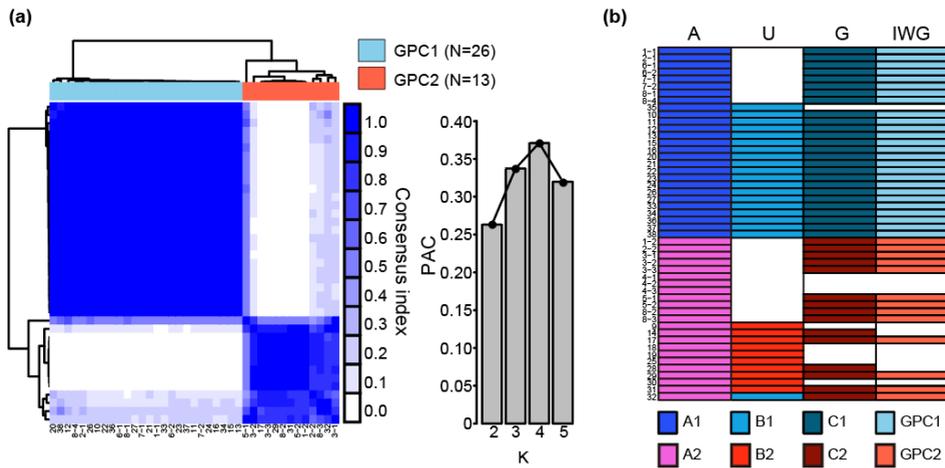


Figure 6. *IDH*-wildtype GBM tumors are separated into two proteomic subtypes. (a) Consensus clustering of *IDH*-wildtype GBM samples (N = 39) based on global proteome data using 1 – Pearson’s correlation coefficient as the distance metric. (Left) Heatmap of the consensus score. (Right) The bar chart indicates the proportion of ambiguous clustering (PAC) for the indicated K values. The number of clusters (K) with the lowest PAC score is considered the optimal cluster number. (b) Comparison of consensus clustering results from different sample composition. A : all glioma samples, U : unique glioma samples, G : GBM samples, IWG : *IDH*-wildtype GBM samples.

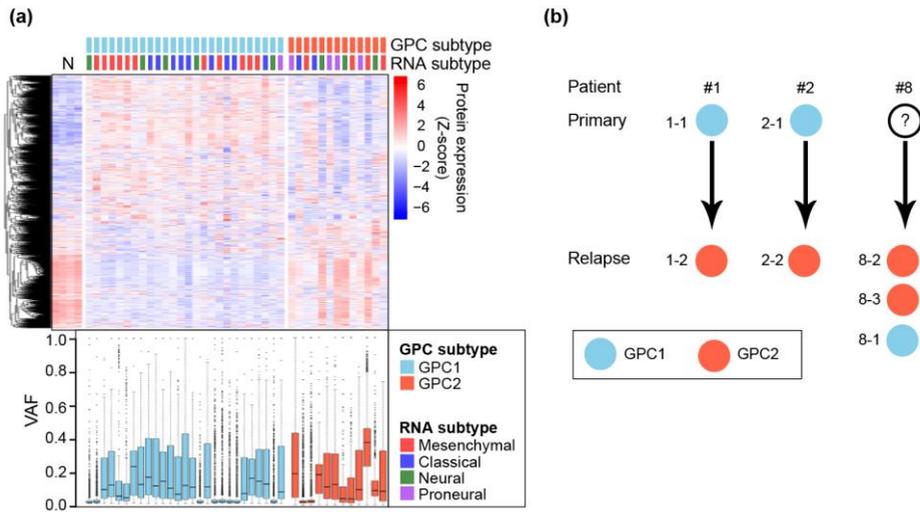


Figure 7. Proteomic signatures and clinical characteristics of GPC subtypes. (a) Characteristics of proteomic subtypes. The columns represent the samples grouped by proteomic subtype. The clustering heatmap represents the Z-score-normalized protein expression levels. The box-and-whisker plot represents the medians (middle line), first quartiles (lower bound line), third quartiles (upper bound line), and the $\pm 1.5 \times$ interquartile ranges (whisker lines) of the variant allele frequency (VAF) of single nucleotide variants (SNVs) per sample. (b) Subtype-switch observed in paired primary and relapse samples. Each column represents a patient, and each circle within a column represents separate primary and relapse tumor samples obtained for that patient. Primary GBM tissue was missing for patient 8.

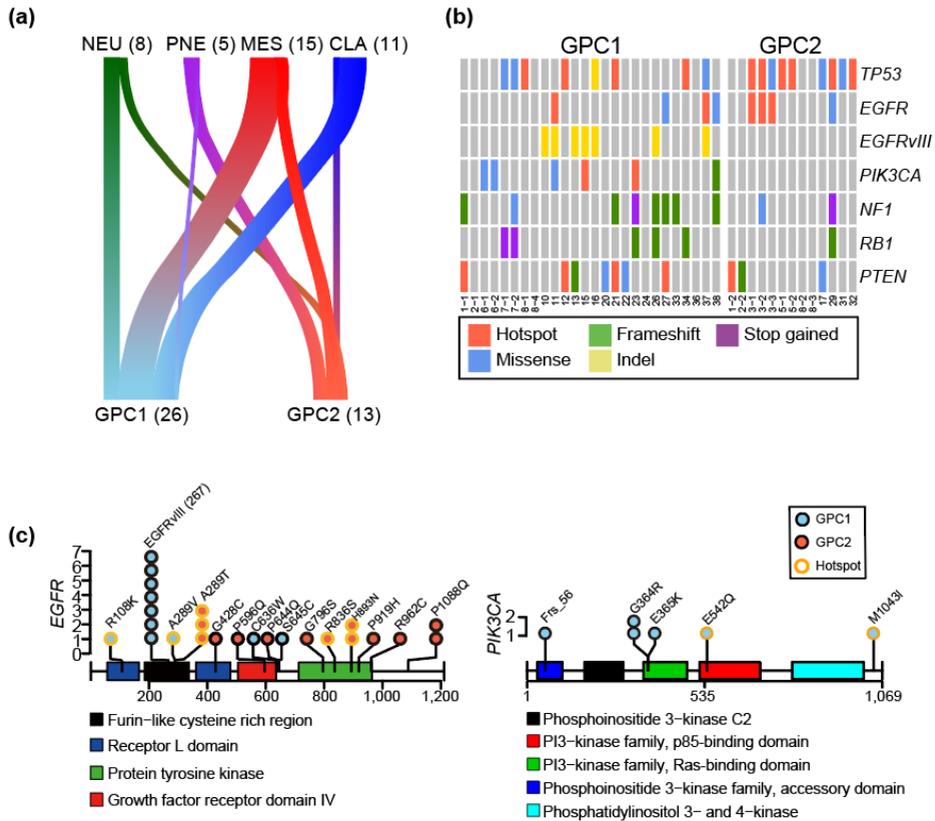


Figure 8. Genomic characteristics of GPC subtypes. (a) The river plot illustrates the association between proteomic and RNA subtypes. The width of the edge between two subtypes is shown in proportion to the number of corresponding samples. CLA : classical, PNE : proneural, NEU : neural, MES : mesenchymal. (b) GBM driver mutations associated with proteome subtypes. The color-coded matrices indicate the cBioPortal-annotated mutation types in the samples grouped by proteomic subtype for the six frequently mutated genes in GBM. (c) Somatic mutations in *EGFRvIII* and *PIK3CA* that were exclusively found in GPC1 and GPC2 tissues. The x axis indicates the position in the gene, and the y axis indicates the frequencies of the mutations. The color in the circle indicates a mutation found either in a GPC1 (blue) or in a GPC2 (red) sample. The outline color of the circle indicates the hotspot mutations. Rectangles with different colors represent protein domains.

3. OXPHOS-related proteins separate proteomic subtypes

To identify the key proteins that characterize the GPC subtypes, I conducted a principal component analysis (PCA) of the global proteome data. The first principal component (PC1) successfully distinguished the two GPCs (Figure 9a; left panel). A gene ontology analysis of the top 10% of proteins (N = 309) with the highest absolute PC1 loading values revealed a striking enrichment of proteins involved in OXPHOS (Figure 9a; right panel). Moreover, the expression of OXPHOS-related proteins was significantly lower in GPC1 than in GPC2. In contrast to the OXPHOS-related proteins, there was no difference in the OXPHOS mRNA levels between the subtypes (Figure 9b). Compared with normal cells, GPC2 tumors expressed similar levels of proteins involved in glycolysis, serine biosynthesis, the tricarboxylic acid (TCA) cycle, glutaminolysis, and OXPHOS, but at significantly higher levels than those found in GPC1 tumors (Figure 9b and 10). This similarity indicates that GPC2 tumors metabolically rely on OXPHOS to gain ATP like normal cells did under normoxia.

In contrast to GPC2 tumors, GPC1 tumors displayed typical characteristics of the Warburg effect. Higher levels of lactate dehydrogenase A (LDHA) and proteins involved in glucose uptake, hexokinase 2 (HK2), the pentose phosphate pathway (PPP), and the one-carbon pathway were observed in GPC1 tumors (Figure 10). Notably, GPC1 expressed higher levels of pyruvate kinase m2 (PKM2) and HK2, whereas GPC2 had an elevated PKM1 (Figure 11a and 11b). In addition, GPC1 tumors exhibited elevated expression of IDH1 protein (the primary producer of NADPH in GBM)^{46,47} beyond the levels found in GPC2 tumors (Figure 11c). The GPC1-activated PPP and one-carbon pathway also generate NADPH, which is a reducing equivalent for tumor cells affected by the Warburg effect⁴⁸. To validate metabolic differences between the subtypes, human-derived GBM cancer cell lines were assigned to either surrogate-GPC1 (sGPC1) or sGPC2 according to their gene expression signatures. Consistent with my proteomic data, sGPC1 cell lines exhibited higher lactate levels (Figure 11d) in the analysis of cancer cell line

encyclopedia metabolomics data⁴⁹. Together, these results suggest that GPC1 tumors are metabolically dependent on the Warburg effect.

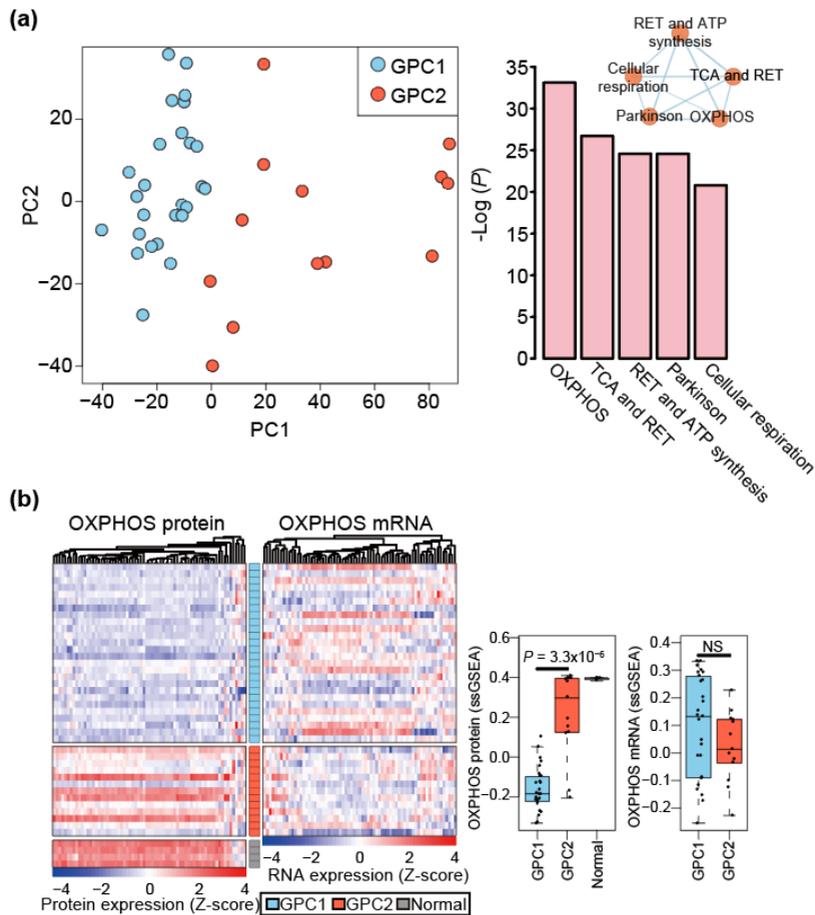


Figure 9. Proteins involved in oxidative-phosphorylation (OXPHOS) are major determinants of *IDH*-wildtype GBM proteomic subtypes. (a) Principal component analysis (PCA) of 39 *IDH*-wildtype GBM samples using global proteome data (left). The statistical significance of the enriched gene sets with the top 10% loading genes ($N = 390$) in PC1 is shown in the right panel. The mutual similarity of the top five gene sets is presented using EnrichmentMap. RET respiratory electron transport. (b) Clustering heatmaps of *IDH*-wildtype GBM tissues with the OXPHOS-related genes based on protein and mRNA expression data (left). The rows represent the tumor samples grouped by proteomic subtype, and the columns represent the genes belonging to the OXPHOS-related gene set. Comparisons of the ssGSEA scores for both proteins and genes are shown on the right. The box-and-whisker plots represent the medians (middle line), first quartiles (lower bound line), third quartiles (upper bound line), and the $\pm 1.5 \times$ interquartile ranges (whisker lines); the raw data are overlaid. P -values were calculated using the two-sided unpaired Wilcoxon rank-sum test. NS : not significant.

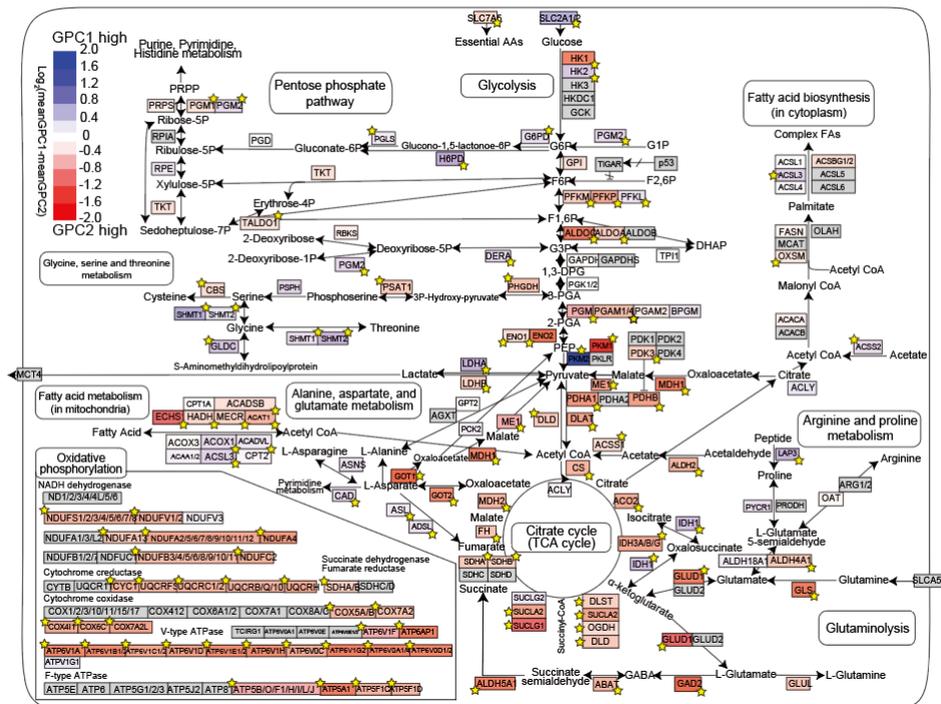


Figure 10. Protein expressions in central carbon metabolism. Pathway view of differentially expressed proteins involved in central carbon metabolism between the two proteome subtypes. The stars indicate a statistically significant difference in protein expression between the two GPC subtypes ($P < 0.05$; two-sided unpaired Student's t-test). Genes with no available protein expression data are shown in gray.

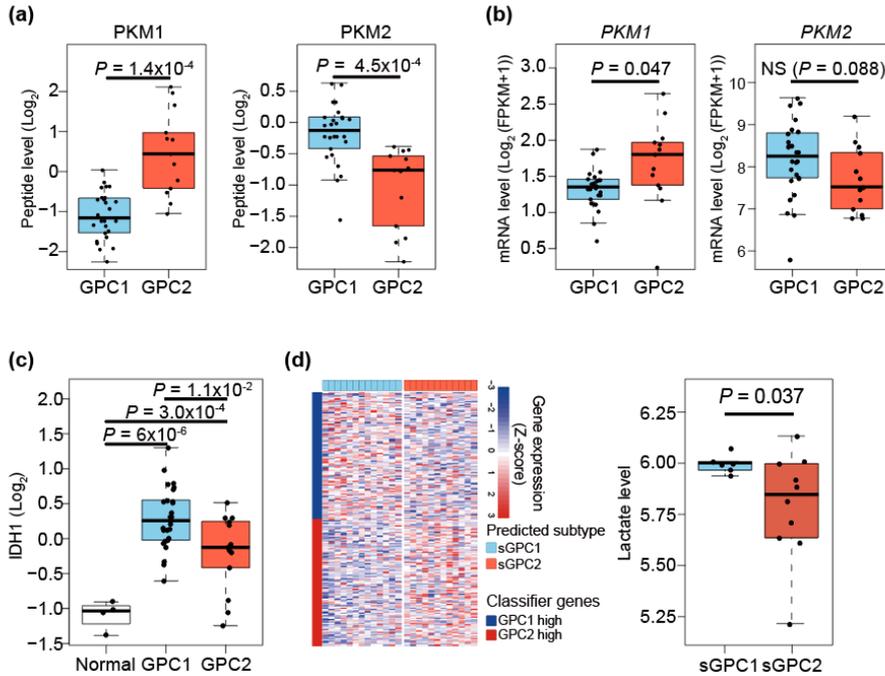


Figure 11. Expression levels of proteins involved in glycolysis and lactate level. (a) Comparison of the peptide expression of PKM1 (exon 9) and PKM2 (exon 10) in GPC subtypes. (b) Comparison of the gene expression of *PKM1* and *PKM2* in GPC subtypes. (c) Comparison of the IDH1 expression between GPC subtypes. (d) Classification of 47 GBM cell lines into gene expression-based surrogate GPC subtypes. The heatmap displays gene expression levels of classifier genes (left). Comparison of lactate levels in GBM cell lines according to sGPC subtype (right). The numbers of samples in sGPC1 and sGPC2 are 6 and 10, respectively. *P*-values were obtained by two-sided unpaired Student's *t*-test. The box-and-whisker plots represent the medians (middle line), first quartiles (lower bound line), third quartiles (upper bound line), and the $\pm 1.5 \times$ interquartile ranges (whisker lines); the raw data are overlaid.

4. GPC-subtype-dependent expression of cell-of-origin markers

Neural stem cell (NSC) is considered as a major source of GBM¹⁸. To understand whether each GPC subtype has a distinct cellular origin, I compared the levels of NSCs (Nestin, Vimentin, CD44), oligodendrocytes (OSP, MOG), and astrocytes (GLUL, GLT-1, GLAST, HepaCAM, ALDH1A1, S100 β) marker proteins that were detected in all GIS. GPC1 tumors displayed significantly high expressions of Nestin, Vimentin, and CD44 (Figure 12a). In GPC1 tumors, I also observed elevations in an active form of cortactin phosphorylated at T364/S368/T401/S405 and its interacting partner, Arp2/3 complex subunits (Figure 12b), which are components not only of filopodia/lamellipodia (cytoplasmic protrusions of migratory cells) but also of invadopodia (invasive protrusions of transformed cells). By contrast, GPC2 tumors were characterized by overexpression of the oligodendrocyte and astrocyte markers OSP, MOG, GLT-1, GLAST, and HepaCAM (Figure 12a). GLUL, ALDH1A1, and S100 β were also relatively highly expressed in GPC2 tumors, but were marginally insignificant (Figure 12a).

Accumulating evidence indicates that GBM stem cells are immune-resistant⁵⁰. Concordantly, GPC1 had increased levels of *CD274* (*PD-L1*, two-way ANOVA $P = 2.02E-6$) and *PDCD1LG2* (*PD-L2*, two-way ANOVA $P = 4.75E-13$) (Figure 13a and 13b). Together, with the observation that recurrent tumors were enriched in GPC2 group (Figure 7b), GPC1 tumors may originate from NSCs and differentiate into GPC2 tumors. However, I cannot exclude the possibility that oligodendrocytes and astrocytes are cellular origins for GPC2 tumors.

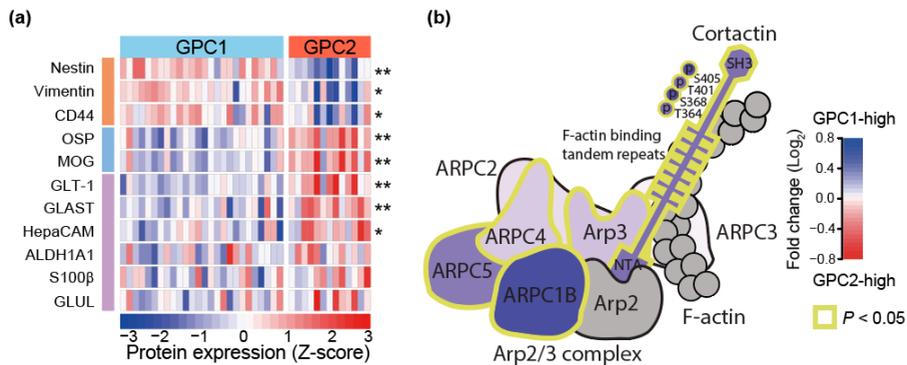
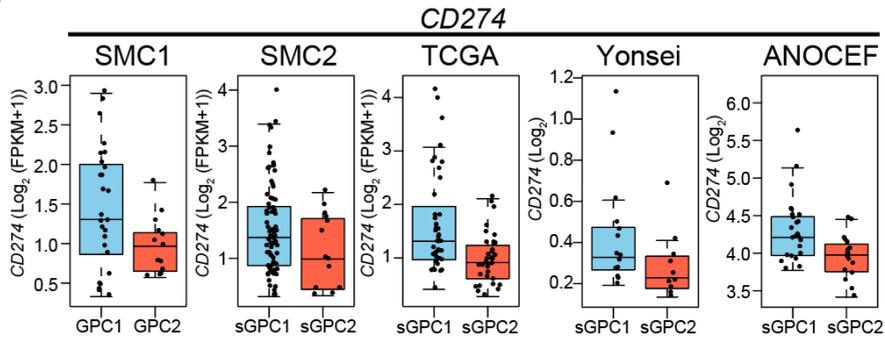


Figure 12. Protein expression of cellular markers and Arp2/3 complex. (a) Differential expression of protein markers for NSCs (orange), oligodendrocytes (blue), and astrocytes (purple) between proteomic subtypes. * $P < 0.05$, ** $P < 0.001$; two-sided unpaired Student's t-test. (b) The cortactin-Arp2/3 complex is elevated in GPC1 tumors. The gray color indicates proteins with no available protein expression data. P -values were calculated by two-sided unpaired Student's t-test.

(a)



(b)

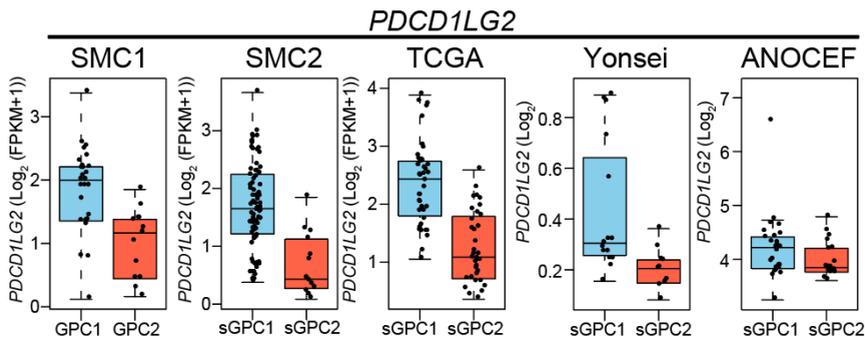


Figure 13. Subtype-specific expression of *CD274* (PD-L1) and *PDCD1LG2* (PD-L2) in the five datasets. (a-b) The box-and-whisker plots represent the medians (middle line), first quartiles (lower bound line), third quartiles (upper bound line), and the ± 1.5 x interquartile ranges (whisker lines). The raw data are overlaid.

5. PHGDH predicts favorable prognosis in *IDH*-wildtype GBM

A previous study reported potential GBM prognostic gene expression biomarkers consisting of 66 favorable and 205 unfavorable genes⁴⁴, however, protein level validation has yet been pursued. I found that only 11 of these markers were validated at the protein level by the univariate Cox regression analysis with the SMC1 cohort, and of those, I identified two biomarkers as favorable for *IDH*-wildtype GBM (phosphoglycerate dehydrogenase, PHGDH, and Raftlin family member 2, RFTN2), and one as unfavorable (FKBP prolyl isomerase 9, FKBP9; Figure 14). Importantly, the protein expression of all three of these markers differed between the two GPC subtypes—FKBP9 was increased in GPC1, whereas PHGDH and RFTN2 were higher in GPC2 (Figure 16a). However, GPC subtypes were not statistically associated with overall survival. (two-sided Log-rank test $P = 0.0548$).

Of the three proteins, PHGDH was the most robust biomarker (univariate Cox $P = 0.0071$) correlated with long-term survivors of *IDH*-wildtype GBM patients in the SMC1 cohort (Figure 14), as well as in other independent data sets at the mRNA-level (Figure 15). Favorable prognosis of PHGDH-high tumors was further validated in 42 independent *IDH*-wildtype GBM tumors assessed by immunohistochemistry on a tumor tissue microarray (SMC-TMA) using an anti-PHGDH antibody (Figure 16b). The good prognosis of the PHGDH-high group suggests that PHGDH may play an anti-invasive role in GBM. Intriguingly, NCT-502, a chemical inhibitor of PHGDH, significantly increased invasion of tumor spheres derived from PHGDH-active GBM cell lines (Figure 17a and 17b) into 3D matrix. Conversely, PHGDH overexpression in PHGDH-deficient GBM cell lines decreased invasion (Figure 17c), suggesting that PHGDH may prolong patient survival by suppressing tumor invasion via its increased enzymatic activity. PHGDH, a rate-determining enzyme in serine biosynthesis that converts 3-phospho-D-glycerate to 3-phosphonoxyypyruvate, is also known to have a promiscuous function with the ability to convert α -ketoglutarate into D-2-

hydroxyglutarate (D-2-HG), similar to the IDH-mutant protein⁵¹. In support of this, I found a positive correlation between PHGDH levels and 2-HG production (Figure 18) by analyzing 878 *IDH*-wildtype cancer cell-line metabolome data⁴⁹.

Taken together, these data indicate that GPC2-associated PHGDH predicts a favorable prognosis in *IDH*-wildtype GBM.

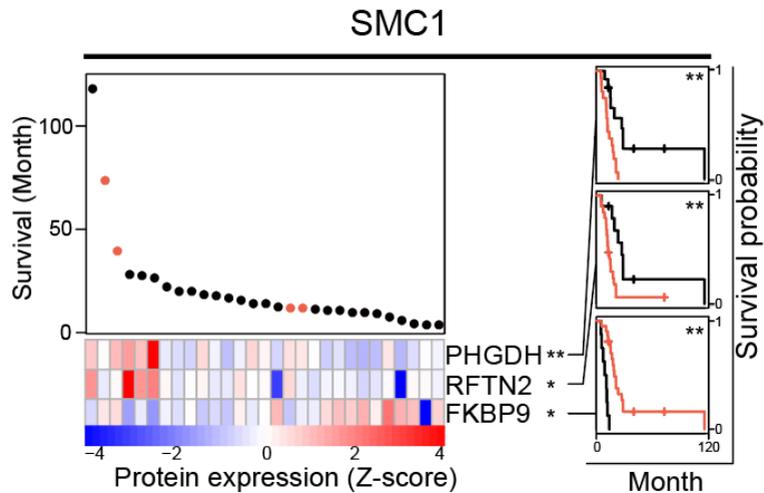


Figure 14. Prognostic biomarker proteins in *IDH*-wildtype GBM. Deceased and surviving patients are denoted by black and red dots, respectively. * $P < 0.05$, ** $P < 0.01$; univariate Cox regression test. Kaplan–Meier (KM) survival curves for *IDH*-wildtype GBM patients ($N = 29$) in the SMC1 cohort were shown on the right for each of the three proteins (black: high expression, red: low expression). ** $P < 0.01$; two-sided log-rank test.

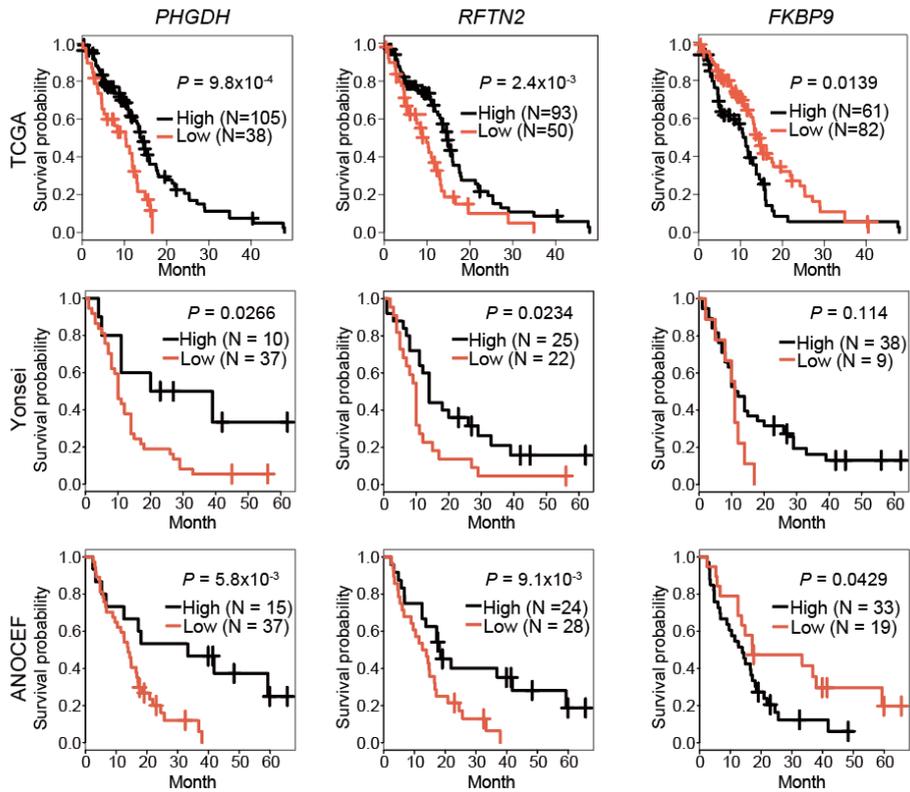


Figure 15. Kaplan-Meier plots for three indicated prognostic markers in the three independent cohorts. The individual cohorts are annotated with their names. TCGA (Upper panels), Yonsei (Middle panels), ANOCEF (lower panels). Patients were classified by the same method used in Figure 14. *P*-values were obtained by the two-sided log-rank test.

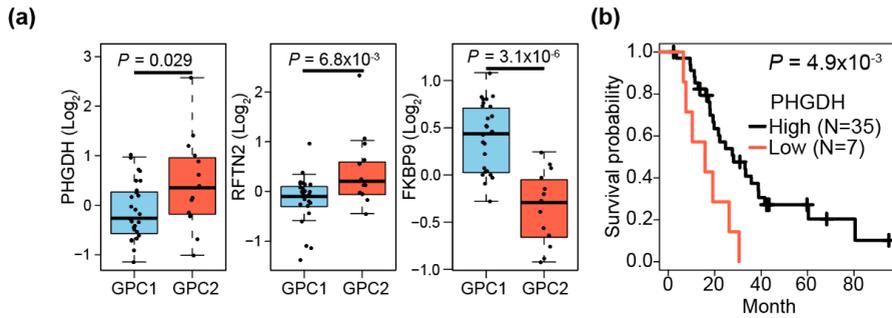


Figure 16. Expression level of PHGDH is associated with patient survival. (a) Box-jitter plots for the abundance of the indicated proteins. The description of the box-and-whisker plots is the same as in Figure 13. Statistical significance of the downregulation of favorable markers (PHGDH and RFTN2) and upregulation of an unfavorable marker (FKBP9) was evaluated by Student's t-test (one-sided unpaired). The number of samples for GPC1 and GPC2 is 26 and 13, respectively. (b) KM survival curves for PHGDH-high vs. low patients in the SMC-TMA cohort.

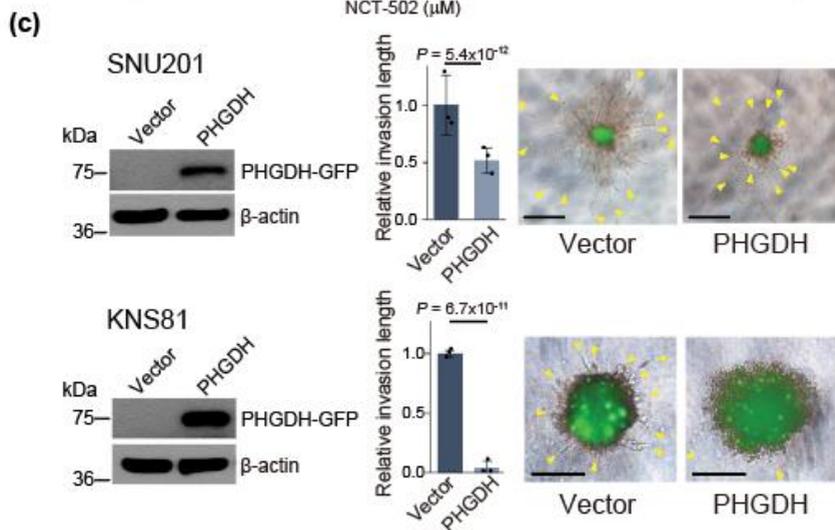
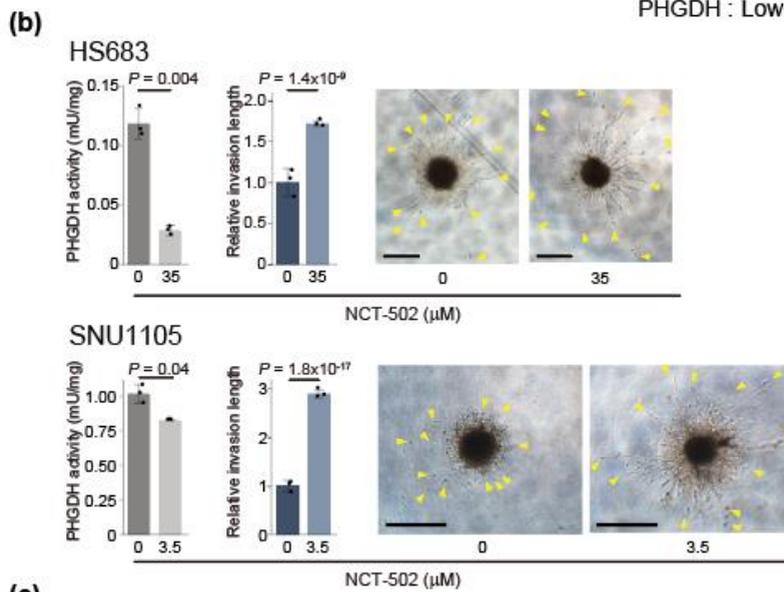
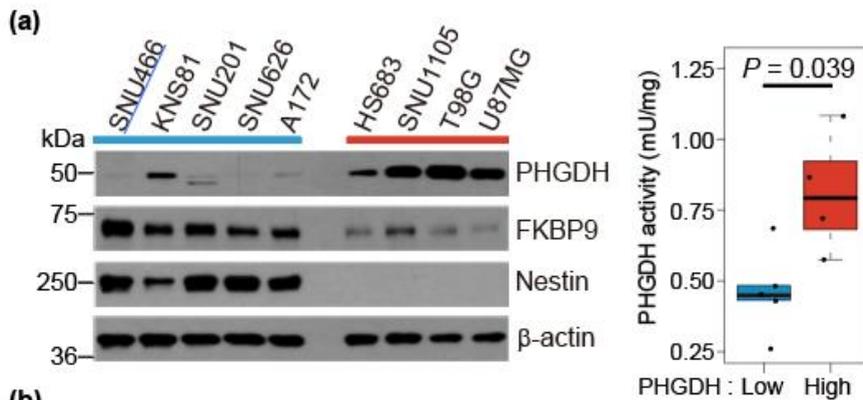


Figure 17. PHGDH suppresses invasion of GBM. (a) Steady state accumulation of the indicated proteins was assessed by immunoblotting of whole cell lysates from the indicated GBM cell lines. (left) Comparison of PHGDH enzymatic activity (right). The box-and-whisker plots represent the medians (middle line), first quartiles (lower bound line), third quartiles (upper bound line), and the ± 1.5 x interquartile ranges (whisker lines). The raw data are overlaid. *P*-values were calculated by two-sided unpaired Student's *t*-test. The numbers of samples in PHGDH-low and high groups are 5 and 4, respectively. (b) PHGDH activity (left), relative invasion lengths (middle), and representative images of 3D invasion (right) of the indicated tumor spheres after treatment of vehicle (DMSO) or NCT-502 for 48 hrs at the indicated concentrations. Two-sided unpaired Student's *t*-test was used to compare PHGDH activity. Two-way ANOVA was used for the comparison of relative invasion length between treatment groups. Error bars indicate \pm SD, *N* = 3. Arrowheads indicate invasive fronts. Scale bar: 500 μ m. (c) PHGDH levels by immunoblotting (left), relative invasion lengths of the tumor spheres (middle panels), and representative images of 3D invasions (right panels) in SNU201 (upper panels) and KNS81 (lower panels)-derived tumor spheres after 96 hrs of transfection of plasmids (empty vector or PHGDH cDNA). Arrowheads indicate invasive fronts. Two-way ANOVA was used for the comparison of relative invasion length between transfection groups. Error bars indicate \pm SD, *n* = 3. Scale bar : 500 μ m (SNU201), 200 μ m (KNS81).

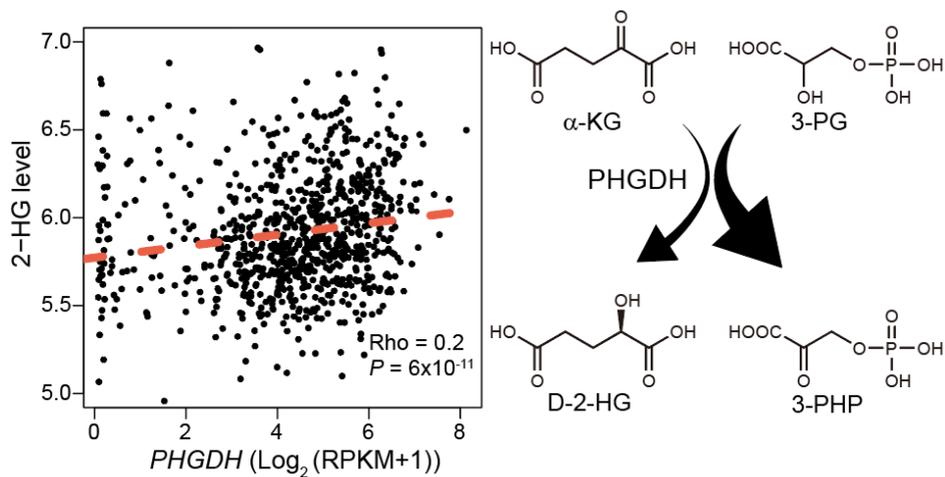


Figure 18. Correlation between *PHGDH* expression and 2-hydroxyglutarate (2-HG) levels in 878 *IDH*-wildtype cancer cell lines. Gene expression and 2-HG abundance data were obtained from the CCLE database. Rho and *P*-values were obtained by Spearman's correlation test. (Right) Schematic drawing of the primary (thick arrow) and promiscuous (thin arrow) function of PHGDH.

6. Spatial and cell-type characteristics of GPC subtypes

Recent single-cell RNA-sequencing analyses have demonstrated that individual tumors are comprised of single-cells with various subtypes and hybrid cellular states. This finding implies that a dominant cellular population might determine the representative subtype of bulk tumors¹⁶. I therefore aimed to apply my GBM proteomic subtype classification to single-cell transcriptome data reported by Darmanis et al.⁴³, which consists of 3,589 single cells from four patient *IDH*-wildtype GBM tumors and provided information regarding the brain cell type (vascular, immune, neuronal, and glial) and three-dimensional location (tumor core, periphery). Of the 3,589 single cells, 357 and 428 cells were classified as sGPC1 and sGPC2, respectively (permutation test $P < 0.05$; Figure 19a). The mapping of these cells on t-distributed stochastic neighbor embedding (t-SNE) coordinates revealed a distinct clustering pattern of two proteome subtypes (Figure 19b). Consistent with my previous findings, NSC markers *CD44* and *VIM* were significantly elevated in single cells of the sGPC1 subtype, whereas oligodendrocyte marker *OLIG2*, *OSP*, and *MOG* and the astrocyte markers *SLC1A2* and *S100B* were highly expressed in the sGPC2 subtype (Figure 19c).

According to the annotation by Darmanis et al.⁴³, the majority of single cells located in the tumor periphery were non-neoplastic cells (95%), whereas the tumor core was largely composed of neoplastic (44%) and myeloid cells (50%). As expected, I found that each of the tumor cores from the four patients comprised a mixture of cells of the two proteome subtypes; however, the ratios of the two subtypes in neoplastic cell populations of the four tumors were highly variable, and made a notable contribution to determining the dominant proteome subtype of the tumor (Figure 20a). By contrast, the surrounding environment primarily determined the proteomic subtype of normal cell populations: 84% of normal cells in the tumor core displayed sGPC1 features, whereas 70% of normal cells in the periphery displayed sGPC2 features (Figure 20b). Because the two subtypes showed differential immune evasion characteristics, I further investigated whether

the neoplastic single cells of each subtype variably expressed immune checkpoint ligands. Indeed, a PD-1 ligand *PD-L1* was upregulated in the neoplastic cells of sGPC1 tumors (Figure 20c).

To further validate the intratumoral heterogeneity of GPC subtypes at a single cell level, I used a tumor microarray (SMC-TMA) of independent *IDH*-wildtype GBM tissues to measure relative expression of PHGDH (good prognostic marker representing GPC2, Figure 14) and Nestin (representing GPC1, Figure 12) by multiplex fluorescent immunohistochemistry that generated reliable signal intensities at a single cell resolution. Consistent with my findings in the proteomic analysis, sGPC1 tumors exhibited a significantly higher fraction of Nestin-positive neoplastic cells, whereas sGPC2 tumors comprised a significantly higher fraction of PHGDH-positive neoplastic cells (Figure 21a and 21b). Intratumoral heterogeneity, observed from the scRNA-seq data, was clearly seen in the multiplex fluorescent immunohistochemistry results. Both Nestin⁺/PHGDH⁻ cells (representing GPC1 subtype) and PHGDH⁺/Nestin⁻ cells (representing GPC2 subtype) were found in all tumor cores, albeit with different ratios matching their sGPC subtypes (i.e., two sGPC1 tumors contained a higher frequency of Nestin⁺ cells), whereas two sGPC2 tumors contained higher frequency of PHGDH⁺ cells (Figure 21a).

These results indicate that GBM tumors comprise cells belonging to both GPC subtypes, and that the ratio of neoplastic cell subtypes affects the overall tumor characteristics. Besides, unlike normal cells in the tumor core which displayed a static sGPC1-like proteomic feature, neoplastic cells in the tumor core exhibited highly diverse GPC subtypes, suggesting that the proteomic subtypes of neoplastic cells are largely determined by cancer intrinsic factors rather than the tumor microenvironment.

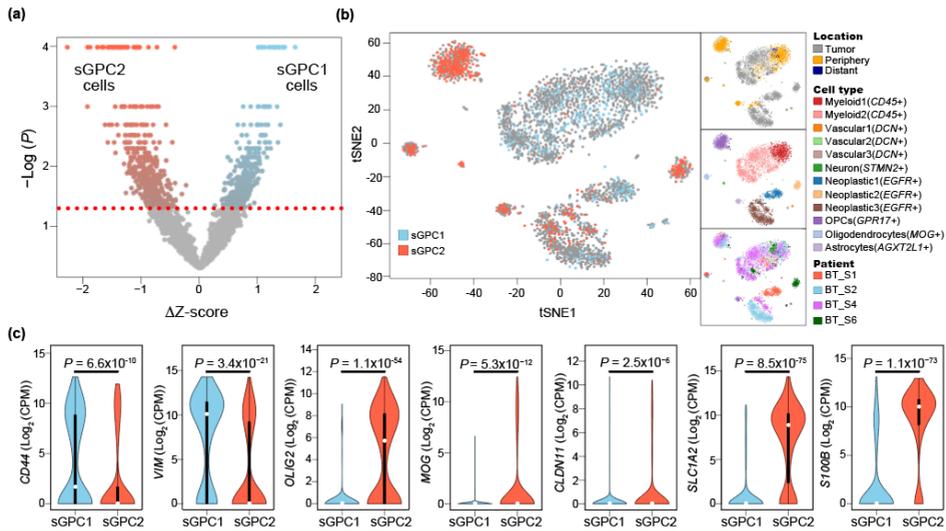


Figure 19. Classification of GBM single cells and cellular marker expressions.

(a) Classification of GBM single cells from the data set reported in Darmanis et al. The red dotted line indicates the statistical threshold used to determine the surrogate GPC (sGPC) subtype of single cells. (b) 2D t-SNE projection of single cells of surrogate-proteomic subtypes. t-SNE coordinates, single-cell annotations describing the location within the 3D tumor mass, cell type, and the patient origin (shown in right side panels) were obtained from Darmanis et al. (c) Subtype-specific mRNA expression of neural stem cell (NSC), oligodendrocyte, and astrocyte markers in single cells. *CD44/VIM*, *OLIG2/MOG/CLDN11*, and *SLC1A2/S100B* represent NSC, oligodendrocyte, and astrocyte markers, respectively. The violin plots represent density distributions.

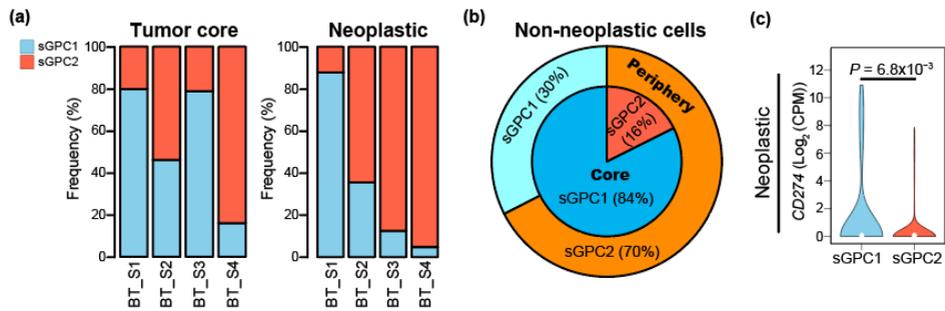


Figure 20. Spatial, cellular, and immuno characteristics of GPC subtypes. (a) Proportion of tumor core single cells (left) and neoplastic single cells (right) of two subtypes in four specimens in the data set generated by Darmanis et al. (b) Proportion of non-neoplastic single cells of two subtypes in the Darmanis et al. data set. (c) Subtype-specific mRNA expression of an immune checkpoint ligand *CD274* (PD-L1) in neoplastic single cells.

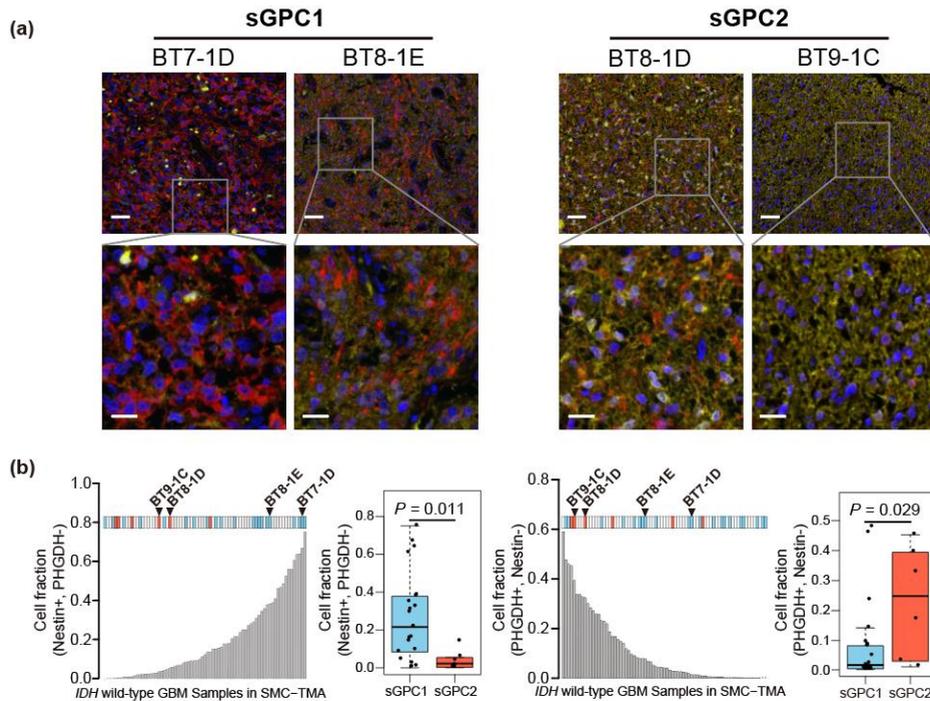


Figure 21. GPC subtypes are determined by dominant population of cell subtypes. (a) Representative images of multiplex fluorescent immunohistochemistry analysis of SMC-TMA samples using PHGDH (yellow) and Nestin (red) antibodies. Nuclei were stained with DAPI (blue). Sample IDs and their respective sGPC-subtypes are indicated above the image. Scale bars: 50 μm (upper panels), 20 μm (lower panels). (b) Fraction of Nestin⁺/PHGDH⁻ (left panel) and PHGDH⁺/Nestin⁻ (right panel) cells in GBM tissues from SMC-TMA cohort. sGPC-subtype for each sample is indicated by color-stripe. Samples used in (a) were denoted by black arrowheads. Comparison of the cell fractions between sGPC1- (N = 21) and sGPC2-subtype samples (N = 6). The box-and-whisker plots represent the medians (middle line), first quartiles (lower bound line), third quartiles (upper bound line), and the ± 1.5 x interquartile ranges (whisker lines). The raw data are overlaid. Statistical significance of sGPC1-subtype dependent elevations of Nestin⁺/PHGDH⁻ cell fractions and sGPC2-subtype dependent elevation of PHGDH⁺/Nestin⁻ cell fractions was evaluated by one-sided unpaired Wilcoxon rank-sum test.

7. Subtype-dependent sensitivities to targeted therapies

To evaluate the clinical utility of proteome data as a predictive marker of drug responses, I retrieved response profiles for 60 targeted agents from 50 PDCs⁵². The 60 drugs covered major oncogenic pathways. Of the 60 drugs tested, I used 51 agents which were cytotoxic to at least one of the PDCs. Although the PDCs had been cultured *in vitro* for several passages for the drug assay, a substantial enrichment of statistically significant correlations was observed between protein biomarkers (measured in tumor tissues) and drug–response phenotypes (assayed in PDCs) (Figure 22). Known drug targets were most strongly correlated with the drug–response, particularly at the protein level (Figure 23a). For example, bortezomib and panobinostat cytotoxicities were significantly correlated with the protein expression levels of the 20S proteasomal subunits and histone deacetylase (HDAC)1/2, respectively (Figure 23b, left panels), which agrees with previous studies showing that target protein expression levels in cancer cells determine the anticancer activities of proteasome and HDAC inhibitors^{53,54}. Interestingly, these correlations were less evident at the mRNA level (Figure 23b, right panels).

I subsequently examined whether any of the 51 targeted agents showed selective cytotoxicity against GPC1 or GPC2 subtype PDCs. Using the median effective dose (ED50) or area under the curve (AUC) values, I identified four GPC1-selective drugs (tandutinib, crizotinib, olaparib, and AZD2014) and two GPC2-selective drugs (erismodegib and canertinib) (Figure 24a). Coherent drug-sensitivity and target-pathway activation relationships for all of these drugs were observed at the protein-levels (Figure 24b): tandutinib (PDGFR inhibitor), PDGFR_Binding; crizotinib (ALK, MET, ROS1 inhibitor), Oncogenesis_by_MET; olaparib (PARP inhibitor), BRCAness score; AZD2014 (mTORC1/2 dual inhibitor), Translational_Initiation; erismodegib (Hedgehog inhibitor), Hedge-hog_GLI_Pathway; and canertinib (pan-ERBB inhibitor), ERBB_Pathway. Taken together, these data suggest that tandutinib, olaparib, crizotinib, and AZD2014 might be a promising targeted therapy for GPC1 tumors

and that erismodegib and canertinib might be more promising for GPC2 tumors.

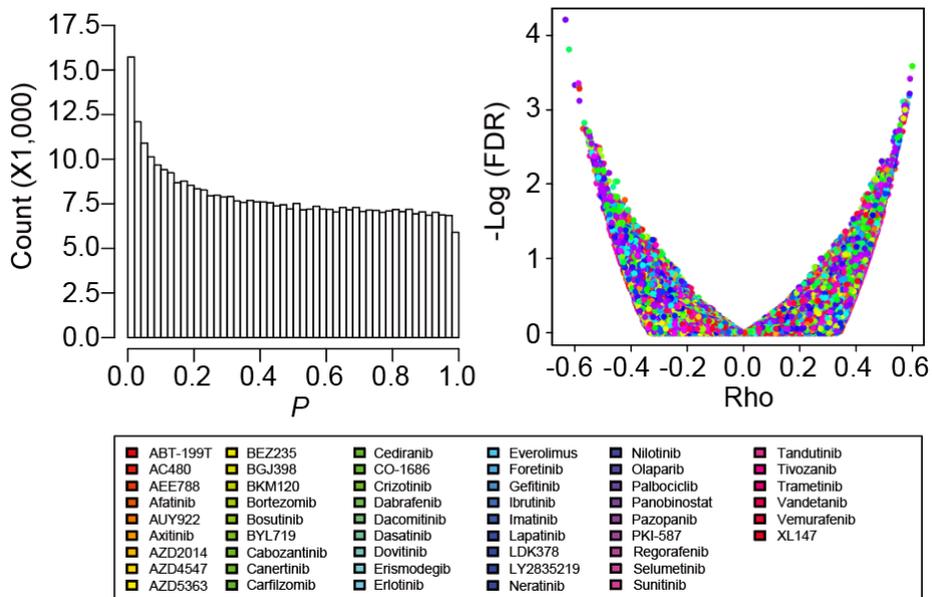


Figure 22. Correlation tests between the drug–response (AUC and ED50) and the abundance of global proteins. (Left) Histogram of P -values. (Right) FDR score distribution. Rho and P -values were obtained by Spearman’s correlation test.

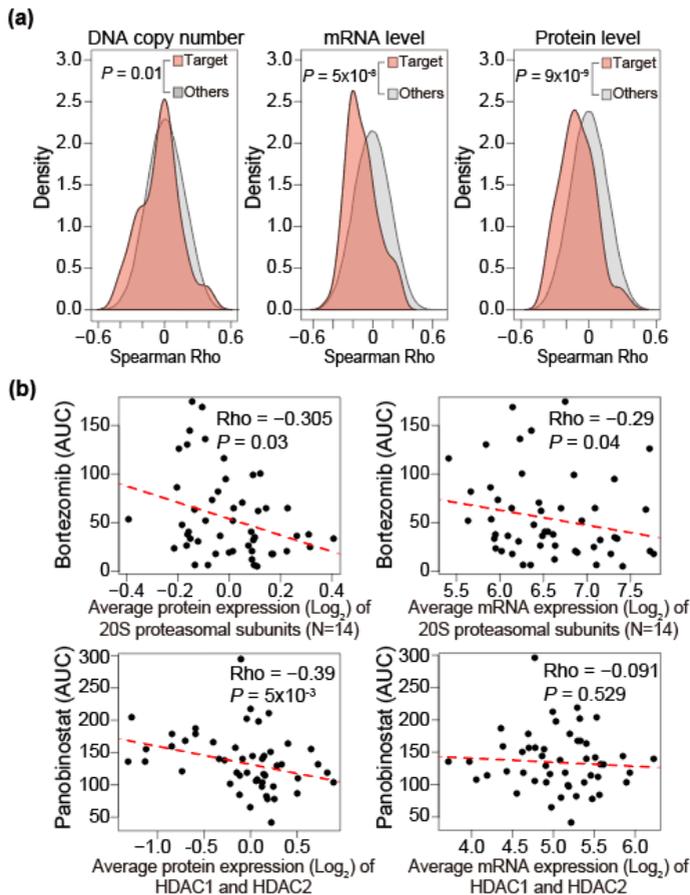


Figure 23. Protein expressions are the most correlative markers for targeted anti-cancer drugs. (a) Density plots for drug response-target correlations at the DNA copy number, RNA and protein levels. Genes with available protein data were used for comparison. P -values were obtained by two-sided unpaired Wilcoxon rank-sum test. (b) Correlations between expression levels of the indicated proteins (or mRNAs) and responses to the indicated pharmacological compounds targeting them. Rho and P -values were obtained by Spearman's correlation test.

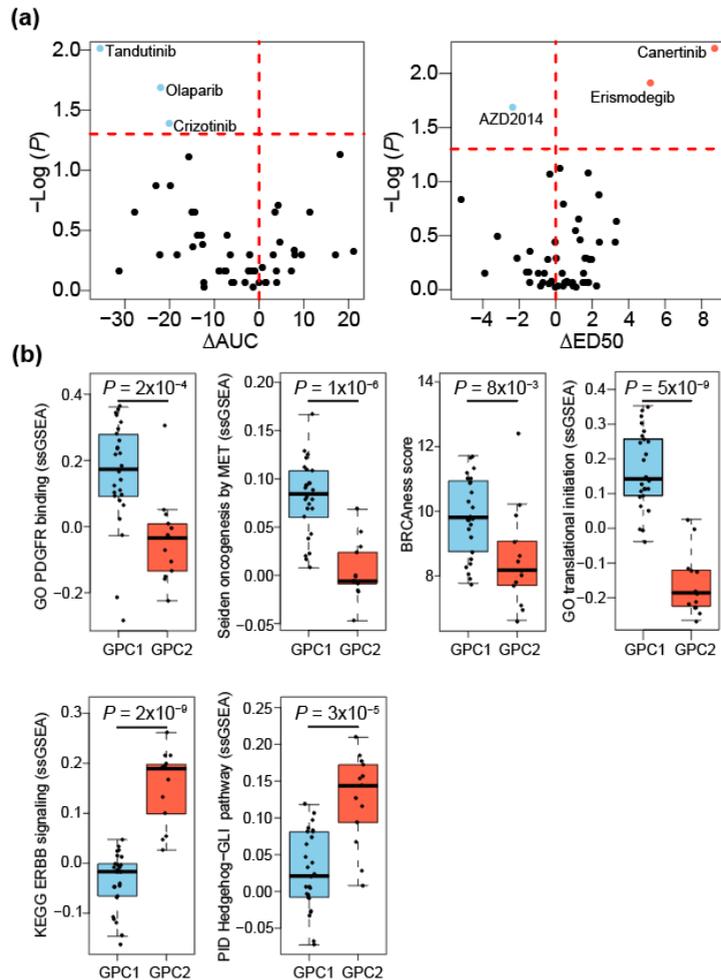


Figure 24. Subtype-specific anti-cancer drugs and activated pathways. (a) GPC1 or GPC2-selective cytotoxicity profile of the indicated targeted therapies to patient-derived tumor cells matched to the proteomic cohort. The x axes represent differences in mean AUC (left) and mean ED50 (right) values between GPC1 and GPC2 PDCs. Statistically significant GPC1 and GPC2-selective drugs were identified by two-sided Kolmogorov–Smirnov test (P -values < 0.05 ; horizontal dashed line) (blue: GPC1 sensitivity, red: GPC2 sensitivity). (b) Subtype-specific differences in relevant gene sets to GPC1-selective (upper panels) and GPC2-selective (lower panels) drugs in c. P -values were obtained with two-sided unpaired Wilcoxon rank-sum test. The box-and-whisker plots represent the medians (middle line), first quartiles (lower bound line), third quartiles (upper bound line) and the $\pm 1.5 \times$ interquartile ranges (whisker lines); the raw data are overlaid. PDGFR: platelet-derived growth factor receptor.

8. Protein markers inform sensitivities to targeted therapies

The majority of the 51 targeted agents used in this study consisted of receptor tyrosine kinases (RTKs), the phosphorylation of which serves as an important marker of their activation statuses. Therefore, I endeavored to uncover phosphoproteomic markers that are associated with specific drug responses using phosphoproteomic data for the 50 gliomas and drug-response data for the matched PDCs. The left-skewed *p*-value distribution that I obtained from the correlation test indicates the enrichment of statistically significant drug-response and phosphoprotein relationships (Figure 25).

As expected, hypersensitivities associated with the phosphorylation-mediated activation of a target protein were observed for several targeted agents. For example, hypersensitivities to afatinib, an EGFR inhibitor, correlated with EGFR-pY1197 (Figure 26a, left panel). Similarly, SRC pS17 and pY419 were associated with sensitivity to bosutinib (Figure 26a, middle and right panels).

For most drugs, however, stronger correlations were found for indirect phosphoproteins or the non-activating phosphorylation sites of the target protein than for activating phosphorylation sites in target proteins. For example, lapatinib, a dual EGFR and HER2 inhibitor showed marked association with EGFR-pT693, a known marker of receptor internalization (Figure 26b, upper left panel). Intriguingly, three other phosphoproteins commonly involved in RTK endocytic recycling—RAB4B-pS193, SNAP91-pT309 and ANK2-pS2516—were even more highly correlated with lapatinib sensitivity (Figure 26b), which suggests a possibility that EGFR recycling activity, rather than EGFR kinase activity, may determine responses to lapatinib. Other correlations that exhibited a high degree of

significance (FDR < 1%) included correlations between bosutinib and TSC2-pS1420, and between AZD4547 (FGFR inhibitor) and TOMM22-pS45 (Figure 26c). The functional impact of these phosphorylation sites and their relationships to the targeted agents are largely unknown.

My final investigation aimed to identify targeted agents selective to the most aggressive *IDH*-wildtype GBMs characterized by low PHGDH expression and high FKBP9 expression in my study (Figure 15). Here, I found that the most significant correlations with both PHGDH and FKBP9 were found for AZD2014 (Figure 27), suggesting that AZD2014 might be a promising targeted therapy for the aggressive *IDH*-wildtype GBM subtype.

In summary, my integrated analysis of pharmaco-proteomic data in patient-matched samples revealed proteomic-subtype-associated and protein biomarker-dependent sensitivities to targeted agents, which underscores the promise of proteome data for GBM therapy in the future.

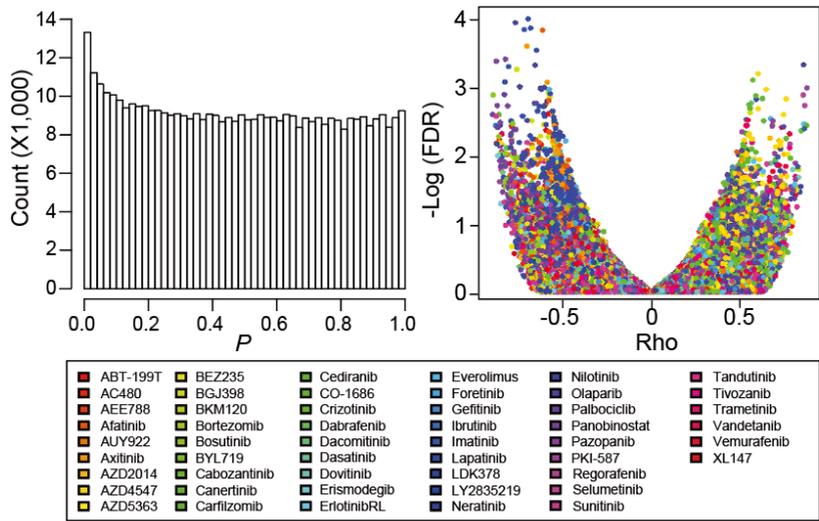


Figure 25. Correlation tests between the drug–response (AUC and ED50) and the phosphorylation level of proteins. (Left) Histogram of P -values. (Right) FDR score distribution. Rho and P -values were obtained by Spearman's correlation test.

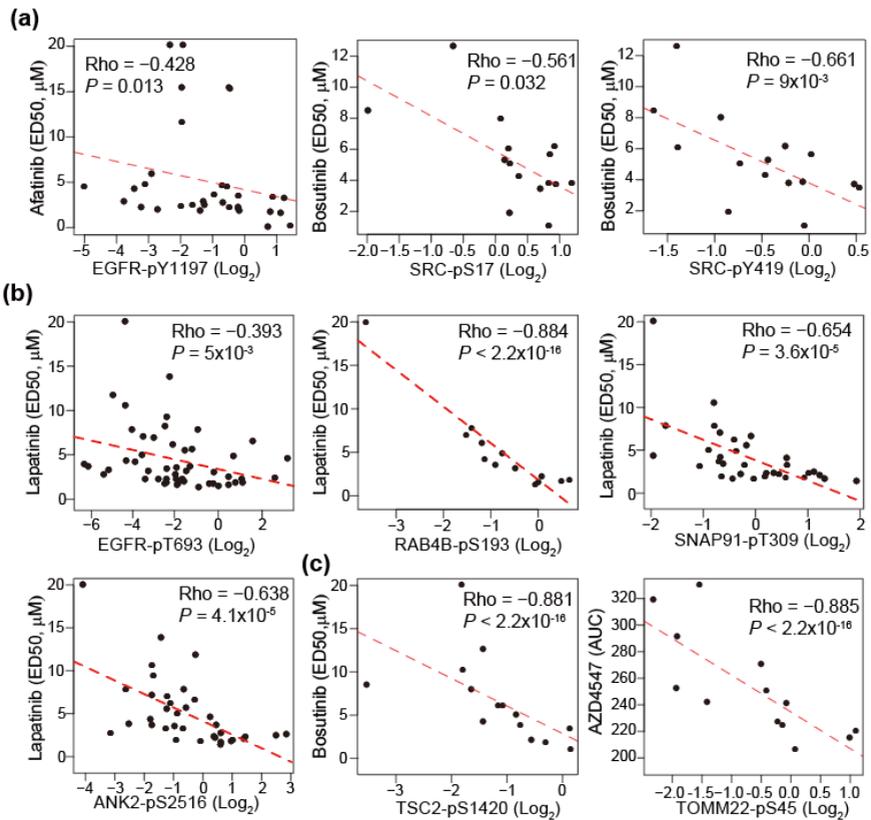


Figure 26. An integrative analysis with pharmacological profiles and phosphoproteome data. (a) Correlation between phosphoprotein levels and the response to the EGFR inhibitor afatinib and SRC inhibitor bosutinib. *P*-values were obtained by Spearman's correlation test. (b) Correlation between phosphoprotein levels and the response to the EGFR-HER2 inhibitor lapatinib. *P*-values were obtained by Spearman's correlation test. (c) Correlation between phosphoprotein levels and the response to bosutinib and FGFR inhibitor AZD4547. *P*-values were obtained by Spearman's correlation test.

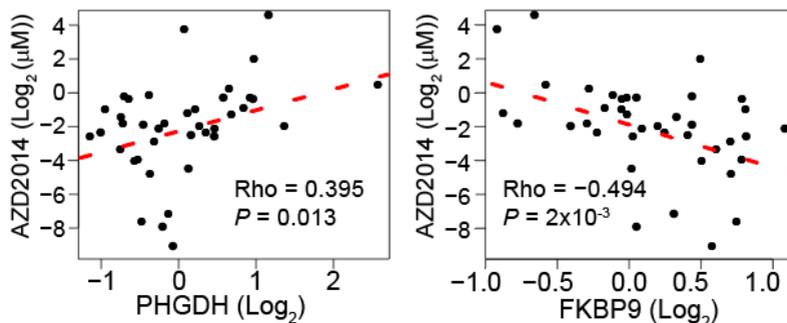


Figure 27. PHGDH and FKBP9 levels are associated with sensitivity to AZD2014. Correlation between the expression levels of the indicated proteins and the response to the indicated pharmacological compound. Rho and *P*-value were obtained by Spearman's correlation test.

9. Screening for pharmacological agents suppressing growth and invasion of GBM cell of origin

GBM is believed to arise from cancerous neural-stem cells or glioma-stem cells (GSCs) in sub-ventricular zone (SVZ)¹⁸. Accumulating studies indicate that GSCs account for drug-resistance and frequent recurrence of GBM. Therefore, targeting GSCs may be a fundamental therapeutics against GBM, particularly the malignant GPC1 tumors.

To identify the pharmacological intervention points and pathways related to glioblastoma-genesis from GSCs, I decided to use a murine GBM model (SVZ-1). The model is characterized by overexpression of human-*EGFRvIII* and CRISPR-Cas9 mediated double knock out of *Trp53* and *Pten*. I embedded SVZ-1 in matrigel and subjected it to a single-concentration (10 μ M) primary drug screening. The chemical library was composed of 6,869 chemical compounds that covered a wide range of genes and biological processes implicated in not only cancer but also other diseases (Figure 28). As a result, 1,425 compounds, covering 233 biological process, induced a greater than 16% decrease in cell invasiveness after 72 hrs of exposure (Figure 29 and Figure 30).

To gain insight into drug-to-pathway relationship, I made drug-set by grouping drugs that share identical targets. Next, I applied a modified single-sample GSEA (ssGSEA) to the measurements of primary screening with the drug-sets containing at least 5 compounds. I found that twenty-nine drug-sets significantly decreased invasion ability of SVZ-1 (Figure 31, Permutation test $P < 0.05$).

Further counter-screenings are needed to exclude poisonous compounds which are cytotoxic to normal stem cell derived spheroids.

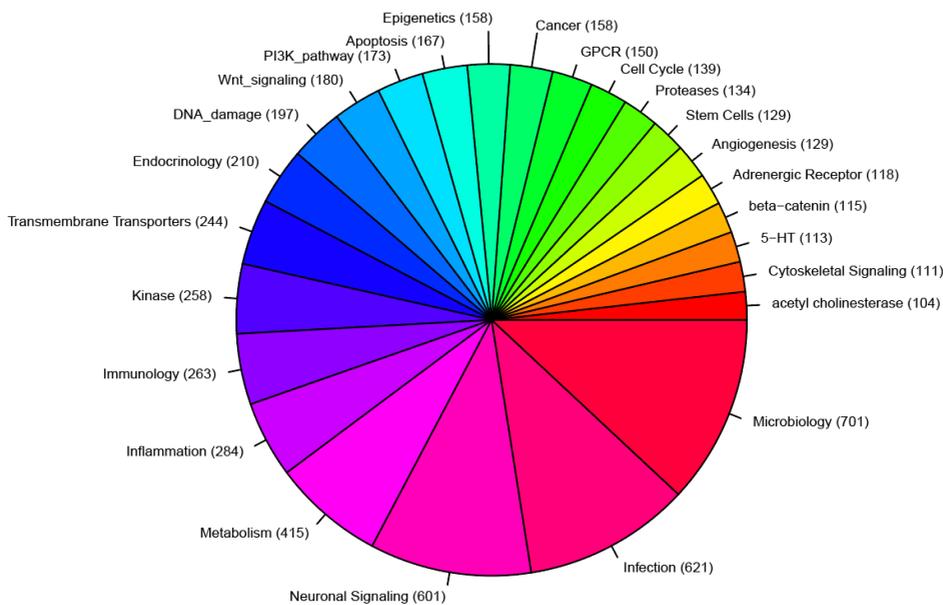


Figure 28. Classification of the 6,869 compounds. Pie chart shows pathways that are targeted by more than 100 compounds. 521 pathways are not shown due to their small number of compounds. The number of compounds with unknown target is 2,487.

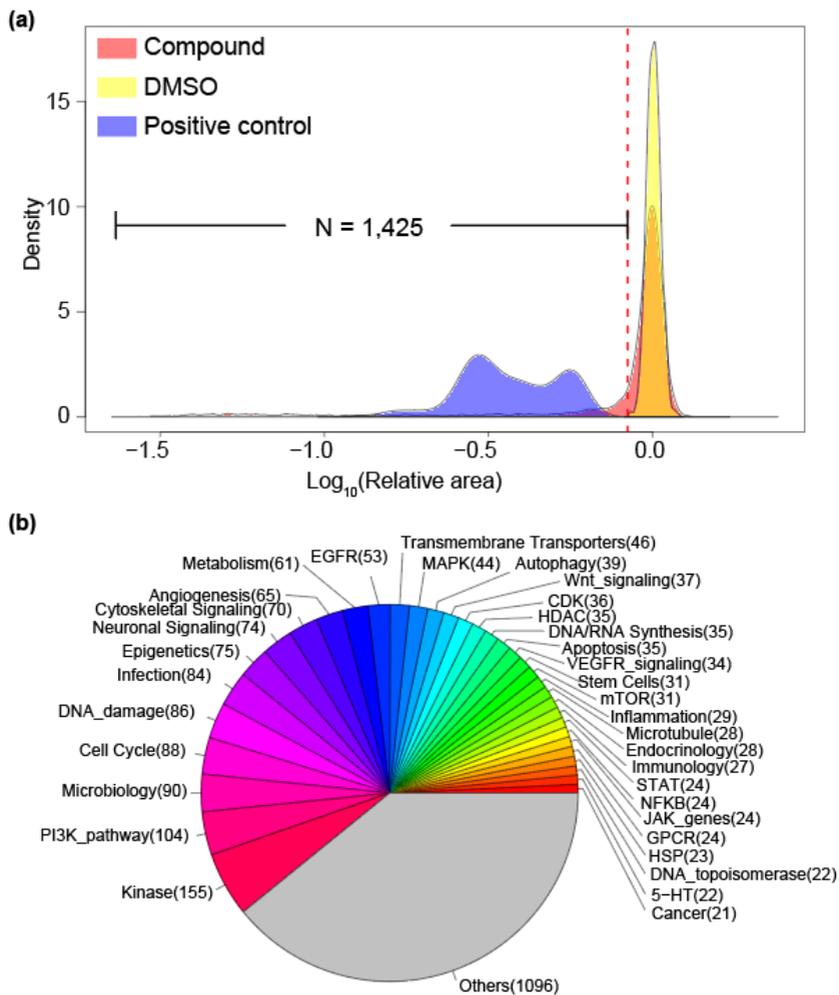


Figure 29. Primary screening results. (a) Overall distribution of relative area normalized to internal DMSO controls. The distribution of chemical compounds, DMSO, and positive controls are colored by red, yellow, and blue, respectively. Red-dotted line indicates cut-off value (0.84). (b) Pie chart represents the individual number of candidates according to target pathways.

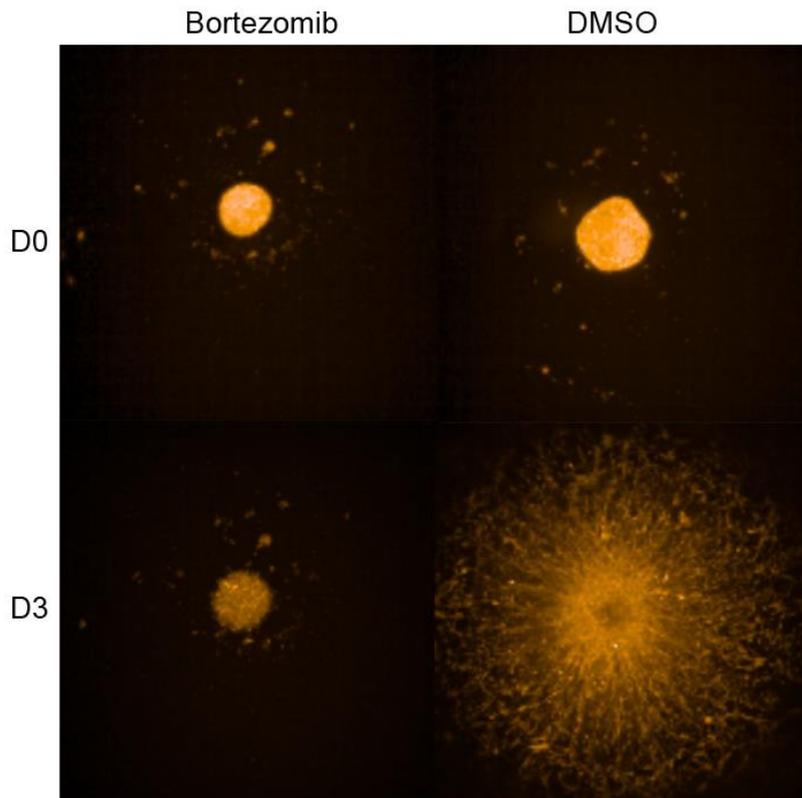


Figure 30. Representative images for primary screening. D0 (upper panel) and D3 (lower panel) represent day0 and day3 respectively.

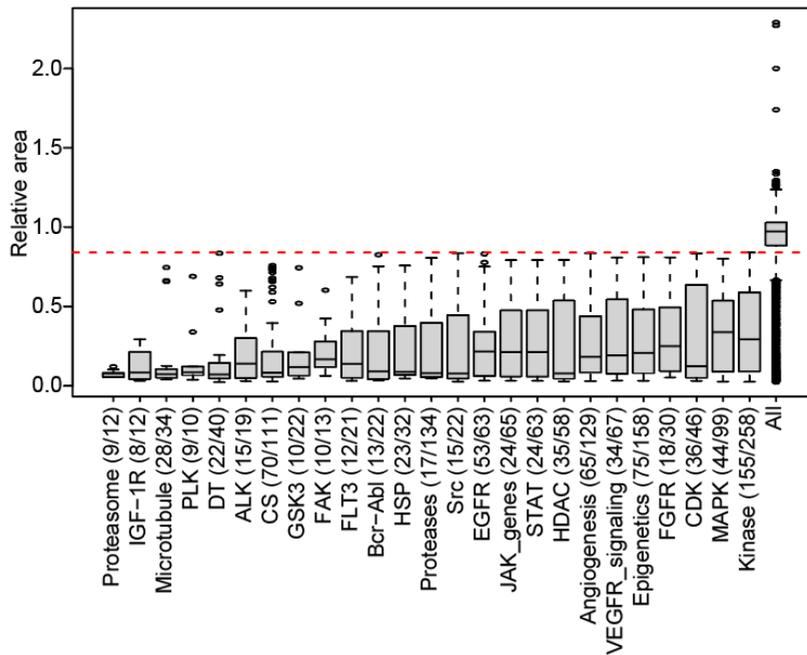


Figure 31. Candidate pathways that significantly suppressed the invasiveness of SVZ-1. The number of candidate compounds and size of the given pathways are in parenthesis. The box-and-whisker plots represent the medians (middle line), first quartiles (lower bound line), third quartiles (upper bound line) and the $\pm 1.5 \times$ interquartile ranges (whisker lines). The red-dotted line indicates cut-off value (0.84).

IV. DISCUSSION

This study employed a large-scale quantitative proteomic approach using LC-MS/MS technology to characterize the intertumoral heterogeneity of 39 *IDH*-wildtype GBM tumors. I found two stable proteomic subtypes of *IDH*-wildtype GBM, GPC1 and GPC2, which are primarily characterized by low or high expression of OXPHOS-related proteins, respectively. GPC1-subtype tumors displayed distinct Warburg-like proteomic features, with increased LDHA, PKM2, and HK2 expression that facilitates lactate production and confers resistance to hypoxic stress in cancer cells^{55,56}. The Warburg effect drives the biosynthesis of nucleotides, lipids, and proteins to support rapid cell proliferation, as well as the disruption of tissue architecture to facilitate tumor motility and immune-cell evasion in the tumor microenvironment⁴⁸. NADPH is a key component in this process, thus the elevated PPP, one-carbon pathway, and IDH1 levels in GPC1 tumors provide additional support for my hypothesis. As I found no direct correlation between the RNA and protein levels of OXPHOS-related proteins (presumably due to their dependence on post-translational proteolysis and protein turnover to control abundance)^{15,57,58}, my proteome-based classification exhibited no significant correlative relationship with previously identified RNA-based subtypes. Therefore, the proteomic subtypes represent previously unrecognized *IDH*-wildtype GBM subgroups. However, some of the samples, particularly in GPC2, exhibited heterogeneous expression patterns, compared to other samples in the subtype, suggesting that increased sample size may lead to additional subtype(s) distinct from the two major GPC subtypes.

The identification of prognostic biomarkers for *IDH*-wildtype GBM remains a challenge. Indeed, among the previously reported 271 gene expression biomarkers that are prognostic for GBM⁴⁴, only three were validated in my *IDH*-wildtype GBM cohort at the protein level. Of these, PHGDH (a rate-limiting enzyme of serine biosynthesis) was elevated in GPC2 tumors of *IDH*-wildtype GBM and

showed the strongest association with long-term survival. IDH-mutant proteins produce the oncometabolite D-2-HG and predict a favorable prognosis in glioma⁹. Similarly, PHGDH is known to have a promiscuous function to generate D-2-HG⁵¹. Thus, further research may be needed to evaluate whether the canonical PHGDH function or its promiscuous function is associated with favorable prognosis in *IDH*-wildtype GBM.

Understanding the mechanistic connections among GPC1 tumor-specific features is important because these tumors carry poor prognostic biomarkers. Nestin is expressed in adult NSCs that reside in the sub-ventricular zone (SVZ) of the brain: it forms a heterodimer with Vimentin during mitosis, promotes the disassembly of intermediate filaments, and supports the survival and renewal of neural progenitor cells. A recent study showed that 56% of human GBM cases originate from SVZ-derived GSCs¹⁸. Thus, I hypothesized that GPC1 subtype might originate from SVZ-derived GSCs. In further support of this hypothesis, I found that GPC1 tumors expressed higher protein levels of phosphorylated cortactin and the Arp2/3 complex. Phosphorylated cortactin activates the Arp2/3 complex to mediate a mechanism by which cancer cells might facilitate actin filamentation and branching while remodeling the extracellular matrix to gain increased motility and invasiveness⁵⁹. GSC maintenance depends on hypoxia inducible factor (HIF) 1 α ⁶⁰ consistently, I found that GPC1 tumors had activated HIF pathway. HIF1 α inhibits cell differentiation by activating signaling pathways driven by Notch, NANOG, TGF β , and SOX2. Metabolically, HIF1 α activates GLUT, PFK1, HK2, and LDHA, thus inducing aerobic glycolysis and leading to increased glucose uptake due to a low ATP yield. Increased glucose uptake and lactate secretion subsequently facilitate immune evasion through immune checkpoint ligands and metabolic competition⁶¹. HIF1 α also activates the one-carbon (folate) pathway to protect cancer stem cells from increased oxidative stress by increasing NADPH (and glutathione) production⁶². This process is consistent with my findings that NADPH-producing IDH1, the PPP and one-carbon pathway were elevated in

GPC1 tumors.

I also demonstrated GBM intratumoral heterogeneity in terms of proteome subtype, although one subtype generally dominated a tumor. This finding is consistent with the recent observation reported in the Ivy Glioblastoma Atlas (<http://glioblastoma.alleninstitute.org>), that the differences in tumor characteristics separated by anatomical location are as large as the differences observed between other tumors⁶³. A subtype switch observed in the longitudinal samples might also highlight intratumoral heterogeneity as a major challenge to successful GBM treatment. My results imply that the two GPC subtypes might need to be controlled simultaneously in GBM treatment to prevent recurrence and have a therapeutic effect.

Predicting an accurate anti-cancer drug response with molecular profiles is one of the most important issue in GBM therapy. I assessed the direct chemical liabilities linked to proteomic information using the previously generated data set with 50 matched PDCs against 51 targeted anticancer drugs. Here, I demonstrated concordant subtype-specific target pathway (BRCAness) activation with olaparib showing efficacy for GPC1 tumor-matched PDCs. Olaparib is under clinical trials with radiotherapy (for MGMT unmethylated patients) and radiotherapy-temozolomide (for MGMT methylated patients) in newly diagnosed GBM⁶⁴. My proteomic classification may guide further patient selection criteria for these efforts. I also demonstrated that AZD2014, a dual mTORC1 and mTORC2 inhibitor, exhibited strong cytotoxicity toward the most aggressive PHGDH-low and FKBP9-high *IDH*-wildtype GBM-derived PDCs. Despite the failure of mTORC1 inhibitors in clinical trials, dual mTORC1/2 inhibitors are increasingly gaining traction in GBM translational research because preclinical evidence indicates that mTORC2 has divergent roles from mTORC1 in facilitating GBM growth, invasiveness, and GSC proliferation^{65,66}, and that mTORC2 activity is selectively enhanced in grade IV tumors⁶⁷. My data are in line with these observations and thus should motivate further clinical studies.

Finally, I conducted a high-throughput drug screening to identify pharmacological intervention points for malignant GPC1 tumors. I found several drug-sets that effectively reduced invasion ability of SVZ-1. For example, proteasome inhibitor-set showed the most significant anti-invasion effect. In line with the elevated signatures related to unfolded protein response in GPC1, this result suggests potential of proteasome inhibitor to treat GPC1 tumors. Currently, a phase III clinical trial (NCT03345095) for GBM is on-going with marizomib, a BBB-crossing second-generation proteasome inhibitor⁶⁸. IGF-1R drug-set is also interesting group. The importance of IGF system in GBM pathogenesis is widely supported because of its roles in apoptosis and cell migration. As normal brain cells barely express IGF-genes, targeting IGF system may have a wide therapeutic window⁶⁹. However, further study should be needed to evaluate cytotoxicity in normal cells, to exclude general toxins, and to find mechanism of action.

In conclusion, my integrated pharmaco-proteogenomic analyses highlight the importance of using proteomic data to understand the connections between GBM cellular origin, oncogenic signaling and metabolic diversity, all of which shape distinct binary molecular states. My data illuminate unique therapeutic vulnerabilities coupled to these binary molecular states and biomarker proteins and suggest potentially effective therapeutic strategies for GBM.

V. CONCLUSION

Pharmaco-proteogenomic analysis with *IDH*-wildtype GBM samples highlights discovered two distinct proteomic subtypes marked by different expression of OXPHOS-related proteins. The GPC1 showed protein signatures associated with the Warburg-effect, neural-stem cell, and poor patient prognosis.

My findings lead to pharmacological-characterization with murine-GBM model mirroring GPC1 tumor. Through the primary screening, I found 1,425 candidates and interesting drug-sets related to GBM biology. I will select the most effective compound by on-going screenings and find its mechanism of action.

Collectively, my results will guide more accurate patient stratification and future directions of GBM therapeutics.

REFERENCES

1. Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med* 2005;352:987-96.
2. Chen B, Wang H, Ge P, Zhao J, Li W, Gu H, et al. Gross total resection of glioma with the intraoperative fluorescence-guidance of fluorescein sodium. *Int J Med Sci* 2012;9:708-14.
3. Osuka S and Van Meir EG. Overcoming therapeutic resistance in glioblastoma: the way forward. *J Clin Invest* 2017;127:415-426.
4. Weller M, Cloughesy T, Perry JR, and Wick W. Standards of care for treatment of recurrent glioblastoma--are we there yet? *Neuro Oncol* 2013;15:4-27.
5. Brennan CW, Verhaak RGW, McKenna A, Campos B, Nounshmehr H, Salama SR, et al. The somatic genomic landscape of glioblastoma. *Cell* 2013;155:462-477.
6. Ceccarelli M, Barthel FP, Malta TM, Sabedot TS, Salama SR, Murray BA, et al. Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell* 2016;164:550-63.
7. McLendon R, Friedman A, Bigner D, Van Meir EG, Brat DJ, M. Mastrogiannis G, et al. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 2008;455:1061-1068.
8. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 2010;17:98-110.
9. Wesseling P and Capper D. WHO 2016 Classification of gliomas. *Neuropathol Appl Neurobiol* 2018;44:139-150.
10. Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. *N Engl J Med* 2005;352:997-1003.
11. Cairncross G, Wang M, Shaw E, Jenkins R, Brachman D, Buckner J, et al. Phase III trial of chemoradiotherapy for anaplastic oligodendroglioma: long-term results of RTOG 9402. *J Clin Oncol* 2013;31:337-43.
12. Tateishi K, Wakimoto H, and Cahill DP. IDH1 Mutation and World Health Organization 2016 Diagnostic Criteria for Adult Diffuse Gliomas: Advances in Surgical Strategy. *Neurosurgery* 2017;64:134-138.
13. Zhang B, Wang J, Wang X, Zhu J, Liu Q, Shi Z, et al. Proteogenomic characterization of human colon and rectal cancer. *Nature* 2014;513:382-7.
14. Mertins P, Mani DR, Ruggles KV, Gillette MA, Clauser KR, Wang P, et al. Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* 2016;534:55-62.

15. Zhang H, Liu T, Zhang Z, Payne SH, Zhang B, McDermott JE, et al. Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell* 2016;166:755-765.
16. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 2014;344:1396-401.
17. Wang Q, Hu B, Hu X, Kim H, Squatrito M, Scarpace L, et al. Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer Cell* 2017;32:42-56.e6.
18. Lee JH, Lee JE, Kahng JY, Kim SH, Park JS, Yoon SJ, et al. Human glioblastoma arises from subventricular zone cells with low-level driver mutations. *Nature* 2018;560:243-247.
19. Li H and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754-60.
20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* (Oxford, England) 2009;25:2078-2079.
21. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43:491-8.
22. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013;31:213-9.
23. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biology* 2016;17:122.
24. Gao J, Chang MT, Johnsen HC, Gao SP, Sylvester BE, Sumer SO, et al. 3D clusters of somatic mutations in cancer reveal numerous rare mutations as functional targets. *Genome Med* 2017;9:4.
25. Chang MT, Asthana S, Gao SP, Lee BH, Chapman JS, Kandath C, et al. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat Biotechnol* 2016;34:155-63.
26. Robinson MD, McCarthy DJ, and Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139-40.
27. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukheim R, and 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:R41.
28. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, and Mesirov JP. GenePattern 2.0. *Nat Genet* 2006;38:500-1.
29. Kwon Y, Ju S, Kaushal P, Lee JW, and Lee C. Neutralizing the Detrimental Effect of an N-Hydroxysuccinimide Quenching Reagent on

- Phosphopeptide in Quantitative Proteomics. *Anal Chem* 2018;90:3019-3023.
30. Kyono Y, Sugiyama N, Imami K, Tomita M, and Ishihama Y. Successive and selective release of phosphorylated peptides captured by hydroxy acid-modified metal oxide chromatography. *J Proteome Res* 2008;7:4585-93.
 31. Thingholm TE, Jørgensen TJ, Jensen ON, and Larsen MR. Highly selective enrichment of phosphorylated peptides using titanium dioxide. *Nat Protoc* 2006;1:1929-35.
 32. Tyanova S, Temu T, and Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* 2016;11:2301-2319.
 33. Madar IH, Lee W, Wang X, Ko S-I, Kim H, Mun D-G, et al. Comprehensive and sensitive proteogenomics data analysis strategy based on complementary multi-stage database search. *International Journal of Mass Spectrometry* 2018;427:11-19.
 34. Park H, Bae J, Kim H, Kim S, Kim H, Mun DG, et al. Compact variant-rich customized sequence database and a fast and sensitive database search for efficient proteogenomic analyses. *Proteomics* 2014;14:2742-9.
 35. Kim S and Pevzner PA. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat Commun* 2014;5:5277.
 36. Şenbabaoğlu Y, Michailidis G, and Li JZ. Critical limitations of consensus clustering in class discovery. *Sci Rep* 2014;4:6207.
 37. Merico D, Isserlin R, Stueker O, Emili A, and Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* 2010;5:e13984.
 38. Türei D, Korcsmáros T, and Saez-Rodriguez J. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat Methods* 2016;13:966-967.
 39. Konstantinopoulos PA, Spentzos D, Karlan BY, Taniguchi T, Fountzilias E, Francoeur N, et al. Gene expression profile of BRCAness that correlates with responsiveness to chemotherapy and with outcome in patients with epithelial ovarian cancer. *J Clin Oncol* 2010;28:3555-61.
 40. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* 2016;44:e71.
 41. Park J, Shim JK, Yoon SJ, Kim SH, Chang JH, and Kang SG. Transcriptome profiling-based identification of prognostic subtypes and multi-omics signatures of glioblastoma. *Sci Rep* 2019;9:10555.
 42. Ducray F, de Reyniès A, Chinot O, Idbaih A, Figarella-Branger D, Colin C, et al. An ANOCEF genomic and transcriptomic microarray study of the response to radiotherapy or to alkylating first-line chemotherapy in

- glioblastoma patients. *Mol Cancer* 2010;9:234.
43. Darmanis S, Sloan SA, Croote D, Mignardi M, Chernikova S, Samghababi P, et al. Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma. *Cell Rep* 2017;21:1399-1410.
 44. Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhori G, et al. A pathology atlas of the human cancer transcriptome. *Science* 2017;357.
 45. Wen Z, Zhong Z, and Darnell JE, Jr. Maximal activation of transcription by Stat1 and Stat3 requires both tyrosine and serine phosphorylation. *Cell* 1995;82:241-50.
 46. Molenaar RJ, Radivoyevitch T, Maciejewski JP, van Noorden CJ, and Bleeker FE. The driver and passenger effects of isocitrate dehydrogenase 1 and 2 mutations in oncogenesis and survival prolongation. *Biochim Biophys Acta* 2014;1846:326-41.
 47. Wahl DR, Dresser J, Wilder-Romans K, Parsels JD, Zhao SG, Davis M, et al. Glioblastoma Therapy Can Be Augmented by Targeting IDH1-Mediated NADPH Biosynthesis. *Cancer Res* 2017;77:960-970.
 48. Liberti MV and Locasale JW. The Warburg Effect: How Does it Benefit Cancer Cells? *Trends Biochem Sci* 2016;41:211-218.
 49. Li H, Ning S, Ghandi M, Kryukov GV, Gopal S, Deik A, et al. The landscape of cancer cell line metabolism. *Nat Med* 2019;25:850-860.
 50. Lathia JD, Mack SC, Mulkearns-Hubert EE, Valentim CL, and Rich JN. Cancer stem cells in glioblastoma. *Genes Dev* 2015;29:1203-17.
 51. Ye D, Guan KL, and Xiong Y. Metabolism, Activity, and Targeting of D- and L-2-Hydroxyglutarates. *Trends Cancer* 2018;4:151-165.
 52. Lee JK, Liu Z, Sa JK, Shin S, Wang J, Bordyuh M, et al. Pharmacogenomic landscape of patient-derived tumor cells informs precision oncology therapy. *Nat Genet* 2018;50:1399-1411.
 53. Arlt A, Bauer I, Schafmayer C, Tepel J, Mürköster SS, Brosch M, et al. Increased proteasome subunit protein expression and proteasome activity in colon cancer relate to an enhanced activation of nuclear factor E2-related factor 2 (Nrf2). *Oncogene* 2009;28:3983-96.
 54. Ropero S and Esteller M. The role of histone deacetylases (HDACs) in human cancer. *Mol Oncol* 2007;1:19-25.
 55. Dong G, Mao Q, Xia W, Xu Y, Wang J, Xu L, et al. PKM2 and cancer: The function of PKM2 beyond glycolysis. *Oncol Lett* 2016;11:1980-1986.
 56. Wolf A, Agnihotri S, Micallef J, Mukherjee J, Sabha N, Cairns R, et al. Hexokinase 2 is a key mediator of aerobic glycolysis and promotes tumor growth in human glioblastoma multiforme. *J Exp Med* 2011;208:313-26.
 57. Jovanovic M, Rooney MS, Mertins P, Przybylski D, Chevrier N, Satija R, et al. Dynamic profiling of the protein life cycle in response to pathogens. *Science* 2015;347:1259038.

58. Ali S and McStay GP. Regulation of Mitochondrial Dynamics by Proteolytic Processing and Protein Turnover. *Antioxidants (Basel)* 2018;7.
59. MacGrath SM and Koleske AJ. Cortactin in cell migration and cancer at a glance. *J Cell Sci* 2012;125:1621-6.
60. Qiang L, Wu T, Zhang HW, Lu N, Hu R, Wang YJ, et al. HIF-1 α is critical for hypoxia-mediated maintenance of glioblastoma stem cells by activating Notch signaling pathway. *Cell Death Differ* 2012;19:284-94.
61. Chang CH, Qiu J, O'Sullivan D, Buck MD, Noguchi T, Curtis JD, et al. Metabolic Competition in the Tumor Microenvironment Is a Driver of Cancer Progression. *Cell* 2015;162:1229-41.
62. Semenza GL. Hypoxia-inducible factors: coupling glucose metabolism and redox regulation with induction of the breast cancer stem cell phenotype. *Embo j* 2017;36:252-259.
63. Puchalski RB, Shah N, Miller J, Dalley R, Nomura SR, Yoon JG, et al. An anatomic transcriptional atlas of human glioblastoma. *Science* 2018;360:660-663.
64. Fulton B, Short SC, James A, Nowicki S, McBain C, Jefferies S, et al. PARADIGM-2: Two parallel phase I studies of olaparib and radiotherapy or olaparib and radiotherapy plus temozolomide in patients with newly diagnosed glioblastoma, with treatment stratified by MGMT status. *Clin Transl Radiat Oncol* 2018;8:12-16.
65. Kahn J, Hayman TJ, Jamal M, Rath BH, Kramp T, Camphausen K, et al. The mTORC1/mTORC2 inhibitor AZD2014 enhances the radiosensitivity of glioblastoma stem-like cells. *Neuro Oncol* 2014;16:29-37.
66. Mecca C, Giambanco I, Bruscoli S, Bereshchenko O, Fioretti B, Riccardi C, et al. PP242 Counteracts Glioblastoma Cell Proliferation, Migration, Invasiveness and Stemness Properties by Inhibiting mTORC2/AKT. *Front Cell Neurosci* 2018;12:99.
67. Alvarenga AW, Machado LE, Rodrigues BR, Lupinacci FC, Sanemastu P, Matta E, et al. Evaluation of Akt and RICTOR Expression Levels in Astrocytomas of All Grades. *J Histochem Cytochem* 2017;65:93-103.
68. Ma L and Diao A. Marizomib, a potent second generation proteasome inhibitor from natural origin. *Anticancer Agents Med Chem* 2015;15:298-306.
69. Maris C, D'Haene N, Trépant AL, Le Mercier M, Sauvage S, Allard J, et al. IGF-IR: a new prognostic biomarker for human glioblastoma. *Br J Cancer* 2015;113:729-37.

ABSTRACT(IN KOREAN)

약물유전단백체적 교모세포종 분석

<지도교수 김 현 석>

연세대학교 대학원 의과학과

오 세 진

교모세포종은 뇌종양 중 가장 짧은 생존기간과 높은 재발율 그리고 극히 제한적인 치료법만 있는 특징을 갖는다. 따라서, 나는 39례의 IDH 야생형 교모세포종 샘플과 11례의 저등급 신경교종 및 4례의 정상 뇌조직 샘플을 활용하여 핵심 발암 기전과 표적가능한 약물작용점을 약물유전단백체적 접근을 통해 찾았다. 나는 2개의 단백질 아형으로 교모세포종이 나뉘며 이것에는 산화적-인산화에 관여하는 단백질들이 중요하단 것을 밝혔다. 특히, GPC1 종양들은 와버그 효과, 나쁜 환자 예후 및 뇌실하대 영역에 있는 신경줄기세포 관련 단백질들의 발현이 높았다. 뇌실하대 영역부터 종양발생 지역으로 이동하는 교모세포종 발암과정에서 중요한 신호전달체계를 파악하기 위해 나는 대용량 약물 스크리닝을 GPC1 아형을 모사한 설치류 유래 교모세포종 모델을 가지고 진행했다. 나의 연구는 교모세포종에 대한 새로운 치료전략을 제시할 것이다.

핵심되는 말 : 교모세포종, 약물유전단백체학, 종양 진화

PUBLICATION LIST

1. **Oh S**, Yeom J, Cho HJ, Kim JH, Yoon SJ, Kim H, et al. Integrated pharmacoproteogenomics defines two subgroups in isocitrate dehydrogenase wild-type glioblastoma with prognostic and therapeutic opportunities. *Nat Commun* 2020;11:3288.
2. Lee Y, Lee CE, **Oh S**, Kim H, Lee J, Kim SB, et al. Pharmacogenomic Analysis Reveals CCNA2 as a Predictive Biomarker of Sensitivity to Polo-Like Kinase I Inhibitor in Gastric Cancer. *Cancers* 2020;12.
3. Lee J, Kim H, Lee JE, Shin S-J, **Oh S**, Kwon G, et al. Selective Cytotoxicity of the NAMPT Inhibitor FK866 Toward Gastric Cancer Cells With Markers of the Epithelial-Mesenchymal Transition, Due to Loss of NAPRT. *Gastroenterology* 2018;155:799-814.e13.
4. **Oh S** and Kim HS. Emerging power of proteomics for delineation of intrinsic tumor subtypes and resistance mechanisms to anti-cancer therapies. *Expert Review of Proteomics* 2016;13:929-939.