Resolving the molecular complexity of brain tumors through machine learning approaches for precision medicine

Sandanaraj, Edwin

2019

Sandanaraj, E. (2019). Resolving the molecular complexity of brain tumors through machine learning approaches for precision medicine. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/136519>

<https://doi.org/10.32657/10356/136519>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**RESOLVING THE MOLECULAR COMPLEXITY OF
BRAIN TUMORS THROUGH MACHINE LEARNING
APPROACHES FOR PRECISION MEDICINE**

**SANDANARAJ EDWIN
SCHOOL OF BIOLOGICAL SCIENCES**

2019

**RESOLVING THE MOLECULAR COMPLEXITY OF
BRAIN TUMORS THROUGH MACHINE LEARNING
APPROACHES FOR PRECISION MEDICINE**

SANDANARAJ EDWIN

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological
University in partial fulfillment of the requirement for the
degree of Doctor of Philosophy

2019

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research done by me except where otherwise stated in this thesis. The thesis work has not been submitted for a degree or professional qualification to any other university or institution. I declare that this thesis is written by myself and is free of plagiarism and of sufficient grammatical clarity to be examined. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

03.10.2019

.....

Date

.....

Sandanaraj Edwin

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it of sufficient grammatical clarity to be examined. To the best of my knowledge, the thesis is free of plagiarism and the research and writing are those of the candidate's except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

03 October 2019

.....

Date

.....

Assoc Prof Andrew Tan

Authorship Attribution Statement

This thesis contains material from two papers published in the following peer-reviewed journals in which I am listed as the co-first author.

Chapter 4 is published as Chong YK*, **Sandanaraj E***, Koh LW, Thangaveloo M, Tan MS, Koh GR, Toh TB, Lim GG, Holbrook JD, Kon OL, Nadarajah M, Ng I, Ng WH, Tan NS, Lim KL, Tang C, Ang BT. ST3GAL1-Associated Transcriptomic Program in Glioblastoma Tumor Growth, Invasion, and Prognosis. *J Natl Cancer Inst.* 2015 Nov 7;108(2).

***Authors contributed equally to this work.** We obtained copyright clearance from Oxford University Press to present the results of published manuscript in this thesis work (License number: 4342451258210).

The contributions of the co-authors are as follows:

- *Dr Tang and A/Prof Ang BT provided the initial project direction and edited the manuscript drafts.*
- *Dr Chong and I arrived at the conclusions of the study findings. This manuscript was drafted by Dr Tang, Dr Chong, A/Prof Ang BT, A/Prof Tan NS and myself.*
- *I designed the computational workflow of the current study with Dr Tang and Prof Holbrook. I conducted the bioinformatics analyses of the study at National Neuroscience Institute and National*

Supercomputing Centre. I executed patient databases mining and predictive learning methodologies of the current work.

- *Clinical correlations of patient database analyses were interpreted by Dr Tang, A/Prof Ang BT, A/Prof Ng, Dr Nadarajah and Dr Ng I.*
- *Dr Chong and Dr Tang worked on the functional evaluations, in vitro and in vivo experiments conducted in the study. Koh LW, Thangaveloo M, Tan MS, Koh GR and Toh TB provided technical assistance to Dr Chong. Prof Lim and Prof Kon provided intellectual insights to our bench experiments.*

Chapter 5 is published as Tan MS*, **Sandanaraj E***, Chong YK, Lim SW, Koh LWH, Ng WH, Tan NS, Tan P, Ang BT, Tang C. Signal transducer and activator of transcription 3-based glioma patient stratification and its application in precision oncology. **Nat Commun. 2019 Aug 9;10(1):3601. doi: 10.1038/s41467-019-11614-x.** * Authors contributed equally to this work.

The contributions of the co-authors are as follows:

- *Dr Tang and A/Prof Ang BT provided the initial project direction and edited the manuscript drafts.*
- *Dr Tan MS and I arrived at the conclusions of the study findings. This manuscript was drafted by Dr Tang, Dr Tan MS, A/Prof Ang BT, A/Prof Tan NS and myself.*
- *I designed the computational workflow of the current study with Dr Tang and Prof Tan P. I conducted the bioinformatics analyses of*

the study at National Neuroscience Institute and National Supercomputing Centre. I executed patient databases mining and predictive learning methodologies of the current work.

- *Clinical correlations of patient database analyses were interpreted by Dr Tang, A/Prof Ang BT and A/Prof Ng WH.*
- *Dr Tan MS and Dr Tang worked on the functional evaluations, in vitro and in vivo experiments conducted in the study. Dr Chong YK, Lim SW and Koh LWH provided technical assistance to Dr Tan MS.*

03.10.2019

.....

Date

.....

Sandanaraj Edwin

Acknowledgements

I would like to express my heartfelt gratitude to my supervisors Dr Carol Tang and A/Prof Ang Beng Ti for their guidance and mentorship throughout my graduate studies. My supervisors extended substantial encouragement, motivation and enthusiasm throughout my PhD journey. Their critical thinking, scientific acumen and amazing presentation skills inspire me constantly and provide strong support to build my research path. I am greatly indebted to my supervisors for their passionate mentorship. I would like to render my special appreciation to my main supervisor A/Prof Andrew Tan for his enthusiastic support and advice during my PhD studies. I am extremely grateful to all my supervisors for their immense help and guidance to carry out this thesis work. Without their encouragement and patience, it would not have been possible to reach this destination.

I thank my colleagues Melanie, Lynnette, Yuk Kien and See Wee for providing supportive functional evidences to the computational predcitions identified in the present study. We always have had several stimulative discussions during the course of my research work. They patiently imparted their knowledge in several aspects of bench experiments and animal models. I also thank Tan Boon (SINAPSE) for his helpful discussions in writing my thesis.

I wish to thank my wife Claudia Shalini, for her supportive words and constant encouragement throughout this entire journey in achieving my dreams. Thanks to my children Judith Sandrea and Jenith Samantha for being patient in the course of my research work. Finally, I thank my

family and friends for their wishes and prayers to achieve this milestone of my research track.

Lastly, I thank National Neuroscience Institute, Singapore Institute for Clinical Sciences and National Supercomputing Centre for providing supportive environment to carry out my research work.

Table of Contents

Acknowledgements	i
Table of Contents	i
List of Figures.....	v
List of Tables	vii
Abbreviations.....	ix
Abstract	1
1 INTRODUCTION	3
1.1 Molecular classification of gliomas	4
1.1.1 Transcriptomic subtypes by Phillips <i>et al.</i> (Phillips <i>et al.</i> , 2006).....	4
1.1.2 Transcriptomic subtypes by The Cancer Genome Atlas (TCGA)	6
1.1.3 Epigenomic molecular subtypes	9
1.1.4 Molecular subgrouping of a large patient cohort analysis	12
1.2 Implementation of the revised WHO classification scheme.....	17
1.3 Brain Tumor Resource in precision medicine.....	19
1.4 Computational approaches in glioma patient stratification	21
1.4.1 Implications of glioma-specific molecular databases	21
1.4.2 Molecular classification of GPCs.....	27
1.4.3 Machine learning methods in genomics.....	28
1.4.4 Predictive approaches associating genomics patterns with clinical phenotypes.....	36
1.4.5 Statistical model validation in clinical medicine	40
1.5 Altered signaling pathways in malignant gliomas	44
1.5.1 Biochemical pathways regulating hyper-proliferation in gliomas	45
1.5.2 Aberrant JAK/STAT regulation in glioma pathogenesis	47
1.6 Gap in knowledge	51
2 MATERIALS AND METHODS	54
2.1 Precision medicine evaluation through the Brain Tumor Resource	54
2.2 High-throughput genomics data quality check and analyses.....	55

2.2.1 National Neuroscience Institute (NNI) glioma molecular reference catalog	57
2.2.2 Microarray data processing	58
2.3 Archiving of public glioma databases	59
2.4 Transcriptomic pattern analysis	60
2.4.1 Functionally-tuned <i>STAT3</i> signature identification	60
2.4.2 <i>ST3GAL1</i> -specific glioma signature analysis	60
2.5 Predictive approaches in glioma patient databases	62
2.5.1 Connectivity map (CMAP) approach	62
2.5.2 Nearest template prediction (NTP) analysis	65
2.5.3 Prediction analysis of microarrays (PAM)	65
2.5.4 Single-sample gene set enrichment analysis (ssGSEA)	66
2.5.5 Cox proportional-hazards regression analysis	66
2.5.6 Goodness-of-fit evaluation analysis	67
2.5.7 Gene Set Enrichment Analysis (GSEA) and pathway evaluation	68
2.5.8 Integrative analysis linking functional pathways with diagnostic imaging data	69
2.6 Protein network analysis using a Bayesian probabilistic model...	69
2.7 Kinome screening analysis in GPCs (A/Prof Tan Nguan Soon from NTU executed PamChip profiling in GPCs)	70
2.8 Relative odds estimation	74
2.9 Experimental validation in glioma-propagating cells (GPCs) (performed by Dr Melanie Tan)	75
2.9.1 Immunoblot analysis in signature-stratified GPCs	75
2.9.2 Viability, sphere assay and dose response experiments in GPCs	76
2.9.3 <i>In vivo</i> evaluations in orthotopic xenograft NOD/SCID (NSG) mouse models	76
RESULTS	77
3 MOLECULAR SUBTYPE PATTERNS OF BRAIN TUMOR RESOURCE	77
3.1 Introduction and objectives	77
3.2 Molecular classification scheme of glioma-propagating cells (GPCs) in NNI and international molecular resources	79
3.2.1 Molecular subtype signatures of GPCs	80

3.2.2 Kinome profiling in GPC molecular subtypes.....	88
3.2.3 Brain molecular subtypes.....	94
3.2.4 Glioma-intrinsic (GI) subtypes in primary tumors	96
3.3 Summary.....	103
4 ST3GAL1-ASSOCIATED TRANSCRIPTOMIC PROGRAM IN GLIOBLASTOMA TUMOR GROWTH, INVASION, AND PROGNOSIS.....	104
4.1 Introduction and objectives.....	104
4.2 Differential binding of peanut agglutinin (PNA) mediated by ST3GAL1 identified GPC fraction with greater proliferation, invasiveness and stem-like properties	104
4.3 ST3GAL1 expression increases with tumor grade and portend poor prognostic in glioma patients.....	107
4.4 Identification of <i>TGFβ</i> as upstream signaling pathway activating ST3GAL1	108
4.5 ST3GAL1-associated gene signature correlates with tumor invasion and glioma prognosis in glioma patients	109
4.6 Identification of <i>FOXM1</i> as a downstream effector mediated by ST3GAL1 signaling	114
4.7 Summary.....	116
5 A GLIOMA-SPECIFIC, FUNCTIONALLY-TUNED <i>STAT3</i> TRANSCRIPTOMIC PROGRAM ADDRESSES THE MOLECULAR MECHANISM OF GLIOMA PATHOGENESIS AND CHEMORESISTANCE	118
5.1 Introduction and objectives.....	119
5.2 Predictive evaluation of the <i>STAT3</i> functionally-tuned signature in patient databases.....	121
5.2.1 Association of clinical phenotypes with the functionally-tuned STAT3 signature.....	124
5.2.2 <i>STAT3</i> signature model evaluation for goodness-of-fit	129
5.2.3 Integrative analysis of functionally-tuned <i>STAT3</i> signature classes with critical genomic aberrations identified in glioma biology	130
5.3 Biological validation of the <i>STAT3</i> functionally-tuned signature	132
5.3.1 <i>STAT3</i> signature stratification of GPCs	133
5.3.2 <i>STAT3</i> signature stratification in orthotopic animal models	136

5.4 Maintenance of <i>STAT3</i> signature status in recurrent GBM patients	137
5.5 Differential protein regulatory network from Bayesian inference in <i>STAT3</i> -stratified patient cohorts.....	138
5.6 <i>STAT3</i> chemoresistance pattern evaluation in glioma patient databases	142
5.6.1 High-throughput kinome assay evaluation in AZD1480-treated tumors.....	143
5.6.2 Altered kinase mechanism in AZD1480 treated GPCs	144
5.7 Summary.....	150
Summary of machine learning approaches applied in this study	152
6 GENERAL DISCUSSION	154
6.1 Discussion.....	154
Preamble	154
6.1.1 Primary tumors: Integrating analytical pipeline for molecular interrogation.....	155
6.1.2 Working with limited pool of patient-derived glioma cell lines	157
6.1.3 Cross-validation of glioma cell information in clinical datasets	160
6.1.4 Application of bioinformatic approaches in our studies	162
6.1.5 A novel kinome analysis pipeline: Probing biochemical interactions	164
6.1.6 Concluding Remarks.....	166
6.2 Future directions	167
6.3 Conclusion	169
7 TABLES	171
8 REFERENCES	194
9 AUTHOR'S PUBLICATIONS	216
10 APPENDICES	221

List of Figures

Figures	Description	Page number
Figure-1.1	Adapted graphical illustration of the molecular classification of gliomas enriched with genomic features (Eckel-Passow <i>et al.</i> , 2015; Kenta Masui, 2016; Louis <i>et al.</i> , 2016).	14
Figure-1.2	Adapted revised WHO classification scheme of gliomas (Louis <i>et al.</i> , 2016).	18
Figure-2.1	Kinome screening analysis pipeline flowchart	74
Figure-3.1	Graphical illustration of NNI Brain Tumor Resource.	79
Figure-3.2	Validation of NNI GPCs with international molecular databases of GPCs.	83
Figure-3.3	Heatmap of the Lottaz signature-based classification of NNI and International GPC transcriptomic databases from Beier, Gunther and Pollard (Beier <i>et al.</i> , 2007; Gunther <i>et al.</i> , 2008; Pollard <i>et al.</i> , 2009).	84
Figure-3.4	Box-whisker distribution of molecular subtype classifier gene expression.	87
Figure-3.5	Active kinase spectrum in mesenchymal cells	92
Figure-3.6	Glioma-intrinsic (GI) subtype classification of NNI primary tumors.	99
Figure-3.7	Mosaic plot representation of molecular subtype enrichment from TCGA, Phillips and glioma-intrinsic (GI) signatures.	100
Figure-3.8	Tumor purity score association with heterogeneity, comparison between NNI and TCGA specimens.	102
Figure-4.1	<i>ST3GAL1</i> -associated gene signature stratifies survival in glioma patients.	111
Figure-4.2	<i>ST3GAL1</i> -high status correlates with increased T2 volume and higher tumor grade in glioma patients; <i>ST3GAL1</i> knockdown prolongs the survival of mouse glioma model.	114
Figure-4.3	The concept map of bioinformatics approaches implemented in evaluating <i>ST3GAL1</i> -mediated glioma invasion.	116
Figure-5.1	Identification of the <i>STAT3</i> functionally-tuned gene signature.	122
Figure-5.2	Top 10 MetaCore pathway maps enriched with <i>STAT3</i> knockdown transcriptomic signature.	124
Figure-5.3	Gene set enrichment analysis (GSEA) on <i>STAT3</i> -stratified patient cohorts.	125
Figure-5.4	Association of <i>STAT3</i> -stratified patient cohorts with clinical phenotypes.	128
Figure-5.5	Enrichment of molecular aberrations and subtypes in <i>STAT3</i> classes in TCGA.	132

Figure-5.6	<i>STAT3</i> functionally-tuned signature stratifies NNI patient cells based on their <i>STAT3</i> status and demonstrates differential responses to known <i>STAT3</i> inhibitors.	135
Figure-5.7	The <i>STAT3</i> -high stratified patient xenograft mouse model demonstrates a favorable response to AZD1480 treatment.	137
Figure-5.8	Differential protein regulatory network in <i>STAT3</i> -stratified classes.	141
Figure-5.9	Identification of sensitive and resistant protein tyrosine kinases and serine/threonine kinases upon AZD1480 treatment in GPCs.	147
Figure-5.10	Integrative analysis of kinome profiles with transcriptomic data using functional module of kinome pipeline.	149
Figure-6.1	Graphical illustration of computational strategies adopted in the thesis.	170
Supplementary Figure-1	Empirical Bayes model adjustment on merged GPC transcriptome databases.	221

List of Tables

Tables	Description	Page number
Table 1	Summary of transcriptomic and epigenomic subtype signatures defining molecular pathways of glioma patients	15
Table 2	Summary of major molecular database resources of glioma patient cohort	26
Table 3	Shrunken class centroids of Lottaz gene signature applied to classify GPCs.	171
Table 4	Comparison of GPC-specific subtype signature addressing prognostic variability.	172
Table 5	Linear regression model identifies up-regulated kinases in mesenchymal GPCs.	173
Table 6	A 6-gene signature stratifies PNA-sorted GPC profiles.	174
Table 7	List of <i>ST3GAL1</i> -associated signature genes with Human Genome U133 PLUS 2.0 probesets used for CMAP analysis.	175
Table 8	Results of multivariate analysis on <i>STAT3</i> signature classes, WHO classes and Glioma-Intrinsic (GI) molecular subtypes.	179
Table 9	Contingency table of <i>ST3GAL1</i> -associated gene signature with Glioma-Intrinsic molecular classification.	181
Table 10	List of functionally-tuned <i>STAT3</i> signature (N=207) identified from <i>STAT3</i> knockdown GPCs and co-expressed genes from patient database.	182
Table 11	Contingency analysis on <i>STAT3</i> stratified classes with GI molecular subtypes.	187
Table 12	Results of univariate and multivariate analysis on <i>STAT3</i> signature classes, WHO classes, Glioma-Intrinsic (GI) molecular subtypes and age of patient covariates in Gravendeel and TCGA databases (Ceccarelli <i>et al.</i> , 2016; Gravendeel <i>et al.</i> , 2009).	188
Table 13	Molecular subtype switching in GBM tumors upon recurrence.	190
Table 14	List of altered protein tyrosine kinases in <i>STAT3</i> -high and -low GPCs upon AZD1480 treatment.	191
Table 15	List of altered serine/threonine kinases in <i>STAT3</i> -high and -low GPCs upon AZD1480 treatment.	192
ST 1	Top 5 stem cell-related gene sets significantly enriched with high <i>ST3GAL1</i> patient groups.	222

ST 2	Top 5 Reactome pathway clusters significantly enriched for the core markers from Dazard <i>et al.</i> , resistance pathway genes.	225
ST 3	List of significantly different genes in <i>ST3GAL1</i> knockdown clones of GPCs.	226
ST 4	List of REMBRANDT samples predicted using <i>ST3GAL1</i> -associated signature.	250
ST 5	List of Gravendeel samples predicted using <i>ST3GAL1</i> -associated signature.	260
ST 6	Enrichment of <i>ST3GAL1</i> -signature stratified patients with cellular invasion marked by T2-zone volume calculated from MRI scans.	269
ST 7	List of Gravendeel samples predicted using functionally-tuned <i>STAT3</i> signature.	272
ST 8	List of TCGA samples predicted using functionally-tuned <i>STAT3</i> signature.	283

Abbreviations

AI	Artificial intelligence
AIC	Akaike information criterion
AIF1	Allograft inflammatory factor 1
AKT1	Serine/Threonine Kinase 1
ANKRD43	Ankyrin repeat domain-containing protein 43
aNSCs	Adult neural stem cells
ATRX	Alpha thalassemia/mental retardation syndrome X- linked
AURK	Aurora kinases
BIC	Bayes information criterion
CD44	CD44 molecule (Indian blood group)
CDF	Cumulative distribution function
CDK	Cyclin-dependent kinase
CDKN2A	Cyclin dependent kinase inhibitor 2A
CDLs	Chromosomal 1p/19q co-deletions
CHI3L1	Chitinase-3-like protein 1
ChIP-Seq	Chromatin immunoprecipitation-sequencing
CL	Classical
ClaNC	Classification to nearest centroids
CMAP	Connectivity map
CNN	Convolutional neural networks
CNS	Central nervous system
Cor	Correlation
CSC	Cancer stem cells
CTNNA1	Catenin Alpha 1
DB	Databases
DBSCAN	Density based spatial clustering of applications with noise
DCX	Doublecortin X
DLDA	Diagonal linear discriminant analysis
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DOCK5	Dedicator of cytokinesis 5
DYRK	Dual-specificity tyrosine phosphorylation-regulated kinase
DYRK1B	Dual specificity tyrosine-phosphorylation-regulated kinase 1B
ECDF	Empirical cumulative distribution functions
EGA	European genome-phenome archive
EGFR	Epidermal growth factor receptor
EMT	Epithelial–mesenchymal transition
ERBB4	Erb-b2 receptor tyrosine kinase 4
ESTIMATE	Estimation of Stromal and Immune cells in Malignant Tumor
EZH2	Enhancer of Zeste Homolog 2
FAK	Focal adhesion kinase
FAS1	Fasciclin 1
FAS2	Fasciclin 2

FDA	Food and Drug Administration
FDR	False discovery rate
fNSCs	fetal neural stem cells
FOXM1	Forkhead box M1
FSCN1	Fascin actin-bundling protein 1
GABRA1	Gamma-aminobutyric Acid Type A Receptor Alpha1 subunit
GAD2	Glutamate Decarboxylase 2
GBM	Glioblastoma
G-CIMP	Glioma-CpG island methylator phenotype
G-DOC	The Georgetown Database of Cancer
GEO	Gene expression omnibus
GFAP	Glial fibrillary acidic protein
GI	Glioma intrinsic
GPC	Glioma-Propagating cells
GSEA	Gene set enrichment analysis
GSVA	Gene set variant analysis
H3	Histone 3
H3K23ac	H3K23 acetylation
H4	Histone 4
HFE	Hemochromatosis
HGG	High-grade glioma
HIF-1	Hypoxia-inducible factor 1
HME	Histone modification enzymes
HPRD	Human Protein Reference Database
HR	Hazard ratio
HRP	IgG horseradish peroxidase
IDH	Isocitrate dehydrogenase
IGF1-R	Insulin-like growth factor 1 receptor
IGFBP2	Insulin-like growth factor-binding protein 2
IHC	Immunohistochemistry
IL-6	Interleukin-6
iNOS	inducible nitric oxide synthase
IR	Ionizing radiation
ISN	International Society of Neuropathology
ITGAM	Integrin subunit alpha M
ITGB1	Integrin beta 1
ITK	Interleukin-2-inducible T-cell kinase
JAK	Janus kinase
JNKs	c-Jun N-terminal kinases
Karn	Karnofsky
KD	Knock-down
KL	Kullback-Leibler
KM	Kaplan-Meier
KS	Kolmogorov-Smirnov
LDA	Linear discriminant analysis

LGALS3	Galectin-3
LGG	Low-grade glioma
LKOCV	Leave-k-fold cross-validation
LOOCV	Leave-one-out cross-validation
LTK	Leukocyte tyrosine kinase
MADM	Mice models using mosaic analysis with double markers
MAL	Myelin and lymphocyte
MAP2	Microtubule associated protein 2
MAST	Microtubule-associated kinase
MDM2	MDM proto-oncogene
MEK	Mitogen-activated extracellular signal-regulated kinase
MERTK	CD44 and Mer tyrosine kinase
MES	Mesenchymal
MET	MET proto-oncogene, receptor tyrosine kinase
MGMT	Methylation of O6-methylguanine-DNA-methyltransferase
MIT	Massachusetts Institute of Technology database
MM	Mismatch
MMP	Matrix metalloproteinases
MR	Magnetic resonance
MRI	Magnetic Resonance Imaging
mRNA	Messenger-RNA
MSigDB	Molecular signature databases
mTOR	Mechanistic Target of Rapamycin
Mut	Mutant
mySQL	My structured query language
NCAM	Neural cell surface adhesion molecules
NCDL	Non-codeletion
NE	Neural
NEFL	Neurofilament Light
NF1	Neurofibromatosis type I
NF-kB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NG2	Neuron-glia antigen
NGS	Next-generation sequencing
NNI	National Neuroscience Institute
NOS	Not otherwise specified
NSCs	Neural stem cells
NSG	NOD scid gamma mice
NT	Non-targeting
NTP	Nearest template prediction
OLIG2	Oligodendrocyte lineage transcription factor 2
OOB	Out-of-bag
OPCs	Oligodendrocyte precursor cells
OPTICS	Ordering points to identify the clustering structure
OS	Overall survival
OSMR	Oncostatin M receptor

PAM	Prediction analysis of microarrays
PC	Principal components
PCA	Principal component analysis
PCNA	Proliferating cell nuclear antigen
PCV	Procarbazine, lomustine, vincristine
PDGF	Platelet-derived growth factor
PDx	Patient-derived xenograft
PECAM	Platelet endothelial cell adhesion molecule
PFS	Progression free survival
PI3K	Phosphoinositide 3 Kinase
PIAS3	Protein inhibitor of activated STAT3
PLK1	Polo-like kinase 1
PM	Perfect-match
PMT	Proneural-to-mesenchymal transition
PN	Proneural
PNA	Peanut agglutinin
PRMTs	Protein arginine methyltransferases
PTEN	Phosphatase and tensin homolog
PTK	Protein tyrosine kinase
PTPRD	Protein tyrosine phosphatase receptor delta
qPCR	Quantitative polymerase chain reaction
RB1	Retinoblastoma gene 1
RDBMS	Relational database management systems
REMBRANDT	Repository of Molecular Brain Neoplasia Data
RFE	Recursive feature elimination
rGBM	Recurrent GBM
RHOF	Ras homolog family member F
RIPA	Radio-immunoprecipitation assay
RNN	Recurrent neural networks
RO	Relative odds
ROS	Reactive-oxygen species
RPPA	Reverse phase protein array
RSK1	Ribosomal protein S6 kinase alpha-1
RTK	Receptor tyrosine kinases
RTK I and II	Receptor tyrosine kinase class I and II
SFKs	Src family tyrosine kinases
SH2	Src homology 2
SLC12A5	Solute Carrier Family 12 Member 5
SOMs	Self-organizing maps
SOX2	SRY (sex determining region Y)-box 2
Srm	SRC-related kinase
ssGSEA	Single sample gene set enrichment analysis
ST3GAL1	ST3 beta-galactoside alpha-2,3-sialyltransferase 1
STAT3	Signal transducer and activator of transcription 3
STK	Serine/Threonine kinase
SYT1	Synaptotagmin 1

TAMs	Tumor-associated macrophages
TCGA	The Cancer Genome Atlas
TERT	Telomerase reverse transcriptase
TFs	Transcription factors
TGFB	Transforming growth factor beta
TMZ	Temozolomide
TNF	Tumor necrosis factor
TOP2A	DNA topoisomerase II alpha
TP53	Tumor protein 53
TRIM24	Tripartite motif-containing 24
UCSF	University of California at San Francisco
ULK	Unc-51 like autophagy activating kinase
VEGF	Vascular endothelial growth factor
VEGFR	Vascular endothelial growth factor receptor
VSN	Variance stabilizing normalization
WHO	World Health Organization
wt	Wild-type

Abstract

Glioblastoma (GBM) tumors are highly aggressive malignant brain tumors and are resistant to conventional therapies. The Cancer Genome Atlas (TCGA) efforts distinguished histologically similar GBM tumors into unique molecular subtypes. The World Health Organization (WHO) has also since incorporated key molecular indicators such as IDH mutations and 1p/19q co-deletions in the clinical classification scheme. The National Neuroscience Institute (NNI) Brain Tumor Resource distinguishes itself as the exclusive collection of patient tumors with corresponding live cells capable of re-creating the full spectrum of the original patient tumor molecular heterogeneity. These cells are thus important to re-create “mouse-patient tumor replicas” that can be prospectively tested with novel compounds, yet have retrospective clinical history, transcriptomic data and tissue paraffin blocks for data mining.

My thesis aims to establish a computational framework for the molecular subtyping of brain tumors using machine learning approaches. The applicability of the empirical Bayes model has been demonstrated in the integration of various transcriptomic databases. We utilize predictive algorithms such as template-based, centroid-based, connectivity map (CMAP) and recursive feature elimination combined with random forest approaches to stratify primary tumors and GBM cells. These subtyping approaches serve as key factors for the development of predictive models and eventually, improving precision medicine strategies. We validate the robustness and clinical relevance of our Brain Tumor Resource by

evaluating two critical pathways for GBM maintenance. We identify a sialyltransferase enzyme (ST3Gal1) transcriptomic program contributing to tumorigenicity and tumor cell invasiveness. Further, we generate a *STAT3* functionally-tuned signature and demonstrate its pivotal role in patient prognosis and chemoresistance. We show that IGF1-R mediates resistance in non-responders to *STAT3* inhibitors. Taken together, our studies demonstrate the application of machine learning approaches in revealing molecular insights into brain tumors and subsequently, the translation of these integrative analyses into more effective targeted therapies in the clinics.

1 INTRODUCTION

Gliomas are primary brain tumors and account for 31% of central nervous system (CNS) tumors. The majority of these intracranial tumors are malignant in nature with poor prognosis despite the existence of improved surgical management and chemotherapeutic interventions. The conventional classification of gliomas primarily evaluates the histological and cellular features of brain tumors. In 2007, the World Health Organization (WHO) proposed three main classes of glioma, namely, pilocytic astrocytomas (prevalent in children), diffused gliomas represented by astrocytomas and oligodendrogliomas, and high-grade astrocytomas called glioblastoma (GBM) (Louis *et al.*, 2007). The histological strategies categorized adult gliomas into astrocytomas (82%), oligodendrogliomas (6%) and brain tumors with mixed-type cellular features (3%) (Dolecek *et al.*, 2012). However, the traditional classification of tumor grades based on histopathological features is woefully inadequate for capturing the spectrum of clinical behavior. Gliomas are markedly heterogeneous tumors and patients often display variable prognosis. Furthermore, histologically similar patients often respond differently to standard-of-care therapies, highlighting the limitation of relying solely on histology to diagnose and subsequently treat patients. Therefore, the new WHO Classification of Tumors of the Central Nervous System (2016) incorporates molecular markers, on top of traditional histologic criteria for brain tumor diagnoses (Louis *et al.*, 2016). With the incorporation of additional molecular information in this revised WHO classification, deeper insights in clinical, experimental and

epidemiological studies will potentially lead to improvements in the lives of patients with brain tumors.

1.1 Molecular classification of gliomas

1.1.1 Transcriptomic subtypes by Phillips *et al.* (Phillips *et al.*, 2006)

Transcriptomic evaluation of naïve, high-grade glioma patients defined three robust molecular subtypes that were termed proneural, proliferative and mesenchymal subtypes (Phillips *et al.*, 2006). A shrunken centroid-based algorithm was parsed into the transcriptome content of primary tumors using 35 gene signatures to delineate glioma-specific molecular subtypes (Tibshirani *et al.*, 2002). The Phillips molecular subtypes of high-grade gliomas established molecular marker enrichment with distinct transcriptomic phenotypes. Proneural subtypes are represented by elevated expression of genes associated with neuroblasts and mature neurons, namely, Oligodendrocyte transcription factor (*OLIG2*), Microtubule Associated Protein 2 (*MAP2*), Doublecortin X (*DCX*), Ectodermal-neural cortex 1, *ENC1* (NeuN), Erb-b2 receptor tyrosine kinase 4 (*ERBB4*), and Glutamate Decarboxylase 2 (*GAD2*). Interestingly, the transcriptomic features of normal brain cells and NSCs resembled the centroids of proneural subtypes. Furthermore, proneural patients were younger and had a better prognosis with standard therapy. In contrast, the proliferative and mesenchymal groups of patients were predominantly elderly (>50 years) and had a poorer prognosis with standard chemotherapy. Of note, the tissue specificity for mesenchymal molecular profiles matched with a different mixture of origins such as

smooth muscle, bone, fetal astrocytes and endothelial cells. The mRNA expression of angiogenesis markers, including platelet endothelial cell adhesion molecule (*PECAM*), vascular endothelial growth factor (*VEGF*), VEGF receptor 1 (*VEGFR1*) and VEGF receptor 2 (*VEGFR2*), was strongly enriched in the mesenchymal subgroup of patients. Similarly, the molecular features of proliferative patients had distinct centroids closer to those of hematopoietic stem cells and proliferative cell lines. The critical genes that were identified as being predictive of proliferative molecular subtypes include proliferating cell nuclear antigen (*PCNA*) and DNA topoisomerase II alpha (*TOP2A*).

The key message from the study was the establishment of molecular determinants using predictive algorithms for high-grade gliomas (HGGs) including grade III and grade IV tumors. Importantly, this was the pioneer study demonstrating that histological grade was superseded by molecular subtypes in addressing the prognostic variability. Treatment-induced molecular subtype switching was presented in the matched primary and recurrent tumors of same patients. Recurrent tumors displayed the tendency towards gaining mesenchymal features such as increased gene expression of *YKL40*, *CD44*, *STAT3* and *VIM*. The 35-gene molecular signature was initially derived from the gene expression features that correlated with overall survival, and subsequently interrogated with agglomerative clustering to associate with cellular pathways. There was a specific need to define the unsupervised gene expression pattern of GBM tumors in a large cohort using integrative

analysis accounting genomics events such as copy number and somatic events data to gain deeper insights.

1.1.2 Transcriptomic subtypes by The Cancer Genome Atlas (TCGA)

The Cancer Genome Atlas (TCGA) effort demonstrated that histologically similar GBM tumors could be molecularly distinguished into four transcriptomic subtypes, with each subtype displaying unique genomic aberrations and clinical outcomes (Verhaak *et al.*, 2010). Proneural (PN), classical (CL), mesenchymal (MES) and neural (NE) subtypes were the four distinct molecular subtypes identified using a consensus pattern-clustering approach. The “Classification to Nearest Centroids” (ClANC) method identified gene classifiers that could optimally delineate gene expression-based subtypes. Integrative analysis using copy number and somatic mutation data combined with molecular expression subtypes revealed deeper genomic insights into the four molecular subtypes. The integration of large amounts of genomic data revealed focal genomic variations associated with molecular subtypes in GBM patients. Interestingly, the classical group of patients had strong enrichment with concurrent genomic aberration of chromosomal 7 gain and chromosomal 10 loss (Bigner *et al.*, 1990). *EGFR* gains were detected in 97% patients of the classical subtype, while a subset of this *EGFR* cohort had a strong prevalence for the somatic mutation (*EGFR* variant III mutation). *TP53* mutation is a common genomic event in GBM patients; interestingly, classical patients were devoid of *TP53* aberrations. In addition, classical patients with *EGFR* gains were mutually exclusive

for the aberrant cell cycle/prosurvival pathway mediated by retinoblastoma 1 (*RB1*) and cyclin dependent kinase inhibitor 2A (*CDKN2A*). Genomic markers associated with developmental pathways, including Notch and Sonic hedgehog (SHH), were significantly activated in patients of the classical subtype. A recent study evaluating subtypes associated with the immune environment showed that the activation of the dendritic cell signature was enriched in the classical subtype patients (Wang *et al.*, 2018).

The mesenchymal group of patients exhibited a strong prevalence of large deletions in the 17q11.2 chromosomal region occupied by the neurofibromatosis type 1 (*NF1*) gene. In addition, the mesenchymal patients exhibited aberrant AKT pathway activation with the depletion of tumor suppressor events driven by *NF1* and *PTEN*. The key expression markers for classifying mesenchymal patient subgroups include chitinase-3-like protein 1 (*CHI3L1*), mesenchymal epithelial transition (*MET*), CD44 and Mer tyrosine kinase (*MERTK*). Notably, the prosurvival nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B) pathway and tumor necrosis pathway were strongly enriched in mesenchymal subtype patients. Interestingly, mesenchymal patients with defective *NF1* were enriched with M2 macrophage signature in the tumor microenvironment, highlighting that the loss of *NF1* enhances macrophage accumulation in mesenchymal tumors (Wang *et al.*, 2018).

Consistent with the Phillips molecular stratification, TCGA further substantiated that proneural patients had a typical enrichment of mutations in *TP53* and *IDH*, with concurrent active expression of

PDGFRA and substantial chromosomal gains in 4q12 (Phillips *et al.*, 2006). Taken together, these findings show that concomitant genomic aberrations such as *PDGFRA* amplifications, IDH and *TP53* mutations were significantly enriched in the proneural subtype. Deactivation of the tumor suppressor gene *CDKN1A* in proneural patients was linked with an activated OLIG2 mechanism. Neural subtype patients were enriched with neuronal markers such as Neurofilament Light (*NEFL*), Gamma-aminobutyric Acid Type A Receptor Alpha1 subunit (*GABRA1*), Synaptotagmin 1 (*SYT1*) and Solute Carrier Family 12 Member 5 (*SLC12A5*). Notably, the recent glioma-intrinsic (GI) classification concluded that the neural subgroup was a non-tumor-specific contamination and was not linked with tumor-specific subtypes (Wang *et al.*, 2018).

The key feature of TCGA molecular signature was to characterize the gene expression subtypes with genomic aberrations and hallmark pathways (The Cancer Genome Atlas Research *et al.*, 2008; Verhaak *et al.*, 2010). TCGA signature includes 840 gene signature, with each molecular subtype (proneural, classical, neural and mesenchymal) delineated by 210 gene features. The ClANC approach was able to distinguish the molecular subtypes of gene expression databases using the proposed molecular signature as described (Dabney, 2006). Despite the fact that the characterization of TCGA molecular subtypes was comprehensive in elucidating molecular pathways, the study did not account for the contribution from tumor microenvironment. The subsequent effort by TCGA resulted in the development of the glioma

intrinsic (GI) signature which factors in IDH aberrations, stromal contamination and transcriptome profiling technologies. The emergence of RNA-Seq technologies unravel deeper insights about full length gene abundance, isoform quantification and fusion transcripts. GI signature represents 50 gene features for each of those three molecular subtypes namely proneural, classical and mesenchymal subtypes (Wang *et al.*, 2018).

1.1.3 Epigenomic molecular subtypes

Epigenetic alterations are global chemical modifications modulating the gene expression level without altering the sequence of DNA molecules. The major epigenetic events encompass DNA methylation, histone modifications and non-coding RNAs' modulations. The consistent evidence for epigenetics inheritance is not yet clear, however both inheritable and non-inheritable epigenetic events act as master regulators of controlling altered phenotypes (Lind and Spagopoulou, 2018). The nuclear chromatin complex is composed of a DNA sequence wrapped with structural octamers known as histone proteins. Biochemical modification of the chromatin complex regulates critical processes of life, such as embryonic development, growth, aging and cancer initiation. Recent epigenomic studies revealed that the landscape of chromatin modifications could alter the somatic mutation profile of several types of cancer, including GBM (Kundaje *et al.*, 2015). The mammalian genome orchestrates the CpG repeats across the genome. These CpG islands are frequently methylated by DNA methyltransferase (DNMT). The methylated islands obstruct

transcriptional initiation machinery and consequently cause genetic repression (Solomon *et al.*, 2008). Studies on global DNA promoter methylation have shown that the transcriptional factor affinity was high in demethylated fragments of genomic regions (Danussi *et al.*, 2018). Unlike somatic mutation-based alterations, epigenetic events are reversible biochemical modifications that occur through inhibition of enzymatic activity; thus, cancers driven by epigenetic events are potentially targetable (Morera *et al.*, 2016).

An unsupervised K-means consensus clustering of promoter DNA methylation data from glioma patient tumors revealed a significant correlation between the glioma-CpG island methylator phenotype (G-CIMP) and the proneural subtype of GBM patients (Noushmehr *et al.*, 2010). The G-CIMP subtype is significantly enriched in patients with low grade glioma with IDH mutations and in younger individuals. Copy number gains of chromosome 7 are the typical characteristic feature of glioma tumors and were randomly distributed across both G-CIMP(+ve) and G-CIMP(-ve) tumors. The distinct copy aberrations associated with G-CIMP(+ve) tumors are amplifications in regions of chromosome 8q and chromosome 10p. The G-CIMP-specific hypermethylation marker panel includes Ankyrin repeat domain-containing protein 43 (*ANKRD43*), Hemochromatosis (*HFE*), Myelin and lymphocyte (*MAL*), Galectin-3 (*LGALS3*), Fasciclin 1 (*FAS1*), Fasciclin 2 (*FAS2*), Ras homolog family member F(*RHOF*) and the hypomethylation marker Dedicator of cytokinesis 5 (*DOCK5*).

The strength of this pioneer epigenomic study by TCGA is to establish the fact that genome-wide promoter methylation contributes to the molecular heterogeneity of GBM patients and, significantly associates with prognosis of GBM patients (Noushmehr *et al.*, 2010). Importantly, the G-CIMP subtypes were able to stratify the proneural GBM patients into two distinct clusters based on the DNA methylation profiles and, these subtypes were significantly correlated with the age of GBM patients.

Adult and pediatric GBM patients display considerable phenotypic heterogeneity. However, the biological and molecular subgrouping of tumors in these patient groups was not characterized until recently. The molecular heterogeneity of GBM tumors emphasizes the need to incorporate transcriptomic data with DNA methylation as well as, mutation status. Integrative analyses of methylation, expression and mutation data revealed the existence of six distinct biological subgroups (Sturm *et al.*, 2012). The pathway-specific mutation subgroups addressed key epigenomic events, including mutations in H3 Histone Family Member 3A (*H3F3A*), which control the global methylation patterns of GBM. The biological subgroups of GBM include IDH, H3F3A.K27, H3F3A.G34, receptor tyrosine kinase class I and II (RTK I and II). The G-CIMP subset was shown to have significant overlap with the IDH mutation group. Of note, *TP53* mutations were distributed at a high frequency across the IDH, K27 and G34 subgroups; however, the prognostic value of *TP53* mutations is highly variable.

IDH and *H3F3A* mutations displayed mutual exclusivity. Patients harboring *TP53* mutations and G34 mutations had better prognosis, whilst

the K27 subgroup had a worse prognosis (Sturm *et al.*, 2012). The epigenomic subgroups (K27 and G34) were present in the majority of pediatric and young GBM patients. RTK I subgroup patients are characterized by *PDGFRA* copy amplification, which overlaps with the proneural subtype. The RTK II group is a typical classical expression group of patients with focal amplification of chromosome 7, including *EGFR* and homozygous loss of the *CDKN2A* locus on chromosome 10. The mesenchymal subgroup of patients contained limited copy-level changes and mimicked methylation patterns of a normal brain.

This epigenomic study addressed the deeper insights of global methylome patterns of adults and pediatric GBM patients corroborating with focal copy aberrations. Importantly, the mutation subgroups correlated with the location of GBM tumors, and the worst prognostic K27 subgroup patients had tumors located at the midline. These efforts of transcriptomic and epigenomic subgrouping lead to identify the glioma patient stratification markers in multi-centre patient cohort analyses and link with prognostic impact of glioma patients for making clinical decisions.

1.1.4 Molecular subgrouping of a large patient cohort analysis

A recent key finding by Eckel-Passow *et al.*, combined critical genomic markers to further refine clinically relevant molecular subgroups enriched with prognostic indicators, acquired somatic mutations and accompanying germline variants (Eckel-Passow *et al.*, 2015). The study screened 1087 glioma patients and distinguished five distinct molecular subgroups based on three molecular aberrations: co-deletion of

chromosome arms in the 1p and 19q regions, metabolic marker IDH mutations inclusive of both *IDH1* and *IDH2* variants, and *hTERT* functional mutations affecting telomerase activity. The combinations of these three vital genomic aberrations were evaluated to decode the clinical phenotype and therapeutic response in 615 low grade gliomas and in 472 GBM patients. This analysis resulted in five molecular subgroups: 1) the triple-positive subgroup; 2) the telomerase reverse transcriptase (*TERT*) and IDH mutation subgroup; 3) the IDH mutation alone subgroup; 4) the triple-negative subgroup; and 5) the *TERT* mutation alone subgroup. Triple-positive and IDH mutation subtypes correlate with the proneural expression pattern. However, the *TP53* and Alpha thalassemia/mental retardation syndrome X-linked (*ATRX*) mutation spectrum uniquely associated with the IDH mutation alone subgroup. The *TERT* and IDH mutation subtypes are prevalent in all expression subgroups, except for the classical subtype. The second worst prognostic clinical subtype is the triple-negative subgroup, which has characteristic features of *EGFR* gain and loss of *PTEN* and *NF1*. The patients in the classical and mesenchymal groups exhibited the *TERT* mutation alone, which has been identified as the worst prognostic clinical subgroup in both low grade glioma and GBM patients. A fraction of the *TERT* mutation alone subgroup exhibited *EGFR* variant III (loss of exons 2 to 7), known as a key mutation for constitutive activation of EGFR. A graphical summary of the molecular classification of gliomas enriched with biomarkers is illustrated in **Figure-1.1**. The key features of molecular signature-based subtypes are summarized in the **Table 1**.

Figure-1.1. Adapted graphical illustration of the molecular classification of gliomas enriched with genomic features (Eckel-Passow *et al.*, 2015; Kenta Masui, 2016; Louis *et al.*, 2016). [§], markers employed in the assessment of diffuse gliomas; [#], genomic characteristics not required for WHO classification.

Table 1. Summary of transcriptomic and epigenomic subtype signatures defining molecular pathways of glioma patients

Molecular signatures	(Phillips <i>et al.</i>, 2006)	(Verhaak <i>et al.</i>, 2010)	(Noushmehr <i>et al.</i>, 2010)	(Sturm <i>et al.</i>, 2012)	(Wang <i>et al.</i>, 2018)
Technology	Gene expression (Affymetrix HGU133A and HGU133B)	Gene expression (Affymetrix HuEx array, Affymetrix U133A array, and Agilent 244K array)	DNA methylation (Illumina Infinium and GoldenGate)	DNA methylation (Illumina HumanMethylation450 BeadChip)	Gene expression (RNA-Sequencing, Affymetrix U133A array)
Patient cohort	Grade III and Grade IV	GBM patients	GBM patients	GBM patients	IDH wt GBM patients
N	Training cohort (n=115); Validation cohort (n=256)	Training cohort (n=200); Validation cohort (n=260)	Initial cohort (n=91); Validation cohort (n=181)	n=210	n=369
Uniqueness of the study	K-means agglomerative clustering was applied to identify three subtypes associated with survival; A novel proliferative subtype was identified; Molecular subtype switching in the recurrent tumor was observed ; Pathways are linked with phenotypes linked with cellular origin	Unsupervised consensus average linkage hierarchical clustering approach defined 4 molecular subtypes; Focal genomic aberrations such as <i>EGFR</i> , <i>NF1</i> and IDH was proposed to identify the molecular subtypes; Provided clinical evidence as EGFR inhibitor was efficacious in stratified patient cohort that paves the way for targeted therapy approaches	Unsupervised consensus clustering approach defined 3 epigenetic subtypes; Large chromosomal aberrations such as Chr8q and Chr10p were correlated with G-CIMP (+) subtypes; 8 DNA methylation markers determined G-CIMP (+) and G-CIMP (-) tumors	Unsupervised k-means consensus clustering delineated six DNA methylome subgroups associated with mutational status, copy-number aberrations and gene expression subtypes; Proposed IHC markers to stratify the methylome subtypes of GBM patients	Unsupervised consensus non-negative matrix factorization determined 3 transcriptomic subtypes; Precisely accounted tumor microenvironment and platform consistency; Comprehensive enrichment of subtypes with Immune micro environment and tumor purity; subtype alteration in recurrent GBM tumors
Gap in knowledge	Supervised evaluation determined the	The tumor microenvironment contribution affecting the molecular subtypes could	Large scale integrative analysis including copy number, somatic mutations	Require multi-centre study to validate the prognostic association	Longitudinal analyses of matched recurrent tumor profiling would shed

	molecular signature focused towards survival pathway, so an unsupervised learning of molecular data could enhance the understanding of molecular subtypes	provide deeper insights about gene expression subtypes	and gene expression evaluation would provide to link epigenomic pathway subtypes	of methylome subtypes in adult and pediatric GBM patients	additional insights on immunotherapy approaches
Data availability	GSE4271	https://portal.gdc.cancer.gov /	https://portal.gdc.cancer.gov /	GSE36278 and GSE36245	https://portal.gdc.cancer.gov

1.2 Implementation of the revised WHO classification scheme

The International Society of Neuropathology (ISN) has recently recognized the importance of molecular features contributing to the heterogeneity of glioma patients, and proposed a four-layered guideline to manage therapeutic strategies, with molecular information as the baseline followed by WHO grading and histological classification as diagnostic criteria (Kenta Masui, 2016). These guidelines were recently amended into the revised WHO classification for CNS tumors (Louis *et al.*, 2016). On top of histological information, the revised WHO classification now includes additional two genomic features in glioma diagnoses, *IDH1/IDH2* mutation and chromosomal *1p/19q* co-deletion, as illustrated in **Figure-1.2**. As such, the current implementation of the WHO classification scheme uses both histological and molecular parameters assessments to establish brain tumor diagnosis.

Revised WHO Classification Scheme

Figure-1.2 Adapted revised WHO classification scheme of gliomas (Louis *et al.*, 2016).

The conventional strategy remains the same as that used to categorize the tumors into astrocytoma, oligoastrocytoma, oligodendroglioma and glioblastoma based on histopathological features. Next, immunohistochemical (IHC) staining evaluates codon 132 in the *IDH1* genomic region and codon 172 in the *IDH2* genomic region, which identifies the primary classes as IDH mutant (mut) and IDH wild-type (wt) tumors. Patients lacking IDH mutations are principally termed IDH wt with the approved conventional classification scheme (Louis *et al.*, 2016). IDH mut tumors are further subjected to evaluation for chromosomal *1p/19q*

co-deletions. Co-occurrence of IDH mutation and chromosomal *1p/19q* co-deletions will be classified as oligodendroglioma. Further IHC evaluation for *ATRX* loss and *TP53* mutation of these IDH mut tumors revealed diffuse astrocytomas. Tumors lacking genetic determinants are classified as not otherwise specified (NOS) for the histological classification (Rogers *et al.*, 2018; Wesseling and Capper, 2018).

1.3 Brain Tumor Resource in precision medicine

Although considerable advances have been made in computational analyses of clinical databases, biological or functional validation of patient data-centric hypotheses remains lacking. This has been largely due to limitations in establishing cellular and animal models that can provide accurate preclinical information. A Brain Tumor Resource should have the capability to recapitulate the patient's original tumor molecular and pathophysiological profiles. Our lab at the National Neuroscience Institute (NNI) has established a tumor resource comprising patient-derived glioma-propagating cells (GPCs,) with matched orthotopic xenograft tumors in immunocompromised mice (NOD-SCID gamma) (Chong *et al.*, 2009). The NNI Brain Tumor Resource distinguishes itself as a collection of patient tumors with corresponding cells that is capable of recreating the full spectrum of the original molecular heterogeneity of patient tumors. Additionally, this resource contains deep sequencing-based genomic and transcriptomic data to facilitate molecular stratification studies. Notably, several studies have highlighted the limitations of relying on commercially procured,

serum-grown glioma cells, which often contain additional cytogenetic aberrations due to culture conditions and extensive *in vitro* passaging (De Witt Hamer *et al.*, 2008; Galli *et al.*, 2004; Joo *et al.*, 2013). Moreover, these cells form xenograft tumors with very different morphologies and gene expression profiles, confounding the ability to test patient stratification approaches.

Our earlier work demonstrated that patient-derived glioma cells grown under serum-free conditions form serially transplantable orthotopic xenografts that phenocopy the patient's original tumor morphology (Chong *et al.*, 2009). We further demonstrated that these cells contain transcriptomic and karyotypic hallmarks of the primary tumor (Tan *et al.*, 2016). We developed a cryopreservation technique called vitrification that preserves these cells at low passages, such that cell viability is optimal upon thawing from liquid nitrogen storage (-196°C). More importantly, we demonstrated that our glioma cells contain transcriptomic patterns and cytogenetic aberrations that are found in the original patient tumors. In turn, each cell line generates orthotopic xenograft tumors that phenocopy the original tumor morphology and transcriptomic profile. In addition, we deep-sequenced our xenograft tumors and compared the transcriptomic data with those of the original patient tumors (Foong *et al.*, 2011; Ng *et al.*, 2012). Our NNI Brain Tumor Resource thus provides an important tool for precision medicine: the ability to prospectively reform xenograft tumors for testing computational predictions (i.e., biological validation) yet have retrospective clinical history for correlative analyses. Collectively, the NNI Brain Tumor Resource merges multi-platform data from patient materials

and seeks to address clinical and imaging phenotypes with molecular data, advancing methods in stratified medicine for glioma therapy.

1.4 Computational approaches in glioma patient stratification

Integrative analysis across a diverse range of molecular information presents huge challenges in terms of data handling and the interpretation of multilevel data. With the advent of next-generation sequencing (NGS) technologies, a large volume of high-resolution genomic information has been generated. The analytical complexity exponentially increases with the volume of genomics data, the noise level of the platforms and the heterogeneity of studies (Azuaje, 2016). The key challenge to studying large genomics databases is implementing appropriate pattern-learning algorithms to generate unequivocal decisions to address molecular heterogeneity.

1.4.1 Implications of glioma-specific molecular databases

Computational predictions are largely dependent on the amount of high-quality data content that feeds into the algorithms. Glial tumors are among those extensively studied, and researchers have archived a massive amount of molecular data in public repositories. The molecular characterization of tumors involves the investigation of primary transcriptomic information to delineate pathway-specific subtypes of gliomas. The subtype patterns have been integrated with somatic aberrations and epigenetic events to find causal driver events. Phillips *et al.* first demonstrated that systematic molecular subtypes are associated

with prognosis in high-grade glioma patients (Phillips *et al.*, 2006). This molecular analysis included the gene expression data from the tumors of 76 high-grade glioma patients. Intriguingly, the subtypes were mapped to detect the molecular mechanisms that are specific to stem cell-associated pathways and lead to developmental regulation. The Phillips molecular database has the largest molecular resources of matched primary and recurrent tumors from high-grade glioma patients. The Repository of Molecular Brain Neoplasia Data (REMBRANDT) integrates a large volume of genomic data and substantial clinical endpoints from 874 glioma tumors (Madhavan *et al.*, 2009). In addition to high-throughput gene expression data, the REMBRANDT database provides rich information about copy number profiles from 834 glioma tumors. Importantly, REMBRANDT has archived diagnostic segment sequences from magnetic resonance (MR) scans of 130 patients. The REMBRANDT database is currently accessible through the Georgetown Database of Cancer (G-DOC) plus platform (<https://gdoc.georgetown.edu>).

The Phillips and REMBRANDT molecular databases lack genetic information about IDH mutations and chromosomal *1p/19q* co-deletions, which hampers the determination of the revised WHO classification scheme for the tumors in these databases. A European database from Gravendeel *et al.* has archived gene expression data from 276 glioma tumors with genetic marker profiling, such as IDH status, chromosomal *1p/19q* co-deletions and *EGFR* amplifications (Gravendeel *et al.*, 2009). The authors identified four gene expression-based molecular clusters using unsupervised pattern-learning approaches and showed

associations with patient prognosis. The Gravendeel molecular database stores critical clinical endpoints, including Karnofsky scores, concomitant medications and prior treatment regimens. The molecular information from the Gravendeel database serves as a major computational platform to retrospectively evaluate the revised WHO classification scheme.

The most comprehensive glioma molecular archive available to date is TCGA's all-glioma patient database (Ceccarelli *et al.*, 2016). TCGA has cataloged rich molecular and clinical information from 1,122 patient tumors of various WHO classification schemes. TCGA was the first study to substantially demonstrate that histology is an inadequate parameter with which to determine the clinical management of glial tumors (Verhaak *et al.*, 2010). Importantly, TCGA established the correlation of gene expression subtypes with genomic aberrations of focal gains and losses. TCGA's molecular evaluation was extended to investigate global DNA methylation of glioma cells and has cataloged 272 instances of glioma-specific DNA methylation (Noushmehr *et al.*, 2010). Of note, global methylation analysis identified a methylation-specific G-CIMP subtype as the key driving event for a good prognostic proneural subgroup. Epigenetic evaluation subsequently identified a correlation of mutation-specific events with epigenetic and expression-based subgroups (Sturm *et al.*, 2012). By integrating the clinical data with molecular aberrations, including germline variations, a large case-control study revealed five biomarker-specific subgroups using IDH and *TERT* mutations with chromosomal *1p/19q* deletion events (Eckel-Passow *et al.*, 2015). Collectively, TCGA is a comprehensive molecular compendium of

glioma resource encompassing DNA variations, focal aberrations, in-depth transcriptomic profiling, catalog of promoter methylation and protein expression of cell-cycle regulatory pathways (Ceccarelli *et al.*, 2016). In addition, TCGA contains transcriptomic profiling from recurrent patient tumors and paints an invaluable molecular portrait of tumor recurrence (Kim *et al.*, 2015).

Molecular evaluation of matched primary and recurrent tumors identified key mutagenic alterations specific to their allelic burden (Suzuki *et al.*, 2015). Allelic burden is the characteristic feature of alteration in the mutant allele frequency. The longitudinal study in glioma tumors identified the allelic burdens that correlate with tumor recurrence and drive clonal selection process. Recurrent tumor profiling is essential to determine the molecular pathway alterations that could support alternative therapeutic strategies to improve prognosis. Recent studies by Samsung Medical Center suggest that there was a strong precedence for molecular subtype switching between matched primary and recurrent tumors (Joo *et al.*, 2013). The Samsung recurrent tumor database provides a transcriptomic profile of 11 matched primary and recurrent tumors. Importantly, a study of recurrent tumor highlights the impact of an altered tumor microenvironment in patient-derived xenograft (PDX) models, and the molecular determinants dictate the aggressive phenotype of patient tumors. Phillips *et al.*, showed using a collection of matched primary and recurrent tumors (N=26), molecular pathway switching towards the mesenchymal subtype upon recurrence. The molecular model accounting for the longitudinal profiles of tumors revealed critical features responsible

for the invasive and recurrent phenotypes of tumors. The introduction of the Ivy Glioblastoma Atlas, which includes transcriptomic signals from cellular tumors, infiltrating tumors, leading edges, microvascular proliferation and pseudopalisading cells, serves as an important molecular resource for dissecting tumors from specific anatomic locations to address the intramolecular heterogeneity of GBM tumors (IvyGBMAtlas, 2018). The brief summary of major molecular databases available for brain tumor patients is catalogued in **Table 2**.

Table 2. Summary of major molecular database resources of glioma patient cohort

Databases	Genentech (Phillips <i>et al.</i> , 2006)	REMBRANDT (Madhavan <i>et al.</i> , 2009)	Gravendeel (Gravendeel <i>et al.</i> , 2009)	Freije (Freije <i>et al.</i> , 2004)	Murat (Murat <i>et al.</i> , 2008)	TCGA (Ceccarelli <i>et al.</i> , 2016)	Samsung (Joo <i>et al.</i> , 2013)	Ivy GBM Atlas (IvyGBMAtlas, 2018)
Number of samples	100	834	276	85	80	1122	58	270
Patient cohort	HGG and LGG	HGG and LGG	HGG and LGG	HGG	GBM	HGG and LGG	GBM	GBM
Molecular profiling	✓	✓	✓	✓	✓	✓	✓	✓
WHO molecular markers	-	-	✓	-	-	✓	-	-
Grade	✓	✓	✓	✓	✓	✓	✓	✓
Histology	✓	✓	✓	✓	✓	✓	✓	✓
Gene expression	✓	✓	✓	✓	✓	✓	✓	✓
Copy number	-	✓	-	-	-	✓	-	✓
DNA methylation	-	-	-	-	-	✓	-	-
Somatic Mutation	-	-	✓	-	-	✓	-	-
MRI data	-	✓	-	-	-	✓	-	-
Recurrent tumors	✓	-	-	-	✓	✓	✓	✓
Overall Survival	✓	✓	✓	✓	✓	✓	✓	
Molecular subtypes	✓	✓	✓	✓	✓	✓	✓	✓

HGG; High-grade Glioma, LGG; Low-grade Glioma, MRI; Magnetic Resonance Imaging

1.4.2 Molecular classification of GPCs

Beside dissecting the molecular subtypes of gliomas, an effective therapeutic strategy will also potentially take into account the targeting of tumorigenic cells driving proliferation and molecular heterogeneity. The origin of gliomas has not been conclusively defined; however, stem-like precursor cells from neural origin were identified in GBM tumors (Canoll and Goldman, 2008; Galli *et al.*, 2004). Animal models established from the stem-like precursor cells from GBM tumors demonstrated the recapitulation of pathophysiology of the original primary tumors (Bachoo *et al.*, 2002; Chong *et al.*, 2009; Verhaak *et al.*, 2010). Our lab has demonstrated the ability to establish GPCs from malignant brain tumors and have maintained these cells in spheroid forms in serum-free media supplemented with growth factors (Chong *et al.*, 2009; Galli *et al.*, 2004). Importantly, we showed that the application of molecular information from primary tumors, GPCs and orthotopic xenograft tumors are critical to evaluate the functional pathways related to gliomagenesis (Koh *et al.*, 2013; Ng *et al.*, 2012; Yeo *et al.*, 2012).

Molecular databases of GPCs have emerged as vital resources for understanding the transcriptional elements driving tumor propagation, differentiation traits and cellular invasion. At NNI, we demonstrated the application of GPCs to identify the functional mechanism attributable to stemness pathways, reactive oxygen species (ROS)-mediated chemoresistance, cell-cycle regulation, tumor suppression mechanisms and cellular invasion (Chong *et al.*, 2016; Foong *et al.*, 2011; Koh *et al.*,

2013; Ng *et al.*, 2012; Yeo *et al.*, 2012). Thus, understanding the molecular characteristics of GPCs is essential for functional validation and deeper insights into key oncogenic processes. As such, there is a growing number of publicly available archives, such as the Gene Expression Omnibus (GEO) and the European Genome-phenome Archive (EGA) (Barrett *et al.*, 2013), contributed by various groups internationally (ours included), cataloging the transcriptional databases of baseline and perturbed GPCs. Lottaz *et al.* established a 24-gene signature that can delineate transcriptional subtypes relevant to stemness pathways by merging three important GPC collections (Beier *et al.*, 2007; Gunther *et al.*, 2008; Lottaz *et al.*, 2010; Pollard *et al.*, 2009). Beier *et al.* used database-derived, genome-wide transcriptomic information from 6 GPC lines derived from primary astrocytic samples. Gunther *et al.* included gene expression data from 8 GPC lines of primary GBM tumors, while the Pollard database archived the mRNA expression data of 9 GPC lines from primary GBM tumors. In NNI, we have a large molecular collection of 32 GPC lines with corresponding patient tumors and xenografts (Chong *et al.*, 2009; Foong *et al.*, 2012; Tan *et al.*, 2016). The current thesis establishes the application of computational analyses on the transcriptomic information from the key databases presented in this section.

1.4.3 Machine learning methods in genomics

Machine learning is the term evolved from its predecessor technology called artificial intelligence (AI). In brief, AI algorithms

conceive a pattern about an entity and its relationship with the associated factors through human-like skills (Bini, 2018; Libbrecht and Noble, 2015). Researchers can provide numerous data relationships from databases until the model is trained in a particular context environment. The trained model has the capability to predict the connections for any *de novo* samples. Concurrently, advanced algorithms establish permuted outcomes in a sensible way through iterative simulations. These computational approaches have deduced successful outcomes in image recognition, natural language processing and data semantics (Grys *et al.*, 2017; Li and Liu, 2015). Genomics involves the search for a small fraction of regulatory elements among a massive number of 3 billion bases of DNA, enabling a vast complexity to be resolved by sophisticated statistical models (Bae *et al.*, 2015).

Machine learning models employ statistical algorithms to study the patterns of rich data structures and translate them into relationship networks connected to key determinant nodes. The clinical application of machine learning approaches has recently gained momentum in the field of genomics (Esteva *et al.*, 2017). The machine learning outcomes from deep algorithms allow physicians and surgeons to make better clinical decisions after connecting myriads of clinical history and diagnostic data from patients. In particular, genomics-guided medicine paves the way for devising next-generation-based treatment strategies by integrating molecular data with clinical diagnostics. Advanced sequencing-based technologies produce deep insights into genetic information with high precision and confidence. Researchers have begun to explore the

implications of big data analytics, which are routinely applied in dynamic data environments such as banking, image recognition, natural language processing and traffic control management. With the advent of next-generation sequencing technologies, genomics evaluation has become a data-intensive science. Thus, it is appropriate to leverage machine learning models in a big data environment to address complex clinical questions (Bini, 2018).

Machine learning approaches represent two functional modules: a) understanding the patterns from training datasets and b) evaluating predictive accuracy in validation cohorts. Some classical machine learning approaches such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have been well-adopted to identify the classifiers for facial and speech recognition patterns. Essentially, these classifying algorithms could be directly applied either in large-scale RNA-Sequencing data to predict the classifiers or in the identification of alterations and variations in whole genome-wide sequencing data. However, neural network algorithms have an inherent sensitivity towards the network structure design, properties of kernels and the number of determinant features. Thus, neural network construction with an inadequate understanding of biological context would lead to erroneous performance of feature determination (Zeng *et al.*, 2016). Hybrid network algorithms have been developed to overcome the issue of parameterization and increase the performance of model quality (Singh *et al.*, 2016).

Another important aspect of the machine learning approach is to learn how the determinants from the data and the acquired model should be adapted to conceive the newly learned information. Class discovery approaches were shown to identify decisive patterns from high-throughput data in patient tumors that were lacking prior contextual information (Golub *et al.*, 1999; Nutt *et al.*, 2003; Vitucci *et al.*, 2011). The approach incorporates a robust feature determination model based on distinct patterns present in the data and subsequently performs cross-validation of independent datasets. Pioneer works on glioma tumors delineated the molecular subtypes using shrunken centroid approaches (Phillips *et al.*, 2006; Tibshirani *et al.*, 2002). The method called the prediction analysis of microarrays (PAM) can estimate a standardized centroid for each subtype identified in a linear discrimination procedure. The centroid represents the mean value of every discriminating feature present in each class divided by the intra-class standard deviation. This centroid is allowed to move closer to the collective centroids for all classes identified in the data. Through this shrinkage process, the classification approach could reduce the noise level of classifiers and determine the features in a context-free situation. The ClaNC is a rank-based approach that was proposed to detect the classes from unknown molecular data on patient tumors (Dabney, 2005; Verhaak *et al.*, 2010). The ClaNC adapts standardized t-statistics to rank the features that are specific to each class. The rank-based class discovery approach more effectively reduces misclassification error rates than does the shrinkage-based evaluation.

Clustering is an unsupervised machine learning approach that is frequently applied to high-dimensional data to understand gross patterns of genome-wide molecular content. Clustering-based algorithms can learn functional data, such as gene expression or DNA methylation data, and to rapidly understand the patterns of molecular complexity associated with prognosis (Gravendeel *et al.*, 2009; Noushmehr *et al.*, 2010). The clustering approach tends to group the data based on similarity matrices identified from the molecular data (Eisen *et al.*, 1998). Two major approaches are utilized in clustering analysis: supervised and unsupervised modes. The supervised clustering approach adheres to reference class vectors, while unsupervised clustering recognizes patterns through specific linkage algorithms. High-dimensional molecular data are often studied using an unsupervised clustering approach to determine the patterns grouped into subtype-specific pathways (Gravendeel *et al.*, 2009; Noushmehr *et al.*, 2010; Verhaak *et al.*, 2010).

The underlying principle for the clustering method is to determine the similarity measure around critical features that are specific to each group. The two important considerations that should be taken into account are complete and partial clustering. Partial clustering classifies the major distinction across the data points. The assumption of a partial clustering model is that some data variables have inherent local noise and distort the connectivity in the data space. In contrast, complete clustering attempts to include all the data points into a defined group (Oyelade *et al.*, 2016). Partial clustering approach is the ideal methodology for high dimensional data analyses such as transcriptomic and epigenomic

profiles. The transcriptomic data space includes ~30,000 full-length gene variables, while the major proportion of those transcripts do not get altered across the samples or expressed in low abundance. Importantly, partial clustering method prevents the assignment of samples into spurious subtypes from the expression data of unrelated genes (Kerr *et al.*, 2008). Transcriptomic analyses apply filtering criteria such as median absolute deviation (MAD) or trim-down low expressed genes based on certain threshold value prior subjecting to learn the subtypes pattern (Verhaak *et al.*, 2010; Wang *et al.*, 2018). The predictive approaches applied in the current thesis interrogate the transcriptome data of tumors and GPCs in order to cluster the samples into published molecular subtypes using partial clustering approach (Cahoy *et al.*, 2008; Lottaz *et al.*, 2010; Wang *et al.*, 2018).

Based on the model construction in grouping the data, the clustering approach is categorized into 4 distinct models, namely, connectivity models, centroid-based models, distribution models, and density models. Connectivity models include a specific clustering approach connecting the similarly featured variables in a noise-free background space. The current thesis exclusively interrogates the connectivity model-based clustering approach in the molecular analysis of glioma patient databases. The most popular hierarchical clustering model is the classical example of a connectivity model-based clustering approach. Hierarchical clustering involves two approaches to establish connectivity among the samples: the agglomerative (bottom-up approach) and top-down approaches. The agglomerative clustering

approach treats each sample as a separate branch of the hierarchical tree and computes similarity matrices across samples. An average (most cases) linkage-based algorithm establishes iterative connections between branches until the dendrogram of aggregated nodes merges into a single tree (Eisen *et al.*, 1998; Sokal, 1958). The top-down approach is initiated from a single tree, and subsequent partition clusters are derived when the distance of the similarity score increases. The hierarchical clustering approach is insensitive for distance matrix implementation and generates reproducible subtypes. The limitation of the hierarchical clustering approach is its inability to scale a large volume of data, as the time complexity for grouping the data is quadratic (Köhn and Hubert, 2015).

The centroid-based model includes the k-means clustering approach, which supports the estimation of the molecular pattern of data, where *a priori* information is available regarding the number of groups. The k-means clustering approach creates random centroids of a specified 'k' number of clusters, and the distance between centroids and data points is iteratively computed. The iterative process is terminated when the centroid distance is no longer altered by the inclusion of neighboring data points (Wu *et al.*, 2003). The centroid-based learning approaches can rapidly identify molecular patterns within a large volume of data, as the time complexity factor follows a linear trend in estimation. The drawback of the k-means clustering approach is that the generation of random centroids tends to identify unstable cluster subtypes during the learning phase.

Enhancement of random initiation algorithms such as k-means clustering, self-organizing maps (SOMs) and model-based Bayesian clustering was proposed by Monti *et al.* (Stefano Monti, 2003). This consensus class discovery approach identifies robust subtypes through resampling-based strategies and estimates a shrinkage pattern for every fraction of features used. The consensus clustering approach enables the identification of the number of optimal subtypes coded in the molecular data in a sample cohort. This result is achievable by estimating the cluster stability during the bootstrapping procedure. In brief, consensus clustering evaluation evaluates the k number of subtypes, while the robustness of subtypes for every k is estimated through the resampling procedure of random feature selection. Thus, the approach provides a consensus across sequential evaluations through multiple resampling procedures. A shrinkage metric is also computed to estimate the empirical cumulative distribution, providing the visual assessment of cluster stability in high-throughput data structures (Wilkerson and Hayes, 2010; Wilkerson *et al.*, 2012). The consensus clustering model has been used to determine the molecular subtypes of GBM, breast, head and neck, and lung cancers (Ceccarelli *et al.*, 2016; Parmar *et al.*, 2015; Prat *et al.*, 2013; Sturm *et al.*, 2012; Verhaak *et al.*, 2010).

The other two clustering approaches include distribution-based clustering approach and density-based clustering that are less often applied in molecular data analyses. The distribution-based clustering approach includes an expectation-maximization algorithm that implies normally distributed cluster subtype patterns, which is quite unlikely in

molecular data evaluations (Xiaowei *et al.*, 1998). Since the data distribution of molecular subtypes does not follow the uniform Gaussian model in patient databases, the distribution-based clustering approach has not been applied in this thesis.

The density-based clustering models include density-based spatial clustering of applications with noise (DBSCAN) and ordering points to identify the clustering structure (OPTICS). Density-based clustering has been successfully used to identify molecular subtypes from patient databases, but the classification sensitivity is weak at discriminating smaller subclasses (Raczynski *et al.*, 2010). The density-based clustering approach was not used in this current thesis for molecular subtyping because of the sensitivity towards minor subtype stratification in patient databases.

1.4.4 Predictive approaches associating genomics patterns with clinical phenotypes

The major scope of machine learning models involves predicting the learned pattern in a new patient cohort through probabilistic inference. These computational approaches transform big data into knowledge that will directly deliver leads for biologists to design experimental strategies and support therapeutic management decisions (Adams, 2015). The findings from consortium efforts such as TCGA, the Mayo glioma case-control series, the UCSF adult glioma study and the Erasmus expression database have demonstrated the significance of pattern learning approaches for deciphering molecular heterogeneity (Gravendeel *et al.*,

2009; Jenkins *et al.*, 2011; 2008; Wrensch *et al.*, 2009). These databases comprise molecular information ranging from patient gene expression, somatic mutations, and copy aberrations to clinical history.

These data archives contain a large amount of information that can further be studied using predictive methodologies such as qualitative enrichment-based approaches, nearest template predictions, shrunken centroids and the classification of nearest centroids (Dabney, 2006; Hoshida, 2010; Lamb *et al.*, 2006; Tibshirani *et al.*, 2002). The predictive algorithms learn the patterns of gene set either in the form of signatures or as pathway-specific information. Subsequently, the gene set patterns will be retrospectively interrogated with patient databases to delineate subtypes associated with functional mechanisms. A typical machine learning model introduced by the Broad Institute, uses a connectivity map-based predictive inference, a form of qualitative enrichment methodologies combined with the Kolmogorov-Smirnov test to estimate the statistical significance of its prediction ability (Lamb *et al.*, 2006). This approach was initially proposed for evaluating transcriptomic signature enrichment with the chemosensitivity of cell lines against a panel of FDA-approved compounds.

Using these predictive learning approaches, our group demonstrated that the molecular content of tumor cells can dictate the disease progression in mouse models, mirroring the patient's original clinical profile (Chong *et al.*, 2016; Foong *et al.*, 2012; Foong *et al.*, 2011; Jin *et al.*, 2017; Koh *et al.*, 2013). Computational algorithms allow us to interrogate and validate molecular signatures derived from a limited but

unique pool of tumor cells from large databases containing molecular patient information. The connectivity map-based approach has been modified to enrich the pathway-specific signature with large expression databases of glioma patients through rank-based pattern matching (Chong *et al.*, 2016; Koh *et al.*, 2013; Yeo *et al.*, 2012).

The nearest shrunken centroid-based classification is a method that identifies classifiers by shrinking the gene features into centroid-based classes (Tibshirani *et al.*, 2002). First, the centroid patterns are learned, and the cross-validation error rate is estimated for the training dataset. The shrinkage threshold allows the algorithm to move the centroid around its predictive ability. When new molecular data of patient tumors are invoked in the trained model, the proximity around the threshold of centroid classes will be determined. Thus, the nearest centroid rule predicts the classification of any sample presented to the model. Nearest shrunken centroid-based approaches were effective in discriminating molecular subtypes in cancer genomics, including brain tumors (Butler *et al.*, 2018; Franks *et al.*, 2018; Lottaz *et al.*, 2010; Phillips *et al.*, 2006).

Gene set enrichment analysis (GSEA) is a classical approach that evaluates a signature of genes that are over- or underrepresented in the rank-based list of phenotype classes. The rank-based sum score is evaluated in the empirical permutation test categorizing two phenotype classes (Subramanian *et al.*, 2005). The GSEA was initially established to determine the pathway or ontology terms that are specifically enriched in two distinct phenotype classes. This methodology was subsequently

enhanced to determine the enrichment score of a signature of genes for a specific sample (Barbie *et al.*, 2009). The empirical cumulative distribution functions (ECDFs) of each sample are computed for rank-normalized genes present in the gene signature and the remaining genes. A sample-specific enrichment score is defined by integrating the variation between ECDFs. This single-sample GSEA (ssGSEA) approach is effective in predicting the class-based molecular subtypes for any given number of signatures (Cahoy *et al.*, 2008; Wang *et al.*, 2018).

By leveraging these predictive approaches, researchers can overcome the limitations of small sample sizes in their own studies and enhance the predictive confidence through validation of independent datasets. Beyond the statistical significance and narrow confidence intervals around the predictions, the critical component is functional validation in a model system with patient-derived materials. The most feasible and effective way is to invoke a complimentary two-way approach, by which the experimental findings of functional evaluations are retrospectively interrogated in large patient databases through computational analyses. This complementary approach allows the conclusive identification of functional evidence in a pipeline encompassing molecular resources of primary tumor and patient-derived tumor-forming cells, the xenograft potential of cells and the clinical data of patients (**Figure-3.1**).

1.4.5 Statistical model validation in clinical medicine

Model validation in statistical terms considers the cross-validation efficiency between prediction and real phenotype-based classes. However, the primary goal of statistical models in clinical medicine is to achieve improved survival and design effective treatment strategies. Substantial advancements in genomics technologies and big data analytics have resulted in high-quality statistical models and defined pattern-based subtypes. Apart from elucidating molecular pathways, these classifications from molecular data need to be corroborated with clinical endpoints such as overall survival (OS), progression-free survival (PFS) and clinical response rates. The *post hoc* evaluation of model validation proceeds by performing a univariate evaluation linking prognostic information of patients. To estimate the function of survival parameters, a nonparametric statistic called the product limit estimator is constructed. A visual representation of the product-limit estimator is displayed as a Kaplan-Meier (KM) curve (Kaplan and Meier, 1958). The distribution of KM survival curves for phenotypic classes was assessed for statistical significance using the log-rank test (Mantel, 1966). However, in clinical evaluations, the number of functional covariates, such as age, Karnofsky quality score and family history of illness, act as potential factors affecting patient prognosis (Bradburn *et al.*, 2003).

A semi-parametric Cox proportional hazards evaluation model estimates the linear trend of hazards regressed with predicted phenotypic classes and other risk factors affecting the hazard rate (Cox, 1972). The Cox proportional hazards model identifies the effect as hazard rate,

otherwise known as the probability of failure (death) occurrence, in a patient cohort. The model implements a multiple linear regression algorithm and computes the logarithm of hazard rates for the included variables against the reference probability. The estimation of Cox proportional coefficients is sensitive for linked covariates within the model, as the algorithm leads to the estimation of spurious probability distributions. The Cox survival model defines the hazard for a covariate as the direct probability for a failure event at a particular period centered around that point in time. As a result, the measure of association for survival is independently correlated with the covariate of interest in the survival model. The outcome of the Cox model computes the coefficients for each covariate defined in the model and the confidence interval of the estimator. The exponential term of the coefficient defines the hazard ratio (HR) addressing the measure of hazard rate over the reference value.

Clinical prediction models must be evaluated for goodness-of-fit, which determines how adequately the model captures the heterogeneity of a survival outcome function. The evaluation of model quality is achieved through maximum likelihood statistics. Akaike information criterion (AIC) and Bayesian information criterion (BIC) score estimations are the two classical measures applied to estimate the performance of model fitting in clinical data analytics (Akaike, 1974; Mohammed *et al.*, 2015; Stone, 1979). AIC is an information theory-based scoring system that estimates the deviation of the probability distribution of a proposed model from the true distribution. The key importance of AIC is to estimate the constant plus an asymptotically unbiased estimator of a function used

for ranking candidate models, which is a variant of the Kullback-Leibler (KL) divergence between the true model and the approximating candidate model (John J. Dziak, 2012; Seghouane and Amari, 2007). Kullback-Leibler divergence is the estimate of dissimilarity measure between two probability model distributions by measuring the entropy level differences. The AIC function mathematically represented as (Venables *et al.*, 2002):

$$AIC = -2 \ln(\hat{L}) + 2p_i$$

Where $\hat{L}$ is the maximum log-likelihood value of model computed from KL divergence; the traditional AIC uses the weight of the parameters as 2, and p_i is the dimension of parameter space used in the model.

The AIC estimation process applies several asymptotic identities, as the true model distribution is unknown in the available parameter space. The approximations might be too liberal for model estimated from smaller sample size, so the traditional AIC scoring approach tends to favor larger models (Akaike, 1974; Wit *et al.*, 2012). The model with lower AIC value corresponds with proximity to true model distribution among the ranked models with different covariates. The significance of AIC score could only addresses about the relative quality of ranking models, not the absolute quality of model. AIC is asymptotically identical to leave-one-out cross-validation (LOOCV) (Stone, 1977).

BIC scoring is defined as the estimate of the posterior probability function for a model being true under Bayesian setup (Busemeyer and Diederich, 2014; John J. Dziak, 2012). BIC scoring estimate is mathematically represented as (Venables *et al.*, 2002):

$$BIC = -2 \ln(\hat{L}) + \ln(n) \times p_i$$

Where $\hat{L}$ is the maximum log-likelihood value of model computed from KL divergence; n is the number of observations in the dataset, and p_i is the dimension of parameter space used in the model. As similar to AIC, the model with least BIC value addresses the model's proximity to true model distribution observed from the dataset. BIC scoring is asymptotically equivalent to leave-k-fold cross-validation (LKOCV) (Shao, 1997).

Both scoring approaches differ in the concept of penalization for selecting a model converging to the true distribution. The penalization strategy of BIC is directly reliant on the product value of sample size and parameters for having logarithmic value of (n) in the estimate function. Hence, even with a sample size of more than 7 (natural logarithmic value of 7 exceeds the value of 2), it tends to penalize the model with more parameters in BIC estimation. The BIC system heavily penalizes the model with the addition of noisy covariates in models with large parameters. Both AIC and BIC scores represent the quality of model fit in addressing the variability. A better-fitted model is represented by a lower score in the model comparison. In application to a real scenario, the AIC score recognizes the explained variability by the covariates, while BIC determines the prediction ability of the covariates (Shmueli, 2010). The AIC estimation suffers with overfitting of the data when the model is built with large parameters, AIC is appropriate in situations where the scope of the model is to arrive highly specific predictions with low false negative calls (John J. Dziak, 2012; McQuarrie and Tsai, 1998). On the other hand,

the BIC is an ideal methodology to arrive sensitive model predictions with low false positive calls (John J. Dziak, 2012).

1.5 Altered signaling pathways in malignant gliomas

TCGA established the genomic profiling of high-grade GBM patients to identify critical structural aberrations behind the molecular heterogeneity of tumors using DNA copy number, gene expression, and promoter methylation profiling technologies (Kim *et al.*, 2011). Integrative analysis of the diverse range of high-content data revealed that GBM patients may be further molecularly distinguishable. TCGA's comprehensive evaluation of genomic aberrations designated three major mutually exclusive signaling pathways: 1) oncogenic activation of *RTK* genes and proliferation through phosphoinositide 3-kinase (*PI3K*) pathway activation and loss of *PTEN*; 2) inhibition of *TP53* through activation of *MDM* proto-oncogene (*MDM2*); and 3) *RB1* inactivation via cyclin-dependent kinase 4 (*CDK4*) activation (Kim *et al.*, 2011). These signaling pathways unraveled a justifiable approach to develop pathway-specific inhibitors such as RTK inhibitors, CDK inhibitors, PI3K inhibitors, AKT inhibitors and mitogen-activated extracellular signal-regulated kinase (MEK) inhibitors. A combined analysis of somatic mutation and methylation profiles explained the pathway mechanism behind temozolomide (TMZ) sensitization. Methylation of the O⁶-methylguanine-DNA-methyltransferase (*MGMT*) gene promoter was determined to contribute to increased sensitivity towards alkylating agents (Hegi *et al.*, 2008; Hirose *et al.*, 2003). Patients with nonfunctional *MGMT* were

significantly mapped with a hyper-mutated phenotype, causing inefficiency in the repair of therapy-induced DNA damage in the tumor, leading to tumor cell death (The Cancer Genome Atlas Research *et al.*, 2008).

1.5.1 Biochemical pathways regulating hyper-proliferation in gliomas

Altered biochemical mechanisms delineate the functional pathways of normal and tumor cells. These biochemical signaling mechanisms provide an enhanced niche by which tumors can retain the epigenetic and transcriptomic alterations that facilitate the proliferation and invasive potential of glioma cells (Wolf *et al.*, 2010). Glycosylation is a critical biochemical alteration in which cell surface moieties are modified with the addition of sugar molecules (oligosaccharides). Almost 50% of the human proteome requires glycosylation-associated post-translational modifications (PTM) for normal cellular processes (Moskal *et al.*, 2009). Glycans are attached to oligosaccharides that mediate cellular functions such as intracellular communication, immunity regulation and oxidative stress (Fukuda, 1995; Haltiwanger and Lowe, 2004). Aberrant glycosylation can stimulate oncogenic signals responsible for cellular migration, metastasis, aggressive phenotypes and invasive potential in a range of tumors (Hakomori, 2002). The mechanism of cellular invasiveness is primarily determined by the interactions between altered moieties and cell surface glycans and subsequently alters the tumor environment (Rorive *et al.*, 2001). Specifically, several cell surface adhesion molecules, such as CD44, neural cell surface adhesion

molecules (NCAMs), and galectins, drive the altered tumor microenvironment through carbohydrate patterning (Jung *et al.*, 2008; Prag *et al.*, 2002; Wiranowska *et al.*, 2006).

Sialylation, a form of cell surface patterning, is the process of adding sialyl residues onto terminal carbohydrate moieties. Cancer cells express aberrant sialylation of glycoproteins mediated by dysregulated sialyltransferases that activate cellular migration, invasion and proliferation (Dall'Olio and Chiricolo, 2001). Several sialyltransferase enzymes have been implicated in regulating cellular machinery in gliomas, but the molecular regulation driving the self-renewal capability in glioma stem cells has not been explored (Moskal *et al.*, 2009). Two specific sialyltransferases, ST3 beta-galactoside alpha-2,3-sialyltransferase 3 (ST3Gal3) and ST6 beta-galactoside alpha-2,6-sialyltransferase 1 (ST6Gal1), are implicated in aggravating glioma progression with robust expression in high-grade tumors (Dawson *et al.*, 2004; Yamamoto *et al.*, 1997). Importantly, previous studies have highlighted that integrin signaling modules altered by sialic acids are associated with tumor progression and cellular malignancy.

A fraction of GBM patients exhibit hyperactivity of the unique sialyltransferase enzyme ST3 beta-galactoside alpha-2,3-sialyltransferase 1 (ST3Gal1), which is responsible for α -2,3-sialylation-mediated cell-cycle processes (Kim *et al.*, 2011). Hyperactivation of ST3Gal1 regulates a range of solid tumor malignancies, such as colon and bladder cancer (Schneider *et al.*, 2001; Videira *et al.*, 2009). ST3Gal1 is localized in the membranes of Golgi bodies and frequently transfers

sialic acid residues to dual-saccharide molecules (Kono *et al.*, 1997). Our group recently established the role of *ST3GAL1* as a significant contributing factor to the activation of tumor cell renewal and invasive potential through a patient-centric approach and experimental validation in GPCs (Chong *et al.*, 2016). The GPC-specific *ST3GAL1* transcriptomic signature was proposed to delineate the *ST3GAL1* pathway, and the signature was an independent predictor of overall survival in glioma patient databases. Moving forward, we utilized patient-centric databases to identify key structural or copy number aberrations defining the ST3Gal1 invasive phenotype in TCGA database (Ceccarelli *et al.*, 2016).

1.5.2 Aberrant JAK/STAT regulation in glioma pathogenesis

Computational mining of high-content genomics data winnowed down the key regulatory pathways that were responsible for chemoresistance, self-renewal and proliferation potential of glioma tumors. The major glioma signaling pathways were identified through the patterns of aberrant transcriptional regulation, however the mechanistic insights of controlling the transcription mechanism are still unexplored. Signal transducer and activator of transcription 3 (*STAT3*), an important transcription factor, is profoundly active in several malignancies, including melanoma, breast, prostate, thyroid and GBM tumors (Bromberg and Darnell, 2000; Rahaman *et al.*, 2002; Yu and Jove, 2004; Yu *et al.*, 2009). *STAT3* hyperactivity is attributed to the phosphorylation mechanism by growth receptors and active cytokine signaling molecules at the Tyr-705 residue (Levy and Darnell, 2002). Phosphorylated *STAT3* (p*STAT3*)

undergoes nuclear migration to trigger *STAT3* downstream target genes or bind with DNA response elements (Bromberg and Darnell, 2000). The interactions between the phosphorylated residue and the Src homology 2 (SH2) domain facilitates the migration of active *STAT3* into the nucleus (Carballo *et al.*, 1999). Aberrant *STAT3* regulation was shown to modulate GBM stem cell function and neural development. Importantly, transient inhibition of *STAT3* significantly affected gliomasphere formation and self-renewal capabilities (Sherry *et al.*, 2009).

Aberrant pSTAT3 expression has been correlated with high grades of glial tumors in recent studies (Brantley *et al.*, 2008; Mizoguchi *et al.*, 2006). However, a separate study using IHC to evaluate pSTAT3 expression in diffuse glioma showed the lack of correlation between pSTAT3 expression and high-grade tumors (Wang *et al.*, 2004). Thus, detecting the activity of *STAT3* using pSTAT3 staining in glioma was inconclusive, implying the need to identify robust markers for *STAT3* regulation (Luwor *et al.*, 2013). The lack of somatic changes or mutations in the *STAT3* genomic region indicates that the active pathways of upstream modulating cascades drive the aberrant transcription of *STAT3* in GBM tumors. The phosphatase protein tyrosine phosphatase receptor delta (*PTPRD*) has an inverse correlation with *STAT3*-associated progression in GBM cell lines (Veeriah *et al.*, 2009). Intriguingly, epigenomic evaluation demonstrated that hypermethylation of the *PTPRD* promoter region was associated with good prognosis in GBM patients. In addition, large, focal copy-level changes in the *PTPRD* gene were frequently detected in GBM tumors (Solomon *et al.*, 2008).

However, direct evidence of *PTPRD* regulating STAT3 activation in patient tumors is still lacking. Shuai *et al.* established that the downstream mechanism of STAT3 was dependent on the activity of a protein inhibitor of activated STAT3 (PIAS3). The *PIAS3* gene is a critical repressor of STAT3 activity by inhibiting the binding of STAT3 to the DNA elements of target genes (Shuai, 2000).

Structural aberrations of the *EGFR* (*EGFRvIII* mutation) induce the persistent activation of STAT3 in GBM tumors. GBM tumors with *EGFRvIII* mutation are tumors with active regulation of the inducible nitric oxide synthase (iNOS) pathway, which is specifically regulated by *STAT3* activation (Puram *et al.*, 2012). A feed-forward signaling pathway linking *EGFRvIII* and the oncostatin M receptor (OSMR) with STAT3 was detected in GBM tumors. The physical binding of OSMR with *EGFRvIII* is critical to driving STAT3-mediated tumorigenesis (Jahani-Asl *et al.*, 2016). Epigenetic regulation through H3K23 acetylation (H3K23ac) and the tripartite motif-containing 24 (TRIM24) mechanism are crucial for STAT3 activity in *EGFRvIII* tumors. The activation of *STAT3* requires the coactivation of H3K23ac and TRIM24 in tumors with active EGFR signaling (Lv *et al.*, 2017).

A strong, positive correlation was observed between a cytokine-mediated pathway regulator, interleukin-6 (IL-6), and the STAT3 signaling pathway in GBM patient samples (Chang *et al.*, 2005; Rolhion *et al.*, 2001). However, the mechanistic association of STAT3 and IL-6 has not been well established in GBM cells. The migratory factors responsible for invasive phenotypes, such as matrix metalloproteinases (MMP), actin-

binding protein (FSCN1) and focal adhesion kinase (FAK), require *STAT3*-mediated transcriptional activation (Li *et al.*, 2010; Lindemann *et al.*, 2011; Senft *et al.*, 2011). In addition, *STAT3* responds to the hypoxic tumor environment by activating invasion factors, including hypoxia-inducible factor 1 (HIF-1) and VEGF, in GBM cells (Keunen *et al.*, 2011). *STAT3* is a downstream transcriptional regulator of the tumor vascularization pathway mediated through VEGFR in GBM tumors (Kang *et al.*, 2010). GBM tumors treated with anti-VEGF therapy develop a *STAT3* resistance pathway, indicating that *STAT3* regulation is critically associated with the tumor microenvironment (de Groot *et al.*, 2012).

Enhancer of zeste homolog 2 (EZH2) serves as the critical epigenetic pathway regulator of glioma patient tumors by methylating the K27 residue of the histone 3 protein (H3K27me). The mechanism of H3K27me is associated with transcriptional repression pathways governing major oncogenic cascades. Interestingly, Kim *et al.* defined a novel role of EZH2 in modulating the hyperphosphorylation of *STAT3*, leading to tumor progression (Kim *et al.*, 2013). This epigenetic regulation paves the way for reversing the *STAT3*-mediated GBM progression that defines a novel epigenetic therapeutic strategy for tumors with high *STAT3* expression levels (*STAT3*-high tumors).

NF- κ B comprises a family of transcription factors that act as master regulators of the inflammatory response. GBM tumors with aberrant PI3K pathway activation and loss of PTEN receive constitutive activation signals from Akt. Consequently, Akt induces the NF- κ B signaling pathway and phosphorylates *STAT3* (Wang *et al.*, 2004). This mechanism had a

strong, positive correlation with insulin-like growth factor-binding protein 2 (IGFBP2)-mediated glioma progression (Holmes *et al.*, 2012). NF- κ B-mediated STAT3 activation is largely dependent on the methylation activity of EZH2 (Kim *et al.*, 2013). Collectively, the inflammatory response pathway (NF- κ B) receives active signaling from the PI3K/Akt pathway in tumors with PTEN loss.

The invasive characteristic of GBMs is in part due to their high migratory potential to invade the surrounding tissue. Proneural-to-Mesenchymal transition (PMT) has been pointed as one of the mechanisms that confer to GBM cells this invasive property. Non-mesenchymal subtypes of tumors typically acquire mesenchymal features at recurrence. Recent studies have implicated STAT3 as a final modulator for the cells switching from proneural to mesenchymal subtype (Segerman *et al.*, 2016). Taken together, these findings show that STAT3 is the ultimate molecular switch for GBM progression affecting prosurvival, angiogenesis and invasive potential. It is important to characterize the molecular heterogeneity of the STAT3 pathway by integrating a patient database with cellular models. Machine learning models would substantially contribute to estimating the predictive strength of STAT3 signals and could potentially contribute to designing inhibitory compounds.

1.6 Gap in knowledge

With the advent of TCGA, our understanding of the complexity of the genetic and molecular landscape of GBM has advanced remarkably.

The extensive molecular platforms established by various groups have generated abundant and voluminous transcriptomic, epigenetic, phosphoproteomic profiles that can be ascribed to GBM tumors to guide future therapeutic decisions. However, the clinical relevance of such molecular characterization is still lacking and the ability to accurately predict clinical outcomes from these multiplexed molecular data remains, till date, the major obstacle to overcome in fully materializing the concept of precision medicine. In this study, I will aim to address the following:

1. I utilized machine learning approaches to first establish gene classifiers using glioma cells of patient origin. Such a method suffers from a statistically underpowered training set, as with all such studies (Levenson *et al.*, 2002). Thus, I determined overlapping gene candidates by interrogating such gene classifiers in primary tumors available through TCGA Rembrandt and other datasets (Madhavan *et al.*, 2009; Verhaak *et al.*, 2010). This allows statistically robust validation. The NNI tumor resource is important for preclinical validation of the predicted hypothesis. As example, I studied two such projects:

2. ST3Gal1 was determined to be the key gene product responsible for glioma cell invasiveness. This study began with a lectin screen of primary glioma cells, or GPCs, as cell surface sialylation is a key determinant of tumor cell infiltration (Dall'Olio and Chiricolo, 2001). Peanut agglutinin stained normal and tumor cells differentially, and when we evaluated the transcriptomic profiles differentiating Peanut-low and – high stained cells, *ST3GAL1* emerged as the top-ranked candidate.

Subsequent analyses proved correlation between *ST3GAL1*-high with MRI scans showing greater T2 zone volume, usually a surrogate marker of tumor cell invasiveness. The *ST3GAL1*-high cohort also associated with the mesenchymal subtype typically representing tumors of poorer prognosis.

3. In the next study, our hypothesis centered on STAT3 activation which precludes the transition of PMT, typically representing recurrent GBM patient tumors. This is important as it forms the foundation of implementing the N-of-1 precision medicine trial where serial monitoring of the patient's tumor allows physicians to assign specific anti-STAT3 therapy at an appropriate stage. We utilized a similar method where a unique *STAT3* gene classifier was co-established by overlapping STAT3-perturbed GPCs with primary tumor clinical datasets. To verify our computational hypothesis, we applied this stratification to our Brain Tumor Resource containing GPCs, orthotopic xenograft tumors and primary tumors. We then selected a mix of *STAT3*-high and -low GPCs to functionally validate our predictions, and further substantiated with matched PDx tumors. We also determined significant correlations to tumor grade, Karnofsky score, and various molecular parameters spelt out in the 2016 WHO classification scheme.

In summary, we believe that this holistic approach that taps into public databases for statistical power, and then verifying in our resource of GPCs, PDx mouse models and primary tumors offers crucial evidence prior to implementation of clinical trials.

2 MATERIALS AND METHODS

2.1 Precision medicine evaluation through the Brain

Tumor Resource

Our external molecular data archive includes data from gene expression microarrays, RNA-Seq summarized values, isoform expression, ChIP-Seq BED Graphs, and key somatic mutation and protein arrays. The raw intensity files from microarrays were downloaded from GEO and processed using R/bioconductor packages (Version 3.9) with standard MAS5 scaling (Barrett *et al.*, 2013; Gautier *et al.*, 2004). The metadata were extracted from the relevant literature and attached to a normalized expression matrix with detection quality metrics. Non-Affymetrix microarray profiles were directly downloaded using R/GEOquery packages (Version 2.50.5) (Davis and Meltzer, 2007). The level 2 and level 3 TCGA data were collated using R/TCGA-assembler packages (Version 2.0.6) (Zhu *et al.*, 2014) and processed using our customized R scripts. The combined low grade glioma (LGG) and GBM molecular databases were created for TCGA RNA-Seq expression databases. The known technical batch effects were controlled through empirical Bayes methods using R/sva package (Version 3.30.1) (Johnson *et al.*, 2007). Principle component analysis (PCA) and unsupervised clustering approaches were applied to the most highly variable genes to ensure that the value distributions were comparable across different databases. ChIP-Seq BED Graphs were downloaded from the Cistrome dataset browser (Liu *et al.*, 2011).

We archived a separate molecular resource for a glioma stem cell collection from international molecular data resources and processed it in a similar fashion as for the tumor molecular profiles of cancer stem cell (CSC) lines derived from adult GBM patients (GSE7181), CSC lines derived from GBM patients (GSE8049) and CSC lines from newly diagnosed GBM patients (GSE15209) (Beier *et al.*, 2007; Gunther *et al.*, 2008; Lottaz *et al.*, 2010; Pollard *et al.*, 2009). We also generated median-centered molecular profiles for the patient tumor resource, which were often readily parsed through our connectivity map (CMAP) pipeline, to detect the molecular stratification. The expression database of recurrent GBM patients from Samsung Medical Center was profiled on Affymetrix Human Exon 1.0 ST Array, and the summarized data were download from GEO (GSE42669) and processed using R/affy packages (Version 1.60.0) (Gautier *et al.*, 2004; Joo *et al.*, 2013). In addition, we gathered the matched recurrent GBM patients' gene expression database from Genentech recurrent tumor collections (GSE4271) and used Affymetrix HGU133A and HGU133B chips for gene profiling (Phillips *et al.*, 2006).

2.2 High-throughput genomics data quality check and analyses

We evaluated the transcriptome measurements that were adequately expressed in our experimental conditions. The detection p -values of the one-color microarray were considered to select the probes with robust expression. We defined a cut-off of 80% for samples that had detectable expression (p -value<0.05) for a specific transcript. The

processed transcripts were normalized using either MAS5 scaling or variance stabilizing normalization (VSN) as described previously (Hubbell *et al.*, 2002; Huber *et al.*, 2002). The chip design of Affymetrix single color microarray quantifies the gene expression from the signals of 11 perfect-match (PM) and mismatch (MM) probes targeting the transcript. This design renders the user to estimate the statistical confidence of gene expression measurements. MAS5 scaling leverages on the PM-MM probes to estimate the background noise and proportionally scale the expression values (Hubbell *et al.*, 2002). MAS5 scaling has been known to be the ideal scaling approach for Affymetrix chip-based microarray profiles (Pepper *et al.*, 2007). MAS5 scaling was recognized as highly selective and sensitive in detecting differentially expressed genes (Pepper *et al.*, 2007). On the other hand, Illumina bead-based microarray technology estimates the gene expression using multiple technical replicates for each probe. The VSN strategy takes the advantage of randomly distributed technical replicates in each array and transform the probe readouts into gene expression values. Studies report that VSN scaling is the recommended approach for normalizing the gene expression data and controlling false-positives of differentially regulated gene expression data of Illumina bead arrays (Du *et al.*, 2008; Lin *et al.*, 2008). The distributions of normalized data were plotted as box-whisker plots to assess the distributions of sample expression and were comparable for differential analysis. In order to understand the heterogeneity in our GBM patients, we performed an unsupervised hierarchical clustering evaluation of the top 10,000 highly variable

transcripts. The 'dist' function from R/stats package (Version 3.7.0) was applied to calculate the Euclidean distance matrix of patient's gene expression data and, the complete agglomeration method of 'hclust' function deployed to establish the relationship of dendrogram construction (Team, 2019). The unsupervised analysis of transcripts allowed us to estimate the novel subtypes in our database cohorts. The PCA evaluation determined the presence of any technical batch effect in our sample cohorts.

2.2.1 National Neuroscience Institute (NNI) glioma molecular reference catalog

Our NNI molecular resource accommodates three layers of information from GPCs, xenograft tumors and matched tumors. Since our lab aims to address the functional mechanism of gliomas, we principally rely on transcriptomic profiles from the NNI tumor resource. The GBM primary tumors from the National Neuroscience Institute (NNI) were obtained with informed consent and de-identified in accordance with the SingHealth Centralised Institutional Review Board A. Our gene expression database was predominantly profiled using Affymetrix GeneChip Microarray. Raw CEL files were processed using a MAS5 scaling algorithm, followed by standard microarray quality checking of profiles on probeset hybridization metrics as recommended by the limma and affy packages (Version 1.60.6) (Gautier *et al.*, 2004; Ritchie *et al.*, 2015). Then, the detection *p*-values and exonic locations of probesets were considered to create high-quality profiles of gene expression resources. This reference molecular resource serves as an important tool

for signature interrogation to identify the enrichment of molecular pathways.

2.2.2 Microarray data processing

We performed linear regression model fitting on simultaneous analyses of the complete expression dataset to detect differentially regulated genes across experimental conditions using the R/limma package (Version 3.38.3) (Ritchie *et al.*, 2015). The method has several advantages, such as sample correlation for replicates that can be incorporated into the model and account for gene dependencies for co-regulated modules. Identification of differential genes was considered the baseline gene signature in our computational analyses. The gene signature was appropriately corrected for multiple correction, and sample correlations were defined in the model for the technical replicates. The statistical significance of genes was estimated based on a false discovery rate (FDR)-adjusted p -value of less than 0.05. We further scored the list of genes depending on the biological question of interest. In the *STAT3* analysis, we wanted to deduce a functionally-tuned set of genes upon knockdown of *STAT3*. The *STAT3* knockdown signature was identified by fitting two statistical parameters (an adjusted p -value < 0.05 and a minimum of 2 log₂ fold change). We further interrogated the *STAT3* knockdown signature for its pattern in the patient database. We evaluated the pairwise Spearman rank correlation of *STAT3* across the patients' transcriptome and yielded a total of 9203 co-expressed probesets ($\rho \pm 0.3 <$) in the REMBRANDT patient database. The *STAT3*

functionally-tuned signature is illustrated in **Figure 4.1**. The dual evaluation of patient profiles and the GPC resource provided a functionally-tuned signature enriched with clinical relevance.

In the *ST3GAL1* analysis, we were interested in the sialylation mechanism behind carbohydrate patterning contributed by *ST3GAL1* knockdown and PNA affinity as integrated factors contributing to the cellular fractions and self-renewal potential of GPCs. We recognized the union of PNA and the *ST3GAL1* knockdown signature as an *ST3GAL1*-associated transcriptomic program. We used the R/lumi package (Version 2.36.0) to process the PNA-associated microarray experiments, which were profiled on Illumina HumanRef-8 v2 microarray chips. We deployed R/lumi packages to process the Illumina microarrays (Du *et al.*, 2008).

2.3 Archiving of public glioma databases

Our data resource of glioma samples contains functional information on patient tumors including gene expression, DNA methylation, copy number, somatic mutation and protein array expression. We implemented genomics data analysis protocols recommended by Biobase functional modules (Huber *et al.*, 2015). The summarized experiments include four layers of information: clinical information, summarized experimental data, annotation information, and assay information. The patient information data are archived in a relational database management systems (RDBMS)-based archive in a MySQL data structure. We adopted R/Bioconductor packages (Version

3.9) for the archival of functional data containers of in-house and public patient tumor databases.

2.4 Transcriptomic pattern analysis

2.4.1 Functionally-tuned *STAT3* signature identification

Brain tumor gene expression drives disease progression and patient survival outcome (Gravendeel *et al.*, 2009; Phillips *et al.*, 2006; Verhaak *et al.*, 2010). As any signaling pathway is better represented by a set of genes than by a single candidate, we first established a gene signature representing the *STAT3* pathway activation status. We identified *STAT3* co-expressed genes in the REMBRANDT patient glioma database (Madhavan *et al.*, 2009) and subsequently narrowed down the gene list by selecting only those candidates that showed inverse expression upon *STAT3* knockdown in 3 patient GPCs (NNI-4, NNI-11 and NNI-12 represent the variability in *STAT3* signaling pathway). The latter approach ensured that only genes modulated by the *STAT3* pathway would be selected for our experimental validation (**Figure-5.1**).

2.4.2 *ST3GAL1*-specific glioma signature analysis

The PNA-sorted GPC profiles were hybridized on Illumina HumanRef-8 v2 BeadChip to study the impact of transcriptome pattern upon PNA affinity. The intensity value was measured using GenomeStudio software. R/lumi packages (Version 2.36.0) were deployed to process the data, and the summarized values were normalized with a variance stabilizing algorithm (Durbin *et al.*, 2002). The non-exonic probes were trimmed off, and the data were quantile

normalized using R/Biobase packages (Version 2.44.0) (Gentleman *et al.*, 2004). We combined the PNA-high and PNA-mid profiles as PNA-high for gene expression analyses. A linear model was fitted to detect the altered transcripts between PNA-high and PNA-low using R/limma packages (Version 3.38.3) (Smyth, 2005). An FDR-adjusted p -value <0.001 was considered to identify the transcripts that were significantly different between the experimental groups.

The *ST3GAL1* knockdown pellets of GPCs were profiled on Affymetrix GeneChip® Human Genome U133 Plus 2.0 Array using a 3' IVT express kit. Raw CEL files were summarized with the MAS5 algorithm, and log₂-scaled and gene expression datasets were created. Principal component analysis was performed to assess detectable batch effects or significant quality defects in our microarray datasets. All data preprocessing analyses were carried out using R/bioconductor packages. A linear model was regressed to assess the differentially expressed genes between knockdown clones and control profiles as described in R/limma packages (Version 3.38.3) (Smyth, 2005). FDR-adjusted p -value <0.05 with a 1.5-log-fold difference was considered to indicate significantly perturbed genes upon *ST3GAL1* knockdown. A subset of differential genes was extracted as the *ST3GAL1* knockdown gene signature by applying a stringent criterion of a 2-log-fold change between knockdown clones and the control profiles.

2.5 Predictive approaches in glioma patient databases

2.5.1 Connectivity map (CMAP) approach

Our predictive analyses evaluate the qualitative enrichment of queried gene signatures against patient transcriptome profiles using a connectivity map (CMAP) approach (Lamb *et al.*, 2006). First, the signature was transformed to directionality according to the $\log_2$ fold change detected in the experimental condition. The queried transcriptome of patients was median scaled to detect the sample deviation from the reference probability distribution. Up- and down-score estimations were calculated for each patient profile, and the running sum was calculated as a connectivity score. The Kolmogorov-Smirnov (KS) test was used to estimate the statistical p -value by identifying the deviation of the queried patient dataset from the empirical Bayes distribution created from random sampling from all patient datasets. We applied 'Bnormal' function from R/wiqid package (Version 0.2.2) (Meredith, 2019) to arrive the empirical distribution by performing 10,000 samplings on patients' transcriptome data with Gibbs sampler in R/base package (Team, 2019). The Gibbs sampler implements Bayesian inference and gamma prior for the precision to achieve randomization in sampling procedure and thus determines the empirical Bayes distribution from transcriptome datasets (Geman and Geman, 1984; Lamb *et al.*, 2006). Patients with positive activation scores tend to have low oncogenic activation (identical to the knockdown signature), while patients with negative activation scores had high oncogenic activation (contrary to *STAT3* knockdown).

The CMAP score from the KS test is the direct measure of connectivity magnitude of molecular signature to the transcriptome pattern of each patient. Mathematically, the CMAP score is the difference of ranked scores from up and down regulated genesets in patient's transcriptomic profile. The normalized CMAP score value ranges from -1 to +1. A null cumulative score will be assigned when both up- and down-regulated genesets score are ranked in the same direction (either positive-positive or negative-negative sign), which implies that the signature genesets are randomly scattered over the transcriptome data of patient and not enriching for signature (Musa *et al.*, 2017). The CMAP approach doesn't permit forced assignment of samples into a specific class, hence the sensitivity (false-positive rate) or specificity assessment are not possible. The use of randomization or permutation-based approach at both instance and reference level provides compelling evidence to control potential false connections between signatures and transcriptomic profile of patient/treatment data (Musa *et al.*, 2017; Zhang and Gant, 2009).

The limitation of the CMAP approach was that ~30% of the patient population was randomly distributed in both up- and down-scoring. The algorithm classified them as 'unclassified,' which posed a limitation when predicting classes in a small sample pool. There is a likelihood of increasing false positives when the samples are fit into a class based on the KS score trend for up- and down-regulated genesets. To overcome this, we applied a flexible shrunken centroid classification approach by measuring cross-validation probabilities from training dataset. The

molecular signature classes (class labels) for the training dataset were identified from CMAP approach. The FDR was controlled by training the shrunken centroid model with optimal number of classifiers having low cross-validation (CV) error. This process is called as 'thresholding' to minimize the CV error. This subset of gene features was used as classifiers for predicting signature classes in smaller sample pool. The shrunken centroid model was executed using R/pamr package (Version 1.48.0) (Tibshirani *et al.*, 2002).

In the present analysis, we trained the model from the transcriptomic pattern of REMBRANDT-stratified classes, which were used to stratify TCGA all-glioma patients. The Phillips molecular classes were predicted in NNI and other tumor databases using the 'shrunken centroid' approach, as described previously (Phillips *et al.*, 2006; Tibshirani *et al.*, 2002). TCGA molecular signature was interrogated using the 'classification of nearest centroid' algorithm to predict for expression classes, as explained previously (Dabney, 2006). We directly applied the *STAT3* and *ST3GAL1* signatures to stratify the NNI tumor and GPCs using the nearest template algorithm implemented in GenePattern server (Hoshida, 2010; Reich *et al.*, 2006). Similarly, the exon array database of recurrent GBM (rGBM) patients was interrogated with the *STAT3* signature using nearest template prediction approaches. The Lottaz molecular signature was applied in our tumor resource to identify molecular subtypes of GPCs using the shrunken centroid algorithm (Lottaz *et al.*, 2010; Tibshirani *et al.*, 2002).

2.5.2 Nearest template prediction (NTP) analysis

Nearest template prediction is a class estimation without requiring corresponding training dataset using prediction confidence measures. This predefined gene signature is queried in the patient tumor databases to base the class predictions on molecular alteration-specific enrichment (Hoshida, 2010). First, we computed Z-score-based transformations for the patient tumor databases. Each gene in the gene set signature was coded with specific subtypes based on its over-representation for a subtype. The parameterized data were keyed in the GenePattern server to perform subtype prediction for the patient databases. GenePattern Server (Version 3.9.2) installed on Windows desktop machine was used to deploy 'NearestTemplatePrediction' module (Version 4) for classifying the patient tumors data (Reich *et al.*, 2006). The stratified classes of patient tumors were further correlated for phenotypic association, such as overall survival, magnetic resonance imaging (MRI) and molecular subtypes.

2.5.3 Prediction analysis of microarrays (PAM)

The PAM approach is a statistical technique for class prediction when training and validation cohort databases are available. We had two important scopes of analysis. First, we trained a PAM algorithm to learn Lottaz signature-based subtypes from the published collection of GPC databases (Lottaz *et al.*, 2010). The trained algorithm was cross-validated to determine the threshold value (δ) with a cross-validation error. Subsequently, the trained PAM model was queried to stratify the validation databases (*de novo* samples) from the NNI and other GPC

databases. The Lottaz signature classes were predicted using the PAM algorithm. The PAM analysis was carried out in R using the pamr package (Version 1.48.0) (Tibshirani *et al.*, 2002).

2.5.4 Single-sample gene set enrichment analysis (ssGSEA)

GI subtype signatures for each subtype were interrogated using ssGSEA with a resampling classification strategy as previously described (Barbie *et al.*, 2009; Wang *et al.*, 2018). Briefly, the empirical cumulative distribution functions (ECDFs) of subtype signatures were estimated from rank-normalized expression values of each sample. The empirical *p*-value was calculated by evaluating the ECDFs of the GI signature and the remaining genes in the transcriptome. To allow an unbiased comparison across three subtype signatures, a random resampling strategy was applied to estimate the robust signature score for each GI subtype. We implied the R/library for the GI subtype stratification scheme in patient databases as documented (Wang *et al.*, 2018). The subtype for each patient was assigned based on the ssGSEA score with the lowest *p*-value.

2.5.5 Cox proportional-hazards regression analysis

Patients with a distinct stratified pathway activation score were evaluated for prognosis using the patient databases. We utilized 5 glioma clinical databases: the Gravendeel (Gravendeel *et al.*, 2009), REMBRANDT (Madhavan *et al.*, 2009), TCGA (all glioma) (Ceccarelli *et al.*, 2016), Freije (Freije *et al.*, 2004) and Nutt (Nutt *et al.*, 2003) databases. First, we evaluated the phenotypic class association with the overall survival of patients using the log-rank test. The Kaplan-Meier

survival curves were plotted to visually assess the survival stratification (Kaplan and Meier, 1958). A linear proportional hazards regression model was adopted to estimate the association of covariates against patient survival (Cox, 1972). The multivariate assessment of survival models included pathway-specific classes, WHO classification schemes, molecular subtypes, age and Karnofsky score. Statistical significance was assessed based on a p -value of less than 0.05 to define the significant association of covariates with survival.

2.5.6 Goodness-of-fit evaluation analysis

Product limit estimator for survival rate were plotted using Kaplan-Meier estimate, and the statistical significance was measured using log-rank tests. In addition, the proportional hazards of the covariates were modeled in Cox proportional hazards analyses using the R/survival package (Version 2.43.3) (Terry TM, 2000). A Wald test p -value of less than 0.05 was considered as a minimal cutoff to determine the statistical significance. We performed multivariate analysis accounting for signature-predicted classes and clinical covariates such as age, histology and Karnofsky score (performance status scale). In addition, a separate multivariate model was examined to evaluate the confounding potential of current signature classes against the established molecular signature classes. The quality of the regression model was estimated using the Akaike information criterion (AIC) score, which is usually adapted in information theory. We calculated the AIC score for univariate analyses evaluating each clinical and molecular covariate. The model with the

lowest AIC score was treated as the baseline model, and the forward inclusion of additional covariates was evaluated for the improvement of baseline model quality. We used a likelihood ratio test to estimate the statistical significance of ΔAIC (difference between the reference model AIC and the updated model's AIC with the inclusion of new covariates).

2.5.7 Gene Set Enrichment Analysis (GSEA) and pathway evaluation

The genome-wide transcriptome pattern of predicted patient classes was analyzed against molecular signature databases (MSigDB) from the Broad Institute using the GSEA Java desktop software application version 2.2.2 (Subramanian *et al.*, 2005; Yoo *et al.*, 2015). In principle, GSEA specifically looks for the enrichment pattern of patient classes with a large collection of signatures from MSigDB to ensure the molecular relevance of signatures or phenotypes. We also leveraged the Reactome pathway database, which archives manually curated pathway annotations. The findings of Reactome pathway cluster analysis were critical in identifying the *ST3GAL1* cluster association with the upstream modulator TGF β 1 (Chong *et al.*, 2016). The core gene nodes enriched in *ST3GAL1*-high patient cohorts were evaluated by Reactome pathway gene clusters using the R/ReactomePA package (Version 1.28.0) (Yu and He, 2016). In addition, we employed GeneGo from MetaCore to identify enriched signatures with highly curated oncogenic pathways.

2.5.8 Integrative analysis linking functional pathways with diagnostic imaging data

MRI files of 43 REMBRANDT patients matched with *ST3GAL1*-signature-predicted classes were downloaded from The Cancer Imaging Archive (<http://cancerimagingarchive.net/>). The image file was analyzed using a SIEMENS syngo fastView DICOM viewer platform. The difference in the volume of the entire lesion and the volume of the tumor alone was calculated as the T2-zone, a surrogate marker for measuring tumor cell invasion. The T2-zone volume was categorized into large and small invasion groups based on the median cut-off. A log-linear model was applied to understand the interactions between T2 volume, *ST3GAL1* status and tumor grade. We used mosaic plots to represent significant associations across the categorical variables using R/vcd (Version 1.4.4) and vcdExtra packages (Version 0.7.1) (Friendly, 1994; Zeileis *et al.*, 2007). All other two-way contingency analyses in the study were analyzed using Fisher's exact test.

2.6 Protein network analysis using a Bayesian probabilistic model

TCGA's reverse-phase protein array (RPPA) data for LGG and GBM were merged and analyzed in view of all gliomas. One hundred fifty-four protein features were detected in both LGG and GBM. Some protein features had multiple antibody complexity, and was resolved by trimming off protein expressions with less variability across the samples. There was no significant batch segregation in the PCA analysis performed on the

final protein array data. We started by identifying the protein features that could best stratify the *STAT3* signature classes using the recursive feature elimination (RFE) approach with a random forest algorithm at each iteration in R/caret (Version 6.0.84) and mlbench packages (Version 2.1.1) (Friedrich Leisch, 2010). Random Forests create an ensemble of subset features that can best classify the *STAT3*-specific patient cohorts, and 135 iterations were performed to identify the top classifiers. A total of 37 protein features were identified from the RFE model to stratify the *STAT3* cohorts. We applied the Bayesian network model to study the interconnections among the top 25 protein classifiers. The BNArray package (Version 1.0.0) has been the most well-studied R package to construct gene regulatory networks from array data (Chen *et al.*, 2006). We independently trained the initial network from the *STAT3*-low and *STAT3*-high patient expression data and estimated the prior probability. A posterior network with Bayes factors was computed with the prior probability learned from the trained model using a greedy search algorithm. An adjacent matrix with Bayes factors defined the parent node and dependencies.

2.7 Kinome screening analysis in GPCs (A/Prof Tan Nguan Soon from NTU executed PamChip profiling in GPCs)

The computational pipeline of kinome screen analysis is represented in a flowchart (**Figure-2.1**). Briefly, we integrated phosphorylated peptide measurements estimated from Pamchip with

transcriptomic data of GPCs using published database resources (Hornbeck *et al.*, 2012; Keshava Prasad *et al.*, 2009; Milacic *et al.*, 2012). Two processing scripts identified the critical kinase regulatory channels in the experimental conditions, namely the biochemical modification and functional modification modules. We first parsed the normalized quantitative readouts of phosphorylation kinetics from PamChip into four major kinase-substrate databases. The mapping analyses accounted for the kinase-substrate associations from Kinexus, PhosphositePlus, Reactome and Human Protein Reference Database (HPRD) (Fabregat *et al.*, 2018; Hornbeck *et al.*, 2012; Keshava Prasad *et al.*, 2009). In total, we shortlisted 2383 kinase-substrates associations from the published databases. We observed that multiple peptide substrates were phosphorylated by specific kinases. We resolved that the multiple kinase-substrates complexity by running a weight estimation algorithm using an in-house developed R-script.

First, we ranked the peptides under each kinase based on their dynamic variations across the experimental conditions. We then constructed peptide clusters for those having a spearman-correlation coefficient of greater than 0.4. The quantitative mean of peptide cluster was calculated by evaluating the ranks for the presence of large number of correlated peptides having strong expression and dynamic variation across the experimental conditions (i, bottom left panel in **Figure-2.1**). To understand the differential kinetics between treated and untreated GPCs or proneural and mesenchymal GPCs, we applied a linear regression model on the quantitative mean summary matrix of kinases (ii, middle

panel in **Figure-2.1**). The differential regression model accounted for the specificity of cellular heterogeneity by measuring consensus correlation statistics as recommended (Ritchie *et al.*, 2015). A regression model with p-value of less than 0.1 or fold-change as a threshold were considered to determine the altered kinases in experimental conditions. A separate regression model was fitted to describe the STK mechanism, as the antibody binding chemistry differed in the technology used rendered a distinct distribution for the STK peptide measurements.

In order to identify upstream kinases that could be responsible for *STAT3*-low cell resistance toward AZD1480 treatment, the identified candidates from the biochemical modification module were interrogated in the functional modification module, which represents the genomics data from published databases and GPC-specific experimental data. The functional module in the pipeline uses a bottom-up approach to track the specific kinases identified from the previous analytic module. We performed database mapping to identify the kinase channels responsible for sensitization and resistance mechanisms of *STAT3* inhibition. We leveraged on The PhosphoAtlas map from the University of California San Francisco (UCSF), which addressed the variant-specific impact of 776 kinase interactions with phosphorylatable target sites (Olow *et al.*, 2016). By applying PhosphoAtlas map, we identified specific transcription factors (TFs) modulated by the upstream kinases (evaluated 6,488 known kinase interactions) (iii, bottom right panel in **Figure-2.1**).

To map the kinase regulatory mechanism with *STAT3* KD transcriptomic pathway, we relied on the transcriptomic factor-based regulatory circuit database from University of Lausanne and Broad Institute (Marbach *et al.*, 2016). The regulatory circuit database catalogues ~1.3 million transcriptional information correlating to 643 transcription factors (**iv**, top right panel in **Figure-2.1**). We overlaid the kinase-TFs maps from PhosphoAtlas on the regulatory circuit database and predicted transcriptional pathway of kinases affected by AZD treatment. The predictive transcriptional network from a bottom-up approach using genomics databases was intersected with *STAT3*-KD transcriptomic data generated in our GPCs. Specifically, the decreased kinases upon AZD1480 treatment were mapped back with downstream target genes (538 down-regulated genes) in *STAT3*-knockdown (KD) GPCs and vice versa. This particular functional module evaluation allows us to understand the transcriptomic network of sensitization and resistance mechanisms in *STAT3* inhibition through kinase regulatory pathways.

Figure-2.1. Kinome screening analysis pipeline flowchart. Computational workflow to prioritize kinase candidates with functional/biological evidence. The workflow includes a biochemical modification module and functional module to integrate phosphorylated kinetics measured from PamChip with *STAT3* knockdown transcriptomic profiles. A linear regression model evaluates the kinase activity between the cells with AZD1480 treatment conditions.

2.8 Relative odds estimation

We sought to understand the clinical association of WHO classification marker IDH status with *STAT3* signature-stratified classes using a relative odds estimation as previously recommended (Wald *et al.*, 1999). First, we calculated the density distribution of CMAP activation score values for both IDH wt and mutant cohorts. Then, the relative odds score was estimated from the proportion of the highest 5th distribution and lowest 5th distribution of *STAT3*-signature score values for IDH phenotypes.

2.9 Experimental validation in glioma-propagating cells (GPCs) (performed by Dr Melanie Tan)

2.9.1 Immunoblot analysis in signature-stratified GPCs

GPCs were lysed with radioimmunoprecipitation assay (RIPA) buffer containing 0.5% sodium deoxycholate, 1% NP-40 detergent, 0.1% sodium dodecyl sulfate, 0.15 M NaCl, and 10 mM Tris-HCl (pH 7.4) with protease and phosphatase inhibitor cocktail tablets (Roche). The dissolved protein (30 µg) was transferred to polyvinylidene difluoride membranes (Millipore). The antibodies utilized were as follows: anti-pSTAT3 (Tyr705; 1:1000; Cell Signaling, #9138), anti-STAT3 (1:1000; Cell Signaling, #9139), and anti-β-actin (1:10000; Sigma Aldrich A5441). Anti-mouse (1:10000; Cell Signaling, #7076) or anti-rabbit (1:10000; Cell Signaling, #7074) IgG horseradish peroxidase (HRP)-linked secondary antibody was used in the experiment. The dilution buffer for antibodies included 5% bovine serum albumin in 10 mM Tris-HCl pH 7.4, 100 mM NaCl, and 0.1% Tween® 20 (Merck). The chemiluminescence detection kit SuperSignal West Pico or Femto (Thermo Scientific) was used to process the membranes. The visualization of protein bands was accomplished using the digital imaging system SYNGENE G Box, iChemiXT. The Quantity One® software (Bio-Rad Laboratories) instrument was used to quantify the protein expression, which was normalized against β-actin levels.

2.9.2 Viability, sphere assay and dose response experiments in GPCs

Gliomasphere-forming ability, a measure of GPC proliferation, was assessed as described in our previous work (Chong *et al.*, 2009). Cell viability was assessed using 10% volume of alamarBlue® (Serotec) at post-treatment days 5 and 10 according to the manufacturer's instructions. The absorbance readings were measured after 16 hours at 570 and 650 nm. For the dose-response studies, we used absorbance readings of 570 and 600 nm. IC₅₀ values were computed from 12-point titration curves ranging from 10⁻⁴ to 10² µM.

2.9.3 *In vivo* evaluations in orthotopic xenograft NOD/SCID (NSG) mouse models

Mice were treated according to the guidelines of the Institutional Animal Care and Use Committee of the National Neuroscience Institute, Singapore. Animal implantation was carried out as previously described (Chong *et al.*, 2016). The stereotaxic injection was administered in the right frontal lobe with cells at a density of 500,000/2 µL of phosphate-buffered saline. The coordinates were +1.0 mm anterior-posterior; +2.0 mm mediolateral; -2.5 mm dorsoventral. Kaplan-Meier analyses were assessed using the log-rank test in GraphPad Prism software.

RESULTS

3 MOLECULAR SUBTYPE PATTERNS OF BRAIN TUMOR RESOURCE

3.1 Introduction and objectives

Brain tumors such as GBMs have a dismal prognosis despite advanced surgical intervention and chemotherapeutic approaches. The mean survival period is 15 months upon diagnosis, highlighting the urgent need for better therapeutic strategies. The conventional classification system of gliomas primarily focused on tumor morphological characteristics such as nuclear pleomorphism, mitotic features, necrosis and vascular endothelial proliferation (Louis *et al.*, 2007). The landmark publication by TCGA in 2010 demonstrated that histologically similar GBM tumors are molecularly distinct, with each subtype corresponding to unique genomic aberrations, gene expression profiles and clinical outcomes (Verhaak *et al.*, 2010), highlighting the limitation of relying solely on histology to diagnose and treat patients.

The majority of these GBM tumors are highly aggressive and infiltrative, and glioma stem-like cells (glioma-propagating cells, GPCs) perpetuate and sustain tumor growth in orthotopic mouse models (Rich and Eyler, 2008). GPCs demonstrated a hierarchical transformation from oligodendrocyte progenitor cells (OPCs) and NSCs in transgenic mouse models (Zheng *et al.*, 2008). In addition, NSCs undergo synchronized somatic alterations in tumor-suppressor pathways, suggesting that the molecular pattern dictates glioma disease progression. Our earlier work showed that patient-derived GPCs contain transcriptomic information that

dictates the primary tumor phenotype (Ng *et al.*, 2012). More importantly, we have demonstrated that GPCs are clinically relevant because they contribute to patient molecular profiles and survival outcomes.

TCGA evaluated the molecular pattern of PDx models with primary tumors using a subtype-specific gene expression signature. Interestingly, the results demonstrated a strong concordance of subtype matching between primary tumors and xenografts established from GPCs (Verhaak *et al.*, 2010). Our group also demonstrated that PDx models from GPCs recapitulate the pathological features of original primary tumors (Chong *et al.*, 2009; Chong *et al.*, 2016; Foong *et al.*, 2012). Thus, a Brain Tumor Resource including GPCs, PDx models and primary tumors serves as a valuable tool to perform precision medicine-based screening. A combined molecular description of GPCs, tumors, and PDx-based models is presented in a limited number of studies (Jin *et al.*, 2017; Lottaz *et al.*, 2010; Rich and Eyler, 2008). We illustrate a flowchart to systematically characterize our NNI Brain Tumor Resource (**Figure-3.1**). The current chapter addresses the computational framework for understanding the critical genomic features of GPCs, primary tumors and orthotopic PDx models in NNI Brain Tumor Resource databases.

Figure-3.1. Graphical illustration of NNI Brain Tumor Resource. Workflow of tissue collection and processing, the creation of a genomic roadmap of patient material, and the implementation of machine learning approaches to understand the molecular characteristics of brain tumors. A molecular similarity map represents a recreation of primary tumors in orthotopic xenograft tumors derived from glioma-propagating cells (GPCs). Each color code denotes unique patient material with matched xenograft tumors. Triangle, xenograft tumor; square, patient tumor.

3.2 Molecular classification scheme of glioma-propagating cells (GPCs) in NNI and international molecular resources

Precision medicine refers to the tailoring of therapeutic strategies based on the unique molecular profile of each individual and involves the application of deep-learning statistical models to derive patterns of molecular classifiers. This approach allows for the classification of patients into subpopulations distinct to their molecular makeup and subsequently, dictating the choice of therapeutic intervention.

3.2.1 Molecular subtype signatures of GPCs

We have previously shown that our GPCs are able to preserve the karyotypic and transcriptomic hallmarks found in the original primary tumor. Importantly, these implanted cells re-establish orthotopic xenograft tumors that phenocopy the original patient tumor pathophysiology (Chong *et al.*, 2009) (**Figure-3.1, Molecular similarity panel**). The NNI Brain Tumor Resource distinguishes itself as the large collection of adult patient tumors in Singapore, with corresponding live cells capable of recreating the full spectrum of the original patient tumor molecular heterogeneity. Importantly, this resource contains deep genomic and transcriptomic data to facilitate molecular stratification studies.

To further understand how our NNI molecular profiles match with international GPC collections, we evaluated the NNI GPC databases (N=32 with one technical replicate) against three publicly available GPC repositories archived in the GEO databases: (GSE7181; N=6), (GSE8049; N=8 with one technical replicate) (GSE15209; N=3 with one technical replicate and N=3 without replicate) (Beier *et al.*, 2007; Gunther *et al.*, 2008; Pollard *et al.*, 2009). We merged the complete gene expression profiles of the GPC repositories mentioned above with our NNI GPC database. All databases were normalized using the standard Affymetrix normalization procedures. We observed that 50% of probesets in the International GPC collection had lower expression levels and those were subsequently scaled with MAS5 scaling factor as recommended (Gautier *et al.*, 2004). The integration of multiple gene expression databases was assessed by evaluating the box plot distribution across

samples. We found that the distribution of GPCs was specific to NNI and international databases. A PCA was subsequently performed on 2500 highly variable probesets in the merged dataset, consistently the sample clusters followed a unique pattern for NNI and international databases, indicating that batch effect was specific to databases. (**Supplementary Figure-1**).

We employed an empirical Bayesian batch effect removal framework across the various GPC databases as well as our NNI resource (Johnson *et al.*, 2007). This statistical model identifies and reduces common technical variabilities among the multiple databases, and potential biological variances are retained (Leek *et al.*, 2019). Empirical Bayes method has substantial advantages in resolving the batch effect problem, as Bayes inferences are better derived from several genes across multiple experimental conditions. More importantly, Bayes method is ideal even for correcting the databases with smaller sample pool and, robust in stabilizing the variances for those genes having very high or very low expression ratios. The basic assumption of the batch effect removal model is that the effect of batch bias tends to affect major proportion of genes causing high variability from increased/decreased expression in certain batches of samples. The model estimates mean (location) and variance (scale) parameters across multiple genes in each database and, shrinks towards the collective mean estimates of multiple batches. These empirical estimates are further applied at each gene level to adjust the batch effect in the merged dataset. The batch adjustments for high dimensional data involve three major steps: 1) standardize the

data, where the model parameters are estimated; 2) Empirical Bayes batch effect parameter estimates using empirical priors, and 3) global adjustment of data for batch effects (Johnson *et al.*, 2007).

The empirical Bayesian batch effect removal model was implemented in the function called 'ComBat' from R/sva package (Version 3.30.1) (Leek *et al.*, 2019). The implementation of 'ComBat' function required to specify the known covariates causing batch effect in the merged gene expression data matrix using a function called 'model.matrix'. Parametric adjustment across samples was set to be true for calculating prior estimates from the data. With the model covariates and parametric estimation, the 'ComBat' function corrected the adjustments for the batch effect across NNI and international GPC databases. A quantile normalization was executed on the batch effect corrected data using 'normalize.quantiles' function implemented in R/preprocessCore package (Version 1.44.0) (Bolstad, 2018). A PCA map is then generated, accounting for a normal variability spread across our NNI GPC collection and that of our international peer collections (**Figure-3.2A**). The explained variance from top two principal components (PC1) and PC2 was 23%, which was estimated from the top variable 2500 probesets. Furthermore, these tumors demonstrate similar enrichment of neural cell lineages previously associated with tumor-initiating and propagating capacity. (**Figure-3.2B**).

Figure-3.2. Validation of NNI GPCs with international molecular databases of GPCs. Comparison of NNI patient-derived glioma cells with international collections based on: **(A)** Top most variable probesets (N=2500) accounted the variance of 23% from the two principal components (PC1) and PC2. Red, NNI (N=32 with one technical replicate); orange, International (GSE7181; N=6), (GSE8049; N=8 with one technical replicate) (GSE15209; N=3 with one technical replicate and N=3 without replicate). N=95 represents combined sample number from NNI GPCs and international resources. (Beier *et al.*, 2007; Gunther *et al.*, 2008; Pollard *et al.*, 2009). **(B)** Neural cell types defined by the Cahoy signature (Cahoy *et al.*, 2008; Lottaz *et al.*, 2010). Cells with a mesenchymal phenotype display a high enrichment score for the cultured astroglia cell type. Color code: dark blue, proneural; aquamarine, mesenchymal. The plots were generated using the base packages of R/stats and R/graphics (Team, 2019).

Lottaz and colleagues established that the GPCs can be molecularly distinguished into two types independent of the culture conditions and growth pattern (Lottaz *et al.*, 2010). We then applied the Lottaz signature to interrogate our merged GPC collection using a shrunken centroid approach (Tibshirani *et al.*, 2002). The transcriptomic database from Lottaz was utilized as our training dataset, and the trained model with high cross-validation accuracy was deployed to classify a new cohort of our study. Following, we ranked the shrunken class centroids

for the molecular subtypes and applied the gene features to classify the molecular subtypes (**Table 3**). This centroid-based classification algorithm defined our cells as distinct proneural or mesenchymal molecular subtypes, as shown in **Figure-3.3**. Most importantly, the mesenchymal GPCs displayed a strong association with the cultured astroglial signature in the single-sample Gene set enrichment analysis (ssGSEA) scoring analysis (**Figure-3.2**) (Barbie *et al.*, 2009). These results are in alignment with published features from TCGA mesenchymal collection (Koh *et al.*, 2013; Verhaak *et al.*, 2010).

Figure-3.3. Heatmap of the Lottaz signature-based classification of NNI (N=32 with one technical replicate) and International GPC transcriptomic databases from Beier, Gunther and Pollard (Beier *et al.*, 2007; Gunther *et*

al., 2008; Pollard *et al.*, 2009). A 24-gene signature stratifies the transcriptomic data of GPCs into proneural and mesenchymal phenotypes. The complete linkage metrics of Euclidean distance estimated from gene expression data was applied to define the hierarchical clustering of molecular subtypes. The plots were generated using the R/pheatmap package (Version 1.0.12) (Kolde, 2019).

To derive a signature that could better classify GPCs, we evaluated the performance of the Lottaz 24-gene signature and Bhat *et al.* conventional markers, which include proneural (*DLL3*, *OLIG2*, *ASCL1*, and *NCAM1*) and mesenchymal (*YKL40*, *SERPINE1*, *TIMP1*, and *TGFBI*) (Bhat *et al.*, 2013). We interrogated signatures into three patient databases using a non-parametric qualitative enrichment approach called connectivity map (CMAP) (Lamb *et al.*, 2006). The performance of these signatures were compared based on Cox proportional hazards regression model and their sensitivity to detect molecular subtypes in three primary glioma patient molecular databases (Ceccarelli *et al.*, 2016; Gravendeel *et al.*, 2009; Madhavan *et al.*, 2009) (**Table 4**). We showed that these two signatures addressed the significant prognostic variability in the log-rank test. The signature by Bhat *et al.* had high specificity, indicated by its high hazard ratio, but its detection rate was less than 20% of the patient cohort, limited its applicability. The Lottaz signature, on the other hand, had lower specificity but greater detection rate. We therefore combined the two signatures to enhance the specificity and detection rate. This composite gene signature [Total=32; 24 gene signature (Lottaz *et al.*, 2010) and 8 marker genes (Bhat *et al.*, 2013)] resulted in an increased detection rate and prognostic prediction as compared to individual signatures. The composite gene signature was able to stratify 54.5%, 66.9% and 39.3%

of the patient tumors in REMBRANDT, Gravendeel and TCGA database, respectively. Prognostic evaluation in patient tumors was also noted to be better in the composite gene signature compared with Lottaz signature [HR=0.178; REMBRANDT, HR=0.29; Gravendeel, HR=0.038; TCGA] (**Table-4**) (Ceccarelli *et al.*, 2016; Gravendeel *et al.*, 2009; Madhavan *et al.*, 2009)

In order to translate the composite gene signature into diagnostic classification markers, we performed regression analysis including normal brain samples and GPCs. Briefly, we fitted a linear model to an independent transcriptomic database from Colorado (GSE86574) that includes normal brain samples and GPCs to identify critical classifiers to distinguish molecular subtypes and normal samples (Amani *et al.*, 2017). We estimated the overall variability and performed pairwise comparison to identify classifier genes (Bhat *et al.*, 2013; Lottaz *et al.*, 2010) that were differentially expressed in proneural and mesenchymal subtypes. Nine proneural (*C1orf61*, *ASCL1*, *ARC*, *NGC*, *ELAVL4*, *DLL3*, *OLIG2*, *SOX2*, and *NCAM1*) and five mesenchymal (*COL1A2*, *PHLDA2*, *CYP1B1*, *ALDH1A3*, and *TGFBI*) markers were identified (**Table 4**).

Lottaz molecular subtypes and normal cells

Figure-3.4. Box-whisker distribution of molecular subtype classifier gene expression. A list of 14 classifiers displays stratified expression levels across proneural and mesenchymal molecular subtypes and normal samples in the merged database of the Colorado (GSE86574) and NNI GPC resources. The box distribution plots were generated using R/graphics and R/beeswarm (Version 0.2.3) packages (Eklund, 2016)

This result is important as it validates the use of subtype classifiers in a non-invasive clinical setting to enable the molecular characterization of tumors. We confirmed the expression level-based stratification of NNI GPC subtypes by qRT-PCR. GBM tumors often do not release detectable circulating tumor cells or soluble biomarkers into the circulatory system. However, previous studies found that large quantities of cancer extracellular vesicles called exosomes exist in the blood of GBM patients, which enables us to understand tumor nature through minimally invasive profiling (Shao *et al.*, 2012; Shao *et al.*, 2015). The identified expression markers in the present study have a potential application in detecting the molecular subtypes of GBM patients using blood samples.

3.2.2 Kinome profiling in GPC molecular subtypes

The Brain Tumor Resource encompassing molecular data, aims to identify the patient cohorts with specific enrichment of key signaling pathways. In this context, we specifically focused on identifying biochemically-active kinase regulatory mechanisms, upregulated in mesenchymal cells. Kinases are well-known druggable targets, and modify up to 30% of all human proteins. The regulatory mechanisms of kinases are often dysregulated in several cancers. In precision medicine

efforts, kinases play an important role in repositioning therapeutic strategies for novel disease indications based on specific genetic aberrations in tumors. Understanding the biochemical mechanism specific to cell type is pivotal for designing targeted therapies and determining toxic effects (Klaeger *et al.*, 2017). Our lab previously demonstrated the efficacy of phosphoinositide 3 kinase/protein kinase B/mechanistic target of rapamycin (PI3K/AKT/mTOR) and Polo-like kinase 1 (PLK1) inhibitors in our screening of GPCs (Foong *et al.*, 2012; Foong *et al.*, 2011). Kinases were shown to contribute towards the mesenchymal transition in wide range of cancers including GBM (Lamouille and Derynck, 2007; Segerman *et al.*, 2016). Thus, identifying the crucial kinase pathways enriched in mesenchymal cells provides the strategy for the design of small-molecule inhibitors targeting aberrant kinase activity.

In our work, we utilized a three-dimensional chip technology called PamChip, to measure the phosphorylation kinetics of PTKs and STKs (Labots *et al.*, 2016). The PamChip profiling of GPCs was carried out by A/Prof Tan Nguan Soon from NTU. Using a novel computational pipeline described in the methods section (**Chapter 2.7**), we performed a differential analysis on a cohort of two proneural and mesenchymal GPC lines using a linear regression model. First, we calculated the quantitative mean summary of the clustered peptides that act as substrates for each kinase. Each quantitative summary considers a weighted rank, taking into account the dynamic variation in peptides and their preference for strong expression. We kept a median fold-change as a threshold to determine

the active kinases of a specific subtype. A separate regression model was fitted to describe the STK mechanism, as the antibody binding chemistry differed in the technology used rendered a distinct distribution of the peptide measurements.

TCGA's comprehensive genomic profiling showed that the RTK/RAS/AKT signaling pathway was aberrantly regulated in 88% of GBM patient tumors, but till date, the clinical efficacy of drugs targeting this pathway remains elusive. Our current study aims to investigate the phosphorylation kinetics of peptide substrates for RTKs and STKs in mesenchymal and proneural GPCs. Our high-throughput kinase profiling identified the top RTKs and STKs altered in mesenchymal cells, as compared to proneural cells (**Figure-3.5** and **Table 5**). A cluster of Src family tyrosine kinases (SFKs), including Lck, PDGFRA, RET, Kit, BLK and CSK, was identified as top-ranked active kinases in mesenchymal cells. These results suggest that kinase activation plays an enhanced role in the progressive phenotype of GBM. Reports emphasized that increased anti-apoptotic and pro-survival mechanisms were attributed to active SFKs, including Lck's conjunction with PI3K signaling pathways (Ahluwalia *et al.*, 2010; Laplante *et al.*, 2006). Despite the strong *in vitro* evidence favoring the role of SRC in glioma progression, the phase 2 clinical trial with Dasatinib targeting SRCs and associated kinases PDGFRA, EPHA2 and KIT did not provide therapeutic benefit in recurrent GBM treatment (Lassman *et al.*, 2015). This clinical trial did not establish biomarker correlations with the efficacy of Dasatinib, which might warrant an in-depth pre-clinical evaluation in a mesenchymal cohort of patient-

derived materials. In addition to SFKs, we observed a strong representation of mitogen-activated protein kinase (MAPK) in mesenchymal cells. Dysregulation of MAPK and PI3K was identified as a genetic hallmark of GBM progression and caused G1/S-defective checkpoints (Verhaak *et al.*, 2010; Vitucci *et al.*, 2013). The activated MAPK pathway promotes GBM progression through HIF1- α - and VEGF-mediated angiogenesis (Lim *et al.*, 2004).

Importantly, a group of CDK family kinases, including CDK-like (CDKL) 1 and 2, and the cell cycle-associated kinase CHED, were active in the mesenchymal phenotype. Crystallographic studies revealed that CDKL proteins were homologous with CDK and MAPK, indicating that the active domains of parent kinases might be potentially conserved (Canning *et al.*, 2018). The hyperactivation of CDKL proteins in mesenchymal cells might be associated with aberrant upstream effects of MAPK family kinases. The CDKL2-induced mesenchymal transition is attributed to the β -catenin signaling pathway in breast cancer progression (Liu *et al.*, 2016). Apart from this cluster of three RTKs, our results revealed collagen-associated kinases such as DDR2, Ca²⁺/calmodulin-dependent kinases HUNK and CMGC family kinases (DYRK1A). The novel association of DDR2 and HUNK kinases with mesenchymal characteristics is yet to be understood. However, these kinases are linked to the canonical cellular migration pathway. Interestingly, the upregulation of DYRK1A in mesenchymal cells suggests that it could be a potential therapeutic intervention, as DYRK1A inhibition suppressed EGFR-mediated hyperproliferation in GBM cells (Pozo *et al.*, 2013). A critical

stemness-associated kinase, JNK1, was upregulated in mesenchymal cells compared with its expression in proneural cells. The activity of JNK1 reflected the increased potential of cellular proliferation and self-renewal in response to cellular stimuli. Therapeutic inhibition of JNK1 was implicated as a viable strategy to control GBM proliferation in both *in vitro* and *in vivo* experiments (Matsuda *et al.*, 2012).

Figure-3.5. Active kinase spectrum in mesenchymal cells. Top STKs (left) and RTKs (right) altered in mesenchymal cells compared to proneural cells.

The PamChip profiling evaluated the distinct mode of phosphorylation kinetics by concentrating serine and threonine residues of peptide substrates (STKs). **Table 5** summarizes the list of top PTKs and STKs identified to be active in mesenchymal cells. We applied the same analytical strategy described in the methods section to detect those aberrantly regulated STKs in mesenchymal cells by comparing the peptide measurements with a proneural group of cells. STK activation

requires dimerization of dual receptors and GTPase modification, while PamChip technology accounts for these reactions in kinetics estimations. The differential STK list includes three major clusters [Aur-family kinase, microtubule-associated kinase (MAST) and autophagy kinase (ULK)], which modulate the proliferation potential mechanism of mesenchymal cells.

The developmentally regulated aurora kinases (AURKB and AURKC) were viewed as promising targets for treating GBM, as the inhibitor compounds reflect robust efficacy in *in vitro* and *in vivo* models of GBM tumors (Lee *et al.*, 2013; Van Brocklyn *et al.*, 2014). The MAST kinase cluster identified in our STK analysis was a novel mechanism related to microtubule association with mesenchymal proliferation that requires functional validation in GPCs. There is evidence presenting anti-microtubule compounds as promising targets for the therapeutic management of GBM (Katsetos *et al.*, 2015). The functional mechanism for anti-microtubule targets was identified to induce microtubule polymerization that affects cell division and mitotic arrest, leading to cell death. In addition, the autophagy kinases ULK1 and ULK2 were upregulated in mesenchymal cohorts, indicating the prevalence of hypoxia-associated mechanisms in mesenchymal profiles (Jawhari *et al.*, 2016). The ULK mechanism was also implicated in temozolomide resistance in GBM cells through the ATM-AMPK pathway, which corroborated the potential mechanism for drug resistance in mesenchymal cells (Zou *et al.*, 2014). The other critical kinase identified in our study was MAPK14 (p38 α), which is elevated in mesenchymal cells

and is involved in a critical pathway of EMT progression through the TGF β -mediated p38 mechanism (Bhowmick *et al.*, 2001). We further observed increased levels of STK33, VRK1, PBK, and Haspin, which were associated with poor prognosis in GBM patients (Dominguez-Brauer *et al.*, 2015; Joel *et al.*, 2015; Rodriguez-Hernandez *et al.*, 2013; Scholl *et al.*, 2009). Taken together, our study was able to catalog the aberrant phosphorylation kinetics of RTK and STK classes. The current results are important because the aberrant, biochemically active kinase mechanism could be the potential molecular switch for progression and should be targeted by candidate compounds.

3.2.3 Brain molecular subtypes

Gene expression drives disease progression and survival mechanisms in glial tumors (Phillips *et al.*, 2006; Verhaak *et al.*, 2010). The molecular characterization of glioma cells largely relies on the quality of tumors collected, which helps to establish prognostic associations through retrospective analyses. TCGA created specific protocols to assess the sample quality and archival procedures. Intriguingly, TCGA observed a considerably low percentage of cells that were viable from the various tumor collection hospitals (Scott *et al.*, 2012; Suh *et al.*, 2013; The Cancer Genome Atlas Research *et al.*, 2008). As with sample quality, the clinical and genetic information databases are pivotal to understand the molecular insights often being interrogated using machine learning models. Our group demonstrated the importance of having expertise in culture and preserving processes such as vitrification of patient-derived

cells (Chong *et al.*, 2009; Lee *et al.*, 2006). These advanced technologies essentially contribute to retaining the karyotypic and transcriptomic hallmarks found in the original primary tumors.

We interrogated the Phillips signature in the gene expression database of NNI primary tumors. Collectively, we categorized our NNI primary tumor collection into proneural (36%), proliferative (14%) and mesenchymal (50%) using the Phillips signature with the centroid-based prediction method (Tibshirani *et al.*, 2002). The results show that the NNI brain tumor database contains diverse signaling pathways of particular brain tumors. The molecular classification of TCGA subtypes included proneural, neural, proliferative and mesenchymal, and each subtype was mapped with specific genomic alterations. Importantly, the authors established the subtype specificity association with therapeutic modalities, which mediate new channels for designing targeted therapies for GBM patients. The critical genomic features mapped with proneural subtypes were IDH and *TP53* mutations and focal gains in *PDGFRA* copy numbers. The proneural group displayed strong expression of SOX-associated cluster genes [*DCX*, *DLL3*, *ASCL1* and *TCF4*] and oligodendrocytic precursor genes [*PDGFRA*, *NKX2-2* and *OLIG2*] (Verhaak *et al.*, 2010). Within our NNI tumor resource, 25% of tumors were of proneural subtype as predicted by the ClaNC method (Dabney, 2006).

The typical feature of *EGFR* copy amplification was present in the classical subtype of patients. The GBM-specific events, chromosomal 7 gain and chromosome 10 loss, were identified in 100% of classical

subtype patients. NNI primary tumors consisted 5% of classical subgroup patients. A mesenchymal patient subtype with poor prognosis had significant enrichment with *NF1* aberrations, including copy number loss and somatic mutations. There were concomitant mutagenic events of *NF1*, and *PTEN* was implicated in the hyper-AKT pathway in mesenchymal tumors. Consistent with the Phillips subtype predictions, in our resource, 50% of NNI primary tumors were mesenchymal in molecular nature. The neural subtype was typified with the gene expression of neuronal markers (*NEFL*, *GABRA1*, *SYT1* and *SLC12A5*); however, no focal copy-level alterations were mapped with the neural subtype group of patients. Of the patients in the NNI primary tumor resource, 17% were classified in the neural group.

3.2.4 Glioma-intrinsic (GI) subtypes in primary tumors

Recently, Wang *et al.* highlighted the contribution of the tumor microenvironment in determining molecular subtype features (Wang *et al.*, 2018). We first interrogated the GI molecular subtype signature in our primary tumor databases using the ssGSEA method (Barbie *et al.*, 2009; Wang *et al.*, 2018). The GI signature uniquely addresses the molecular heterogeneity of glioma tumors by eliminating the contributions from tumor-associated host cells and the tumor microenvironment. The GI subtypes are identifiable using 50 specific gene clusters, and each gene set represents proneural, classical and mesenchymal subtypes. Next, NNI primary tumor database was interrogated with GI signature using the weighted ssGSEA ranking method coded as R-libraries archived by the

author (Wang *et al.*, 2018). Our primary tumor collection at NNI exhibited a distribution of 25% proneural, 35% classical and 40% mesenchymal patients (**Figure-3.6**). The molecular subtypes using 3 different signatures (GI signature, Phillips classification and TCGA subtypes) had significant overlap across the molecular subtypes. The log-linear regression model estimated that the subtypes were significantly correlated (Pearson residual p -value<2.2e-16) (**Figure-3.7**).

The refinement of signature by accounting for tumor microenvironmental and tumor-intrinsic cells delineated three robust molecular subtypes across different technology platforms. A 3-dimensional mosaic plot analysis addresses important insights (**Figure-3.7**). For instance, TCGA signature stratified 5% (N=3) of our tumors as classical, whereas 36% (N=20) of our tumors were sub-grouped as classical by GI signature; the latter being more closely associated with the percentage distribution in public datasets as evidenced by *EGFR* abnormalities (*EGFR* amplification and gain-of-function mutations drive the classical subgroup). These results suggest that TCGA's signature underestimates the mesenchymal contribution typically represented by *EGFR* abnormalities. In contrast, the proneural subtype by the GI signature has greater concordance (86%) with TCGA's proneural tumors, while the rest of those 14% of tumors were classified as neural by TCGA. Similarly, 82% of mesenchymal tumors by the GI signature was robustly classified as mesenchymal subtype by TCGA signature. These results indicate a marginal proportion of subgroup mismatching across these three signature-based subtypes in our NNI tumors. I believe that whether

any classification scheme is more accurate cannot be established from bioinformatics analyses, such as here, alone. We must recognize that the original TCGA signature does not bear any prognostic correlation. However, the GI classification has significant prognostic implications. More importantly, when we align the major somatic mutations with the GI classification, there is a greater match of these mutations with lineage-tracing animal models involving *TP53*, *PTEN*, *EGFR* etc. that implicate tumor-initiation and -sustenance roles (Alcantara Llaguno *et al.*, 2009; Furnari *et al.*, 2007; Liu *et al.*, 2015; Wang *et al.*, 2018; Zheng *et al.*, 2008). Overall, tumor progression and recurrence is a complex process. While mutations analyzed in clinical databases offer important insight, their driver or passenger roles can only be tested using transgenic animal models.

Figure-3.6. Glioma-intrinsic (GI) subtype classification of NNI primary tumors (N=28 with one technical replicate). Heatmap representation of GI subtypes with subtype-specific expression patterns. The complete linkage metrics of Euclidean distance estimated from gene expression data was applied to define the hierarchical clustering of molecular subtypes. The heatmap was generated using the R/pheatmap package (Version 1.0.12) (Kolde, 2019).

Interestingly, GI-signature-based analysis revealed that the previously determined neural (NE) molecular subtype was purely a result of non-tumor-specific contamination. Tumor microenvironment evaluation thus explains the rationale for the absence of focal genetic alternations in

the neural subtype. In parallel, the neural subtype was shown as to be correlated with tumor margin phenotypes, which had a considerably low number of tumor cells (Brennan *et al.*, 2013; Sturm *et al.*, 2012).

Figure-3.7. Mosaic plot representation of molecular subtype enrichment from TCGA, Phillips and glioma-intrinsic (GI) signatures. All three molecular signatures (TCGA, Phillips, and GI) display consistent overlap across the subtypes (Pearson residual p -value $< 2.2e^{-16}$). The mosaic plot was drawn on the categorical distribution of molecular subtypes in NNI primary tumor resource using R/vcdExtra package (Version 0.7.1) (Zeileis *et al.*, 2007)

Taken together, the evidence shows that evaluation of tumor purity using gene expression signatures is critical in understanding the

molecular characteristics of tumors. Carter *et al.* proposed a computational approach to assess tumor purity using a scoring method called ABSOLUTE, which is based on somatic alterations of the tumor genome from exome data (Carter *et al.*, 2012). By combining stromal and immunological features from gene expression data, Yoshihara *et al.* identified an algorithm called Estimation of Stromal and Immune cells in Malignant Tumor (ESTIMATE) to calculate the purity of tumors (Yoshihara *et al.*, 2013). In principle, the algorithm was built on the premise that the bulk of non-tumor components present in resected tumors is made up of stromal and immune cells

We then set out to assess the tumor purity based on the gene expression features in our NNI primary tumor resource. The interrogation of the ESTIMATE algorithm on the NNI tumor database calculated three different score estimates using the ssGSEA approach (Barbie *et al.*, 2009): the stromal enrichment score, immune score and collective ESTIMATE score (**Figure-3.8A**). The ESTIMATE score of each sample was fitted on the regression curve plotted on the ABSOLUTE scores established from TCGA cancer datasets to calculate the tumor purity (Carter *et al.*, 2012). The ESTIMATE score displays a strong negative correlation with tumor purity, which implies that a lower score reflects minimal contamination from non-tumor components. Our NNI primary tumors had a mean purity of 62%, which is within a comparable range with the observation from largest TCGA tumor collections (59%). In addition, the tumor purity of our NNI resource ranges from 15% to 99%. We sought to understand the association of the purity score with

molecular subtypes of the primary tumor. As established by recent studies, mesenchymal tumors have an increased presence of tumor-associated macrophages (TAMs), such as microglia and circulatory monocytes (Gabrusiewicz *et al.*, 2016; Ye *et al.*, 2012). In accordance with previous reports, we observed a strong correlation of tumor purity with GI subtypes in NNI primary tumors. Mesenchymal tumors had significantly lower purity scores and exhibited mesenchymal enrichment with TAM components. Importantly, we also evaluated microglial marker expression in matched GPCs and primary tumors classified by GI signature. *ITGAM* and *AIF1* expression were significantly higher in mesenchymal tumors than in non-mesenchymal tumors, and the association was lacking in GPCs (**Figure-3.8B**).

Figure-3.8. Tumor purity score association with heterogeneity, comparison between NNI specimens. (A) The tumor purity score significantly correlates with GI subtypes. The mesenchymal subtype maps with low ESTIMATE scores, suggesting heterogeneous infiltration of tumor macrophages and microglial cells in the microenvironment. Evaluation of microglial marker expression of **(Bi)** *ITGAM* and **(ii)** *AIF1* which is exclusively present in mesenchymal tumors. The box distribution plots were generated using R/graphics and R/beeswarm (Version 0.2.3) packages (Eklund, 2016)

3.3 Summary

Taken together, this chapter highlights the clinical relevance and utility of our NNI Brain Tumor Resource (Chong *et al.*, 2016; Koh *et al.*, 2013; Ng *et al.*, 2012; Tan *et al.*, 2019; Xu *et al.*, 2018; Yeo *et al.*, 2012). Recent studies suggest that several molecular parameters outperform current clinical indicators in accounting for patient survival profiles, thus underscoring the significance of inter-tumoral heterogeneity in responsiveness to targeted therapeutic strategies (Brennan *et al.*, 2013; Ceccarelli *et al.*, 2016; Noushmehr *et al.*, 2010; Sturm *et al.*, 2012). We now show here in this chapter, using machine learning approaches, how our patient-derived glioma-propagating cells (GPCs) cluster closely with international collections based on transcriptomic profiles and furthermore utilizing predictive algorithms, how pathway signatures interrogate into patient's gene expression data to address the molecular heterogeneity. Importantly, employing the ESTIMATE algorithm, we demonstrate that our NNI primary tumors generate concordant tumor purity scores to that of TCGA's collection and are distinct from stromal environment contributions, suggesting the acquisition of proper tumor material during surgery. These results therefore indicate the NNI Brain Tumor Resource as a molecularly well-annotated biobank with capability in establishing precision medicine-based therapeutic approaches, imaging modalities and defining novel biomarkers to monitor in real-time disease progression where repeat surgeries may not always be feasible.

4 ST3GAL1-ASSOCIATED TRANSCRIPTOMIC PROGRAM IN GLIOBLASTOMA TUMOR GROWTH, INVASION, AND PROGNOSIS.

4.1 Introduction and objectives

Literature has shown that several sialyltransferase enzymes are responsible for the regulation of cellular machinery in glioma, but the molecular mechanisms driving self-renewal capability in glioma stem-like cells remain unexplored (Moskal *et al.*, 2009). Hyper-activation of a unique sialyltransferase enzyme, ST3 beta-galactoside alpha-2,3-sialyltransferase 1 (ST3Gal1) has been implicated in a range of solid tumor malignancies, such as colon and bladder cancer (Schneider *et al.*, 2001; Videira *et al.*, 2009). We recently established the role of ST3Gal1 in mediating GPC self-renewal and invasive potential using a patient-centric approach followed by experimental validation *in vitro* and *in vivo* (Chong *et al.*, 2016). Importantly, we also showed that the *ST3GAL1* transcriptomic signature stratifies glioma patients, and the transcriptomic signature performs as an independent predictor of overall survival.

4.2 Differential binding of peanut agglutinin (PNA) mediated by *ST3GAL1* identified GPC fraction with greater proliferation, invasiveness and stem-like properties

Dr Chong YK performed PNA sorting experiments, transcriptomic profiling of the GPCs and functional characterization of computational predictions in the study. Our group recently identified that the GPC fraction with negative/low binding for lectin peanut agglutinin (PNA)

exhibited increased cellular proliferation and invasiveness (Chong *et al.*, 2016; Salner *et al.*, 1982). We further performed GSEA on the PNA-sorted transcriptome, derived using microarray expression analysis on glioma cells sorted based on high and low PNA binding, with stemness-related gene signatures and found that PNA-low fraction cells had strong enrichment (normalized enrichment score=0.81) with stemness traits (Patel *et al.*, 2014; Subramanian *et al.*, 2005). The top 5 stemness genes associated with the transcriptome of PNA-low cells were *SCG2*, *CCND2*, *ENO2*, *ARRDC3*, and *PCMTD2*. This finding further corroborated with the *in vitro* studies that PNA-low binding cell fractions indeed exhibit increased stemness property as evidenced by their increased self-renewal property using clonogenic assay.

In addition, PNA sorted microarray expression analysis identified *ST3GAL1* as the most differentially expressed gene (**Table 6**; p -value=1.50e⁻⁰⁹). Next, we interrogated the PNA-associated signature in patient databases to evaluate the prognostic role in glioma patients using CMAP pipeline described in **Chapter 2.5.1**. Patients with PNA-high-like transcriptomes had significantly better overall survival than those with PNA-low-like profiles in both REMBRANDT (log-rank p -value=9.60e⁻⁰⁵) and Gravendeel (log-rank p -value=0.004) patient databases. Importantly, we observed an inverse association of *ST3GAL1* mRNA expression with PNA affinity. Over-expression of ST3Gal1 protein in GPCs and subsequent PNA staining confirmed the observation that ST3Gal1 drives the differential PNA binding pattern.

Indeed, PNA, an exogenous lectin isolated from peanut, has been used to detect T-cell maturation (Pereira *et al.*, 1976; Reisner *et al.*, 1976; Rose and Malchiodi, 1981). The PNA-high phenotype is caused by its high binding affinity to non-sialylated glycan molecules, which are strongly expressed in thymocytes in the cortical region. However, mature thymocytes tend to express higher levels of ST3Gal1, which sialylate core 1 O-glycans and hinder the binding of PNA, giving rise to the PNA-low phenotype (Despont *et al.*, 1975; Gillespie *et al.*, 1993; Priatel *et al.*, 2000; Wu *et al.*, 1997). The transcriptomic evaluation of PNA affinity suggests that the molecular mechanism of PNA implicated in *ST3GAL1* signaling might reveal prognostic insights and stemness traits relevant to glioma proliferation.

Studies suggest that an increase in the sialylation of glycoproteins caused by up-regulation of sialyltransferases is frequently found in cancer cells (Dall'Olio and Chiricolo, 2001; Kannagi *et al.*, 2004; Mahal *et al.*, 2001). Furthermore, increased sialylation also promotes adhesion, motility, and invasiveness *in vitro* and enhances tumor metastasis *in vivo* (Hedlund *et al.*, 2008; Schneider *et al.*, 2001). Moreover, the vast repertoire of sugar moieties and their linkages are likely to confer unique immunogenic profiles to cancer cells. In these studies, the depletion of sialic acid residues on the cell surface abrogated metastatic potential, while over-expression of the candidate sialyltransferase gene was associated with poor prognosis in malignant gliomas (Yamamoto *et al.*, 1997). These findings suggest that inhibition of sialyltransferases may be

useful for the treatment of brain tumors, which are notorious for their highly infiltrative nature.

4.3 *ST3GAL1* expression increases with tumor grade and portend poor prognostic in glioma patients

Next, we sought to elucidate the role of sialyltransferase ST3Gal1 in glioma progression, and identify patient groups likely to benefit from ST3Gal1 inhibitor treatment. We probed further into *ST3GAL1* mRNA expression level in REMBRANDT and Gravendeel databases (Gravendeel *et al.*, 2009; Madhavan *et al.*, 2009). Increased *ST3GAL1* expression was associated with higher-grade tumors (Fisher's test p -value<0.001) and poorer prognostic mesenchymal subtype patients (t-test p -value<0.001) in 5 major glioma patient databases (Gravendeel *et al.*, 2009; Madhavan *et al.*, 2009; Murat *et al.*, 2008; Noushmehr *et al.*, 2010; Phillips *et al.*, 2006), therefore confirming *ST3GAL1* as a prognostic indicator of glioma progression. Next, we implemented GSEA on stem-cell-related gene sets to identify whether the *ST3GAL1* transcription level was associated with stemness pathways (Subramanian *et al.*, 2005). We extracted 132 stemness signatures from MSigDB and looked for enrichments in patient groups with high, intermediate and low expression of *ST3GAL1* gene in the REMBRANDT and Gravendeel patient databases (**Supplementary Table 1**). We observed that 80% (N=107 gene sets) of stemness signatures were positively enriched with high *ST3GAL1* expression group. Gene set from Dazard *et al.*, comprising of stemness-related genes up-regulated upon response to UVB irradiation

(Dazard *et al.*, 2003), was the top-ranked pathway associated with *ST3GAL1*-high expression groups. Collectively, our results indicate that high *ST3GAL1* expression portends worst prognosis, and correlates with mesenchymal features.

4.4 Identification of *TGFβ* as upstream signaling pathway activating *ST3GAL1*

To determine the key upstream mechanism driving the aberrant *ST3GAL1* expression, we performed functional annotation clustering. Reactome pathway analysis evaluates pathway association on queried gene nodes and estimates the strength of mapped nodes against the pathway background (Yu and He, 2016). Briefly, we extracted the *ST3GAL1*-associated stemness gene nodes enriched in the signature by Dazard *et al.* (Dazard *et al.*, 2003) and Reactome pathway evaluation identified five key pathways, including chemokine receptor pathways (q -value=0.005), signal transduction (q -value<0.005), disease-associated (q -value=0.01), *TGFβ*-signaling (q -value=0.01) and *TGFβ*-signaling activation of SMADs (q -value=0.01) (**Supplementary Table 2**). We specifically relied on a disease-associated pathway cluster in which the enriched nodes comprised *ST3GAL1* along with *TGFB1*, *HBEGF*, *IGFBP6*, *ATF3*, *CYBA* and *ITPR1*. Interestingly, *TGFB1*, a key signaling molecule in mesenchymal GBM tumors, was among key (Lottaz *et al.*, 2010; Ng *et al.*, 2012; Phillips *et al.*, 2006; Verhaak *et al.*, 2010). We further identified the positive association of *ST3GAL1* and *TGFB1* in public glioma repositories (REMBRANDT all glioma_{cor}=0.31, p -value=2.29e⁻¹⁰ Gravendeel all glioma_{cor}=0.50, p -value=3.63e⁻²⁰). We also

observed a significant positive correlation of p-Smad2 (the downstream effector of TGF β) and ST3Gal1 in 19 GBM tumors via immunohistochemical staining ($\text{cor}=0.86$, $p=1.0\text{e}^{-04}$); (Chong *et al.*, 2016). To further understand the role of TGF β in *ST3GAL1* transcription, we identified p-Smad2 binding sites at -1274 to -1266 base pairs of the *ST3GAL1* region. In support of computational predictions, we showed that exogenous introduction of TGF β 1 resulted in activation of *ST3GAL1* transcription, which, in turn, was abrogated by the TGF β inhibitor SB431542. In summary, our patient-centric analysis and subsequent validation in GPCs confirmed that *TGFB1*, which typifies the mesenchymal subtype, binds to the *ST3GAL1* promoter and activates its transcription.

4.5 *ST3GAL1*-associated gene signature correlates with tumor invasion and glioma prognosis in glioma patients

To understand the contribution of *ST3GAL1* in glioma cell maintenance, we performed lentiviral-mediated *ST3GAL1* knockdown in two NNI GPC lines: NNI-11 and NNI-8. Interestingly, we noticed that *ST3GAL1* mitigation significantly affects the gliomasphere-forming ability and viability of GPCs *in vitro*. In addition, the mRNA expression of stemness markers such as *OLIG2*, *POU3F2*, *SALL2*, and *SOX2* was significantly reduced upon *ST3GAL1* knockdown (Chong *et al.*, 2016). Our functional experiments implicated *ST3GAL1* in inducing the self-renewal traits of GPCs. To understand the downstream regulators of the *ST3GAL1* pathway, we employed microarray profiling of non-targeting

(NT) control and *ST3GAL1* knockdown GPCs. The functional experiments for *ST3GAL1* characterization on the GPCs were performed by Dr Chong YK. A linear regression model was fitted to discern the differential gene list between knockdown cells and NT control cells. A gene list of 394 probesets was significantly modulated by *ST3GAL1* knockdown (adjusted p -value<0.05; **Supplementary Table 3**). The *ST3GAL1* knockdown gene signature was derived by further winnowing down the list with a 2-log fold change between knockdown clones and the control profiles (N=86 probesets). We arrived at the *ST3GAL1*-associated signature by merging the PNA signature and *ST3GAL1* knockdown signature (N=107 probesets; **Table 7**).

We interrogated the *ST3GAL1*-associated signature with glioma patient databases using a CMAP approach. Interestingly, a distinct pattern of association for the *ST3GAL1* knockdown signature was detected within glioma patient databases (Ceccarelli *et al.*, 2016; Gravendeel *et al.*, 2009; Madhavan *et al.*, 2009). The *ST3GAL1* signature stratified patient survival in the REMBRANDT (N=186), Gravendeel (N=156) and TCGA (N=252) databases. (**Figure-4.1A; Table 8 and, Supplementary Tables 4 and 5**). Patients with activated *ST3GAL1* signaling were associated with shorter survival (HR=2.47, 95% CI=1.72 to 3.55, REMBRANDT $p=1.92e^{-08}$; HR=2.89, 95% CI=1.94 to 4.30, Gravendeel $p=1.05e^{-11}$; HR=5.92, 95% CI=3.61 to 9.67, TCGA $p=2.55e^{-15}$). Importantly, *ST3GAL1*-high patients had significant enrichment with the poor prognostic mesenchymal and classical subtypes (**Figure-4.1B**, GI classification). In contrast, the proneural molecular subgroup, a

phenotype commonly associated with IDH mutations, mapped with *ST3GAL1*-low patient classes. The molecular subtype enrichment with *ST3GAL1* patient classes is summarized in **Table 9**.

Figure-4.1. *ST3GAL1*-associated gene signature stratifies survival in glioma patients. (A) The *ST3GAL1*-associated gene signature stratifies survival in (i) REMBRANDT (hazard ratio [HR]=2.47, 95% confidence interval [CI]=1.72 to 3.55, $p=1.92e^{-08}$); (ii) Gravendeel (HR=2.89, 95% CI=1.94 to 4.30, $p=1.05e^{-11}$); (iii) TCGA (HR=5.90, 95% CI=3.61 to 9.67, $p=2.55e^{-15}$) clinical glioma databases. Kaplan-Meier curves were drawn using R/survival package (Version 2.43.3) (Terry TM, 2000). (B) *ST3GAL1*-high/low patient cohorts enrich the glioma-intrinsic (GI) molecular classification scheme specified by proneural, classical, mesenchymal subtypes and WHO classifications [IDH wild-type (wt), IDH mutant with chromosomal *1p/19q* co-deletion (IDH-mut-CDL) and IDH mutant with chromosomal *1p/19q* non-codeletion (IDH-mut-NCDL)] in (i) REMBRANDT (tumor grades in the instances when WHO classes are not

available), (ii) Gravendeel and (iii) TCGA databases. The corresponding heatmaps of the *ST3GAL1*-associated gene signature candidates are shown. The heatmaps were generated using the R/pheatmap package (Version 1.0.12) (Kolde, 2019). The complete linkage metrics of Euclidean distance estimated from gene expression data was applied to define the hierarchical clustering of *ST3GAL1*-high/low patient classes.

Next, we assessed the prognostic contribution of the *ST3GAL1* signature, in addition to molecular subtypes and WHO status, using multivariate analysis, in 3 major patient databases. (Ceccarelli *et al.*, 2016; Gravendeel *et al.*, 2009; Madhavan *et al.*, 2009). The *ST3GAL1*-associated signature was an independent predictor of survival after adjusting for age, GI molecular subtypes and WHO classification scheme (grades were applied for the REMBRANDT database) in the Gravendeel and REMBRANDT databases. In TCGA, we observed that *ST3GAL1*-associated signature was an alternative predictor for survival in the multivariate model, whereas age and WHO classes were independent predictors for overall survival (**Table 8**).

In glioma patients, we demonstrated evidence for the association of *ST3GAL1* classes with T2-weighted MRI images, a proxy marker for weighing tumor cell invasion (McMillan *et al.*, 2009). We collected pre-surgical MR multi-sequence images from 130 REMBRANDT patients stored in Georgetown University's Georgetown Database of Cancer (GDOC) system (Clark *et al.*, 2013; Madhavan *et al.*, 2009). Next, we mapped the *ST3GAL1* phenotype identified from the CMAP approach with cellular invasion matrices analyzed from MRI images. Recently, T2-weighted signatures have been enriched with molecular subtypes,

reflecting the emerging interest in integrating molecular mechanisms with glioma management (Rathore *et al.*, 2018). In the current evaluation, *ST3GAL1*-high patients displayed larger T2 volume, correlating with a significant increase in tumor cell invasion (Pearson residuals p -value=1.46e⁻⁰⁴; **Supplementary Table 6**). Collectively, we show the *ST3GAL1*-phenotype association with radiologic results and emphasize the role and relevance of *ST3GAL1* in tumor cell invasion (**Figure-4.2A**).

As we have demonstrated above the prognostic value of the *ST3GAL1*-associated signature its contribution to tumor invasion, we sought to determine whether ST3Gal1 plays a role in brain tumor formation. We stereotactically implanted GPCs transduced with either NT control vector or 2 different clones of *ST3GAL1* knockdown vectors (C1, C2) (**Figure-4.2Bi**). We employed 2 GPC lines to evaluate the molecular mechanism associated with *ST3GAL1* abrogation. Notably, mice in the control vector group succumbed to tumor formation within 60 days post-implantation in both molecular phenotypes. In contrast, we observed that the median survival differences of mice implanted with *ST3GAL1* knockdown in both GPC lines was extended significantly (**Figure-4.2Bii**), suggesting a role of ST3Gal1 in gliomagenesis and mitigation of ST3Gal1 mechanism could potentially improve the survival of GBM patients. The *in vivo* validation experiments for *ST3GAL1* knockdown was performed by Dr Carol Tang and Dr Chong YK.

Figure-4.2. *ST3GAL1*-high status correlates with increased T2 volume and higher tumor grade in glioma patients; *ST3GAL1* knockdown prolongs the survival of mouse glioma model. (Ai) Magnetic resonance imaging (MRI) shows that tumor invasion (high/low) correlated with *ST3GAL1* status. (ii) The mosaic plot shows the existence of enriched correlation among *ST3GAL1* status, tumor grades, and tumor T2-zone. Dark purple boxes represent enriched interaction criteria; $p=1.46 \times 10^{-4}$. The mosaic plot was drawn on the molecular subtypes using R/vcdExtra package (Version 0.7.1) (Zeileis *et al.*, 2007) (Bi) The coronal sections of sh*ST3GAL1* and nontargeting clones demonstrate the tumorigenicity of NNI-11 GPCs reduced with *ST3GAL1* mitigation. The scale bar denotes 500 μ m. The survival plots of mice implanted (ii) NNI-11 and (iii) NNI-21 displays significant survival benefits upon *ST3GAL1* knockdown. NNI-11 (groups of 8 in 3 arms: non-targeting, C1, and C2) and NNI-21 (groups of 6 in 3 arms: non-targeting, C1, and C2) are shown. *** $p < 0.001$.

4.6 Identification of *FOXM1* as a downstream effector mediated by *ST3GAL1* signaling

To understand the downstream mechanism of *ST3GAL1* knockdown profiles, we interrogated the MetaCore pathway collection to look for *ST3GAL1* effector genes. The MetaCore pathway evaluation of *ST3GAL1* differential genes was mapped with cell-cycle-associated

mechanisms (90%) and progesterone-mediated maturation pathway (10%). Notably, we detected a significant enrichment of the FOX family of transcription factor *FOXM1*-associated pathways with the *ST3GAL1* signature (enrichment p -value= $2.25e^{-05}$). Among the downstream gene nodes enriched were *FOXM1*, *CDK1(p34)*, *cyclin B2*, *CDC25C*, *KNSL1*, and *p90RSK1*. Subsequently, we estimated the pairwise correlation for *ST3GAL1* mRNA expression with FOX family pathway nodes in patient databases. Of these, *FOXM1* had a significant correlation with *ST3GAL1* in both the REMBRANDT (p -value= $1.41e^{-52}$) and Gravendeel databases (p -value= $6.56e^{-22}$). Interestingly, the role of *FOXM1*-mediated transcription has been implicated as a pivotal modulator for the epithelial-to-mesenchymal transition (EMT) in solid malignancies (Wang *et al.*, 2015). Indeed, our *in vitro* experiments confirmed that *ST3GAL1* knockdown reduced FoxM1 expression and over-expression of exogenous *FOXM1* partially rescued gliomasphere formation, invasive capability, and viability of *ST3GAL1* knockdown GPCs. Collectively, our study demonstrated that *ST3GAL1* knockdown phenotypes are mediated by FoxM1. The concept map of bioinformatics analyses undertaken for the study is summarized in **Figure-4.3**.

Figure-4.3. The concept map of bioinformatics approaches implemented in evaluating *ST3GAL1*-mediated glioma invasion. PNA, peanut agglutinin; GPC, glioma-propagating cells; FDR, false discovery rate; MSigDB, molecular signatures database (concept map reproduced from (Chong *et al.*, 2016)).

4.7 Summary

We demonstrated the use of a bioinformatics approach to identify a novel *ST3GAL1*-mediated invasion pathway of GBM tumors by integrating patients' molecular databases with functional data from

patient-derived materials. We showed that the *ST3GAL1*-associated transcriptomic program portends poor prognosis in glioma patients and enriches for higher tumor grades of the mesenchymal molecular classification. We further identified TGF β , a signaling pathway that is often aberrantly activated in mesenchymal molecular subtype of glioma patients, directly triggers *ST3GAL1* transcription and activity. In addition, we molecularly profiled GPCs with perturbations of the gene and examined core activation programs. We provide evidence that ST3Gal1 regulates FoxM1, a key stem cell self-renewal and tumor-propagating gene. Collectively, our study provides biologically significant insight into the role of TGF β pathway-mediated ST3Gal1 regulation in promoting GPC self-renewal and tumorigenicity.

5 A GLIOMA-SPECIFIC, FUNCTIONALLY-TUNED *STAT3* TRANSCRIPTOMIC PROGRAM ADDRESSES THE MOLECULAR MECHANISM OF GLIOMA PATHOGENESIS AND CHEMORESISTANCE

In our ST3Gal1 work, we alluded to the role of the non-canonical TGF β activation in driving the mesenchymal phenotype of GPCs. The transcriptional role of TGF β has been shown to be pivotal in orchestrating GPCs for the transdifferentiation of highly invasive mesenchymal phenotype cells, depending on the tumor microenvironment (Joseph *et al.*, 2014). The PMT process requires specific phosphorylation cascades to be activated and intriguingly, oncogenic. JAK was recognized as a constitutive activator of TGF β receptors by its phosphorylation of Ser/Thr residues and subsequently, driving active SMADs and STATs transcriptional channels (Tang *et al.*, 2017). Most importantly, *STAT3* was identified as a final modulating switch for the radiation-induced PMT feature of GBM cells (Segerman *et al.*, 2016). Taken together, the transcriptional pathway *STAT3* and TGF β pathway-mediated *ST3GAL1* could potentially play key regulatory roles in contributing to the chemoresistant and invasive phenotypes of GBM tumors, and better understanding of the mechanisms of these pathways would allow for a clearer delineation of the molecular mechanisms of glioma patient subtypes.

5.1 Introduction and objectives

The clinical relevance of distinct molecular subtypes highlights key signaling pathways that contribute to glioma heterogeneity. In light of molecular evidence addressing the prognostic impact, the WHO revised the classification scheme for brain tumors to include molecular data (Louis *et al.*, 2016). To date, although TCGA and other consortia efforts have associated critical genomic aberrations with the heterogeneity of glial tumors, no functional role has been ascribed. More importantly, together with other groups, we have demonstrated that cultured patient-derived GPCs remain clinically relevant as they contain transcriptomic patterns and cytogenetic aberrations that mirrors the original tumors. These cells are capable of generating orthotopic tumors that phenocopies the original tumor morphology and transcriptomic profiles (Chong *et al.*, 2009; Foong *et al.*, 2011; Verhaak *et al.*, 2010). Our NNI Brain Tumor Resource thus provides an essential tool for precision medicine with the ability to prospectively reform tumor xenografts for testing computational predictions (with biological validation) while utilizing a retrospective clinical history for correlative analyses.

Recent findings from TCGA emphasized the need to select patient cohorts for specific targeting based on their molecular regulatory pathways. Moreover, current chemotherapeutic strategies are not always effective across all histologically identical brain tumor patients. This concept thus represents a paradigm shift in future diagnosis of brain tumor patients. Based on our earlier compound screening, we identified several small molecule inhibitors, including JAK/STAT pathway inhibitors,

to be effective in abrogating GPC growth and survival (Foong *et al.*, 2012). The JAK/STAT signaling pathway has been implicated in several malignancies of the blood and central nervous system (Hedvat *et al.*, 2009; Jamieson *et al.*, 2006), specifically in polycythemia vera, where the hematopoietic stem cells are characterized by a hyperactivating *JAK2* V617F mutation (Jamieson *et al.*, 2006). The use of JAK/STAT inhibitors effectively targets aberrant, self-renewing hematopoietic stem cells, subsequently leading to tumor growth cessation (Geron *et al.*, 2008; Levine and Gilliland, 2008; Morgan and Gilliland, 2008). Interestingly, our preliminary analysis indicated the absence of such mutation in patient glioma samples, therefore, other distinct molecular events leading to STAT3 upregulation in gliomas remain to be identified.

STAT3 is activated by phosphorylation of Tyr705, promoting cytosolic dimerization, nuclear translocation, and DNA binding (Darnell *et al.*, 1994). STAT3 activation is mediated by cytokines through the JAK family kinases (Schindler and Darnell, 1995), or via growth factor receptors such as c-MET and EGFR (Boccaccio *et al.*, 1998; Quesnelle *et al.*, 2007). Src family kinases have also been implicated in STAT3 activation (Bowman *et al.*, 2001). A growing body of evidence has demonstrated an essential role for autocrine and paracrine cytokine loops in driving the aberrant activation of STAT3 in human cancers. In particular, interleukin 6 (IL-6) signaling has been implicated in tumorigenesis via the JAK family kinases (Berishaj *et al.*, 2007; Gao *et al.*, 2007; Lam *et al.*, 2008). These observations provide a molecular basis for constitutive STAT3 activation in solid tumor types. Importantly, STAT3

activation has been associated with the PMT shift of GBM tumors, resulting in enhanced tumor cell invasion and, consequently, poorer prognosis (Segerman *et al.*, 2016). We therefore propose to address patient heterogeneity by developing a molecular stratification method that identifies patient cohorts amenable to JAK/STAT inhibition therapy. We will identify STAT3-associated prognostic markers that could better reflect patient chemoresistance profiles and utilize our preclinical mouse model to facilitate future drug testing.

5.2 Predictive evaluation of the *STAT3* functionally-tuned signature in patient databases

The preliminary evaluation of *STAT3* mRNA expression confirmed the prognostic significance and molecular heterogeneity across three major glioma patient databases REMBRANDT, Gravendeel and TCGA all-glioma archives (Ceccarelli *et al.*, 2016; Gravendeel *et al.*, 2009; Madhavan *et al.*, 2009). To establish the functional mechanism behind *STAT3*-driven tumorigenicity in GPCs, we evaluated the gene expression microarray data of lentiviral-mediated *STAT3* knockdown in three NNI GPCs. Microarray profiling of *STAT3* knockdown GPCs were carried out by Dr Melanie Tan. A linear regression model detected a total of 992 probesets that were significantly modulated (adjusted p -value<0.01) between *STAT3* knockdown GPCs and the non-targeting vector control (Ritchie *et al.*, 2015). We treated this differential list of genes as a GPC-specific *STAT3* knockdown signature. In parallel, we derived a *STAT3* co-expressed gene module from REMBRANDT glioma patient database.

The intersection of the *STAT3* co-expressed gene module with the *STAT3* knockdown signature as illustrated in **Figure-5.1**, generated our *STAT3* functionally-tuned signature (N=207). Specifically, we identified for genes from the co-expressed gene module which showed an inverse expression upon *STAT3* knockdown. We verified the functional evidence of *STAT3* mitigation by a reduction of viability and gliomasphere-forming potential. *STAT3* was the most significantly downregulated gene in the functionally-tuned signature (adjusted p -value=1.11e⁻⁰⁹), along with several critical *STAT3* downstream target genes (**Table 10**).

Figure-5.1. Identification of the *STAT3* functionally-tuned gene signature. *STAT3* co-expressed genes from patient database (middle panel) display inverse expression upon *STAT3* knockdown in 3 patient glioma cells (left panel).

MetaCore pathway enrichment analysis on our *STAT3* signature underscored vital signaling pathways such as IL-17-induced signaling, EGFR development pathway, and cell adhesion role of cyclin-dependent kinase (CDK5) (**Figure-5.2**). The core nodes affected in IL-17-induced signaling pathway were *STAT3*, p90RSK (*RPS6KA3*) and p38MAPK (*MAPK14*), indicating that MAPK14 and *RPS6KA3* were the downstream

nodes directly affected upon *STAT3* alteration. The second notable pathway was the *EGFR* development pathway, with the classical cohort and a portion of the mesenchymal patient cohort having a strong representation of *EGFR* gain (Eckel-Passow *et al.*, 2015). The knockdown signature nodes, including *MAPK14* and *STAT3* itself, were the major players in the *EGFR* development pathway. Our results suggest that *STAT3* inhibition could modulate *EGFR* regulation through MAP kinases, which warrants further validation in cellular experiments. Cell adhesion through the *CDK5* pathway has been shown to be a crucial mechanism regulating malignancies in the central nervous system (Zhu *et al.*, 2011). *STAT3* modulation affects alpha-catenin (*CTNNA1*) and integrin beta 1 (*ITGB1*) in the cellular adhesion pathway (Carrasco-Garcia *et al.*, 2018; Paul *et al.*, 2013). Our microarray results identified established critical downstream pathway genes of *STAT3* and also revealed novel gene sets upon functional knockdown in GPCs.

Figure-5.2. Top 10 MetaCore pathway maps enriched with *STAT3* knockdown transcriptomic signature. The X-axis refers to enrichment *p*-values.

5.2.1 Association of clinical phenotypes with the functionally-tuned *STAT3* signature

The *STAT3* functionally-tuned signature was interrogated with REMBRANDT, Gravendeel, and TCGA all-glioma patient databases (Gravendeel *et al.*, 2009; Madhavan *et al.*, 2009; Verhaak *et al.*, 2010). The modified CMAP approach of our *STAT3* signature, as explained in the methodology **Section 2.5.1**, could stratify 88% of REMBRANDT, 86% of Gravendeel (Kolmogorov-Smirnov enrichment test *p*-value<0.05) and

89% of TCGA patient cohorts (**Supplementary Tables 7 and 8**). Patients with a positive value of normalized activation score represents a similar association for the *STAT3* knockdown signature (*STAT3*-low) and the opposite for a negative score (*STAT3*-high). We applied this predictive computational approach successfully and demonstrated its capability in our earlier projects (Chong *et al.*, 2016; Koh *et al.*, 2013; Ng *et al.*, 2012; Yeo *et al.*, 2012). Gene set enrichment analysis (GSEA) evaluation confirmed that the *STAT3* gene signature was enriched in JAK/STAT and cytokine signaling modules archived in molecular signature databases (MSigDB, **Figure-5.3**) (Subramanian *et al.*, 2005). Our results suggest that the *STAT3* functionally-tuned signature has clinical relevance in patient databases and is consistent with established signaling pathways.

Figure-5.3. Gene set enrichment analysis (GSEA) on *STAT3*-stratified patient cohorts. GSEA reveals canonical JAK/STAT signaling modules enriched with *STAT3*-high patients from the (A) REMBRANDT and (B) Gravendeel databases. The GSEA plots were generated using GSEA Java desktop software application v2.2.2 from Broad Institute.

We sought to understand the molecular subtype mechanisms with *STAT3*-stratified patient cohorts. The poor prognostic classical and mesenchymal subtypes were significantly associated with *STAT3*-high patient cohorts in the REMBRANDT, Gravendeel and TCGA databases (**Table 11**). We observed that 92% of mesenchymal patients overlapped with *STAT3*-high tumor patients (**Figure-5.4A**), suggesting that *STAT3* plays a vital role in mediating the PMT, tumor necrosis and inflammatory infiltration mechanisms. The mechanism of *STAT3* has been implicated as a final modulating switch for radiation-induced PMT events, consequently impacting interleukin-associated necrotic events, as seen in the pathway enrichment analysis in the current study (Segerman *et al.*, 2016). In addition, the concurrent aberrations of *NF1* and *PTEN* association with *STAT3* activation might warrant additional investigations into the PMT mechanism in GBM tumors. Intriguingly, 77% of the classical subgroup patients were identified in the *STAT3*-high patient cohort. The classical cohort represents a unique map of substantial copy number changes in chromosome 7 gain and chromosome 10 loss. Half of the classical subgroup patients harboring the *EGFRvIII* mutation and its constitutive mechanism has been shown to maintain *STAT3* activity irrespective of EGF signaling (de la Iglesia *et al.*, 2009; de la Iglesia *et al.*, 2008). Recently, *STAT3* was shown to modulate a feed-forward mechanism through the oncostatin receptor (OSMR), which is responsible for *EGFRvIII*-mediated activation in GBM pathogenesis (Jahani-Asl *et al.*, 2016). Collectively, molecular subtype enrichment with

STAT3 score reveals critical pathway connections with *STAT3*-associated glioma progression.

Patients predicted for *STAT3*-high molecular program also had a significant association with IDH wt status (Gravendeel, 71.67%, p -value=2.30e⁻¹⁰; and TCGA, 94.71%, p -value<2.2e⁻¹⁶; Fisher's exact test). We evaluated whether the *STAT3* signature can be applied as a diagnostic measure to integrate with IDH mutation status, a WHO classification marker, using relative odds bases statistics as previously described (Wald *et al.*, 1999). We calculated the density distribution of CMAP activation score values for both IDH wt and mutant cohorts. The relative odds score was then estimated from the proportion of the highest fifth distribution and lowest fifth distribution of *STAT3* signature score values for IDH phenotypes. The estimated relative odds at a 5% false positive rate were 4.5, indicating that patients with a negative *STAT3* score (highest fifth) are nearly five times likely to be IDH wt than those with a positive *STAT3* score (in the lowest fifth). This finding emphasizes the clinical relevance of the *STAT3* score with the WHO diagnostic classification marker of IDH mutation status (**Figure-5.4B**).

Our patient database analysis revealed that patients with *STAT3*-high pathway activation were associated with poor prognosis, while a significant survival benefit was observed in the *STAT3*-low patient cohort (**Figure-5.4C**). We observed consistent prognostic impact in the Freije and Nutt glioma databases (data not shown) (Freije *et al.*, 2004; Nutt *et al.*, 2003). Adopting a multivariate Cox regression model, we demonstrate that the *STAT3* signature was an independent factor for predicting

survival variability in glioma patients after accounting for diagnostic and molecular predictors in glioma management (**Table 12**). In summary, we provide compelling evidence that the functionally-tuned *STAT3* signature was able to address the molecular heterogeneity in glioma tumors through *STAT3* mechanism *in vitro* and patient databases.

Figure-5.4. Association of *STAT3*-stratified patient cohorts with clinical phenotypes. (A) *STAT3*-high patient cohorts were enriched in mesenchymal and classical molecular subtypes, with IDH-wild-type status. *STAT3*-low tumors, in contrast, comprised mostly low-grade gliomas (LGGs), IDH-mutant (*1p19q* co-deleted and non-co-deleted) and proneural molecular subtypes. The annotation barplots were generated using the R/pheatmap package (Version 1.0.12) (Kolde, 2019). **(B)** The density distribution of IDH phenotypes shows a strong

association ($RO_{Q1-5}=4.5$) with *STAT3* signature scores. The R/ggplot2 package (Version 3.2.1) was used to draw the frequency distribution plot (Wickham, 2016). **(C)** An enrichment of *STAT3* pathway activation defined the poor prognosis patients (*STAT3*-high, 0.69 years) while patients of *STAT3*-low survived significantly longer (3.69 years). **(Di)** A combined model including Karnofsky (Karn) score with *STAT3* and WHO classes represents the best statistical model to account for the variability in patient survival using the Bayesian Information Criterion (BIC) method. **(Dii)** The NNI *STAT3* signature performs better than the existing 'Alvarez' *STAT3* signature for glioma patient prognosis (Alvarez *et al.*, 2007). Kaplan-Meier curves were drawn using R/survival package (Version 2.43.3) (Terry TM, 2000). The complete linkage metrics of Euclidean distance estimated from gene expression data was applied to define the hierarchical clustering of *ST3GAL1*-high/low patient classes.

5.2.2 *STAT3* signature model evaluation for goodness-of-fit

A goodness-of-fit evaluation estimates the quality of the model by addressing the variability with the inclusion of covariates. We leveraged on an information theory measure, called Bayesian Information Criterion (BIC) scoring, to identify the best representation of the statistical model addressing survival variability. We calculated the BIC score for the combination of clinical and molecular variables identified in our regression analysis. A stepwise covariate model was able to assess the covariate strength contributing to survival variability. The model combination was generated by adding covariates to the best univariate model. As shown in **Figure-5.4Di**, the WHO classification was the best univariate model addressing overall survival variability in patient databases. We found that a combination of the *STAT3*, WHO classes and Karnofsky score (patient's functional status) was the best statistical model accounting for variability in patient survival (**Figure-5.4Di**, shortest bar). The rule of thumb to detect

a superior model is the observation of a drop in the BIC score of ≥ 10 compared to the reference model (Kass and Raftery, 1995). Furthermore, we compared the performance of our NNI *STAT3* signature with that published by Alvarez et al., which was derived from *STAT3* co-expressed genes from various solid tumors (Alvarez *et al.*, 2007). Similarly, the NNI functionally-tuned *STAT3* gene signature out-performed the Alvarez signature in addressing prognostic variability (**Figure-5.4Dii**).

5.2.3 Integrative analysis of functionally-tuned *STAT3* signature classes with critical genomic aberrations identified in glioma biology

STAT3-high stratified patients had enriched somatic aberrations and mesenchymal molecular subtypes identified in TCGA evaluation (**Figure-5.5**). IDH mutations, including *IDH1* and *IDH2*, an important hallmark of the WHO classification scheme, were enriched in 93.7% (N=164) of *STAT3*-low patient classes (p -value $< 2.2 \times 10^{-16}$; **Figure-5.5**). IDH mutant tumors are deprived of NADPH production, which leads to ROS production affecting *STAT3* phosphorylation (Baldewpersad Tewarie *et al.*, 2013). This ROS-based mechanism is mediated through the ERK-MAPK pathway, which may contribute to the role of IDH mutation in *STAT3*-low patients (Souza *et al.*, 2009). In the clinic, glial tumors with *1p/19q* CDLs indicate that such patients are amenable to conventional procarbazine, lomustine, vincristine (PCV) therapy, compared to patients lacking the co-deletion status (Cairncross *et al.*, 1998). The integrative analysis on mapping the co-deletion parameter with *STAT3* stratification indicated that all *STAT3*-high patients lacked *1p/19q* CDL. Our data

highlight that this cohort of patients lacking CDL status can now potentially benefit from STAT3 inhibition therapy. The GBM-specific concomitant aberrations in chromosomal 7 gain and 10 loss were highly prevalent in *STAT3*-high cohorts, which is generally associated with EGFR-mediated classical subtype enrichment in *STAT3*-high patients (Verhaak *et al.*, 2010).

Pan-glioma analysis identified six methylation subtypes and four gene expression subtypes from 932 DNA methylation profiles and 667 RNA sequencing datasets (Ceccarelli *et al.*, 2016). The *STAT3*-high patients were significantly mapped with IDH wt-associated methylation subtypes (LGm4, LGm5, and LGm6; 98%, 97% and 67%, respectively). Thus, the integrative evaluation with methylation events suggested that genome-wide DNA hypomethylation might be associated with active STAT3 signals driving glioma pathogenesis. Intriguingly, the GBM-like protein cluster (K1) had significant enrichment in the *STAT3*-high patients, who also exhibited elevated protein activities of IGFBP2, pEGFR, pAKT, cyclins, HSP70, and PKC. The active protein clusters associated with the STAT3 mechanism warrant functional validation to understand how the protein machinery modifies the STAT3 pathway in glioma patients (**Figure-5.5**). Collectively, our data underscore that the diagnostic and treatment characteristics are well correlated with the STAT3 pathway, which represents the prospective application of the *STAT3* signature in the clinical management of glioma tumors.

Figure-5.5 Enrichment of molecular aberrations and subtypes in *STAT3* classes in TCGA. The key hallmark genomic aberrations showed significant enrichment with *STAT3*-stratified cohorts. The *STAT3*-high patients had significant overlap with IDH wt, non-codeletion (NCDL) status for chromosomal 1p/19q, chromosomal 7 gain with 10 loss (chr7G10L), co-gains of telomere shortening status, GI mesenchymal (MES) and classical (CL) subtype, pan-expression subtype (LGr4), pan-methylation subtypes (LGm4, LGm5 and LGm6) and GBM-like protein subtypes (K1). Fisher's exact test p -value<0.0001 for all aberrations is shown. Wt, wild-type; mut, mutant; meth, methylation; No-meth, no methylation; WHO, World Health Organization; RPPA, reverse phase protein array. The oncogram plot was generated using the R/pheatmap package (Version 1.0.12) (Kolde, 2019)

5.3 Biological validation of the *STAT3* functionally-tuned signature

A gene signature is an enriched genomic representation of biological phenotype (Abu-Farha *et al.*, 2008). It can be interrogated with molecular databases to discern signature similarities of a transcriptomic query profile using pattern-matching algorithms (Dabney, 2006; Hoshida, 2010; Lamb *et al.*, 2006; Tibshirani *et al.*, 2002). This evaluation is an

essential element in clinical database analyses to identify the role of molecular mechanisms for prognosis and chemoresistance.

5.3.1 *STAT3* signature stratification of GPCs

Dr Melanie Tan performed GPCs culture experiments and drug treatment profiling of STAT inhibitors. We interrogated the *STAT3* functionally-tuned gene signature in our NNI tumor resource, for which we created reference molecular profiling of our NNI GPCs, xenografts and matched primary tumors. The predictive analysis on our first batch of NNI GPCs (N=11) were performed with hierarchical clustering of *STAT3* signature and identified two distinct cluster patterns. Immunoblot analysis of pSTAT3 levels confirmed that the two clustered GPCs had differential expression patterns (**Figure-5.6A**). Specifically, the pSTAT3/STAT3 protein level expressed a distinct pattern as aligned to the predicted high or low groups of *STAT3* expression. The interrogation of the *STAT3* signature on the expanded GPC cohort (N=42, inclusive of replicates) was executed using a shrunken centroid algorithm with cross-validation metrics (Tibshirani *et al.*, 2002).

To obtain the training dataset for the shrunken centroid approach, we used a connectivity map (CMAP) method to delineate our GPCs while using the *STAT3* signature as a reference (Lamb *et al.*, 2006). The CMAP categorized ~30% of the sampling cohort as ambiguous cases as the activation scores were distributed randomly across the sample space of two classes. The *STAT3* signature pattern of predicted classes from the connectivity map was then trained in the shrunken centroid algorithm, and

the trained model could better predict *STAT3* classes for the new GPCs of interest (**Figure-5.6B**).

We verified the differences in chemosensitivity between the *STAT3*-high and *STAT3*-low GPC cohorts using biochemical assays to validate our predictions (**Figure-5.6C-E**). The pharmacological inhibition of *STAT3* for *in vitro* assays was assessed using three *STAT3* small molecule inhibitors: AZD1480, Stattic, and WP1066 (Hedvat *et al.*, 2009; Iwamaru *et al.*, 2006; Stechishin *et al.*, 2013). Consistent with our computational predictions, the stratified GPCs displayed significant variability with regards to the self-renewal capability and viability of cells (**Figure-5.6C**). *STAT3*-high GPCs showed a greater response associated with reduced glioma-forming ability and cell viability than the *STAT3*-low GPCs (**Figure-5.6D**). Of note, the half-maximal inhibitory concentration (IC_{50}) experimental profiles showed that *STAT3*-high GPCs exhibited increased sensitivity compared to that observed in the *STAT3*-low GPCs for all three *STAT3* inhibitors tested (**Figure-5.6E**). Collectively, the *in vitro* assays confirmed that the *STAT3* functionally-tuned signature was able to capture the molecular heterogeneity attributed to a *STAT3* mechanism affecting the viability of GPCs.

Figure-5.6. STAT3 functionally-tuned signature stratifies NNI patient cells based on their STAT3 status and demonstrates differential responses to known STAT3 inhibitors. (A) Immunoblot evaluation of phospho-STAT3 levels displays variability in stratified cell lines. STAT3-high cell lines (NNI-21, NNI-24, NNI-12) showed elevated phospho-STAT3 compared to STAT3-low cell lines (NNI-11, NNI-20, NNI-23). **(B)** Box-whisker plot showing the distribution of STAT3 mRNA expression across NNI patient cells. **(C)** Pharmacological inhibition with STAT3 inhibitors **(i)** AZD1480; **(ii)** Stattic; **(iii)** WP1066 affects dose-dependent viability and **(D)** gliomaspHERE-forming ability in STAT3-high cell lines (NNI-21, NNI-24) and STAT3-low cell lines (NNI-20, NNI-23). ***, $p < 0.001$ versus DMSO. **(E)** Seven glioma patient cell lines were treated with **(i)** AZD1480; **(ii)** Stattic; and **(iii)** WP1066, and their corresponding IC_{50} values were determined.

Consistent with our bioinformatics data prediction, *STAT3*-high cell lines that included NNI-12, NNI-21, NNI-24 were sensitive to *STAT3* inhibition therapy, while NNI-4, NNI-11, NNI-20 and NNI-23, identified as *STAT3*-low cell lines, were chemoresistant.

5.3.2 *STAT3* signature stratification in orthotopic animal models

Dr Melanie Tan and Dr Carol Tang conducted animal experiment studies for *in vivo* validation. We established a preclinical mouse model to provide evidence for the mechanism of *STAT3* small molecule inhibitors. From our previous study and TCGA efforts, we recognized that the orthotopic mouse models from patient-derived material (PDx models) recreated the original patient tumor pathophysiology and retained the molecular portrait of original tumors (Chong *et al.*, 2009; Koh *et al.*, 2013; Ng *et al.*, 2012). In mice implanted with *STAT3*-high NNI-24 GPCs, AZD1480 treatment resulted in a dose-dependent extension of survival and prolonged tumor latency (**Figure-5.7A**). In contrast, mice bearing *STAT3*-low NNI-20 GPCs did not receive significant treatment benefit from AZD1480 (**Figure-5.7B**). In summary, we provide substantial evidence that the preclinical mouse model recapitulates the bioinformatics predictions of patient-specific responses based on their *STAT3* pathway activation defined by our gene signature.

Figure-5.7. The *STAT3*-high stratified patient xenograft mouse model demonstrates a favorable response to AZD1480 treatment. The *STAT3* functionally-tuned gene signature was applied to stratify the patient xenograft model: *STAT3*-high (NNI-24) and *STAT3*-low (NNI-20) **(A)** NNI-24 xenograft model displayed significant prolonged survival and tumor latency upon treatment with AZD1480 compared to **(B)** NNI-20 xenograft model, which received no survival benefit from AZD1480 treatment, consistent with their status as *STAT3*-high and *STAT3*-low, respectively.

5.4 Maintenance of *STAT3* signature status in recurrent GBM patients

Tumor recurrence is a considerable obstacle in the therapeutic management of glioma patients. Recurrent GBM (rGBM) tumors undergo significant genomic changes with regards to spatiotemporal evolution (Johnson *et al.*, 2014; Kim *et al.*, 2015; Mazor *et al.*, 2016). Recent literature has emphasized the role of *STAT3* in molecular subtype transition, particularly revealing *STAT3* activation as the prime mechanism for mesenchymal transition (Ozawa *et al.*, 2014). These findings prompted us to explore the utility of our *STAT3* functionally-tuned gene signature's contribution for subtype switching in rGBM patients. In confirmation with our earlier BIC analysis (**Figure-5.4D(i)**), the *STAT3* signature was a critical element for understanding the prognostic impact

of glioma patients. Therefore, we interrogated the *STAT3* signature with a matched recurrent patient tumor database from Samsung Medical Center and Genentech (Joo *et al.*, 2013; Phillips *et al.*, 2006).

The *STAT3* signature analysis in the matched recurrent tumor database revealed that the *STAT3*-high cohort regardless of molecular subtype, can undergo switching to either *STAT3*-high or *STAT3*-low classes upon recurrence. To select the responsive cohort for *STAT3* inhibition therapy, the subjects must be validated for *STAT3*-high status and mesenchymal subtype through serial molecular profiling. Interestingly, mesenchymal *STAT3*-high tumor tends to maintain its *STAT3* status upon recurrence (86%) (yellow highlight in **Table 13**). A subset of patients belonging to the mesenchymal subtype with *STAT3*-low profile at first presentation had switched to *STAT3*-high status when the tumor recurred. In contrast, the non-mesenchymal patients exhibited random switching of their *STAT3* status upon recurrence. The log-linear model identified a strong precedence of mesenchymal and *STAT3* status association that contributes to tumor recurrence (log-linear p -value=0.018).

5.5 Differential protein regulatory network from Bayesian inference in *STAT3*-stratified patient cohorts

STAT3 signaling is indispensable for cellular proliferation and self-renewal of GBM-stem cells (Sherry *et al.*, 2009). To explore *STAT3* activation status in the stratified patient cohort, we leveraged on protein expression array information from 138 cell cycle associated proteins

extracted from TCGA reverse phase protein array (RPPA) resource for LGG and GBM samples. We aimed to: (i) identify protein network in *STAT3*-stratified patients; (ii) assess the interconnections of the identified proteins. First, we applied a recursive feature elimination (RFE) algorithm to determine the key protein expression features delineate *STAT3*-stratified patient cohorts.

RFE algorithm determines the protein features by evaluating the magnitude of decrease in classification error with the exclusion of each protein feature. This process will be executed iteratively until the model reaches the least classification error. Random forest nodes are constructed at each iteration process with 2/3rd of dataset for training the classifier to predict in the testing dataset (out-of-bag samples; OOB). The OOB error rate was the mean square error of predictor variables arrived from the trees generated in the prediction process. The OOB error rate and percentage of variance accounted by the predictors are the significant factors to determine the performance of random forest (Bhat *et al.*, 2013; Darst *et al.*, 2018). The RFE classification methodology on RPPA data was implemented using R/caret (Version 6.0.84) and R/mlbench packages (Version 2.1.1) (Friedrich Leisch, 2010). We first used 'rfeControl' function to create control object including feature nodes for random forest approach. The recursive elimination method was performed using 'rfe' function to determine the top protein features. This evaluation identified 37 protein features with an out-of-bag (OOB) error rate of 6%. A subset of deterministic protein features included EGFR, ERBB, MAPK8/JNK, IGFBP2, PTEN, CDKNB, and CCNB1, belonging to

the critical nodes of three major GBM signaling pathways (Verhaak *et al.*, 2010).

We applied a Bayesian probabilistic model to the top 25 proteins featured. The Bayesian regulatory network construction in R/BNArray package (Version 1.0.0) uses the following main functions available in R/deal package (Version 1.2.39) (Bottcher and Dethlefsen, 2018). The 'network' function in R/deal package was used to learn parameter priors from the trained protein expression network. Subsequently, the joint prior distribution was calculated using 'jointprior' function. The 'getnetwork' function connected the regulatory relationship of protein nodes using the estimated joint prior probabilities from previous steps. The final regulatory network with highest Bayes score was built by 'autosearchEx' function using greedy search algorithm with random restarts to identify the high confidence direction of edges. The workflow for the Bayesian regulatory network construction was adapted from the published protocols by Chen *et al.*, (Chen *et al.*, 2006).

The Bayesian analysis revealed that the regulatory connection was variable between *STAT3*-high and *STAT3*-low patient cohorts. The gross pattern analysis showed that proteins identified in *STAT3*-high had greater dependencies (connections) among themselves than in *STAT3*-low classes. *STAT5A* showed higher probability in influencing multiple protein nodes, including *BAD*, *PGR*, *ERBB2*, *EIF4G1*, *ANXA1*, and *RPS6P2* in *STAT3*-high patient class, while *ANXA1* was the only protein regulation dependent on *STAT5A* in the *STAT3*-low cohort. In addition, *EGFR* acted as a parent node along with *CCNB1*, *PGR*, *CDKN1B*, and

SERPINE1 to activate IGFBP2 in *STAT3*-high patients (**Figure-5.8A**). In contrast, *STAT3*-low patients had alternate triggering nodes of MAPK8, CCNB1, and SERPINE1 to modulate IGFBP2 as shown in **Figure-5.8B**. Furthermore, PTEN and NF2 had greater dependencies in regulating KIT and RPS6P2. Unsupervised clustering of 196 protein antibodies revealed GBM-like and LGG-like clusters in the protein array effort by TCGA (Ceccarelli *et al.*, 2016). We observed that *STAT3*-high patients have active enrichment with IGFBP2, EGFR, CCNB1, PTEN, and annexin, resembling GBM-like cluster. In summary, the protein array analyses on patient databases predicted protein regulatory network in the *STAT3*-stratified patient cohorts using a probabilistic model.

Figure-5.8. Differential protein regulatory network in *STAT3*-stratified classes. Bayesian analysis of the top 25 protein classifiers evaluating the feature dependence in both *STAT3* patient cohorts. **(A)** EGFR was designated the top parent node establishing differential connections with CDKNs, leading to the regulation of IGFBP2 in the *STAT3*-high cohort. **(B)** In contrast, PTEN-dependent nodes were active in *STAT3*-low patients. The BNArray package (Version 1.0.0) was deployed to construct gene regulatory networks from protein array data (Chen *et al.*, 2006).

5.6 STAT3 chemoresistance pattern evaluation in glioma patient databases

Our integrative analyses using gene expression and protein arrays highlight the relevance of employing such bioinformatics analyses to study the complex network contributing to chemoresistance due chiefly to the heterogeneous nature of GBM. Similarly, we showed that a subset of GPCs displayed resistance to STAT3 inhibition (**Figure-5.6C-E**). We aimed to elucidate the mechanism by evaluating whether any biochemically active kinase regulatory is mediating the resistance upon STAT3 inhibition in GPCs.

There are over 500 protein kinase genes which constitutes ~2% of the human genome. Alteration in function of these protein kinases may cause pathological changes within the cells giving rise to cancer. More than 200 kinase inhibitors have been tested or in testing in over 2,500 clinical trials. In 2011, more than 300 clinical trials were started, with 19 in phase III, expanding upon the 14 approved kinase inhibitors. Receptor tyrosine kinase (RTK)/PI3K, JAK, NF- κ B and Notch are key kinase pathways addressed in large patient cohort studies for treatment in GBM patients (Joo *et al.*, 2012; Kauffmann-Zeh *et al.*, 1997; Sharma *et al.*, 2010; Sibilio *et al.*, 2000). However, strategies to overcome drug-induced resistance have not been precisely addressed. We hence applied genome-wide kinome profiling to AZD1480-treated GPCs to understand the responsive and resistance mechanisms developed upon STAT3 inhibition therapy.

5.6.1 High-throughput kinome assay evaluation in AZD1480-treated tumors

A/Prof Tan Nguan Soon from NTU conducted the kinome profiling of AZD1480 treated GPCs. We performed activity-based kinase assay profiling on two distinct cohorts of our GPCs stratified by the functionally-tuned *STAT3* signature. We employed a three-dimensional peptide array system, PamChip (PamGene technology), which allows us to characterize the phosphorylated peptide dynamics of protein tyrosine kinase (PTKs) and serine/threonine kinase (STKs). The high throughput technology allows sample lysate to pass through the porous membrane coded with kinase substrates, in which the substrate phosphorylation status by active kinases can be subsequently quantified.

We developed a novel computational pipeline by integrating multimodal data structures from phosphorylation and transcription factor databases (**Figure-2.1**). The details of the analytical protocols are described in the methods section. In brief, we cataloged kinase-substrate information from Kinexus, PhosphoSite, Reactome and human protein reference databases (Hornbeck *et al.*, 2012; Keshava Prasad *et al.*, 2009; Milacic *et al.*, 2012). Each substrate from the PamChip measurements were parsed through the merged kinase-substrate database. The combined experimental data with the kinase-substrate matrix were interrogated with a summarizing algorithm, which estimates the pairwise correlation for each kinase-substrate cluster. A quantitative summary of the kinase-substrate cluster was calculated for the weighted-rank peptide cluster. The peptide cluster with strong kinase activity and dynamic range

received a high weightage score. The unique kinase-substrate matrix was assessed using a linear regression model to identify differential kinases between AZD1480- and DMSO-treated GPCs. A p -value of less than 0.1 was considered statistically significant to determine the kinases altered by STAT3 inhibitor treatment. Subsequently, we mapped the STAT3 inhibitor-altered kinase profiles with *STAT3* knockdown transcriptomic profiles using two gene-specific databases (Marbach *et al.*, 2016; Olow *et al.*, 2016).

5.6.2 Altered kinase mechanism in AZD1480 treated GPCs

Kinome evaluation determined the sensitive and resistant kinases affected by AZD1480 treatment. The current analysis evaluated the kinases aberrantly phosphorylated at tyrosine residues and serine/threonine residues. When the GPCs were exposed to AZD1480 treatment, AZD1480's prime targets JAK2/JAK3 activity were invariably affected, implied that the analytical approach identified the true-positive hits (**Table 14**). Importantly, leukocyte tyrosine kinase (LTK) was significantly inhibited by AZD1480 treatment in our analysis, which was known as essential component of prosurvival pathway signature in GBM patients (Freije *et al.*, 2004). Consistent with our studies, mutagenic plasmid library evaluating AZD1480 targets uncovered ALK/LTK, FGFR, RET and TRK based on the specific inhibitory mechanism (Gudernova *et al.*, 2017). Besides, angiogenic kinase receptor KDR was the second primary kinase module abrogated by AZD1480 treatment. The role of KDR in GBM proliferation is well established in glioma biology. However, the association of AZD1480 inhibition resulting in a substantial reduction

of angiogenic KDR has previously been identified (Joensuu *et al.*, 2005; Steiner *et al.*, 2004). In addition, our high-throughput kinome profiling underscored EphA3, Fes, and IRR as those critical responsive kinases for AZD1480 treatment. Of which, EphA3 and Fes were characterized to enhance GBM proliferation in previous studies and thus, supports to reposition AZD1480 targeting additional pathway insights (Day *et al.*, 2013; Zinn *et al.*, 2015). On the contrary, we observed a surge of critical kinase pathways including interleukin-2-inducible T-cell kinase (ITK), dual specificity tyrosine-phosphorylation-regulated kinase 1B (DYRK1B), SRC-related kinase (Srm), Kit and EGFR activation. The AZD1480-mediated inhibition displayed significant survival improvement in EGFR-amplified xenograft models, indicating that activated EGFR pathway required for JAK/STAT inhibition (McFarland *et al.*, 2011). Both of our GPCs used in the current evaluation had normal copies of EGFR, which addresses the rationale for AZD1480-mediated EGFR activation in the present analysis. Other resistance kinases listed in the present evaluation are novel and the functional mechanism for the AZD1480-linked resistance pathway is yet to be explored.

The enzymatic activity on the OH group of serine or threonine residues of substrates attributes for 25% of kinase regulatory mechanism. The current PamChip technology performs two antibody reactions for increased specificity to precisely determine STK activity in experimental conditions. The present kinome profiling measured the phosphorylation kinetics for 144 STKs. Applying the analytical pipeline on peptide measurement, we identified AZD1480 treatment modulated STKs in

GPCs. STAT3 pathway inhibition through AZD1480 affected significant kinase clusters including JNKs, Aurora, CDKs and ICKs (**Table 15**).

We did not observe distinct mechanism for the active STKs in *STAT3*-high GPCs. However, *STAT3*-low GPCs showed distinct responsive kinases include AurA, AurB, KSR, Nek6, LRRK2, and CDK9. The list of sensitive and resistant kinases is illustrated as a graphical representation in **Figure-5.9**. Several of the identified STKs were in agreement with AZD1480 mechanism reported by others (Couto *et al.*, 2012; Derenzini *et al.*, 2011; Zhang *et al.*, 2015).

Next, we sought to understand the PTK mechanism of AZD1480 in *STAT3*-signature stratified cohorts of NNI GPCs. We fitted two independent linear models to understand the AZD1480 altered kinases in *STAT3*-low and -high GPCs. Interestingly, we detected 67.7% of modulated kinases decreased in *STAT3*-high GPCs, while 27% of kinases were inhibited in *STAT3*-low GPCs. Among the critical responsive kinases identified in *STAT3*-high GPCs were ERK, RAF, Ror, Ret, EphA4/10, and Syk. Contrary to the observation, ERK, RAF1, Ret and EphA10 kinases uniquely displayed significant resistance in *STAT3*-low GPCs (**Figure-5.9**, right panel). AZD1480-mediated resistance in the ERK pathway through a negative-feedback loop was previously addressed in Hodgkin lymphoma studies (Derenzini *et al.*, 2011). The pathways including RAF and Ret were shown to be mitigated by AZD1480 treatment in thyroid cell lines (Couto *et al.*, 2012; Yang *et al.*, 2013). The primary scope of our kinome profiling evaluation in the current *STAT3* work is to identify the chemoresistance modulation by *STAT3* inhibition

therapy. Interestingly, we identified insulin-like growth factor 1 receptor (IGF1-R), the topmost kinase elevated exclusively in *STAT3*-low GPCs. The RPPA evaluation on TCGA patient databases was also in supportive of IGFBP2, a precursor pathway for IGF1-R, which demonstrated altered variation in *STAT3*-low and –high patients. The interaction between the C-terminal domain of IGF1-R and *STAT3* has been known as a critical function for *STAT* mechanism (Zhang *et al.*, 2006). We further proceeded to perform dual inhibition of both *STAT3* and IGF1-R pathway to overcome IGF1-R mediated resistance in *STAT3*-low animal models. Our orthotopic animal model experiments confirmed that dual inhibition of *STAT3* and IGF1-R had significantly prolonged the survival and extended tumor latency.

Figure-5.9. Identification of sensitive and resistant protein tyrosine kinases and serine/threonine kinases upon AZD1480 treatment in GPCs. Biochemical modification module identifies significantly modulated kinases in *STAT3*-stratified cells (*STAT3*-high/low).

We further mapped the sensitive and resistance kinases with *STAT3* knockdown transcriptome pathway gene nodes using kinase-substrate-transcription factor databases (Marbach *et al.*, 2016; Olow *et al.*, 2016). The current result is crucial to characterize how the downstream transcriptome pathways are regulated by kinase channel, upon treatment with AZD1480. Functional modification of kinome profiling analysis revealed that AZD1480 specifically inhibited JAK2 kinase, which affected a transcription factor cluster that includes GTF2I, STAT5A, and STAT3 (**Figure-5.10A**, right panel). The downstream target genes mapped with *STAT3* knockdown pathway were *GTF2I* (N=161), *STAT5A* (N=95) and *STAT3* (N=35). Intriguingly, GPC's resistance to AZD1480 treatment were contributed by EGFR and Kit, with both kinases activating the STAT1 transcription factor, triggering the up-regulation of 12 genes from the *STAT3* knockdown transcriptome (**Figure-5.10A**, left panel).

By integrating with *STAT3* knockdown gene expression data, we deciphered the critical STK pathway and the substrate transcription factors affecting *STAT3* mechanism (**Figure-5.10B**). *STAT3* inhibition using AZD1480 affected the phosphorylation cascades of AKTs, AURKC, ATR, CDK and MAPKs, which subsequently mitigate the recruitment of 99 transcription factors that deactivate the transcription process for *STAT3* downstream targets. On the other hand, a cluster of kinases developed resistance, which included CAMK2A, PAK1, PRKAs, RPS6KA1 and SGK1 activating 37 transcription factors and up-regulatory cascades of *STAT3* knockdown transcriptome.

A Modulation of PTKs and TFs in AZD1480-treated GBM cells

B Modulation of STKs and TFs in AZD1480-treated GBM cells

Figure-5.10. Integrative analysis of kinome profiles with transcriptomic data using functional module of kinome pipeline. (A) Functional modification module integrates the tyrosine kinases and (B) serine/threonine kinases with *STAT3* knockdown transcriptome to identify the modulation of kinases upon AZD1480 treatment.

5.7 Summary

Collectively, our study provides firm basis for the utility of *STAT3* inhibitors for GBM treatment. Using machine learning models, we established a *STAT3* functionally-tuned gene signature. The *STAT3* gene signature was derived from linear regression model that accounts for the intra-block correlation of cellular clones within the regression model to estimate the differential gene expression features (Ritchie *et al.*, 2015). The variance between gene expression measurements were adjusted with global variance using empirical Bayes model to reduce the number of false positives in differentially expressed gene signatures.

The rank-based CMAP prediction approach was able to precisely delineate the *STAT3* signature classes in patient tumor databases with high statistical confidence. The CMAP approach has unique advantage to control potential false connections with the use of randomization or permutation-based approach at both signatures and transcriptomic profile of patient/treatment data (Musa *et al.*, 2017; Zhang and Gant, 2009). The agglomerative clustering using complete linkage of Euclidean distance matrices of gene signature pattern was applied to see the *STAT3* signature classes in patient tumor databases. We showed that the *STAT3* signature had prognostic relevance and patients with active *STAT3* signature patterns were enriched for mesenchymal and classical subtype, typically associated with poor prognosis. BIC scoring approach was able to identify the top ranking Cox regression model including WHO subgroup information, Karnofsky score and *STAT3* signature classes that strongly addressed the prognostic variability of patient tumors.

Using cellular and animal models, we demonstrated that *STAT3*-high GBM tumors showed a favorable response to AZD1480, while *STAT3*-low tumors developed chemoresistance. The construction of protein regulatory networks was performed using Bayesian modeling approaches by estimating joint probability distributions of key protein features identified from random forest-recursive feature elimination (RF-RFE) model. As kinases represent an attractive therapeutic target in major pharmaceutical pipelines, we established an approach of using computational predictions to measure the phosphorylation levels of 144 kinases in *STAT3* signature-stratified GBM cells treated with *STAT3* inhibitor using the PamChip technology. We developed a novel computational pipeline for kinome assay data by integrating phospho-chemical interactions with functional genomic data through kinase-substrate databases. Our approach identified IGF1-R as a top-ranking tyrosine kinase uniquely elevated in *STAT3*-low tumors upon treatment with AZD1480. These results suggest a potential therapeutic strategy utilizing a dual inhibitor approach to sensitize the *STAT3*-low GBM patient subgroup. A critical element of precision medicine studies does not solely focus on *STAT3* responders alone (*STAT3*-high), but also evaluates therapeutic options to bring about a curative outcome for non-responders (*STAT3*-low).

Summary of machine learning approaches applied in this study

The current thesis applies the following pattern learning approaches to understand the molecular profiles of primary tumors and patient-derived materials. Agglomerative hierarchical clustering is the most common method to recognize the subtype patterns of gene expression and DNA methylation data (Ceccarelli *et al.*, 2016; Noushmehr *et al.*, 2010; Verhaak *et al.*, 2010; Wang *et al.*, 2018). The pathway signatures were interrogated to learn the molecularly distinguishable sub-groups of NNI tumors and public patient tumor databases using hierarchical clustering method (**Figure-3.2B**, **Figure-3.3**, **Figure-3.6**, **Figure-4.1** and **Figure-5.1**). The complete linkage metrics of Euclidean distance estimated from gene expression data was applied to determine the hierarchical clustering of molecular subtypes. The most frequently applied TCGA molecular subtypes signature pattern was first trained to learn the centroids of proneural, classical, neural and mesenchymal subtypes. Subsequently, the trained model with low classification error was interrogated to define the subtypes of new patient data using classification of nearest centroid algorithm (**Figure-3.7**) (Dabney, 2006). Similarly, we adapted shrunken centroid approach to train the pattern of gene features for classifying Phillips molecular subtypes for assigning the subtypes of *de novo* tumor data (**Figure-3.7**) (Tibshirani *et al.*, 2002).

The current work presents the application of supervised learning method to determine the classifiers of molecular phenotypes such as

ST3GAL1 and *STAT3* signatures. For this purpose, we used specific linear regression model, which has been effectively used to identify differential pattern of gene features from high-throughput expression data (**Figure-3.5**, **Figure-4.1**, **Figure-5.9** and **Figure-5.10**). The advantage of this linear regression function is to learn the feature information through empirical Bayes approaches, which estimate posterior variance from the underlying data to define the biological or technical heterogeneity (Ritchie *et al.*, 2015). In addition, the protein array data was learnt using recursive feature elimination (RFE) approach to identify the features that could accurately classify the *STAT3*-specific patient cohorts using iterative learning processes (**Figure-5.8**).

6 GENERAL DISCUSSION

6.1 Discussion

Preamble

Glioblastoma was the first cancer to be profiled by TCGA (The Cancer Genome Atlas Research *et al.*, 2008; Verhaak *et al.*, 2010). Major findings concluded that gene expression dictates tumor progression and disease outcome. These findings suggest a major shift in treatment paradigm. Physicians must now be informed of the molecular patterns of GBM tumors prior to implementation of therapeutic strategies. Indeed, the 2016 revised WHO classification scheme incorporates several molecular parameters in addition to current clinical indicators. Although such molecular parameters currently have no influence on treatment decision, their inclusion warrants deeper investigation into the design and conduct of clinical trials, where targeted therapy is expected to bring about more long-lasting and curative outcome (Louis *et al.*, 2016). The NNI Brain Tumor Resource forms a core element of precision oncology-driven studies (Chong *et al.*, 2009; Ng *et al.*, 2012). Patient-derived GPCs and PDx mouse models offer researchers the capability to re-form the patient's original tumor morphology and molecular patterns, so as to facilitate experimental validation prior to initiation of clinical trials. However, integrating the massive research and clinical data archived in public and NNI databases is not a trivial task, as similar data may be acquired using very different platforms. Thus, in my thesis, I have demonstrated through various methods how this can be achieved with robust and meaningful outcome.

6.1.1 Primary tumors: Integrating analytical pipeline for molecular interrogation

Machine learning models embrace strong potential to resolve the complex data into critical features that can be applied to unravel insights of disease phenotypes (Zou *et al.*, 2019). The predictive models have been shown to interrogate the trained signatures in patient tumor data and classify them into pathway-specific subtypes with prognosis correlation (Verhaak *et al.*, 2010; Wang *et al.*, 2018). We applied a nearest shrunken centroid approach to stratify our patient tumors and GPCs using previous Phillips and Lottaz molecular signatures (Lottaz *et al.*, 2010; Phillips *et al.*, 2006). This approach is essentially a modified diagonal linear discriminant analysis (DLDA) suitable for identifying classifiers from high dimensional data such as microarrays. The general rule of DLDA is to estimate the distance between gene expression readouts and centroids of class vectors. The cross-validation accuracy of feature selection is achieved by reducing the FDR through LOOCV in the training dataset (Tibshirani *et al.*, 2002). The limitation of this approach was previously reported as a class imbalance problem that resulted in poor stratification for minor subtypes (Blagus and Lusa, 2010). The current training datasets in both cases of Phillips and Lottaz molecular subtype stratification have uniform distribution of subtypes. We achieved a classification accuracy of 86-94% in our cross-validated evaluation.

Another alternative machine learning model called 'ClANC' was applied in the current thesis to interrogate the TCGA 840-gene signature to identify GBM-specific molecular subtypes (Verhaak *et al.*, 2010). The

ClaNC approach is also a modified linear discriminant analysis (LDA) procedure in which a regular t-statistics is applied to rank each gene feature by its capability to differentiate the class vectors. The misclassification error rates are computed using 5-fold cross validation in the training datasets. The unique advantage of ClaNC was found to be simpler, interpretable with significantly lower error rate to predict the subtypes in the simulated and real microarray datasets compared with either shrunken centroid approaches or other LDA models (Dabney, 2005, 2006). The proneural, neural, classical and mesenchymal subtypes were predictable using 210 genes for each class with the CV error rate of ~10% in the current training process using TCGA GBM microarray database. The limitation of this ClaNC approach lies in the signature being derived from a chip-based platform, which precludes the application of the TCGA signature in a platform-independent manner, such as interrogating the RNA-seq databases of patient tumors using the trained model.

To achieve the robust identification of molecular subtypes across different gene expression platforms, a glioma-intrinsic (GI) subtype signature was identified through consensus non-negative matrix factorization approach (Wang *et al.*, 2018). The GI gene signature representing each proneural, classical and mesenchymal subtype was achieved by measuring positive silhouette width in the core training TCGA IDH wild-type database. We adapted a rank-based single sample GeneSet Enrichment Analysis (ssGSEA) method to predict the GI subtypes in our tumor resource using resampling procedures as defined by Wang *et al.* (Wang *et al.*, 2018). The resampling methodology was

proposed to overcome the cross-platform measurements of gene readouts to compare the empirical cumulative distribution functions (ECDF) of gene signatures and the random distribution from remainder genes. The resampling procedure-based subtype assignment performed well with adjusting distribution measures to student T-distribution for the prediction database of less than 50 samples to arrive at the consistent predictions. The resampling-based ssGSEA prediction methodology is a promising method to identify the pathway-specific subtypes across assay platforms, and perform well by accounting for the virtual distributions derived from the sampling cohort.

The current subtype classification approaches are applicable for the identification of precise molecular features and enable us to interrogate features with or without the presence of trained datasets. The present work also leverages on the classification approaches such as template-based prediction and connectivity map-based class prediction to interrogate the molecular signatures identified from GPCs with genomic perturbations. These approaches are directly relevant at validating functional pathway signatures established from gene perturbations or small molecule inhibition assays in clinical databases.

6.1.2 Working with limited pool of patient-derived glioma cell lines

Preclinical validation using a cellular platform or animal model is limited due to the small number of biological replicates, as with all such approaches. In such small datasets, identifying a classifier by training a

model with minimal CV error rate is not a meaningful attempt. As such, the first step is to determine the critical gene features by performing supervised analysis between perturbed and non-perturbed microarray profiles. Several parametric and non-parametric methods such as regression models, FDR-based scoring methods and Bayesian models are routinely applied to define the molecular features between the experimental conditions (Efron and Tibshirani, 2002; Ritchie *et al.*, 2015; Tusher *et al.*, 2001). The functional experiments often establish correlated gene expression profiles between perturbed experimental conditions, such as a particular clone of cells subjected to two experimental conditions. It is important that the feature extraction algorithms account for this intra-block correlation between the perturbed and non-perturbed conditions, by fitting random effect in the model before estimating differential expression features. In the present work, we applied a linear regression model that allows specification of the intra-block correlation as random effects inside the regression model estimation for differential gene expression features (Ritchie *et al.*, 2015). The variance between gene expression measurements are minimized by scaling them towards global variance using empirical Bayes model to reduce the number of false positives in differentially expressed gene signatures. The estimation of global variance is often affected by the number of samples present in the experimental cohort (Wei *et al.*, 2004).

To overcome the limitation posed by a small sample pool of GPCs, we attempted to merge the databases of international stem-like cell collections (Beier *et al.*, 2007; Gunther *et al.*, 2008; Pollard *et al.*, 2009). We

detected considerable batch effects correlated with each database across the sample cohorts. The batch effect issue was effectively neutralized using the empirical Bayes method proposed by Johnson *et al.*, (Johnson *et al.*, 2007). The posterior probability for the adjustment of expression level variance across batches was well-corrected using the Bayes modeling approach. More importantly, this Bayes method was appropriate for correcting the databases of both smaller and larger sample pools. The model estimates statistical parameters such as mean and variance across multiple genes in each database, and determines the collective mean estimates of multiple batches. The PCA-based visualization is commonly performed to evaluate the batch effect corrections for high dimensional data (Lazar *et al.*, 2013; Reese *et al.*, 2013). The limitation of the PCA method is the inconclusive detection of batch effect adjustment of databases having large intra-variability of samples (Nyamundanda *et al.*, 2017). Nevertheless, the batch effect adjustment using the empirical Bayes method with R/ComBat function from R/sva package (Version 3.30.1) was shown to outperform six different variance correction approaches evaluated in genomic studies (Chen *et al.*, 2011) (Leek *et al.*, 2019).

In addition, we subjected the batch-corrected gene expression profiles in agglomerative hierarchical clustering to complete linkage metrics of Euclidean distance estimated from gene expression data. This clustering evaluation understanding of the presence of molecularly distinct subtype pattern. The classical studies from TCGA and other pioneer efforts executed the agglomerative clustering on highly variable

molecular features to construct the dendrogram pattern of patient tumors (Murat *et al.*, 2008; Noushmehr *et al.*, 2010; Verhaak *et al.*, 2010; Wang *et al.*, 2018). The hierarchical clustering approach tends to be sensitive at producing stable cluster subtypes when the sample size is smaller, and largely variable with the selection of distance measures and linkage methods. A bootstrapping-based consensus clustering was extensively applied to improve the cluster stability of molecular features through the resampling procedure of random feature selection structures (Wilkerson and Hayes, 2010; Wilkerson *et al.*, 2012). These clustering and variance-based approaches are applied to visualize the signature patterns and batch variability adjustments across patient tumor databases in the current work.

6.1.3 Cross-validation of glioma cell information in clinical datasets

The emergence of large public clinical databases such as TCGA (The Cancer Genome Atlas Research *et al.*, 2008) and REMBRANDT (Madhavan *et al.*, 2009) necessitates the development of computational pipelines that have: (1) Capability to interrogate multi-dimensional clinical data (tumor grade, WHO classification scheme, patient age, treatment decision, Karnofsky score), and (2) Ability to link those data back to the experimental platform; i.e. cross-validate clinical data with genetic perturbations in isogeneic cell lines and animal models. My work incorporates the Connectivity Map (CMAP) approach, first developed by the Broad Institute (Lamb *et al.*, 2006), which ranks patterns of

association between transcriptomic programs in patient cells with molecular and clinical parameters in clinical databases. This enrichment program is unique because it offers a means to merge data from multiple platforms and rank patients, such that the likely responders can be identified according to their predicted levels of sensitivity.

An alternative approach utilizes the Nearest Template Prediction (NTP) approach to stratify patient tumors and GPCs into signature-specific classes. In clinical scenarios, unlike the previous CMAP approach, no patient should ideally be assigned a neutral score as that leads to indecision on treatment approach. The NTP method is flexible and effective at overcoming the limitations of class prediction analyses such as changes in technology platforms, lack of large training datasets and unavailability of complete signature in the evaluation cohort (Hoshida, 2010). The algorithm entails a signature geneset with directionality of measurement such as fold-change, and the templates of each subtype are created by standardizing gene expression level. The cosine distance value estimates the proximity of gene expression measurements of each tumor to the subtype template. The statistical confidence is estimated by the nominal p-value based on the distance measurement from null distribution created by random signatures through resampling the whole transcriptome 1,000 times. The NTP-based prediction is a flexible pipeline for the class assignment of any number of subtypes with prediction confidence p-value.

6.1.4 Application of bioinformatic approaches in our studies

In our *STAT3* study, we developed a functionally-tuned gene signature representative of pathway activation, by integrating candidates that are co-expressed with *STAT3* in patient databases, but have inverse association with *STAT3* knockdown in our NNI patient cells. This strategy identifies gene candidates that are *STAT3*-regulated yet have prognostic outcome. The functionally-tuned *STAT3* signature was derived by fitting a regression model adjusted with empirical Bayes variance estimated from the *STAT3* knockdown (KD) expression profile in GPCs, compared with the non-targeting vector control. Subsequently, the rank-based CMAP approach interrogates the signature in patient transcriptomic databases. This strategy is meaningful and particularly important because it combines the greater statistical power of large public clinical databases with biological validation in the unavoidably smaller set of cell lines and PDx models. We previously demonstrated the robustness of our cell lines and PDx tumors at recapitulating the patient's original tumor pathophysiology and molecular profile (Chong *et al.*, 2009).

Liu *et al.* adopted a similar approach in probing the clinical relevance of purified CD44+/CD24- cells in breast cancer databases (Liu *et al.*, 2007). In addition, my thesis highlights the utility of the *STAT3* functionally-tuned gene signature in ranking patients and subsequently identifying matched cell lines based on their sensitivity to *STAT3* small molecule inhibitors. This information is particularly important in the clinic where treatment decisions are assessed for their likelihood of success. Furthermore, the *STAT3* gene signature performed as an independent

prognostic predictor after accounting for other clinical variables. The Cox proportional hazard model was fitted with covariates such as age, Karnofsky score, molecular subtype signatures and *STAT3*-signature classes associated with overall survival. We applied the Bayesian Information Criterion (BIC) scoring approach to address the prognostic (overall survival) variability in patient databases. BIC scoring is a parsimony approach that attempts to identify the ranking of models by applying a penalty scoring for complex parameterization (Shao, 1997). Another similar approach called AIC has also been applied to rank the regression models which differ in the penalty scoring (Akaike, 1974; Stone, 1977). A better-fitted model is represented by a lower score in the model comparison. The AIC estimation overfits the data when the model is built with large parameters; hence, AIC is appropriate in situations where the objective of the model is to achieve a highly sensitive one with low false negative calls (John J. Dziak, 2012; McQuarrie and Tsai, 1998). The BIC is ideal for the identification of a model with high specificity that predicts low false positive calls (John J. Dziak, 2012). The present work achieved a model with the lowest BIC score, including WHO molecular subgroups, Karnofsky score and *STAT3* signature classes, strongly predictive of survival heterogeneity in glioma patient databases.

We further elucidated the mechanism of differential protein connections by analysing the protein expression database of TCGA tumors. A RFE algorithm with random forest implementation was interrogated to determine the key protein features delineating *STAT3*-high and *STAT3*-low patient cohorts. The feature extraction process is

executed by calculating classification error upon the exclusion of a specific feature. Random forest (RF) nodes are constructed at each iteration process with 2/3rd of dataset for training the classifier to predict the classes in the testing dataset (out-of-bag samples; OOB) (Darst *et al.*, 2018). The limitation of the RF-RFE approach is that the presence of correlated variables potentially affects the estimation of the classification error and thus impacts the feature determination. By using the joint prior probabilities, we established Bayesian regulatory networks for *STAT3*-high and -low patient tumor groups. The workflow for the Bayesian regulatory network construction was adapted from the published protocols by Chen *et al.*, (Chen *et al.*, 2006). The protein regulatory network in the *STAT3*-high patient group is driven by EGFR associations with CDKNs; on the contrary, PTEN-dependent regulatory connections were exclusive in the *STAT3*-low patient cohort. The key parental node connections in *STAT3*-high and -low protein networks were in strong alignment with GBM-like and LGG-like protein clusters of patients, respectively (Ceccarelli *et al.*, 2016).

6.1.5 A novel kinome analysis pipeline: Probing biochemical interactions

Kinases are the most extensively studied protein drug targets with several candidates in clinical trials. They are involved in all aspects of cellular function and drive many hallmark phenotypes of tumor biology. Moreover, kinases can be specifically targeted via domain-specific modulation using small molecules or monoclonal antibodies. Our study

utilizes a kinome screen of STAT3 drug-inhibited isogenic GPCs, with the intention of assigning such biochemical activation to our earlier computationally derived *STAT3* functionally-tuned gene signature. We will map the kinases triggered by the drug inhibition, with its corresponding transcription factors and target genes. We propose a novel computational pipeline by integrating multimodal genomic data with kinase dynamics measured in a three-dimensional chip. Importantly, the proposed pipeline is able to detect differentially regulated kinase pathways upon STAT3 inhibition treatment. Specifically, we identified the top-ranked kinase pathway, IGF1-R, that was up-regulated in *STAT3*-low GBM cells upon treatment with AZD1480, and validated this both *in vitro* and *in vivo*. Literature has previously demonstrated that the altered IGF-1R pathway modulates key oncogenes such as *VEGF*, *MMP2*, *TWIST*, *TIMP1* and *HIF1A* via the downstream *STAT3* transcriptional pathway (Azar *et al.*, 2011; Chakrabarty and Kondratick, 2006; Chatterjee *et al.*, 2004; Fukushima *et al.*, 2007; Miyake *et al.*, 2005; Pažanin *et al.*, 2011; Wang *et al.*, 2003). Our findings corroborated the mechanistic association between STAT3 and IGF1-R at regulating drug-mediated resistance. We thus propose a combination therapeutic strategy comprising STAT3 and IGF1-R inhibitors that potentially benefit *STAT3*-low patients. In addition, integrative analysis of our kinome screen data along with *STAT3* knockdown data revealed a comprehensive kinase regulatory mechanism. Collectively, we were able to define significant kinases that act as therapeutic targets in both *STAT3*-high and -low patient cohorts.

6.1.6 Concluding Remarks

My thesis has shown that through various machine learning approaches, multi-dimensional molecular and clinical information on different platforms can be integrated for analysis in preclinical studies. Such studies are particularly effective at establishing proof-of-concept biological validation using limited cellular and animal resources, yet deriving statistical confidence from computational predictions in large clinical datasets. This empowers the testing of precision oncology-driven therapeutic strategies. Our work was also featured in BioCentury which has garnered significant industry interest, and will also be elaborated in an upcoming issue of Asia Pacific Biotech News.

6.2 Future directions

My PhD thesis is primarily focused on constructing a computational framework to characterize brain tumors using machine learning approaches. The patient stratification methods presented in this thesis have been published in 4 manuscripts within the period of my PhD candidature (Chong *et al.*, 2016; Ma *et al.*, 2016; Tan *et al.*, 2016; Xu *et al.*, 2018). Our work characterized the molecular subtypes of the NNI Brain Tumor Resource to serve the goals for precision medicine strategies. We constructed the molecular resource as a backbone for our computational models to characterize the altered signaling pathways of brain tumors. Next, we will pursue the following objectives to delve deeper into the complex mechanisms of disease progression and recurrence, as well as to refine computational parameters to profile and integrate genomic analyses, for the translation into more meaningful clinical therapy and outcomes for brain tumor patients.

1. We will develop a molecular 'reference baseline profile' for our primary and recurrent tumors using deep RNA-Sequencing measurements. Next-generation sequence profiling will provide us with deeper molecular insights such as structural variations and aberrant splice forms in tumors and cells. These data will serve as an important database infrastructure to identify stratified patients amenable for precision medicine strategies.
2. Proneural-to-Mesenchymal transition (PMT) is the key phenomenon driving the hyper-proliferation of brain tumor cells.

There is an urgent need to understand the mechanisms driving molecular switching. We will develop a computational model to evaluate the molecular determinants that are causal for molecular subtype switching by comparing subtype-switched and non-switched tumors.

3. We recognize that the functionally-tuned *STAT3* signature stratifies the patients who could benefit from *STAT3* inhibition therapy in our patient tumor databases. Using integrative analysis of somatic mutations from TCGA collections, we will identify genomic variants in patients with an active *STAT3* transcriptomic program. The role of these variants in mediating *STAT3* signaling will be functionally tested. We identified several kinase candidates that were uniquely altered by *STAT3* inhibition therapy. Next, we will implement a Bayesian probabilistic model to determine the topological alterations of kinase modifications. This evaluation is important as it serves to detect a master dependency node (lead kinase) to be targeted for the *STAT3* signature-stratified patient cohorts.

6.3 Conclusion

We demonstrated the capability of NNI Brain Tumor Resource to evaluate precision medicine goals. Importantly, the computational approaches were leveraged to identify the molecular signature enrichment in patient tumors and cells. Machine learning models ensured that the molecular profiles of NNI resource were well integrated with international Brain Tumor Resources. Our NNI Brain Tumor Resource platform coupled with international tumor databases identified the role of *STAT3* and *ST3GAL1* signaling pathways in aggressive brain tumors. These signatures were able to identify patients associated with aberrant molecular pathways and stratify them into responsive cohorts and predict survival outcomes. The high-throughput kinome screening detected IGF1-R resistance mechanism that developed in non-responders of STAT3 inhibition therapy. Our computational strategies, as illustrated in **Figure-6.1**, serve as an important tool for developing precision medicine strategies to refine each patient's needs, minimize chemotherapeutic side effects and consequently improve the patient's survival outcome and quality of life.

Figure-6.1. Graphical illustration of computational strategies adopted in the thesis. Analytical approaches are centered around NNI Brain Tumor Resources by generating proof-of-concepts (POCs) from patient database analysis from in-house and overseas databases. We start with pattern analysis of molecular data from transcriptomes, DNA sequences and methylome profiles and subsequently adopt predictive learning from molecular patterns associated with aberrant pathways. This process has led us to generate functional biomarkers validated in our GPCs and mouse models. Based on the functional evidence, we derive precision medicine strategies for our patients' care using the machine learning pipeline.

7 TABLES

Table 3. Shrunk class centroids of Lottaz gene signature applied to classify patient cells in National Neuroscience Institute (NNI) Brain Tumor Resource (Lottaz *et al.*, 2010)

ID	Symbol	Centroid order	Proneural	Mesenchymal	<i>p</i> -value
205103_at	C1orf61	1	Yes	No	0.01
209988_s_at	ASCL1	2	Yes	No	0.01
209987_s_at	ASCL1	3	Yes	No	0.01
210090_at	ARC	4	Yes	No	0.01
39966_at	NGC	5	Yes	No	0.01
222301_at	C1orf61	6	Yes	No	0.01
206051_at	ELAVL4	7	Yes	No	0.01
213768_s_at	ASCL1	8	Yes	No	0.01
219107_at	BCAN	9	Yes	No	0.01
205737_at	KCNQ2	10	Yes	No	0.01
213609_s_at	SEZ6L	11	Yes	No	0.01
202403_s_at	COL1A2	12	No	Yes	0.01
205358_at	GLUR2	13	Yes	No	0.03
219537_x_at	DLL3	14	Yes	No	0.03
209803_s_at	PHLDA2	15	No	Yes	0.01
205613_at	SYT17	16	Yes	No	0.05
202437_s_at	CYP1B1	17	No	Yes	0.01
203180_at	ALDH1A3	18	No	Yes	0.01
202436_s_at	CYP1B1	19	No	Yes	0.01
213484_at	Unknown	20	Yes	No	0.03
201506_at	TGFBI	21	No	Yes	0.01
221047_s_at	MARK1	22	Yes	No	0.04
204298_s_at	LOX	23	No	Yes	0.04
219229_at	SLCO3A1	24	No	Yes	0.04
211564_s_at	PDLIM4	25	No	Yes	0.04
204915_s_at	SOX-11	26	Yes	No	0.04
217744_s_at	PERP	27	No	Yes	0.04
210845_s_at	PLAUR	28	No	Yes	0.04
211924_s_at	PLAUR	29	No	Yes	0.04

Table 4. Comparison of Glioma-propagating cells (GPCs)-specific subtype signature addressing prognostic variability in patient databases.

Databases	REMBRANDT		
	HR	Log-rank <i>p</i>-value	% detected
Bhat signature	0.097	<0.0001	43 (14.8%)
Lottaz	0.29	<0.0001	146 (50.3%)
Composite signature	0.178	<0.0001	158 (54.5%)
Databases	Gravendeel		
	HR	Log-rank <i>p</i>-value	% detected
Bhat signature	0.081	<0.0001	42 (15.4%)
Lottaz	0.281	<0.0001	146 (53.7%)
Composite signature	0.29	<0.0001	182 (66.9%)
Databases	TCGA		
	HR	Log-rank <i>p</i>-value	% detected
Bhat signature	0	0.995	12 (3.1%)
Lottaz	0.06	<0.0001	106 (27%)
Composite signature	0.038	<0.0001	154 (39.3%)

*Mesenchymal patient cohort was considered as reference category to estimate the coefficient in survival model; HR = Hazard ratio.

Table 5. Linear regression model identifies up-regulated kinases in mesenchymal Glioma-propagating cells (GPCs)

Protein Tyrosine Kinases		Serine/Threonine Kinases	
Gene	Log ₂ -Foldchange	Gene	Log ₂ -Foldchange
Lck	0.377944954	p38[alpha]	1.112226934
PDGFR[alpha]	0.259393207	PLK3	1.112226934
ASK/MAP3K5	0.252074189	ULK1	0.985920692
MAP2K7	0.252074189	VRK1	0.935142162
MEKK6/MAP3K6	0.252074189	MAST1	0.873878393
Ret	0.192293459	STK33	0.848148999
Kit	0.186382156	MAST4	0.816541393
BLK	0.181117362	AurB/Aur1	0.81154772
CDKL1	0.169836536	CDKL4	0.758199475
CDKL2	0.169836536	AurC/Aur3	0.758057389
CHED	0.169836536	Haspin	0.758057389
CRK7	0.169836536	PBK	0.758057389
DYRK1A	0.169836536	ULK2	0.758057389
HUNK	0.169836536		
JNK1	0.169836536		
JNK2	0.169836536		
JNK3	0.169836536		
Nek1	0.169836536		
DDR2	0.166600932		
SEK1/MAP2K4	0.16609883		
MEK1/MAP2K1	0.151174633		
MEK2/MAP2K2	0.151174633		
CSK	0.150769326		

Table 6. A 6-gene signature stratifies PNA-sorted GPC profiles

Probe ID	Symbol	Log ₂ Fold change (PNA-high vs PNA-low)	<i>p</i> -value	Adjusted <i>p</i> -value
580437	ST3GAL1	-0.55883947	1.50e ⁻⁰⁹	2.74e ⁻⁰⁵
6660300	GLS	-0.219833202	8.95e ⁻⁰⁸	0.0008
630450	IFIT2	0.760032374	1.39e ⁻⁰⁷	0.0008
4200025	KCNN2	0.338191913	5.32e ⁻⁰⁷	0.0024
4050091	RBM9	-0.316788815	6.90e ⁻⁰⁷	0.0025
3180682	FAM100B	-0.284297141	1.69e ⁻⁰⁶	0.0051

*Two-sided *p*-values were calculated in moderated t-statistics in linear models for microarray data.

Adjusted values were calculated by Benjamini and Hochberg's method.

PNA, Peanut Agglutinin; GPC, glioma-propagating cell.

Table 7. List of *ST3GAL1*-associated signature genes with Human Genome U133 PLUS 2.0 probesets used for CMAP analysis

Probe ID	Entrez gene ID	Gene symbol	Log ₂ -fold change	<i>p</i> -value*	Adjusted <i>p</i> -value†
219978_s_at	51203	NUSAP1	-2.18	<0.001	<0.001
218039_at	51203	NUSAP1	-2.01	<0.001	<0.001
239688_at	8243	SMC1A	-2.00	<0.001	<0.001
226096_at	252995	FNDC5	2.00	<0.001	0.005
1554242_a_at	1690	COCH	-2.23	<0.001	0.002
227296_at	113655	MFSD3	-2.40	<0.001	<0.001
200660_at	6282	S100A11	-2.11	<0.001	<0.001
210448_s_at	5026	P2RX5	-2.19	<0.001	<0.001
219804_at	79933	SYNPO2L	-2.29	<0.001	<0.001
200644_at	65108	MARCKSL1	-2.47	<0.001	<0.001
223082_at	30011	SH3KBP1	-2.53	<0.001	<0.001
1554168_a_at	30011	SH3KBP1	-2.48	<0.001	<0.001
1556429_a_at	93594	WDR67	-3.07	<0.001	<0.001
1560648_s_at	7259	TSPYL1	-2.03	<0.001	<0.001
203379_at	6195	RPS6KA1	-2.20	<0.001	<0.001
208657_s_at	10801	Sep-09	-2.17	<0.001	<0.001
222403_at	23788	MTCH2	-2.27	<0.001	<0.001
217489_s_at	3570	IL6R	-2.16	<0.001	0.004
200791_s_at	8826	IQGAP1	-2.38	<0.001	<0.001
210815_s_at	10203	CALCRL	2.06	0.003	0.01
225368_at	28996	HIPK2	-2.19	<0.001	<0.001
225115_at	28996	HIPK2	-2.06	<0.001	<0.001
225097_at	28996	HIPK2	-2.14	<0.001	<0.001
202094_at	332	BIRC5	-2.29	<0.001	<0.001
202095_s_at	332	BIRC5	-2.34	<0.001	<0.001
205046_at	1062	CENPE	-2.04	<0.001	<0.001
231120_x_at	5570	PKIB	-2.26	<0.001	0.004
218585_s_at	51514	DTL	-2.08	<0.001	<0.001

207742_s_at	2649	NR6A1	2.48	<0.001	<0.001
232124_at	729085	FAM198A	-2.11	<0.001	0.006
212005_at	26099	C1orf144	-2.82	<0.001	<0.001
239002_at	259266	ASPM	-2.13	<0.001	<0.001
1552843_at	10861	SLC26A1	-2.31	<0.001	<0.001
201851_at	6455	SH3GL1	-2.13	<0.001	<0.001
222256_s_at	100137047 /// 100137049 /// 8681	JMJD7 /// JMJD7- PLA2G4B /// PLA2G4B	-2.01	<0.001	0.003
224846_at	92799	SHKBP1	-2.10	<0.001	0.002
228846_at	4084	MXD1	2.13	<0.001	<0.001
201890_at	6241	RRM2	-2.67	<0.001	<0.001
209773_s_at	6241	RRM2	-2.88	<0.001	<0.001
226408_at	8463	TEAD2	-2.08	<0.001	<0.001
1555950_a_at	1604	CD55	-2.22	0.001	0.007
212155_at	149603	RNF187	-2.25	<0.001	<0.001
211122_s_at	6373	CXCL11	2.15	<0.001	0.005
225972_at	169200	TMEM64	-2.12	<0.001	<0.001
225033_at	6482	ST3GAL1	-2.00	<0.001	<0.001
203154_s_at	10298	PAK4	-2.06	<0.001	<0.001
214793_at	1849	DUSP7	-2.35	<0.001	<0.001
203213_at	983	CDK1	-2.36	<0.001	<0.001
216396_s_at	9538	EI24	-2.27	<0.001	<0.001
205565_s_at	2395	FXN	-2.10	<0.001	<0.001
225836_s_at	83695	C12orf32	-2.14	<0.001	<0.001
225837_at	83695	C12orf32	-2.00	<0.001	<0.001
212739_s_at	4833	NME4	-2.00	<0.001	<0.001
228205_at	7086	TKT	-3.17	<0.001	<0.001
1553601_a_at	259236	TMIE	-2.45	<0.001	<0.001
1556357_s_at	157697	ERICH1	-2.08	0.002	0.01
236481_at	2184	FAH	-2.27	<0.001	<0.001
1554140_at	79819	WDR78	2.52	<0.001	<0.001
208561_at	10060	ABCC9	-2.13	0.004	0.02

231358_at	83876	MRO	-2.77	<0.001	<0.001
206023_at	10874	NMU	-2.13	<0.001	<0.001
209408_at	11004	KIF2C	-2.18	<0.001	<0.001
1559265_at	387640	SKIDA1	-2.24	0.001	0.009
228403_at	375704	ENHO	-2.11	<0.001	<0.001
206884_s_at	8796	SCEL	2.41	<0.001	0.001
230830_at	123264	OSTBETA	2.14	<0.001	0.002
201710_at	4605	MYBL2	-2.80	<0.001	<0.001
221795_at	4915	NTRK2	-2.04	<0.001	<0.001
221796_at	4915	NTRK2	-2.42	<0.001	<0.001
1556368_at	5261	PHKG2	2.06	0.001	0.008
1556369_a_at	5261	PHKG2	2.04	<0.001	0.003
201517_at	22916	NCBP2	-2.11	<0.001	<0.001
232574_at	64131	XYLT1	-2.12	<0.001	0.005
204709_s_at	9493	KIF23	-2.10	<0.001	<0.001
208998_at	7351	UCP2	-2.48	<0.001	<0.001
230521_at	84904	C9orf100	-2.12	<0.001	0.003
209680_s_at	3833	KIFC1	-2.59	<0.001	<0.001
241436_at	6340	SCNN1G	-2.08	0.004	0.02
225113_at	8540	AGPS	-2.10	<0.001	<0.001
230974_at	11269	DDX19B	-2.41	<0.001	<0.001
1555801_s_at	151126	ZNF385B	-2.00	0.007	0.03
210121_at	8707	B3GALT2	2.11	<0.001	<0.001
208396_s_at	5136	PDE1A	2.22	<0.001	<0.001
200606_at	1832	DSP	-2.02	<0.001	0.006
237137_at	677766	SCARNA2	2.45	<0.001	<0.001
232976_at	100033429 /// 100033431 /// 100033432 /// 692236 /// 727708	SNORD116-17 /// SNORD116-19 /// SNORD116-20 /// SNORD116-21 /// SNORD116@	2.14	<0.001	<0.001
208322_s_at	6482	ST3GAL1	-0.56	<0.001	<0.001
215874_at	6482	ST3GAL1	-0.56	<0.001	<0.001

225033_at.1	6482	ST3GAL1	-0.56	<0.001	<0.001
225034_at	6482	ST3GAL1	-0.56	<0.001	<0.001
1559452_a_at	6482	ST3GAL1	-0.56	<0.001	<0.001
220116_at	3781	KCNN2	0.34	<0.001	0.002
213901_x_at	23543	RBFOX2	-0.32	<0.001	0.003
201492_s_at	6171	RPL41	-0.32	<0.001	0.003
212104_s_at	23543	RBFOX2	-0.32	<0.001	0.003
216215_s_at	23543	RBFOX2	-0.32	<0.001	0.003
203159_at	2744	GLS	-0.22	<0.001	<0.001
244796_at	2744	GLS	-0.22	<0.001	<0.001
203157_s_at	2744	GLS	-0.22	<0.001	<0.001
221510_s_at	2744	GLS	-0.22	<0.001	<0.001
203158_s_at	2744	GLS	-0.22	<0.001	<0.001
223079_s_at	2744	GLS	-0.22	<0.001	<0.001
211414_at	2744	GLS	-0.22	<0.001	<0.001
226757_at	3433	IFIT2	0.76	<0.001	<0.001
217502_at	3433	IFIT2	0.76	<0.001	<0.001
224783_at	283991	FAM100B	-0.28	<0.001	0.005
224785_at	283991	FAM100B	-0.28	<0.001	0.005

*Two-sided *p*-value calculated in moderated t-statistics in linear models for microarray data.

†Adjusted *p*-values by Benjamini and Hochberg's method.

Table 8. Results of multivariate analysis on *ST3GAL1* signature classes, WHO classes and Glioma-Intrinsic (GI) molecular subtypes

Covariates	Univariate analysis			Multivariate analysis		
	HR (95% CI)	SE	Pr(> z)*	HR (95% CI)	SE	Pr(> z)*
REMBRANDT						
<i>ST3GAL1</i> -high	2.587 (1.837-3.644)	0.175	5.37e ⁻⁰⁸	2.068 (1.432-2.986)	0.187	1.00e ⁻⁰⁴
GI-Classical	2.155 (1.383-3.357)	0.226	0.0007	2.183 (1.392-3.425)	0.23	0.0006
GI-Mesenchymal	2.687 (1.814-3.98)	0.201	8.28e ⁻⁰⁷	2.189 (1.445-3.316)	0.212	0.0002
AGE	1.0157 (1.005-1.027)	0.006	4.00e ⁻⁰³	1.016 (1.004-1.028)	0.006	9.00e ⁻⁰³
Gravendeel						
<i>ST3GAL1</i> -high	3.393 (2.347-4.905)	0.188	8.15e ⁻¹¹	2.186 (1.346-3.551)	0.247	1.58e ⁻⁰³
WHO-IDH Mut-NCDL	1.898 (0.974-3.700)	0.34	0.0596	1.889 (0.928-3.843)	0.362	0.07947
WHO-IDH WT	3.151 (1.674-5.932)	0.322	3.00e ⁻⁰⁴	2.425 (1.207-4.873)	0.356	0.01287
GI-Classical	2.057 (1.356-3.124)	0.213	0.0007	1.416 (0.854-2.348)	0.257	0.177
GI-Mesenchymal	2.209 (1.455-3.356)	0.2132	2.00e ⁻⁰⁴	1.374 (0.806-2.343)	0.272	0.243
AGE	1.033 (1.02-1.048)	0.007	1.21e ⁻⁰⁶	1.031 (1.016-1.047)	0.007	7.15e ⁻⁰⁵
TCGA						
<i>ST3GAL1</i> -high	5.92 (3.614-9.697)	0.252	1.63e ⁻¹²	1.371 (0.744-2.526)	0.312	0.312
WHO-IDH Mut-NCDL	1.007 (0.335-3.024)	0.561	0.99	1.449 (0.467-4.5)	0.578	0.521
WHO-IDH WT	15.064 (5.823-38.974)	0.485	2.23e ⁻⁰⁸	10.376 (3.546-30.367)	0.548	0.0000195
GI-Classical	2.197 (1.309-3.688)	0.264	2.80E-03	0.601 (0.336-1.074)	0.297	0.0855
GI-Mesenchymal	2.452 (1.429-4.211)	0.276	1.10e ⁻⁰³	0.978 (0.540-1.771)	0.303	0.9416
AGE	1.074 (1.057-1.093)	0.008	<2.2e ⁻¹⁶	1.054 (1.033-1.075)	0.0102	3.25e ⁻⁰⁷

**ST3GAL1*-low patients were treated as the reference to estimate the coefficient in Cox regression model; WHO-IDH Mut-CDL (Co-deleted for chromosomal *1p/19q*) patients were treated as the reference to estimate the coefficient in Cox regression model; GI (Glioma Intrinsic)-Proneural patients were treated as the reference to estimate the coefficient in Cox regression model. CI = confidence interval; WHO = World Health Organization; HR = Hazard ratio; IDH = Isocitrate dehydrogenase; Mut = Mutation; NCDL = non-codeleted for chromosomal *1p/19q*; $\Pr(<|z|)$ = two-sided Wald test *p*-value; SE = standard error of coefficient.

Table 9. Contingency table of *ST3GAL1*-associated signature classes with Glioma-Intrinsic molecular classification

REMBRANDT				
<i>ST3GAL1</i> status	Mesenchymal	Classical	Proneural	Total
<i>ST3GAL1</i> -low	24 (22.64%)	18 (17.0%)	64 (60.38%)	106 (100%)
<i>ST3GAL1</i> -high	35 (43.75%)	23 (28.75%)	22 (27.5%)	80 (100%)
X-squared = 19.927, df = 2, <i>p</i> -value = 4.708e ⁻⁰⁵				
Gravendeel				
<i>ST3GAL1</i> status	Mesenchymal	Classical	Proneural	Total
<i>ST3GAL1</i> -low	20 (22.73%)	17 (19.32%)	51 (57.96%)	88 (100%)
<i>ST3GAL1</i> -high	26 (38.24%)	28 (41.18%)	14 (20.59%)	68 (100%)
X-squared = 22.336, df = 2, <i>p</i> -value = 1.412e ⁻⁰⁵				
TCGA				
<i>ST3GAL1</i> status	Mesenchymal	Classical	Proneural	Total
<i>ST3GAL1</i> -low	29 (24.17%)	32 (26.67%)	59 (49.17%)	120 (100%)
<i>ST3GAL1</i> -high	33 (26.19%)	60 (47.62%)	33 (26.19%)	132 (100%)
X-squared = 15.991, df = 2, <i>p</i> -value = 0.00034				

df, degrees of freedom; X-squared, chi-squared

Table 10. List of functionally-tuned STAT3 signature (N=207) identified from STAT3 knockdown GPCs and co-expressed genes from patient database.

Probe ID	Gene symbol	Log ₂ fold change (knockdown vs NT)	p-value	Adjusted p-value
205353_s_at	PEBP1P2	0.433296765	1.67E-05	0.001743919
210825_s_at	PEBP1	0.442394112	6.26E-06	0.000868573
205353_s_at	PEBP1	0.433296765	1.67E-05	0.001743919
217491_x_at	COX7CP1	1.047442341	6.63E-13	7.24E-09
212964_at	HIC2	0.782347662	2.99E-07	9.23E-05
227064_at	ANKRD40	1.071426888	6.14E-06	0.000860402
213234_at	KIAA1467	0.909560041	1.90E-08	1.25E-05
205353_s_at	RETSAT	0.433296765	1.67E-05	0.001743919
203387_s_at	TBC1D4	0.495299346	9.38E-06	0.001146284
210201_x_at	BIN1	1.259061855	1.31E-09	1.79E-06
214643_x_at	BIN1	0.578410723	4.45E-06	0.000670277
210202_s_at	BIN1	0.574198906	1.35E-07	5.10E-05
202931_x_at	BIN1	0.905217007	4.96E-08	2.40E-05
214439_x_at	BIN1	1.237094138	4.21E-10	8.83E-07
221810_at	RAB15	0.570770067	7.32E-07	0.000179691
59697_at	RAB15	0.6078659	1.53E-05	0.001627199
201674_s_at	AKAP1	0.427032672	1.80E-05	0.001813
209380_s_at	ABCC5	0.488877013	5.46E-05	0.004102576
228454_at	LCOR	0.716439336	3.64E-07	0.000108373
226520_at	LCOR	0.711460444	5.19E-07	0.000143898
238974_at	C2orf69	0.354853672	0.000105882	0.006623759
238890_at	BRWD1	0.529133088	1.40E-05	0.001519851
230296_at	C16orf52	0.527585515	1.59E-09	1.98E-06
222880_at	AKT3	0.442970745	0.000118808	0.007191216
212607_at	AKT3	0.663056326	8.07E-06	0.001034896
209845_at	MKRN1	0.878168821	1.02E-05	0.00121941
213304_at	FAM179B	0.445533968	2.99E-05	0.00262377
211383_s_at	WDR37	0.321256758	8.55E-05	0.005709813
212050_at	WIPF2	0.293485722	2.26E-05	0.002129344
226033_at	USP31	0.348069858	2.71E-06	0.000468537
218862_at	ASB13	0.302990815	4.47E-05	0.003499296
223184_s_at	AGPAT3	0.655549583	4.15E-05	0.003330501
212114_at	ATXN7L3B	0.753989001	3.39E-08	1.78E-05
225957_at	CREBRF	0.557724594	0.000184028	0.009963001
217491_x_at	COX7C	1.047442341	6.63E-13	7.24E-09
201134_x_at	COX7C	1.210242166	1.14E-13	2.40E-09
219175_s_at	SLC41A3	0.510114508	1.15E-06	0.000248324
224931_at	SLC41A3	0.477907138	3.49E-08	1.82E-05
222494_at	FOXN3	0.283877313	1.27E-05	0.001408797

231969_at	STOX2	0.502052836	1.70E-05	0.001757107
205052_at	AUH	0.99041228	1.48E-07	5.45E-05
219093_at	PID1	0.803943854	5.30E-05	0.003997974
212655_at	ZCCHC14	0.532781896	0.000174755	0.009630579
225511_at	GPRC5B	0.379615887	1.73E-05	0.001769406
211475_s_at	BAG1	0.527496975	5.79E-06	0.00082353
226680_at	IKZF5	0.461945003	1.37E-06	0.000289101
37566_at	KIAA1045	0.365948316	0.000104529	0.006576747
226310_at	RICTOR	0.517924196	1.23E-05	0.001386839
226312_at	RICTOR	0.62017486	2.27E-08	1.40E-05
228248_at	RICTOR	0.517367522	1.02E-05	0.00121941
224495_at	SNORD118	-0.568950232	7.91E-06	0.001015928
231896_s_at	DENR	-0.496587511	1.80E-07	6.24E-05
218622_at	NUP37	-0.412470015	0.000139572	0.008134963
220239_at	KLHL7	-0.953057006	3.16E-12	2.16E-08
203664_s_at	POLR2D	-0.362518088	8.94E-05	0.005866287
202663_at	WIPF1	-0.72324048	7.34E-05	0.005078614
1554451_s_at	DNAJC14	-0.298195513	1.40E-05	0.001519851
225837_at	RHNO1	-0.852648658	2.59E-05	0.002354766
220358_at	BATF3	-0.268712986	3.31E-05	0.002816552
224783_at	UBALD2	-0.562831502	1.23E-05	0.001386839
225872_at	SLC35F5	-0.702897734	2.79E-05	0.00248779
222519_s_at	IFT57	-0.485287361	3.47E-05	0.002915154
205061_s_at	EXOSC9	-0.357316048	8.93E-05	0.005866287
222602_at	UBA6	-0.630284325	6.29E-08	2.89E-05
222601_at	UBA6	-0.479435608	1.24E-06	0.000265086
239413_at	CEP152	-0.470898387	1.11E-05	0.001298372
202253_s_at	DNM2	-0.321716645	3.46E-05	0.002909123
224714_at	MKI67IP	-0.221325861	0.000175689	0.009650242
217738_at	NAMPT	-0.516889292	8.49E-06	0.001077299
220199_s_at	AIDA	-1.00372432	6.37E-06	0.000876941
224617_at	PTBP3	-0.267390012	3.28E-06	0.000529742
225592_at	PPP1R18	-0.422623732	1.90E-05	0.001886765
219083_at	SHQ1	-0.262055432	5.82E-05	0.004286846
203024_s_at	C5orf15	-0.624591476	3.10E-06	0.000517693
225080_at	MYO1C	-0.744080211	1.81E-08	1.22E-05
225592_at	NRM	-0.422623732	1.90E-05	0.001886765
218131_s_at	GATAD2A	-0.431793696	0.000160996	0.009045734
202716_at	PTPN1	-0.482761247	1.44E-05	0.00155989
218474_s_at	KCTD5	-0.346923516	2.20E-06	0.000402599
1552470_a_at	ABHD11	-0.334488179	2.79E-05	0.00248779
1553678_a_at	ITGB1P1	-0.498832565	7.03E-05	0.004949843
225308_s_at	TANC1	-0.370186201	1.83E-06	0.000358235
209130_at	SNAP23	-0.566059891	0.00015134	0.008672756
202076_at	BIRC2	-0.466272661	5.89E-05	0.004327914

210844_x_at	CTNNA1	-0.408342655	3.55E-06	0.000560909
200764_s_at	CTNNA1	-0.373769988	8.58E-06	0.001082429
200765_x_at	CTNNA1	-0.420775093	1.05E-05	0.001245242
232231_at	RUNX2	-0.585408616	1.89E-06	0.000366498
238590_x_at	TMEM107	-0.457089117	8.56E-05	0.005709813
239824_s_at	TMEM107	-0.467439181	0.000113984	0.00700223
224495_at	TMEM107	-0.568950232	7.91E-06	0.001015928
213012_at	NEDD4	-0.815041138	8.72E-11	2.65E-07
208749_x_at	FLOT1	-0.716672114	1.64E-06	0.000330039
222446_s_at	BACE2	-0.452072584	7.90E-05	0.005389695
212949_at	NCAPH	-0.483401536	2.46E-05	0.002279434
225297_at	HAUS1	-0.222105321	0.000101285	0.006402158
202350_s_at	MATN2	-0.678512064	4.84E-06	0.000707225
225406_at	TWSG1	-0.96002712	3.01E-05	0.002636179
204017_at	KDEL3	-0.745956098	2.25E-08	1.40E-05
207265_s_at	KDEL3	-0.502976395	7.46E-07	0.000181052
203843_at	RPS6KA3	-0.872759755	6.03E-05	0.004393941
238606_at	ZNF747	-0.313718652	7.58E-05	0.005209986
204531_s_at	BRCA1	-0.375033919	9.08E-05	0.00591885
228234_at	TICAM2	-0.547371653	3.23E-07	9.92E-05
213141_at	PSKH1	-0.493616067	2.61E-07	8.42E-05
212262_at	QKI	-0.513125157	1.42E-06	0.000294669
208828_at	POLE3	-0.471787456	1.03E-06	0.000229381
228234_at	TMED7-TICAM2	-0.547371653	3.23E-07	9.92E-05
239231_at	ZNF101	-0.348344439	4.23E-06	0.000648808
202043_s_at	SMS	-0.700475345	3.25E-05	0.002780939
202695_s_at	STK17A	-0.350944507	4.29E-05	0.003399304
203823_at	RGS3	-0.655681581	0.000150751	0.008648043
212900_at	SEC24A	-0.336548654	2.81E-05	0.002497189
219838_at	TTC23	-0.340813525	4.59E-05	0.003571744
212415_at	SEPT6	-0.711330002	9.00E-05	0.00588946
208689_s_at	RPN2	-0.329444537	5.52E-05	0.004132446
213399_x_at	RPN2	-0.311282628	0.000183774	0.009963001
201776_s_at	EFCAB14	-0.63282897	1.57E-07	5.57E-05
228273_at	PRR11	-0.22943255	8.96E-05	0.005866287
218718_at	PDGFC	-0.521990685	6.06E-05	0.004404599
213113_s_at	SLC43A3	-0.499561147	5.74E-06	0.000817851
210692_s_at	SLC43A3	-0.407818545	8.94E-06	0.00110968
225538_at	ZCCHC9	-0.525692121	6.93E-05	0.004890429
209307_at	SWAP70	-0.924465151	7.32E-07	0.000179691
209306_s_at	SWAP70	-0.740490966	6.88E-07	0.000177278
235489_at	RHOJ	-0.545250331	1.46E-05	0.001580498
209344_at	TPM4	-0.605468838	1.58E-05	0.001672422
235177_at	METTL21A	-0.353911844	2.37E-05	0.002209913

203967_at	CDC6	-0.39014975	5.81E-05	0.004286846
217926_at	MRI1	-0.50991417	3.15E-06	0.000522419
217926_at	C19orf53	-0.50991417	3.15E-06	0.000522419
223065_s_at	STARD3NL	-0.685511323	8.53E-08	3.46E-05
1053_at	RFC2	-0.300698133	2.47E-05	0.002289575
201125_s_at	ITGB5	-0.710229499	6.01E-05	0.004382592
1553978_at	MEF2BNB-MEF2B	-0.568404669	7.07E-06	0.000948173
209198_s_at	SYT11	-0.83003113	3.38E-11	1.23E-07
224858_at	ZDHHC5	-0.534040372	1.63E-05	0.001713985
1553978_at	MEF2BNB	-0.568404669	7.07E-06	0.000948173
224749_at	ITFG3	-0.444375221	5.36E-05	0.004034133
231823_s_at	SH3PXD2B	-0.720720947	2.19E-08	1.40E-05
223024_at	AP1M1	-0.447245519	8.01E-07	0.00019022
209387_s_at	TM4SF1	-0.714078244	7.63E-05	0.005237147
200907_s_at	PALLD	-0.425345733	0.000105071	0.006595677
200897_s_at	PALLD	-0.449899798	6.38E-05	0.004588076
227493_s_at	KIAA1143	-0.567336092	8.64E-05	0.005729516
200001_at	CAPNS1	-0.85259993	3.06E-06	0.000513471
201924_at	AFF1	-0.531039743	5.53E-05	0.004134901
201114_x_at	PSMA7	-0.657910857	8.69E-07	0.000202878
216088_s_at	PSMA7	-0.755081417	3.71E-06	0.000580328
201095_at	DAP	-0.604241	4.21E-05	0.003365398
235609_at	BRIP1	-0.546501014	1.19E-08	9.65E-06
203344_s_at	RBBP8	-0.65704224	8.62E-05	0.005729516
212115_at	HN1L	-0.218924918	9.22E-05	0.005986392
211945_s_at	ITGB1	-0.450259304	5.56E-05	0.00414682
1553678_a_at	ITGB1	-0.498832565	7.03E-05	0.004949843
200004_at	EIF4G2	-0.256327971	1.69E-06	0.000337765
218815_s_at	TMEM51	-0.374214572	1.61E-05	0.001699196
204027_s_at	METTL1	-0.432260327	3.14E-06	0.000522419
212294_at	GNG12	-0.42807063	2.65E-07	8.42E-05
202944_at	NAGA	-0.866544715	9.85E-07	0.000223648
1555041_a_at	NAGA	-0.309702492	5.99E-05	0.004376945
202943_s_at	NAGA	-0.326773537	8.21E-05	0.005547013
224675_at	MESDC2	-0.582250851	7.53E-07	0.000181052
226519_s_at	AGXT2L2	-0.441047662	0.000103461	0.006517994
203445_s_at	CTDSP2	-0.987781015	1.20E-06	0.000258176
208735_s_at	CTDSP2	-0.562503754	8.50E-06	0.001077299
221773_at	ELK3	-1.957343884	1.35E-10	3.50E-07
212005_at	SZRD1	-0.283819637	0.000113567	0.006989855
219119_at	NAA38	-0.246383118	1.50E-05	0.001609269
212923_s_at	PXDC1	-0.366157539	3.16E-06	0.000522419
218070_s_at	GMPPA	-0.379481284	0.000118505	0.007191019
210719_s_at	HMG20B	-0.534857994	9.16E-10	1.47E-06

218025_s_at	ECI2	-0.424193994	2.58E-05	0.002354766
209709_s_at	HMMR	-0.743896588	3.86E-08	1.97E-05
207165_at	HMMR	-0.742132798	2.06E-09	2.37E-06
218092_s_at	AGFG1	-0.383646807	2.18E-05	0.002060779
225415_at	DTX3L	-1.03771168	8.11E-05	0.005497901
201128_s_at	ACLY	-0.436348759	7.09E-08	3.08E-05
201127_s_at	ACLY	-0.291642941	7.13E-05	0.005004752
211561_x_at	MAPK14	-0.303502273	9.34E-05	0.006040609
218847_at	IGF2BP2	-0.412588523	2.53E-05	0.002331414
205573_s_at	SNX7	-1.115182619	1.24E-05	0.001387865
202006_at	PTPN12	-0.271249741	6.32E-05	0.004552959
219863_at	HERC5	-0.65950278	7.05E-05	0.004955015
218076_s_at	ARHGAP17	-0.388457923	2.34E-06	0.000424747
227038_at	SGMS2	-0.396559122	1.49E-06	0.000306174
204451_at	FZD1	-0.281400201	9.43E-05	0.006085641
208992_s_at	STAT3	-2.29344946	1.32E-13	2.40E-09
225289_at	STAT3	-2.241324845	1.01E-11	4.58E-08
208991_at	STAT3	-2.414521098	2.03E-14	1.11E-09
226925_at	ACPL2	-0.699390584	7.28E-05	0.005052033
208829_at	TAPBP	-0.326491451	0.00011513	0.00704676
209476_at	TMX1	-0.285869969	1.68E-05	0.001747912
208948_s_at	STAU1	-0.46951292	7.66E-06	0.00099212
219924_s_at	ZMYM6	-0.60779049	9.07E-05	0.005917134
219924_s_at	ZMYM6NB	-0.60779049	9.07E-05	0.005917134
226122_at	PLEKHG1	-0.688659964	1.16E-05	0.001333511
229893_at	FRMD3	-0.469241058	5.71E-05	0.004238124
224752_at	C7orf73	-0.280896418	5.84E-05	0.004300475
224751_at	C7orf73	-0.605150532	5.12E-10	9.99E-07
222871_at	KLHDC8A	-0.98319795	4.09E-09	3.99E-06
207467_x_at	CAST	-1.040908778	2.57E-08	1.49E-05
208908_s_at	CAST	-1.06842078	2.26E-08	1.40E-05
202149_at	NEDD9	-0.985723556	1.74E-07	6.04E-05
60471_at	RIN3	-0.291845305	1.71E-05	0.001757107
219457_s_at	RIN3	-0.257013893	4.84E-05	0.003715644
201342_at	SNRPC	-0.390825429	4.50E-07	0.0001286

*Two-sided *p*-value calculated in moderated t-statistics in linear models for microarray data.

†Adjusted *p*-values by Benjamini and Hochberg's method.

Table 11. Contingency analysis on *STAT3* stratified classes with GI molecular subtypes

REMBRANDT			
STAT3/GI subtypes	STAT3-high	STAT3-low	Fisher's Exact <i>p</i> -value
Proneural	20 (18.519%)	88 (81.481%)	< 2.2e ⁻¹⁶
Classical	52 (72.222%)	20 (27.778%)	
Mesenchymal	72 (84.706%)	13 (15.294%)	
Gravendeel			
STAT3/GI subtypes	STAT3-high	STAT3-low	Fisher's Exact <i>p</i> -value
Proneural	9 (9.278%)	88 (90.722%)	< 2.2e ⁻¹⁶
Classical	52 (76.471%)	16 (23.529%)	
Mesenchymal	67 (91.781%)	6 (8.219%)	
TCGA			
STAT3/GI subtypes	STAT3-high	STAT3-low	Fisher's Exact <i>p</i> -value
Proneural	36 (24%)	114 (76%)	< 2.2e ⁻¹⁶
Classical	78 (67.241%)	38 (32.759%)	
Mesenchymal	61 (72.619%)	23 (27.381%)	

Table 12. Results of univariate and multivariate analysis on *STAT3* signature classes, WHO classes, Glioma-Intrinsic (GI) molecular subtypes and age of patient covariates in Gravendeel and TCGA databases (Ceccarelli *et al.*, 2016; Gravendeel *et al.*, 2009)

Covariates	Univariate analysis			Multivariate analysis		
	HR (95% CI)	SE	Pr(> z)*	HR (95% CI)	SE	Pr(> z)*
Gravendeel						
<i>STAT3</i> -High	4.434 (3.255-6.040)	0.158	<2.2e ⁻¹⁶	3.963 (2.278-6.897)	0.283	1.11e ⁻⁰⁶
WHO-IDH Mut-NCDL	2.033 (1.189-3.476)	0.274	0.0095	2.795 (1.591-4.908)	0.287	0.0003
WHO-IDH WT	2.874 (1.761-4.692)	0.250	2.42e ⁻⁰⁵	2.169 (1.255-3.748)	0.279	0.0055
GI-Classical	1.932 (1.372-2.720)	0.175	0.0002	0.725 (0.444-1.184)	0.251	0.1987
GI-Mesenchymal	2.429 (1.740-3.39)	0.170	1.86e ⁻⁰⁷	0.754 (0.437-1.302)	0.279	0.3112
AGE	1.045 (1.033-1.056)	0.006	2.00e ⁻¹⁵	1.042 (1.029-1.056)	0.007	4.37e ⁻¹⁰
TCGA						
<i>STAT3</i> -High	19.367 (10.91 -34.38)	0.293	<2.2e ⁻¹⁶	3.163 (1.080-9.263)	0.548	0.0357
WHO-IDH Mut-NCDL	1.250 (0.574-2.719)	0.397	0.574	1.329 (0.596-2.964)	0.409	0.4871
WHO-IDH WT	28.404 (12.299-65.602)	0.427	4.66e ⁻¹⁵	5.775 (1.472-22.652)	0.697	0.0119
GI-Classical	2.357 (1.551-3.580)	0.213	5.85e ⁻⁰⁵	0.774 (0.481-1.246)	0.243	0.2920
GI-Mesenchymal	2.774 (1.808-4.256)	0.218	2.99e ⁻⁰⁶	1.189 (0.731-1.932)	0.248	0.4850
AGE	1.074 (1.059-1.089)	0.007	<2.2e ⁻¹⁶	1.041 (1.023-1.059)	0.009	7.78e ⁻⁰⁶

**STAT3*-low patient cohort was considered a reference category to estimate the

coefficient in Cox regression model; WHO-IDH Mut-CDL (co-deleted for chromosome *1p/19q*) patients were treated as the reference to estimate the coefficient in Cox regression model; GI (Glioma Intrinsic)-Proneural patients were treated as the reference to estimate the coefficient in Cox regression model. CI = confidence interval; WHO = World Health Organization; HR = Hazard ratio; IDH = Isocitrate dehydrogenase; Mut = Mutation; NCDL = non-codeleted for chromosome *1p/19q*; $\Pr(<|z|)$ = two-sided Wald test *p*-value; SE = standard error of coefficient.

Table 13. Molecular subtype switching in GBM tumors upon recurrence.

The distribution of matched recurrent patients was analyzed for subtype switching using the Samsung Medical Center and Genentech databases (Joo *et al.*, 2013; Phillips *et al.*, 2006).

Molecular classes	First presentation	Recurrence	Log-likelihood Ratio statistics <i>p</i> -value
Mesenchymal (N=11; 35%)	STAT3-high (N=7; 64%)	STAT3-high (N=6; 86%)	0.018
		STAT3-low (N=1; 14%)	
	STAT3-low (N=4; 36%)	STAT3-high (N=4; 100%)	
Classical (N=11; 35%)	STAT3-high (N=7; 64%)	STAT3-high (N=4; 57%)	
		STAT3-low (N=3; 43%)	
	STAT3-low (4; N=36%)	STAT3-high (N=3; 75%)	
		STAT3-low (N=1; 25%)	
Proneural (N=13; 30%)	STAT3-high (N=4; 31%)	STAT3-high (N=3; 75%)	
		STAT3-low (N=1; 25%)	
	STAT3-low (N=9; 69%)	STAT3-high (N=2; 22%)	
		STAT3-low (N=7; 78%)	

Table 14. List of altered protein tyrosine kinases in *STAT3*-high and -low GPCs upon AZD1480 treatment

STAT3-high			STAT3-low		
Kinases	Log₂ Fold change	p-value	Kinases	Log₂ Fold change	p-value
ERK1	-0.13	0.000855	Srm	0.597229	0.021536
ERK2	-0.13	0.000855	DDR2	0.58974	0.030416
RAF1	-0.13	0.000855	DYRK1B	0.580768	0.031197
BLK	0.116378	0.003183	CTK	-0.45091	0.038848
ROR2	-0.13288	0.009185	Kit	0.643374	0.04008
JAK3	-0.07216	0.009442	IGF1R	0.334539	0.042187
JAK3~b	-0.07216	0.009442	ITK	0.401508	0.045133
Ret	-0.08399	0.014765	Mer	0.647419	0.046863
Lyn	0.06046	0.020579	ZAP70	-0.32716	0.049778
Ron	0.071592	0.022478	FGFR4	0.378161	0.050056
EphA4	-0.1646	0.023182	FAK	0.65263	0.050294
EphA10	-0.07543	0.03562	EGFR	0.392426	0.050318
Syk	-0.14001	0.038167	ASK/MAP3K5	0.415857	0.051123
Srm	0.065889	0.049258	MAP2K7	0.415857	0.051123
PKA[alpha]	0.046967	0.096575	MEKK6/MAP3K6	0.415857	0.051123
			SEK1/MAP2K4	0.415857	0.051123
			Brk	-0.33914	0.05334
			Tyro3/Sky	0.344678	0.053668
			ALK	-0.41229	0.056698
			FGFR1	0.402501	0.056858
			Ret	0.407246	0.057328
			DDR1	0.365469	0.058653
			InSR	-0.33601	0.059367
			Yes	-0.39535	0.062791
			Src	-0.27137	0.063148
			ERK1	0.398635	0.067246
			ERK2	0.398635	0.067246
			RAF1	0.398635	0.067246
			EphA10	0.385644	0.068755
			Ron	0.367742	0.069071
			Fes	-0.33334	0.070047
			PYK2	0.70148	0.070952
			TXK	0.732158	0.077581
			CSK	0.755434	0.084521
			JAK3	-0.30078	0.092282
			JAK3~b	-0.30078	0.092282
			MEK2/MAP2K2	0.41729	0.09632

Table 15. List of altered serine/threonine kinases in *STAT3*-high and -low GPCs upon AZD1480 treatment

<i>STAT3</i> -high			<i>STAT3</i> -low		
Kinases	Log ₂ Fold change	<i>p</i> -value	Kinases	Log ₂ Fold change	<i>p</i> -value
Lmr2	0.541331	0.013697	Lmr2	0.538119	0.000875
DYRK1B	-0.29761	0.021537	KSR	-0.57812	0.000884
SGK1	0.335221	0.030639	Nek6	-0.4904	0.001318
DYRK1A	-0.34054	0.033164	IRAK2	0.830925	0.001454
PKC[delta]	0.226021	0.03768	AurB/Aur1	-0.49334	0.00176
Akt2/PKB[beta]	-0.33427	0.039188	AurC/Aur3	-0.46329	0.002419
TBK1	-0.23773	0.044998	Haspin	-0.46329	0.002419
IKK[epsilon]	-0.28506	0.046246	PBK	-0.46329	0.002419
CDKL2	-0.25217	0.04691	ULK2	-0.46329	0.002419
HIPK2	-0.24595	0.047821	VRK1	-0.46329	0.002419
HIPK3	-0.24595	0.047821	PKA[alpha]	0.488368	0.003992
PKA[alpha]	0.375586	0.050728	PKC[alpha]	0.547474	0.004011
ALK	0.264399	0.053149	AurA/Aur2	-0.44537	0.004029
CDKL1	0.264399	0.053149	CDKL3	0.49377	0.005883
MAK	0.264399	0.053149	SNRK	-0.39285	0.008754
ZAP70	0.264399	0.053149	PSKH1	0.492275	0.010335
ROCK1	-0.28401	0.058563	CTK	0.361494	0.01263
PKC[alpha]	0.458112	0.069798	PYK2	0.361494	0.01263
RSK3	0.254815	0.070297	Brk	-0.37003	0.014856
PKC[beta]	-0.22949	0.071587	HIPK2	-0.34468	0.014903
PKN1/PRK1	-0.24098	0.078304	HIPK3	-0.34468	0.014903
Sgk085	-0.2118	0.078522	DYRK1B	-0.35083	0.017601
CDK2	-0.21478	0.081193	PKC[zeta]	-0.31723	0.024637
JNK1	-0.2252	0.084145	RSK3	0.331949	0.025418
JNK3	-0.2252	0.084145	MSK1~b	-0.30908	0.026541
ICK	-0.27561	0.085663	OSR1	0.357291	0.027048
SGK3	-0.21637	0.087307	PKN1/PRK1	-0.30844	0.028246
Akt1/PKB[alpha]	-0.33504	0.090899	IKK[epsilon]	-0.28728	0.034856
Yes	0.231415	0.0909	LRRK2	-0.29886	0.036921
CDKL3	0.533748	0.091639	ATR	-0.28829	0.037103
PKG1	-0.20398	0.095344	CDKL2	-0.2779	0.042651
STK33	-0.21011	0.095611	PRKY	-0.28072	0.044662
JNK2	-0.21796	0.097686	PKG2	-0.28494	0.045362
AAK1	0.264773	0.099027	Nek9	-0.27137	0.049926
			CDK9	-0.27285	0.050859
			Nek8	0.268948	0.052646
			JNK1	-0.26781	0.053037
			JNK3	-0.26781	0.053037
			MSK1	-0.26504	0.056134

			Yes	0.259054	0.059126
			JNK2	-0.2595	0.05956
			SGK1	0.264784	0.064847
			PINK1	-0.25802	0.067124
			PRKX	-0.25739	0.067709
			PAK1	0.256187	0.070622
			Pim2	0.256187	0.070622
			TBK1	-0.23438	0.080952
			CDK2	-0.2327	0.083682
			DYRK1A	-0.28546	0.086145
			MAST4	-0.23523	0.090885
			PKG1	-0.23272	0.09341
			DAPK3	-0.28789	0.095921
			Akt2/PKB[beta]	-0.25368	0.09792

8 REFERENCES

1. Abu-Farha, M., Lambert, J.P., Al-Madhoun, A.S., Elisma, F., Skerjanc, I.S., Figeys, D., 2008. The tale of two domains: proteomics and genomics analysis of SMYD2, a new histone methyltransferase. *Mol Cell Proteomics* 7, 560-572.
2. Adams, J.U., 2015. Genetics: Big hopes for big data. *Nature* 527, S108-109.
3. Ahluwalia, M.S., de Groot, J., Liu, W.M., Gladson, C.L., 2010. Targeting SRC in glioblastoma tumors and brain metastases: rationale and preclinical studies. *Cancer Lett* 298, 139-149.
4. Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19, 716-723.
5. Alcantara Llaguno, S., Chen, J., Kwon, C.H., Jackson, E.L., Li, Y., Burns, D.K., Alvarez-Buylla, A., Parada, L.F., 2009. Malignant astrocytomas originate from neural stem/progenitor cells in a somatic tumor suppressor mouse model. *Cancer Cell* 15, 45-56.
6. Alvarez, J.V., Mukherjee, N., Chakravarti, A., Robe, P., Zhai, G., Chakladar, A., Loeffler, J., Black, P., Frank, D.A., 2007. A STAT3 Gene Expression Signature in Gliomas is Associated with a Poor Prognosis. *Transl Oncogenomics* 2, 99-105.
7. Amani, V., Donson, A.M., Lummus, S.C., Prince, E.W., Griesinger, A.M., Witt, D.A., Hankinson, T.C., Handler, M.H., Dorris, K., Vibhakar, R., Foreman, N.K., Hoffman, L.M., 2017. Characterization of 2 Novel Ependymoma Cell Lines With Chromosome 1q Gain Derived From Posterior Fossa Tumors of Childhood. *J Neuropathol Exp Neurol* 76, 595-604.
8. Azar, W.J., Azar, S.H., Higgins, S., Hu, J.F., Hoffman, A.R., Newgreen, D.F., Werther, G.A., Russo, V.C., 2011. IGFBP-2 enhances VEGF gene promoter activity and consequent promotion of angiogenesis by neuroblastoma cells. *Endocrinology* 152, 3332-3342.
9. Azuaje, F., 2016. Computational models for predicting drug responses in cancer research. *Brief Bioinform.*
10. Bachoo, R.M., Maher, E.A., Ligon, K.L., Sharpless, N.E., Chan, S.S., You, M.J., Tang, Y., DeFrances, J., Stover, E., Weissleder, R., Rowitch, D.H., Louis, D.N., DePinho, R.A., 2002. Epidermal growth factor receptor and Ink4a/Arf: convergent mechanisms governing terminal differentiation and transformation along the neural stem cell to astrocyte axis. *Cancer cell* 1, 269-277.
11. Bae, B.I., Jayaraman, D., Walsh, C.A., 2015. Genetic changes shaping the human brain. *Dev Cell* 32, 423-434.
12. Baldewpersad Tewarie, N.M., Burgers, I.A., Dawood, Y., den Boon, H.C., den Brok, M.G., Klunder, J.H., Koopmans, K.B., Rademaker, E., van den Broek, H.B., van den Bersselaar, S.M., Witjes, J.J., Van Noorden, C.J., Atai, N.A., 2013. NADP⁺-dependent IDH1 R132 mutation and its relevance for glioma patient survival. *Med Hypotheses* 80, 728-731.
13. Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., Frohling, S., Chan, E.M., Sos, M.L., Michel, K., Mermel, C., Silver, S.J., Weir, B.A., Reiling, J.H., Sheng, Q., Gupta, P.B., Wadlow, R.C., Le, H., Hoersch, S., Wittner, B.S., Ramaswamy, S., Livingston, D.M., Sabatini, D.M., Meyerson, M., Thomas, R.K., Lander, E.S., Mesirov, J.P., Root, D.E., Gilliland, D.G., Jacks, T., Hahn, W.C., 2009. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* 462, 108-112.
14. Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C.L., Serova, N., Davis, S., Soboleva, A., 2013. NCBI

- GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* 41, D991-995.
15. Beier, D., Hau, P., Proescholdt, M., Lohmeier, A., Wischhusen, J., Oefner, P.J., Aigner, L., Brawanski, A., Bogdahn, U., Beier, C.P., 2007. CD133(+) and CD133(-) glioblastoma-derived cancer stem cells show differential growth characteristics and molecular profiles. *Cancer Res* 67, 4010-4015.
 16. Berishaj, M., Gao, S.P., Ahmed, S., Leslie, K., Al-Ahmadie, H., Gerald, W.L., Bornmann, W., Bromberg, J.F., 2007. Stat3 is tyrosine-phosphorylated through the interleukin-6/glycoprotein 130/Janus kinase pathway in breast cancer. *Breast Cancer Res* 9, R32.
 17. Bhat, K.P.L., Balasubramaniyan, V., Vaillant, B., Ezhilarasan, R., Hummelink, K., Hollingsworth, F., Wani, K., Heathcock, L., James, J.D., Goodman, L.D., Conroy, S., Long, L., Lelic, N., Wang, S., Gumin, J., Raj, D., Kodama, Y., Raghunathan, A., Olar, A., Joshi, K., Pelloso, C.E., Heimberger, A., Kim, S.H., Cahill, D.P., Rao, G., Den Dunnen, W.F.A., Boddeke, H., Phillips, H.S., Nakano, I., Lang, F.F., Colman, H., Sulman, E.P., Aldape, K., 2013. Mesenchymal differentiation mediated by NF-kappaB promotes radiation resistance in glioblastoma. *Cancer Cell* 24, 331-346.
 18. Bhowmick, N.A., Zent, R., Ghiassi, M., McDonnell, M., Moses, H.L., 2001. Integrin beta 1 signaling is necessary for transforming growth factor-beta activation of p38MAPK and epithelial plasticity. *J Biol Chem* 276, 46707-46713.
 19. Bigner, S.H., Mark, J., Bigner, D.D., 1990. Cytogenetics of human brain tumors. *Cancer Genet Cytogenet* 47, 141-154.
 20. Bini, S.A., 2018. Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? *J Arthroplasty*.
 21. Blagus, R., Lusa, L., 2010. Class prediction for high-dimensional class-imbalanced data. *BMC Bioinformatics* 11, 523.
 22. Boccaccio, C., Ando, M., Tamagnone, L., Bardelli, A., Michieli, P., Battistini, C., Comoglio, P.M., 1998. Induction of epithelial tubules by growth factor HGF depends on the STAT pathway. *Nature* 391, 285-288.
 23. Bolstad, B., 2018. preprocessCore: A collection of pre-processing functions.
 24. Bottcher, S.G., Dethlefsen, C., 2018. deal: Learning Bayesian Networks with Mixed Variables.
 25. Bowman, T., Broome, M.A., Sinibaldi, D., Wharton, W., Pledger, W.J., Sedivy, J.M., Irby, R., Yeatman, T., Courtneidge, S.A., Jove, R., 2001. Stat3-mediated Myc expression is required for Src transformation and PDGF-induced mitogenesis. *Proc Natl Acad Sci U S A* 98, 7319-7324.
 26. Bradburn, M.J., Clark, T.G., Love, S.B., Altman, D.G., 2003. Survival Analysis Part II: Multivariate data analysis – an introduction to concepts and methods. *British Journal of Cancer* 89, 431-436.
 27. Brantley, E.C., Nabors, L.B., Gillespie, G.Y., Choi, Y.H., Palmer, C.A., Harrison, K., Roarty, K., Benveniste, E.N., 2008. Loss of Protein Inhibitors of Activated STAT-3 Expression in Glioblastoma Multiforme Tumors: Implications for STAT-3 Activation and Gene Expression. *Clinical Cancer Research* 14, 4694-4704.
 28. Brennan, C.W., Verhaak, R.G., McKenna, A., Campos, B., Noshmehr, H., Salama, S.R., Zheng, S., Chakravarty, D., Sanborn, J.Z., Berman, S.H., Beroukhi, R., Bernard, B., Wu, C.J., Genovese, G., Shmulevich, I., Barnholtz-Sloan, J., Zou, L., Vegesna, R., Shukla, S.A., Ciriello, G., Yung, W.K., Zhang, W., Sougnez, C., Mikkelsen, T., Aldape, K., Bigner, D.D., Van Meir, E.G., Prados, M., Sloan, A., Black, K.L., Eschbacher, J., Finocchiaro, G., Friedman, W., Andrews, D.W., Guha, A., Iacocca, M., O'Neill, B.P., Foltz, G., Myers, J., Weisenberger, D.J., Penny, R., Kucherlapati, R., Perou, C.M., Hayes, D.N., Gibbs, R., Marra, M., Mills, G.B., Lander, E., Spellman, P., Wilson, R., Sander, C., Weinstein, J., Meyerson, M., Gabriel, S., Laird, P.W., Haussler, D., Getz,

- G., Chin, L., Network, T.R., 2013. The somatic genomic landscape of glioblastoma. *Cell* 155, 462-477.
29. Bromberg, J., Darnell, J.E., Jr., 2000. The role of STATs in transcriptional control and their impact on cellular function. *Oncogene* 19, 2468-2473.
 30. Busemeyer, J.R., Diederich, A., 2014. Estimation and Testing of Computational Psychological Models. 49-61.
 31. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., Satija, R., 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* 36, 411-420.
 32. Cahoy, J.D., Emery, B., Kaushal, A., Foo, L.C., Zamanian, J.L., Christopherson, K.S., Xing, Y., Lubischer, J.L., Krieg, P.A., Krupenko, S.A., Thompson, W.J., Barres, B.A., 2008. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J Neurosci* 28, 264-278.
 33. Cairncross, J.G., Ueki, K., Zlatescu, M.C., Lisle, D.K., Finkelstein, D.M., Hammond, R.R., Silver, J.S., Stark, P.C., Macdonald, D.R., Ino, Y., Ramsay, D.A., Louis, D.N., 1998. Specific genetic predictors of chemotherapeutic response and survival in patients with anaplastic oligodendrogliomas. *J Natl Cancer Inst* 90, 1473-1479.
 34. Canning, P., Park, K., Goncalves, J., Li, C., Howard, C.J., Sharpe, T.D., Holt, L.J., Pelletier, L., Bullock, A.N., Leroux, M.R., 2018. CDKL Family Kinases Have Evolved Distinct Structural Features and Ciliary Function. *Cell reports* 22, 885-894.
 35. Canoll, P., Goldman, J.E., 2008. The interface between glial progenitors and gliomas. *Acta Neuropathol* 116, 465-477.
 36. Carballo, M., Conde, M., El Bekay, R., Martín-Nieto, J., Camacho, M.a.J., Monteseirín, J., Conde, J., Bedoya, F.J., Sobrino, F., 1999. Oxidative Stress Triggers STAT3 Tyrosine Phosphorylation and Nuclear Translocation in Human Lymphocytes. *Journal of Biological Chemistry* 274, 17580-17586.
 37. Carrasco-Garcia, E., Auzmendi-Iriarte, J., Matheu, A., 2018. Integrin alpha7: a novel promising target in glioblastoma stem cells. *Stem Cell Investig* 5, 2.
 38. Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., Beroukhi, R., Pellman, D., Levine, D.A., Lander, E.S., Meyerson, M., Getz, G., 2012. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol* 30, 413-421.
 39. Ceccarelli, M., Barthel, F.P., Malta, T.M., Sabedot, T.S., Salama, S.R., Murray, B.A., Morozova, O., Newton, Y., Radenbaugh, A., Pagnotta, S.M., Anjum, S., Wang, J., Manyam, G., Zoppoli, P., Ling, S., Rao, A.A., Grifford, M., Cherniack, A.D., Zhang, H., Poisson, L., Carlotti, C.G., Jr., Tirapelli, D.P., Rao, A., Mikkelsen, T., Lau, C.C., Yung, W.K., Rabadan, R., Huse, J., Brat, D.J., Lehman, N.L., Barnholtz-Sloan, J.S., Zheng, S., Hess, K., Rao, G., Meyerson, M., Beroukhi, R., Cooper, L., Akbani, R., Wrensch, M., Haussler, D., Aldape, K.D., Laird, P.W., Gutmann, D.H., Network, T.R., Nounshmehr, H., Iavarone, A., Verhaak, R.G., 2016. Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell* 164, 550-563.
 40. Chakrabarty, S., Kondratik, L., 2006. Insulin-like growth factor binding protein-2 stimulates proliferation and activates multiple cascades of the mitogen-activated protein kinase pathways in NIH-OVCAR3 human epithelial ovarian cancer cells. *Cancer Biol Ther* 5, 189-197.
 41. Chang, C.Y., Li, M.C., Liao, S.L., Huang, Y.L., Shen, C.C., Pan, H.C., 2005. Prognostic and clinical implication of IL-6 expression in glioblastoma multiforme. *J Clin Neurosci* 12, 930-933.

42. Chatterjee, S., Park, E.S., Soloff, M.S., 2004. Proliferation of DU145 prostate cancer cells is inhibited by suppressing insulin-like growth factor binding protein-2. *Int J Urol* 11, 876-884.
43. Chen, C., Grennan, K., Badner, J., Zhang, D., Gershon, E., Jin, L., Liu, C., 2011. Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods. *PloS one* 6, e17238.
44. Chen, X., Chen, M., Ning, K., 2006. BNArray: an R package for constructing gene regulatory networks from microarray data by using Bayesian network. *Bioinformatics* 22, 2952-2954.
45. Chong, Y.K., Toh, T.B., Zaiden, N., Poonepalli, A., Leong, S.H., Ong, C.E., Yu, Y., Tan, P.B., See, S.J., Ng, W.H., Ng, I., Hande, M.P., Kon, O.L., Ang, B.T., Tang, C., 2009. Cryopreservation of neurospheres derived from human glioblastoma multiforme. *Stem Cells* 27, 29-39.
46. Chong, Y.K., Sandanaraj, E., Koh, L.W., Thangaveloo, M., Tan, M.S., Koh, G.R., Toh, T.B., Lim, G.G., Holbrook, J.D., Kon, O.L., Nadarajah, M., Ng, I., Ng, W.H., Tan, N.S., Lim, K.L., Tang, C., Ang, B.T., 2016. ST3GAL1-Associated Transcriptomic Program in Glioblastoma Tumor Growth, Invasion, and Prognosis. *J Natl Cancer Inst* 108, djv326.
47. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., Prior, F., 2013. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging* 26, 1045-1057.
48. Couto, J.P., Almeida, A., Daly, L., Sobrinho-Simoes, M., Bromberg, J.F., Soares, P., 2012. AZD1480 blocks growth and tumorigenesis of RET- activated thyroid cancer cell lines. *PloS one* 7, e46869.
49. Cox, D.R., 1972. Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)* 34, 187-220.
50. Dabney, A.R., 2005. Classification of microarrays to nearest centroids. *Bioinformatics* 21, 4148-4154.
51. Dabney, A.R., 2006. ClaNC: point-and-click software for classifying microarrays to nearest centroids. *Bioinformatics* 22, 122-123.
52. Dall'Olio, F., Chiricolo, M., 2001. Sialyltransferases in cancer. *Glycoconj J* 18, 841-850.
53. Danussi, C., Bose, P., Parthasarathy, P.T., Silberman, P.C., Van Arnem, J.S., Vitucci, M., Tang, O.Y., Heguy, A., Wang, Y., Chan, T.A., Riggins, G.J., Sulman, E.P., Lang, F., Creighton, C.J., Deneen, B., Miller, C.R., Picketts, D.J., Kannan, K., Huse, J.T., 2018. Atrx inactivation drives disease-defining phenotypes in glioma cells of origin through global epigenomic remodeling. *Nat Commun* 9, 1057.
54. Darnell, J.E., Jr., Kerr, I.M., Stark, G.R., 1994. Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science* 264, 1415-1421.
55. Darst, B.F., Malecki, K.C., Engelman, C.D., 2018. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC Genet* 19, 65.
56. Davis, S., Meltzer, P.S., 2007. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 23, 1846-1847.
57. Dawson, G., Moskal, J.R., Dawson, S.A., 2004. Transfection of 2,6 and 2,3-sialyltransferase genes and GlcNAc-transferase genes into human glioma cell line U-373 MG affects glycoconjugate expression and enhances cell death. *J Neurochem* 89, 1436-1444.
58. Day, B.W., Stringer, B.W., Al-Ejeh, F., Ting, M.J., Wilson, J., Ensbey, K.S., Jamieson, P.R., Bruce, Z.C., Lim, Y.C., Offenhauser, C., Charmsaz, S., Cooper, L.T., Ellacott, J.K., Harding, A., Leveque, L., Inglis, P., Allan, S., Walker, D.G., Lackmann, M., Osborne, G., Khanna, K.K., Reynolds, B.A.,

- Lickliter, J.D., Boyd, A.W., 2013. EphA3 maintains tumorigenicity and is a therapeutic target in glioblastoma multiforme. *Cancer Cell* 23, 238-248.
59. Dazard, J.E., Gal, H., Amariglio, N., Rechavi, G., Domany, E., Givol, D., 2003. Genome-wide comparison of human keratinocyte and squamous cell carcinoma responses to UVB irradiation: implications for skin and epithelial cancer. *Oncogene* 22, 2993-3006.
 60. de Groot, J., Liang, J., Kong, L.Y., Wei, J., Piao, Y., Fuller, G., Qiao, W., Heimberger, A.B., 2012. Modulating antiangiogenic resistance by inhibiting the signal transducer and activator of transcription 3 pathway in glioblastoma. *Oncotarget* 3, 1036-1048.
 61. de la Iglesia, N., Puram, S.V., Bonni, A., 2009. STAT3 regulation of glioblastoma pathogenesis. *Curr Mol Med* 9, 580-590.
 62. de la Iglesia, N., Konopka, G., Puram, S.V., Chan, J.A., Bachoo, R.M., You, M.J., Levy, D.E., Depinho, R.A., Bonni, A., 2008. Identification of a PTEN-regulated STAT3 brain tumor suppressor pathway. *Genes Dev* 22, 449-462.
 63. De Witt Hamer, P.C., Van Tilborg, A.A., Eijk, P.P., Sminia, P., Troost, D., Van Noorden, C.J., Ylstra, B., Leenstra, S., 2008. The genomic profile of human malignant glioma is altered early in primary cell culture and preserved in spheroids. *Oncogene* 27, 2091-2096.
 64. Derenzini, E., Lemoine, M., Buglio, D., Katayama, H., Ji, Y., Davis, R.E., Sen, S., Younes, A., 2011. The JAK inhibitor AZD1480 regulates proliferation and immunity in Hodgkin lymphoma. *Blood Cancer J* 1, e46.
 65. Despont, J.P., Abel, C.A., Grey, H.M., 1975. Sialic acids and sialyltransferases in murine lymphoid cells: indicators of T cell maturation. *Cell Immunol* 17, 487-494.
 66. Dolecek, T.A., Propp, J.M., Stroup, N.E., Kruchko, C., 2012. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2005-2009. *Neuro Oncol* 14 Suppl 5, v1-49.
 67. Dominguez-Brauer, C., Thu, K.L., Mason, J.M., Blaser, H., Bray, M.R., Mak, T.W., 2015. Targeting Mitosis in Cancer: Emerging Strategies. *Mol Cell* 60, 524-536.
 68. Du, P., Kibbe, W.A., Lin, S.M., 2008. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* 24, 1547-1548.
 69. Durbin, B.P., Hardin, J.S., Hawkins, D.M., Rocke, D.M., 2002. A variance-stabilizing transformation for gene-expression microarray data. *Bioinformatics* 18 Suppl 1, S105-110.
 70. Eckel-Passow, J.E., Lachance, D.H., Molinaro, A.M., Walsh, K.M., Decker, P.A., Sicotte, H., Pekmezci, M., Rice, T., Kosel, M.L., Smirnov, I.V., Sarkar, G., Caron, A.A., Kollmeyer, T.M., Praska, C.E., Chada, A.R., Halder, C., Hansen, H.M., McCoy, L.S., Bracci, P.M., Marshall, R., Zheng, S., Reis, G.F., Pico, A.R., O'Neill, B.P., Buckner, J.C., Giannini, C., Huse, J.T., Perry, A., Tihan, T., Berger, M.S., Chang, S.M., Prados, M.D., Wiemels, J., Wiencke, J.K., Wrensch, M.R., Jenkins, R.B., 2015. Glioma Groups Based on 1p/19q, IDH, and TERT Promoter Mutations in Tumors. *N Engl J Med* 372, 2499-2508.
 71. Efron, B., Tibshirani, R., 2002. Empirical bayes methods and false discovery rates for microarrays. *Genet Epidemiol* 23, 70-86.
 72. Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D., 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95, 14863-14868.
 73. Eklund, A., 2016. beeswarm: The Bee Swarm Plot, an Alternative to Stripchart.
 74. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S., 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115-118.
 75. Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., Milacic, M., Roca, C.D., Rothfels,

- K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Viteri, G., Weiser, J., Wu, G., Stein, L., Hermjakob, H., D'Eustachio, P., 2018. The Reactome Pathway Knowledgebase. *Nucleic Acids Res* 46, D649-D655.
76. Foong, C.S., Sandanaraj, E., Brooks, H.B., Campbell, R.M., Ang, B.T., Chong, Y.K., Tang, C., 2012. Glioma-Propagating Cells as an In Vitro Screening Platform: PLK1 as a Case Study. *J Biomol Screen* 17, 1136-1150.
 77. Foong, C.S., Ng, F.S., Phong, M., Toh, T.B., Chong, Y.K., Tucker-Kellogg, G., Campbell, R.M., Ang, B.T., Tang, C., 2011. Cryopreservation of cancer-initiating cells derived from glioblastoma. *Front Biosci (Schol Ed)* 3, 698-708.
 78. Franks, J.M., Cai, G., Whitfield, M.L., 2018. Feature specific quantile normalization enables cross-platform classification of molecular subtypes using gene expression data. *Bioinformatics* 34, 1868-1874.
 79. Freije, W.A., Castro-Vargas, F.E., Fang, Z., Horvath, S., Cloughesy, T., Liao, L.M., Mischel, P.S., Nelson, S.F., 2004. Gene expression profiling of gliomas strongly predicts survival. *Cancer Res* 64, 6503-6510.
 80. Friedrich Leisch, E.D., 2010. mlbench: Machine Learning Benchmark Problems.
 81. Friendly, M., 1994. Mosaic displays for multi-way contingency tables. *Journal of the American Statistical Association* 89, 190-200.
 82. Fukuda, M., 1995. Carbohydrate-dependent cell adhesion. *Bioorg Med Chem* 3, 207-215.
 83. Fukushima, T., Tezuka, T., Shimomura, T., Nakano, S., Kataoka, H., 2007. Silencing of insulin-like growth factor-binding protein-2 in human glioblastoma cells reduces both invasiveness and expression of progression-associated gene CD24. *The Journal of biological chemistry* 282, 18634-18644.
 84. Furnari, F.B., Fenton, T., Bachoo, R.M., Mukasa, A., Stommel, J.M., Stegh, A., Hahn, W.C., Ligon, K.L., Louis, D.N., Brennan, C., Chin, L., DePinho, R.A., Cavenee, W.K., 2007. Malignant astrocytic glioma: genetics, biology, and paths to treatment. *Genes Dev* 21, 2683-2710.
 85. Gabrusiewicz, K., Rodriguez, B., Wei, J., Hashimoto, Y., Healy, L.M., Maiti, S.N., Thomas, G., Zhou, S., Wang, Q., Elakkad, A., Liebelt, B.D., Yaghi, N.K., Ezhilarasan, R., Huang, N., Weinberg, J.S., Prabhu, S.S., Rao, G., Sawaya, R., Langford, L.A., Bruner, J.M., Fuller, G.N., Bar-Or, A., Li, W., Colen, R.R., Curran, M.A., Bhat, K.P., Antel, J.P., Cooper, L.J., Sulman, E.P., Heimberger, A.B., 2016. Glioblastoma-infiltrated innate immune cells resemble M0 macrophage phenotype. *JCI Insight* 1.
 86. Galli, R., Binda, E., Orfanelli, U., Cipelletti, B., Gritti, A., De Vitis, S., Fiocco, R., Foroni, C., Dimeco, F., Vescovi, A., 2004. Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma. *Cancer Res* 64, 7011-7021.
 87. Gao, S.P., Mark, K.G., Leslie, K., Pao, W., Motoi, N., Gerald, W.L., Travis, W.D., Bornmann, W., Veatch, D., Clarkson, B., Bromberg, J.F., 2007. Mutations in the EGFR kinase domain mediate STAT3 activation via IL-6 production in human lung adenocarcinomas. *J Clin Invest* 117, 3846-3856.
 88. Gautier, L., Cope, L., Bolstad, B.M., Irizarry, R.A., 2004. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20, 307-315.
 89. Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell* 6, 721-741.
 90. Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A.J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J.Y., Zhang, J., 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5, R80.

91. Geron, I., Abrahamsson, A.E., Barroga, C.F., Kavalerchik, E., Gotlib, J., Hood, J.D., Durocher, J., Mak, C.C., Noronha, G., Soll, R.M., Tefferi, A., Kaushansky, K., Jamieson, C.H., 2008. Selective inhibition of JAK2-driven erythroid differentiation of polycythemia vera progenitors. *Cancer Cell* 13, 321-330.
92. Gillespie, W., Paulson, J.C., Kelm, S., Pang, M., Baum, L.G., 1993. Regulation of alpha 2,3-sialyltransferase expression correlates with conversion of peanut agglutinin (PNA)+ to PNA- phenotype in developing thymocytes. *J Biol Chem* 268, 3801-3804.
93. Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., Lander, E.S., 1999. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286, 531-537.
94. Gonda, D.D., Cheung, V.J., Muller, K.A., Goyal, A., Carter, B.S., Chen, C.C., 2014. The Cancer Genome Atlas expression profiles of low-grade gliomas. *Neurosurg Focus* 36, E23.
95. Gravendeel, L.A., Kouwenhoven, M.C., Gevaert, O., de Rooi, J.J., Stubbs, A.P., Duijm, J.E., Daemen, A., Bleeker, F.E., Bralten, L.B., Kloosterhof, N.K., De Moor, B., Eilers, P.H., van der Spek, P.J., Kros, J.M., Sillevs Smitt, P.A., van den Bent, M.J., French, P.J., 2009. Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology. *Cancer Res* 69, 9065-9072.
96. Grys, B.T., Lo, D.S., Sahin, N., Kraus, O.Z., Morris, Q., Boone, C., Andrews, B.J., 2017. Machine learning and computer vision approaches for phenotypic profiling. *J Cell Biol* 216, 65-71.
97. Gudernova, I., Balek, L., Varecha, M., Kucerova, J.F., Kunova Bosakova, M., Fafulek, B., Palusova, V., Uldrijan, S., Trantirek, L., Krejci, P., 2017. Inhibitor repurposing reveals ALK, LTK, FGFR, RET and TRK kinases as the targets of AZD1480. *Oncotarget* 8, 109319-109331.
98. Gunther, H.S., Schmidt, N.O., Phillips, H.S., Kemming, D., Kharbanda, S., Soriano, R., Modrusan, Z., Meissner, H., Westphal, M., Lamszus, K., 2008. Glioblastoma-derived stem cell-enriched cultures form distinct subgroups according to molecular and phenotypic criteria. *Oncogene* 27, 2897-2909.
99. Hakomori, S., 2002. Glycosylation defining cancer malignancy: new wine in an old bottle. *Proc Natl Acad Sci U S A* 99, 10231-10233.
100. Haltiwanger, R.S., Lowe, J.B., 2004. Role of glycosylation in development. *Annu Rev Biochem* 73, 491-537.
101. Hedlund, M., Ng, E., Varki, A., Varki, N.M., 2008. alpha 2-6-Linked sialic acids on N-glycans modulate carcinoma differentiation in vivo. *Cancer Res* 68, 388-394.
102. Hedvat, M., Huszar, D., Herrmann, A., Gozgit, J.M., Schroeder, A., Sheehy, A., Buettner, R., Proia, D., Kowolik, C.M., Xin, H., Armstrong, B., Beberitz, G., Weng, S., Wang, L., Ye, M., McEachern, K., Chen, H., Morosini, D., Bell, K., Alimzhanov, M., Ioannidis, S., McCoon, P., Cao, Z.A., Yu, H., Jove, R., Zinda, M., 2009a. The JAK2 Inhibitor, AZD1480, Potently Blocks Stat3 Signaling and Oncogenesis in Solid Tumors. *Cancer cell* 16, 487-497.
103. Hegi, M.E., Liu, L., Herman, J.G., Stupp, R., Wick, W., Weller, M., Mehta, M.P., Gilbert, M.R., 2008. Correlation of O6-methylguanine methyltransferase (MGMT) promoter methylation with clinical outcomes in glioblastoma and clinical strategies to modulate MGMT activity. *J Clin Oncol* 26, 4189-4199.
104. Hirose, Y., Kreklau, E.L., Erickson, L.C., Berger, M.S., Pieper, R.O., 2003. Delayed repletion of O6-methylguanine-DNA methyltransferase resulting in failure to protect the human glioblastoma cell line SF767 from temozolomide-induced cytotoxicity. *J Neurosurg* 98, 591-598.
105. Holmes, K.M., Annala, M., Chua, C.Y., Dunlap, S.M., Liu, Y., Hugen, N., Moore, L.M., Cogdell, D., Hu, L., Nykter, M., Hess, K., Fuller, G.N., Zhang, W., 2012. Insulin-like growth factor-binding protein 2-driven glioma progression is

- prevented by blocking a clinically significant integrin, integrin-linked kinase, and NF-kappaB network. *Proc Natl Acad Sci U S A* 109, 3475-3480.
106. Hornbeck, P.V., Kornhauser, J.M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V., Sullivan, M., 2012. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res* 40, D261-270.
 107. Hoshida, Y., 2010. Nearest template prediction: a single-sample-based flexible class prediction with confidence assessment. *PloS one* 5, e15543.
 108. Hubbell, E., Liu, W.M., Mei, R., 2002. Robust estimators for expression analysis. *Bioinformatics* 18, 1585-1592.
 109. Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A., Vingron, M., 2002. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* 18 Suppl 1, S96-104.
 110. Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K.D., Irizarry, R.A., Lawrence, M., Love, M.I., MacDonald, J., Obenchain, V., Oles, A.K., Pages, H., Reyes, A., Shannon, P., Smyth, G.K., Tenenbaum, D., Waldron, L., Morgan, M., 2015. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods* 12, 115-121.
 111. IvyGBMAtlas, 2018. Anatomic Gene Expression Atlas Fuels Glioblastoma Discovery. *Cancer Discov*.
 112. Iwamaru, A., Szymanski, S., Iwado, E., Aoki, H., Yokoyama, T., Fokt, I., Hess, K., Conrad, C., Madden, T., Sawaya, R., Kondo, S., Priebe, W., Kondo, Y., 2006. A novel inhibitor of the STAT3 pathway induces apoptosis in malignant glioma cells both in vitro and in vivo. *Oncogene* 26, 2435-2444.
 113. Jahani-Asl, A., Yin, H., Soleimani, V.D., Haque, T., Luchman, H.A., Chang, N.C., Sincennes, M.C., Puram, S.V., Scott, A.M., Lorimer, I.A., Perkins, T.J., Ligon, K.L., Weiss, S., Rudnicki, M.A., Bonni, A., 2016. Control of glioblastoma tumorigenesis by feed-forward cytokine signaling. *Nature neuroscience* 19, 798-806.
 114. Jamieson, C.H., Gotlib, J., Durocher, J.A., Chao, M.P., Mariappan, M.R., Lay, M., Jones, C., Zehnder, J.L., Lilleberg, S.L., Weissman, I.L., 2006. The JAK2 V617F mutation occurs in hematopoietic stem cells in polycythemia vera and predisposes toward erythroid differentiation. *Proc Natl Acad Sci U S A* 103, 6224-6229.
 115. Jawhari, S., Ratinaud, M.H., Verdier, M., 2016. Glioblastoma, hypoxia and autophagy: a survival-prone 'menage-a-trois'. *Cell Death Dis* 7, e2434.
 116. Jenkins, R.B., Wensch, M.R., Johnson, D., Fridley, B.L., Decker, P.A., Xiao, Y., Kollmeyer, T.M., Rynearson, A.L., Fink, S., Rice, T., McCoy, L.S., Halder, C., Kosel, M.L., Giannini, C., Tihan, T., O'Neill, B.P., Lachance, D.H., Yang, P., Wiemels, J., Wiencke, J.K., 2011. Distinct germ line polymorphisms underlie glioma morphologic heterogeneity. *Cancer Genet* 204, 13-18.
 117. Jin, X., Kim, L.J.Y., Wu, Q., Wallace, L.C., Prager, B.C., Sanvoranart, T., Gimple, R.C., Wang, X., Mack, S.C., Miller, T.E., Huang, P., Valentim, C.L., Zhou, Q.G., Barnholtz-Sloan, J.S., Bao, S., Sloan, A.E., Rich, J.N., 2017. Targeting glioma stem cells through combined BMI1 and EZH2 inhibition. *Nat Med* 23, 1352-1361.
 118. Joel, M., Mughal, A.A., Grieg, Z., Murrell, W., Palmero, S., Mikkelsen, B., Fjerdingsstad, H.B., Sandberg, C.J., Behnan, J., Glover, J.C., Langmoen, I.A., Stangeland, B., 2015. Targeting PBK/TOPK decreases growth and survival of glioma initiating cells in vitro and attenuates tumor growth in vivo. *Mol Cancer* 14, 121.

119. Joensuu, H., Pupa, M., Sihto, H., Tynnen, O., Nupponen, N.N., 2005. Amplification of genes encoding KIT, PDGFR α and VEGFR2 receptor tyrosine kinases is frequent in glioblastoma multiforme. *J Pathol* 207, 224-231.
120. John J. Dziak, D.L.C., Stephanie T. Lanza, Runze Li, 2012. Sensitivity and specificity of information criteria. In: Dziak, J. (Ed.), *The Methodology Center. The Pennsylvania State University, College of Health and Human Development*, pp. 1-30.
121. Johnson, B.E., Mazar, T., Hong, C., Barnes, M., Aihara, K., McLean, C.Y., Fouse, S.D., Yamamoto, S., Ueda, H., Tatsuno, K., Asthana, S., Jalbert, L.E., Nelson, S.J., Bollen, A.W., Gustafson, W.C., Charron, E., Weiss, W.A., Smirnov, I.V., Song, J.S., Olshen, A.B., Cha, S., Zhao, Y., Moore, R.A., Mungall, A.J., Jones, S.J., Hirst, M., Marra, M.A., Saito, N., Aburatani, H., Mukasa, A., Berger, M.S., Chang, S.M., Taylor, B.S., Costello, J.F., 2014. Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science* 343, 189-193.
122. Johnson, W.E., Li, C., Rabinovic, A., 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8, 118-127.
123. Joo, K.M., Kim, J., Jin, J., Kim, M., Seol, H.J., Muradov, J., Yang, H., Choi, Y.L., Park, W.Y., Kong, D.S., Lee, J.I., Ko, Y.H., Woo, H.G., Lee, J., Kim, S., Nam, D.H., 2013. Patient-specific orthotopic glioblastoma xenograft models recapitulate the histopathology and biology of human glioblastomas in situ. *Cell reports* 3, 260-273.
124. Joo, K.M., Jin, J., Kim, E., Ho Kim, K., Kim, Y., Gu Kang, B., Kang, Y.J., Lathia, J.D., Cheong, K.H., Song, P.H., Kim, H., Seol, H.J., Kong, D.S., Lee, J.I., Rich, J.N., Lee, J., Nam, D.H., 2012. MET signaling regulates glioblastoma stem cells. *Cancer Res* 72, 3828-3838.
125. Joseph, J.V., Conroy, S., Tomar, T., Eggens-Meijer, E., Bhat, K., Copray, S., Walenkamp, A.M., Boddeke, E., Balasubramanyan, V., Wagemakers, M., den Dunnen, W.F., Kruij, F.A., 2014. TGF- β is an inducer of ZEB1-dependent mesenchymal transdifferentiation in glioblastoma that is associated with tumor invasion. *Cell Death Dis* 5, e1443.
126. Jung, T.Y., Jung, S., Ryu, H.H., Jeong, Y.I., Jin, Y.H., Jin, S.G., Kim, I.Y., Kang, S.S., Kim, H.S., 2008. Role of galectin-1 in migration and invasion of human glioblastoma multiforme cell lines. *J Neurosurg* 109, 273-284.
127. Kang, S.H., Yu, M.O., Park, K.J., Chi, S.G., Park, D.H., Chung, Y.G., 2010. Activated STAT3 regulates hypoxia-induced angiogenesis and cell migration in human glioblastoma. *Neurosurgery* 67, 1386-1395; discussion 1395.
128. Kannagi, R., Izawa, M., Koike, T., Miyazaki, K., Kimura, N., 2004. Carbohydrate-mediated cell adhesion in cancer metastasis and angiogenesis. *Cancer Sci* 95, 377-384.
129. Kaplan, E.L., Meier, P., 1958. Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association* 53, 457.
130. Kass, R.E., Raftery, A.E., 1995. Bayes Factors. *Journal of the American Statistical Association* 90, 773-795.
131. Katsetos, C.D., Reginato, M.J., Baas, P.W., D'Agostino, L., Legido, A., Tuszynski, J.A., Draberova, E., Draber, P., 2015. Emerging microtubule targets in glioma therapy. *Semin Pediatr Neurol* 22, 49-72.
132. Kauffmann-Zeh, A., Rodriguez-Vician, P., Ulrich, E., Gilbert, C., Coffey, P., Downward, J., Evan, G., 1997. Suppression of c-Myc-induced apoptosis by Ras signalling through PI(3)K and PKB. *Nature* 385, 544-548.
133. Kenta Masui, P.S.M., Guido Reifenberger, 2016. Molecular classification of gliomas. In: Weller, M.S.B.a.M. (Ed.), *Handbook of Clinical Neurology*. Elsevier, pp. 97-120.
134. Kerr, G., Ruskin, H.J., Crane, M., Doolan, P., 2008. Techniques for clustering gene expression data. *Comput Biol Med* 38, 283-293.

135. Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., Balakrishnan, L., Marimuthu, A., Banerjee, S., Somanathan, D.S., Sebastian, A., Rani, S., Ray, S., Harrys Kishore, C.J., Kanth, S., Ahmed, M., Kashyap, M.K., Mohmood, R., Ramachandra, Y.L., Krishna, V., Rahiman, B.A., Mohan, S., Ranganathan, P., Ramabadran, S., Chaerkady, R., Pandey, A., 2009. Human Protein Reference Database--2009 update. *Nucleic Acids Res* 37, D767-772.
136. Keunen, O., Johansson, M., Oudin, A., Sanzey, M., Rahim, S.A., Fack, F., Thorsen, F., Taxt, T., Bartos, M., Jirik, R., Miletic, H., Wang, J., Stieber, D., Stuhr, L., Moen, I., Rygh, C.B., Bjerkvig, R., Niclou, S.P., 2011. Anti-VEGF treatment reduces blood supply and increases tumor cell invasion in glioblastoma. *Proc Natl Acad Sci U S A* 108, 3749-3754.
137. Kim, E., Kim, M., Woo, D.H., Shin, Y., Shin, J., Chang, N., Oh, Y.T., Kim, H., Rheey, J., Nakano, I., Lee, C., Joo, K.M., Rich, J.N., Nam, D.H., Lee, J., 2013. Phosphorylation of EZH2 activates STAT3 signaling via STAT3 methylation and promotes tumorigenicity of glioblastoma stem-like cells. *Cancer Cell* 23, 839-852.
138. Kim, J., Lee, I.H., Cho, H.J., Park, C.K., Jung, Y.S., Kim, Y., Nam, S.H., Kim, B.S., Johnson, M.D., Kong, D.S., Seol, H.J., Lee, J.I., Joo, K.M., Yoon, Y., Park, W.Y., Lee, J., Park, P.J., Nam, D.H., 2015. Spatiotemporal Evolution of the Primary Glioblastoma Genome. *Cancer Cell* 28, 318-328.
139. Kim, T.M., Huang, W., Park, R., Park, P.J., Johnson, M.D., 2011. A developmental taxonomy of glioblastoma defined and maintained by MicroRNAs. *Cancer Res* 71, 3387-3399.
140. Klaeger, S., Heinzlmeir, S., Wilhelm, M., Polzer, H., Vick, B., Koenig, P.A., Reinecke, M., Ruprecht, B., Petzoldt, S., Meng, C., Zecha, J., Reiter, K., Qiao, H., Helm, D., Koch, H., Schoof, M., Canevari, G., Casale, E., Depaolini, S.R., Feuchtinger, A., Wu, Z., Schmidt, T., Rueckert, L., Becker, W., Huenges, J., Garz, A.K., Gohlke, B.O., Zolg, D.P., Kayser, G., Vooder, T., Preissner, R., Hahne, H., Tonisson, N., Kramer, K., Gotze, K., Bassermann, F., Schlegl, J., Ehrlich, H.C., Aiche, S., Walch, A., Greif, P.A., Schneider, S., Felder, E.R., Ruland, J., Medard, G., Jeremias, I., Spiekermann, K., Kuster, B., 2017. The target landscape of clinical kinase drugs. *Science* 358.
141. Koh, L.W., Koh, G.R., Ng, F.S., Toh, T.B., Sandanaraj, E., Chong, Y.K., Phong, M., Tucker-Kellogg, G., Kon, O.L., Ng, W.H., Ng, I.H., Clement, M.V., Pervaiz, S., Ang, B.T., Tang, C.S., 2013. A distinct reactive oxygen species profile confers chemoresistance in glioma-propagating cells and associates with patient survival outcome. *Antioxidants & redox signaling* 19, 2261-2279.
142. Köhn, H.-F., Hubert, L.J., 2015. Hierarchical Cluster Analysis. 1-13.
143. Kolde, R., 2019. pheatmap: Pretty Heatmaps.
144. Kono, M., Ohyama, Y., Lee, Y.C., Hamamoto, T., Kojima, N., Tsuji, S., 1997. Mouse beta-galactoside alpha 2,3-sialyltransferases: comparison of in vitro substrate specificities and tissue specific expression. *Glycobiology* 7, 469-479.
145. Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., Amin, V., Whitaker, J.W., Schultz, M.D., Ward, L.D., Sarkar, A., Quon, G., Sandstrom, R.S., Eaton, M.L., Wu, Y.C., Pfenning, A.R., Wang, X., Claussnitzer, M., Liu, Y., Coarfa, C., Harris, R.A., Shores, N., Epstein, C.B., Gjoneska, E., Leung, D., Xie, W., Hawkins, R.D., Lister, R., Hong, C., Gascard, P., Mungall, A.J., Moore, R., Chuah, E., Tam, A., Canfield, T.K., Hansen, R.S., Kaul, R., Sabo, P.J., Bansal, M.S., Carles, A., Dixon, J.R., Farh, K.H., Feizi, S., Karlic, R., Kim, A.R., Kulkarni, A., Li, D., Lowdon, R., Elliott, G., Mercer, T.R., Neph, S.J., Onuchic, V., Polak, P., Rajagopal, N., Ray, P., Sallari, R.C., Siebenthall, K.T., Sinnott-Armstrong, N.A., Stevens, M., Thurman, R.E., Wu, J., Zhang, B., Zhou, X., Beaudet, A.E., Boyer,

- L.A., De Jager, P.L., Farnham, P.J., Fisher, S.J., Haussler, D., Jones, S.J., Li, W., Marra, M.A., McManus, M.T., Sunyaev, S., Thomson, J.A., Tlsty, T.D., Tsai, L.H., Wang, W., Waterland, R.A., Zhang, M.Q., Chadwick, L.H., Bernstein, B.E., Costello, J.F., Ecker, J.R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B., Stamatoyannopoulos, J.A., Wang, T., Kellis, M., 2015. Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317-330.
146. Labots, M., Gotink, K.J., Dekker, H., Azijli, K., van der Mijn, J.C., Huijts, C.M., Piersma, S.R., Jimenez, C.R., Verheul, H.M., 2016. Evaluation of a tyrosine kinase peptide microarray for tyrosine kinase inhibitor therapy selection in cancer. *Exp Mol Med* 48, e279.
 147. Lam, L.T., Wright, G., Davis, R.E., Lenz, G., Farinha, P., Dang, L., Chan, J.W., Rosenwald, A., Gascoyne, R.D., Staudt, L.M., 2008. Cooperative signaling through the signal transducer and activator of transcription 3 and nuclear factor- κ B pathways in subtypes of diffuse large B-cell lymphoma. *Blood* 111, 3701-3713.
 148. Lamb, J., 2007. The Connectivity Map: a new tool for biomedical research. *Nat Rev Cancer* 7, 54-60.
 149. Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N., Reich, M., Hieronymus, H., Wei, G., Armstrong, S.A., Haggarty, S.J., Clemons, P.A., Wei, R., Carr, S.A., Lander, E.S., Golub, T.R., 2006. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929-1935.
 150. Lamouille, S., Derynck, R., 2007. Cell size and invasion in TGF-beta-induced epithelial to mesenchymal transition is regulated by activation of the mTOR pathway. *J Cell Biol* 178, 437-451.
 151. Laplante, P., Raymond, M.A., Labelle, A., Abe, J., Iozzo, R.V., Hebert, M.J., 2006. Perlecan proteolysis induces an alpha2beta1 integrin- and Src family kinase-dependent anti-apoptotic pathway in fibroblasts in the absence of focal adhesion kinase activation. *J Biol Chem* 281, 30383-30392.
 152. Lassman, A.B., Pugh, S.L., Gilbert, M.R., Aldape, K.D., Geinoz, S., Beumer, J.H., Christner, S.M., Komaki, R., DeAngelis, L.M., Gaur, R., Youssef, E., Wagner, H., Won, M., Mehta, M.P., 2015. Phase 2 trial of dasatinib in target-selected patients with recurrent glioblastoma (RTOG 0627). *Neuro Oncol* 17, 992-998.
 153. Lazar, C., Meganck, S., Taminiau, J., Steenhoff, D., Coletta, A., Molter, C., Weiss-Solis, D.Y., Duque, R., Bersini, H., Nowe, A., 2013. Batch effect removal methods for microarray gene expression data integration: a survey. *Brief Bioinform* 14, 469-490.
 154. Lee, J., Kotliarova, S., Kotliarov, Y., Li, A., Su, Q., Donin, N.M., Pastorino, S., Purow, B.W., Christopher, N., Zhang, W., Park, J.K., Fine, H.A., 2006. Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. *Cancer cell* 9, 391-403.
 155. Lee, P.Y., Chen, C.L., Lin, Z.Z., Cheng, A.L., Chen, E.I., Whang-Peng, J., Huang, C.Y., 2013. The Aurora kinases inhibitor VE-465 is a novel treatment for glioblastoma multiforme. *Oncology* 84, 326-335.
 156. Leek, J.T., Johnson, W.E., Parker, H.S., Fertig, E.J., Jaffe, A.E., Storey, J.D., Zhang, Y., Torres, L.C., 2019. sva: Surrogate Variable Analysis.
 157. Levenson, A.S., Kliakhandler, I.L., Svoboda, K.M., Pease, K.M., Kaiser, S.A., Ward, J.E., 3rd, Jordan, V.C., 2002. Molecular classification of selective oestrogen receptor modulators on the basis of gene expression profiles of breast cancer cells expressing oestrogen receptor alpha. *Br J Cancer* 87, 449-456.

158. Levine, R.L., Gilliland, D.G., 2008. Myeloproliferative disorders. *Blood* 112, 2190-2198.
159. Levy, D.E., Darnell, J.E., Jr., 2002. Stats: transcriptional control and biological impact. *Nat Rev Mol Cell Biol* 3, 651-662.
160. Li, R., Li, G., Deng, L., Liu, Q., Dai, J., Shen, J., Zhang, J., 2010. IL-6 augments the invasiveness of U87MG human glioblastoma multiforme cells via up-regulation of MMP-2 and fascin-1. *Oncol Rep* 23, 1553-1559.
161. Li, Y., Liu, H., 2015. Learning Semantic Tags from Big Data for Clinical Text Representation. *AMIA Jt Summits Transl Sci Proc* 2015, 461-465.
162. Libbrecht, M.W., Noble, W.S., 2015. Machine learning applications in genetics and genomics. *Nat Rev Genet* 16, 321-332.
163. Lim, J.H., Lee, E.S., You, H.J., Lee, J.W., Park, J.W., Chun, Y.S., 2004. Ras-dependent induction of HIF-1 α 785 via the Raf/MEK/ERK pathway: a novel mechanism of Ras-mediated tumor promotion. *Oncogene* 23, 9427-9431.
164. Lin, S.M., Du, P., Huber, W., Kibbe, W.A., 2008. Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Res* 36, e11.
165. Lind, M.I., Spagopoulou, F., 2018. Evolutionary consequences of epigenetic inheritance. *Heredity (Edinb)* 121, 205-209.
166. Lindemann, C., Hackmann, O., Delic, S., Schmidt, N., Reifenberger, G., Riemenschneider, M.J., 2011. SOCS3 promoter methylation is mutually exclusive to EGFR amplification in gliomas and promotes glioma cell invasion through STAT3 and FAK activation. *Acta neuropathologica* 122, 241-251.
167. Liu, F., Gu, L.N., Shan, B.E., Geng, C.Z., Sang, M.X., 2016. Biomarkers for EMT and MET in breast cancer: An update. *Oncol Lett* 12, 4869-4876.
168. Liu, F., Hon, G.C., Villa, G.R., Turner, K.M., Ikegami, S., Yang, H., Ye, Z., Li, B., Kuan, S., Lee, A.Y., Zanca, C., Wei, B., Lucey, G., Jenkins, D., Zhang, W., Barr, C.L., Furnari, F.B., Cloughesy, T.F., Yong, W.H., Gahman, T.C., Shiau, A.K., Cavenee, W.K., Ren, B., Mischel, P.S., 2015. EGFR Mutation Promotes Glioblastoma through Epigenome and Transcription Factor Network Remodeling. *Mol Cell* 60, 307-318.
169. Liu, R., Wang, X., Chen, G.Y., Dalerba, P., Gurney, A., Hoey, T., Sherlock, G., Lewicki, J., Shedden, K., Clarke, M.F., 2007. The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med* 356, 217-226.
170. Liu, T., Ortiz, J.A., Taing, L., Meyer, C.A., Lee, B., Zhang, Y., Shin, H., Wong, S.S., Ma, J., Lei, Y., Pape, U.J., Poidinger, M., Chen, Y., Yeung, K., Brown, M., Turpaz, Y., Liu, X.S., 2011. Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol* 12, R83.
171. Lottaz, C., Beier, D., Meyer, K., Kumar, P., Hermann, A., Schwarz, J., Junker, M., Oefner, P.J., Bogdahn, U., Wischhusen, J., Spang, R., Storch, A., Beier, C.P., 2010. Transcriptional profiles of CD133+ and CD133- glioblastoma-derived cancer stem cell lines suggest different cells of origin. *Cancer Res* 70, 2030-2040.
172. Louis, D.N., Ohgaki, H., Wiestler, O.D., Cavenee, W.K., Burger, P.C., Jouvet, A., Scheithauer, B.W., Kleihues, P., 2007. The 2007 WHO classification of tumours of the central nervous system. *Acta neuropathologica* 114, 97-109.
173. Louis, D.N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., Ellison, D.W., 2016. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta neuropathologica* 131, 803-820.
174. Luwor, R.B., Styli, S.S., Kaye, A.H., 2013. The role of Stat3 in glioblastoma multiforme. *J Clin Neurosci* 20, 907-911.
175. Lv, D., Li, Y., Zhang, W., Alvarez, A.A., Song, L., Tang, J., Gao, W.Q., Hu, B., Cheng, S.Y., Feng, H., 2017. TRIM24 is an oncogenic transcriptional co-activator of STAT3 in glioblastoma. *Nat Commun* 8, 1454.

176. Ma, N.K., Lim, J.K., Leong, M.F., Sandanaraj, E., Ang, B.T., Tang, C., Wan, A.C., 2016. Collaboration of 3D context and extracellular matrix in the development of glioma stemness in a 3D model. *Biomaterials* 78, 62-73.
177. Madhavan, S., Zenklusen, J.C., Kotliarov, Y., Sahni, H., Fine, H.A., Buetow, K., 2009. Rembrandt: helping personalized medicine become a reality through integrative translational research. *Mol Cancer Res* 7, 157-167.
178. Mahal, L.K., Charter, N.W., Angata, K., Fukuda, M., Koshland, D.E., Jr., Bertozzi, C.R., 2001. A small-molecule modulator of poly-alpha 2,8-sialic acid expression on cultured neurons and tumor cells. *Science* 294, 380-381.
179. Mantel, N., 1966. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep* 50, 163-170.
180. Marbach, D., Lamparter, D., Quon, G., Kellis, M., Kutalik, Z., Bergmann, S., 2016. Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. *Nat Methods* 13, 366-370.
181. Matsuda, K., Sato, A., Okada, M., Shibuya, K., Seino, S., Suzuki, K., Watanabe, E., Narita, Y., Shibui, S., Kayama, T., Kitanaka, C., 2012. Targeting JNK for therapeutic depletion of stem-like glioblastoma cells. *Sci Rep* 2, 516.
182. Mazor, T., Pankov, A., Song, J.S., Costello, J.F., 2016. Intratumoral Heterogeneity of the Epigenome. *Cancer Cell* 29, 440-451.
183. McFarland, B.C., Ma, J.Y., Langford, C.P., Gillespie, G.Y., Yu, H., Zheng, Y., Nozell, S.E., Huszar, D., Benveniste, E.N., 2011. Therapeutic potential of AZD1480 for the treatment of human glioblastoma. *Mol Cancer Ther* 10, 2384-2393.
184. McMillan, K.M., Ehtesham, M., Stevenson, C.B., Edgeworth, M.L., Thompson, R.C., Price, R.R., 2009. T2 detection of tumor invasion within segmented components of glioblastoma multiforme. *J Magn Reson Imaging* 29, 251-257.
185. McQuarrie, A.D.R., Tsai, C.-L., 1998. Regression and Time Series Model Selection. World Scientific.
186. Meredith, M., 2019. wqid: Quick and Dirty Estimates for Wildlife Populations.
187. Milacic, M., Haw, R., Rothfels, K., Wu, G., Croft, D., Hermjakob, H., D'Eustachio, P., Stein, L., 2012. Annotating cancer variants and anti-cancer therapeutics in reactome. *Cancers (Basel)* 4, 1180-1211.
188. Miyake, H., Hara, I., Yamanaka, K., Muramaki, M., Gleave, M., Eto, H., 2005. Introduction of insulin-like growth factor binding protein-2 gene into human bladder cancer cells enhances their metastatic potential. *Oncology reports* 13, 341-345.
189. Mizoguchi, M., Betensky, R.A., Batchelor, T.T., Bernay, D.C., Louis, D.N., Nutt, C.L., 2006. Activation of STAT3, MAPK, and AKT in Malignant Astrocytic Gliomas. *Journal of Neuropathology and Experimental Neurology* 65, 1181-1188.
190. Mohammed, E.A., Naugler, C., Far, B.H., 2015. Emerging Business Intelligence Framework for a Clinical Laboratory Through Big Data Analytics. 577-602.
191. Morera, L., Lubbert, M., Jung, M., 2016. Targeting histone methyltransferases and demethylases in clinical trials for cancer therapy. *Clin Epigenetics* 8, 57.
192. Morgan, K.J., Gilliland, D.G., 2008. A role for JAK2 mutations in myeloproliferative diseases. *Annu Rev Med* 59, 213-222.
193. Moskal, J.R., Kroes, R.A., Dawson, G., 2009. The glycobiology of brain tumors: disease relevance and therapeutic potential. *Expert Rev Neurother* 9, 1529-1545.
194. Mukthavaram, R., Ouyang, X., Saklecha, R., Jiang, P., Nomura, N., Pingle, S.C., Guo, F., Makale, M., Kesari, S., 2015. Effect of the JAK2/STAT3 inhibitor SAR317461 on human glioblastoma tumorspheres. *J Transl Med* 13, 269.
195. Murat, A., Migliavacca, E., Gorlia, T., Lambiv, W.L., Shay, T., Hamou, M.F., de Tribolet, N., Regli, L., Wick, W., Kouwenhoven, M.C., Hainfellner, J.A., Heppner, F.L., Dietrich, P.Y., Zimmer, Y., Cairncross, J.G., Janzer, R.C.,

- Domany, E., Delorenzi, M., Stupp, R., Hegi, M.E., 2008. Stem cell-related "self-renewal" signature and high epidermal growth factor receptor expression associated with resistance to concomitant chemoradiotherapy in glioblastoma. *J Clin Oncol* 26, 3015-3024.
196. Musa, A., Ghoraie, L.S., Zhang, S.D., Glazko, G., Yli-Harja, O., Dehmer, M., Haibe-Kains, B., Emmert-Streib, F., 2017. A review of connectivity map and computational approaches in pharmacogenomics. *Brief Bioinform* 18, 903.
 197. Ng, F.S., Toh, T.B., Ting, E.H., Koh, G.R., Sandanaraj, E., Phong, M., Wong, S.S., Leong, S.H., Kon, O.L., Tucker-Kellogg, G., Ng, W.H., Ng, I., Tang, C., Ang, B.T., 2012. Progenitor-like traits contribute to patient survival and prognosis in oligodendroglial tumors. *Clin Cancer Res* 18, 4122-4135.
 198. Nouchmehr, H., Weisenberger, D.J., Diefes, K., Phillips, H.S., Pujara, K., Berman, B.P., Pan, F., Pelloso, C.E., Sulman, E.P., Bhat, K.P., Verhaak, R.G., Hoadley, K.A., Hayes, D.N., Perou, C.M., Schmidt, H.K., Ding, L., Wilson, R.K., Van Den Berg, D., Shen, H., Bengtsson, H., Neuvial, P., Cope, L.M., Buckley, J., Herman, J.G., Baylin, S.B., Laird, P.W., Aldape, K., 2010. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* 17, 510-522.
 199. Nutt, C.L., Mani, D.R., Betensky, R.A., Tamayo, P., Cairncross, J.G., Ladd, C., Pohl, U., Hartmann, C., McLaughlin, M.E., Batchelor, T.T., Black, P.M., von Deimling, A., Pomeroy, S.L., Golub, T.R., Louis, D.N., 2003. Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. *Cancer Res* 63, 1602-1607.
 200. Nyamundanda, G., Poudel, P., Patil, Y., Sadanandam, A., 2017. A Novel Statistical Method to Diagnose, Quantify and Correct Batch Effects in Genomic Studies. *Sci Rep* 7, 10849.
 201. Olow, A., Chen, Z., Niedner, R.H., Wolf, D.M., Yau, C., Pankov, A., Lee, E.P., Brown-Swigart, L., van 't Veer, L.J., Coppe, J.P., 2016. An Atlas of the Human Kinome Reveals the Mutational Landscape Underlying Dysregulated Phosphorylation Cascades in Cancer. *Cancer Res* 76, 1733-1745.
 202. Oyelade, J., Isewon, I., Oladipupo, F., Aromolaran, O., Uwoghien, E., Ameh, F., Ahas, M., Adebisi, E., 2016. Clustering Algorithms: Their Application to Gene Expression Data. *Bioinform Biol Insights* 10, 237-253.
 203. Ozawa, T., Riester, M., Cheng, Y.K., Huse, J.T., Squatrito, M., Helmy, K., Charles, N., Michor, F., Holland, E.C., 2014. Most human non-GCIMP glioblastoma subtypes evolve from a common proneural-like precursor glioma. *Cancer Cell* 26, 288-300.
 204. Parmar, C., Leijenaar, R.T.H., Grossmann, P., Rios Velazquez, E., Bussink, J., Rietveld, D., Rietbergen, M.M., Haibe-Kains, B., Lambin, P., Aerts, H.J.W.L., 2015. Radiomic feature clusters and Prognostic Signatures specific for Lung and Head & Neck cancer. *Scientific Reports* 5.
 205. Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., Louis, D.N., Rozenblatt-Rosen, O., Suva, M.L., Regev, A., Bernstein, B.E., 2014. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344, 1396-1401.
 206. Paul, I., Bhattacharya, S., Chatterjee, A., Ghosh, M.K., 2013. Current Understanding on EGFR and Wnt/beta-Catenin Signaling in Glioma and Their Possible Crosstalk. *Genes Cancer* 4, 427-446.
 207. Pažanin, L., Vučić, M., Čupić, H., Plavec, D., Krušlin, B., 2011. IGFBP-2 expression, angiogenesis and pseudopalisades in glioblastoma. *Translational Neuroscience* 2, 219-224.
 208. Pepper, S.D., Saunders, E.K., Edwards, L.E., Wilson, C.L., Miller, C.J., 2007. The utility of MAS5 expression summary and detection call algorithms. *BMC Bioinformatics* 8, 273.

209. Pereira, M.E., Kabat, E.A., Lotan, R., Sharon, N., 1976. Immunochemical studies on the specificity of the peanut (*Arachis hypogaea*) agglutinin. *Carbohydr Res* 51, 107-118.
210. Phillips, H.S., Kharbanda, S., Chen, R., Forrest, W.F., Soriano, R.H., Wu, T.D., Misra, A., Nigro, J.M., Colman, H., Soroceanu, L., Williams, P.M., Modrusan, Z., Feuerstein, B.G., Aldape, K., 2006. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* 9, 157-173.
211. Pollard, S.M., Yoshikawa, K., Clarke, I.D., Danovi, D., Stricker, S., Russell, R., Bayani, J., Head, R., Lee, M., Bernstein, M., Squire, J.A., Smith, A., Dirks, P., 2009. Glioma stem cell lines expanded in adherent culture have tumor-specific phenotypes and are suitable for chemical and genetic screens. *Cell Stem Cell* 4, 568-580.
212. Pozo, N., Zahonero, C., Fernandez, P., Linares, J.M., Ayuso, A., Hagiwara, M., Perez, A., Ricoy, J.R., Hernandez-Lain, A., Sepulveda, J.M., Sanchez-Gomez, P., 2013. Inhibition of DYRK1A destabilizes EGFR and reduces EGFR-dependent glioblastoma growth. *J Clin Invest* 123, 2475-2487.
213. Prag, S., Lepekhn, E.A., Kolkova, K., Hartmann-Petersen, R., Kawa, A., Walmod, P.S., Belman, V., Gallagher, H.C., Berezin, V., Bock, E., Pedersen, N., 2002. NCAM regulates cell motility. *J Cell Sci* 115, 283-292.
214. Prat, A., Adamo, B., Fan, C., Peg, V., Vidal, M., Galván, P., Vivancos, A., Nuciforo, P., Palmer, H.G., Dawood, S., Rodón, J., Cajal, S.R.y., Campo, J.M.D., Felip, E., Tabernero, J., Cortés, J., 2013. Genomic Analyses across Six Cancer Types Identify Basal-like Breast Cancer as a Unique Molecular Entity. *Scientific Reports* 3.
215. Priatel, J.J., Chui, D., Hiraoka, N., Simmons, C.J., Richardson, K.B., Page, D.M., Fukuda, M., Varki, N.M., Marth, J.D., 2000. The ST3Gal-I sialyltransferase controls CD8⁺ T lymphocyte homeostasis by modulating O-glycan biosynthesis. *Immunity* 12, 273-283.
216. Puram, S.V., Yeung, C.M., Jahani-Asl, A., Lin, C., de la Iglesia, N., Konopka, G., Jackson-Grusby, L., Bonni, A., 2012. STAT3-iNOS Signaling Mediates EGFRvIII-Induced Glial Proliferation and Transformation. *J Neurosci* 32, 7806-7818.
217. Quesnelle, K.M., Boehm, A.L., Grandis, J.R., 2007. STAT-mediated EGFR signaling in cancer. *J Cell Biochem* 102, 311-319.
218. Raczyński, L., Wozniak, K., Rubel, T., Zaremba, K., 2010. Application of Density Based Clustering to Microarray Data Analysis. *International Journal of Electronics and Telecommunications* 56.
219. Rahaman, S.O., Harbor, P.C., Chernova, O., Barnett, G.H., Vogelbaum, M.A., Haque, S.J., 2002. Inhibition of constitutively active Stat3 suppresses proliferation and induces apoptosis in glioblastoma multiforme cells. *Oncogene* 21, 8404-8413.
220. Rathore, S., Akbari, H., Rozycki, M., Abdullah, K.G., Nasrallah, M.P., Binder, Z.A., Davuluri, R.V., Lustig, R.A., Dahmane, N., Bilello, M., O'Rourke, D.M., Davatzikos, C., 2018. Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond IDH1. *Sci Rep* 8, 5087.
221. Reese, S.E., Archer, K.J., Therneau, T.M., Atkinson, E.J., Vachon, C.M., de Andrade, M., Kocher, J.P., Eckel-Passow, J.E., 2013. A new statistic for identifying batch effects in high-throughput genomic data that uses guided principal component analysis. *Bioinformatics* 29, 2877-2883.
222. Regad, T., 2015. Targeting RTK Signaling Pathways in Cancer. *Cancers (Basel)* 7, 1758-1784.
223. Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., Mesirov, J.P., 2006. GenePattern 2.0. *Nat Genet* 38, 500-501.

224. Reisner, Y., Linker-Israeli, M., Sharon, N., 1976. Separation of mouse thymocytes into two subpopulations by the use of peanut agglutinin. *Cell Immunol* 25, 129-134.
225. Rich, J.N., Eyler, C.E., 2008. Cancer stem cells in brain tumor biology. *Cold Spring Harb Symp Quant Biol* 73, 411-420.
226. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43, e47.
227. Rodriguez-Hernandez, I., Vazquez-Cedeira, M., Santos-Briz, A., Garcia, J.L., Fernandez, I.F., Gomez-Moreta, J.A., Martin-Vallejo, J., Gonzalez-Sarmiento, R., Lazo, P.A., 2013. VRK2 identifies a subgroup of primary high-grade astrocytomas with a better prognosis. *BMC Clin Pathol* 13, 23.
228. Rogers, T.W., Tsui, A., Gonzales, M., 2018. Re-classification of gliomas by the 2016 revision of the who classification of CNS tumours. *Pathology* 50, S123.
229. Rolhion, C., Penault-Llorca, F., Kemeny, J.L., Lemaire, J.J., Jullien, C., Labit-Bouvier, C., Finat-Duclos, F., Verrelle, P., 2001. Interleukin-6 overexpression as a marker of malignancy in human gliomas. *J Neurosurg* 94, 97-101.
230. Rorive, S., Belot, N., Decaestecker, C., Lefranc, F., Gordower, L., Micik, S., Maurage, C.A., Kaltner, H., Ruchoux, M.M., Danguy, A., Gabius, H.J., Salmon, I., Kiss, R., Camby, I., 2001. Galectin-1 is highly expressed in human gliomas with relevance for modulation of invasion of tumor astrocytes into the brain parenchyma. *Glia* 33, 241-255.
231. Rose, M.L., Malchiodi, F., 1981. Binding of peanut lectin to thymic cortex and germinal centres of lymphoid tissue. *Immunology* 42, 583-591.
232. Salner, A.L., Obbagy, J.E., Hellman, S., 1982. Differing stem cell self-renewal of lectin-separated murine bone marrow fractions. *J Natl Cancer Inst* 68, 639-641.
233. Schindler, C., Darnell, J.E., Jr., 1995. Transcriptional responses to polypeptide ligands: the JAK-STAT pathway. *Annu Rev Biochem* 64, 621-651.
234. Schneider, F., Kemmner, W., Haensch, W., Franke, G., Gretschel, S., Karsten, U., Schlag, P.M., 2001. Overexpression of sialyltransferase CMP-sialic acid:Galbeta1,3GalNAc-R alpha6-Sialyltransferase is related to poor patient survival in human colorectal carcinomas. *Cancer Res* 61, 4605-4611.
235. Scholl, C., Frohling, S., Dunn, I.F., Schinzel, A.C., Barbie, D.A., Kim, S.Y., Silver, S.J., Tamayo, P., Wadlow, R.C., Ramaswamy, S., Dohner, K., Bullinger, L., Sandy, P., Boehm, J.S., Root, D.E., Jacks, T., Hahn, W.C., Gilliland, D.G., 2009. Synthetic lethal interaction between oncogenic KRAS dependency and STK33 suppression in human cancer cells. *Cell* 137, 821-834.
236. Schworer, R., Schmidt, R.R., 2002. Efficient sialyltransferase inhibitors based on glycosides of N-acetylglucosamine. *J Am Chem Soc* 124, 1632-1637.
237. Scott, C.T., Caulfield, T., Borgelt, E., Illes, J., 2012. Personal medicine -the new banking crisis. *Nat Biotech* 30, 141-147.
238. Segerman, A., Niklasson, M., Haglund, C., Bergstrom, T., Jarvius, M., Xie, Y., Westermark, A., Sonmez, D., Hermansson, A., Kastemar, M., Naimaie-Ali, Z., Nyberg, F., Berglund, M., Sundstrom, M., Hesselager, G., Uhrbom, L., Gustafsson, M., Larsson, R., Fryknas, M., Segerman, B., Westermark, B., 2016. Clonal Variation in Drug and Radiation Response among Glioma-Initiating Cells Is Linked to Proneural-Mesenchymal Transition. *Cell reports* 17, 2994-3009.
239. Seghouane, A.K., Amari, S., 2007. The AIC criterion and symmetrizing the Kullback-Leibler divergence. *IEEE Trans Neural Netw* 18, 97-106.
240. Senft, C., Priester, M., Polacin, M., Schroder, K., Seifert, V., Kogel, D., Weissenberger, J., 2011. Inhibition of the JAK-2/STAT3 signaling pathway impedes the migratory and invasive potential of human glioblastoma cells. *J Neurooncol* 101, 393-403.

241. Shao, H., Chung, J., Balaj, L., Charest, A., Bigner, D.D., Carter, B.S., Hochberg, F.H., Breakefield, X.O., Weissleder, R., Lee, H., 2012. Protein typing of circulating microvesicles allows real-time monitoring of glioblastoma therapy. *Nat Med* 18, 1835-1840.
242. Shao, H., Chung, J., Lee, K., Balaj, L., Min, C., Carter, B.S., Hochberg, F.H., Breakefield, X.O., Lee, H., Weissleder, R., 2015. Chip-based analysis of exosomal mRNA mediating drug resistance in glioblastoma. *Nat Commun* 6, 6999.
243. Shao, J., 1997. An asymptotic theory for linear model selection. *Statistica Sinica* 7, 221-242.
244. Sharma, S.V., Lee, D.Y., Li, B., Quinlan, M.P., Takahashi, F., Maheswaran, S., McDermott, U., Azizian, N., Zou, L., Fischbach, M.A., Wong, K.K., Brandstetter, K., Wittner, B., Ramaswamy, S., Classon, M., Settleman, J., 2010. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* 141, 69-80.
245. Sherry, M.M., Reeves, A., Wu, J.K., Cochran, B.H., 2009. STAT3 is required for proliferation and maintenance of multipotency in glioblastoma stem cells. *Stem Cells* 27, 2383-2392.
246. Shmueli, G., 2010. To Explain or to Predict? *Statistical Science* 25, 289-310.
247. Shuai, K., 2000. Modulation of STAT signaling by STAT-interacting proteins. *Oncogene* 19, 2638-2644.
248. Sibilio, M., Fleischmann, A., Behrens, A., Stingl, L., Carroll, J., Watt, F.M., Schlessinger, J., Wagner, E.F., 2000. The EGF receptor provides an essential survival signal for SOS-dependent skin tumor development. *Cell* 102, 211-220.
249. Singh, R., Lanchantin, J., Robins, G., Qi, Y., 2016. DeepChrome: deep-learning for predicting gene expression from histone modifications. *Bioinformatics* 32, i639-i648.
250. Skropeta, D., Schworer, R., Haag, T., Schmidt, R.R., 2004. Asymmetric synthesis and affinity of potent sialyltransferase inhibitors based on transition-state analogues. *Glycoconj J* 21, 205-219.
251. Smyth, G.K., 2005. limma: Linear Models for Microarray Data. 397-420.
252. Sokal, R.R., 1958. A statistical method for evaluating systematic relationship. *University of Kansas Science Bulletin* 28, 1409-1438.
253. Solomon, D.A., Kim, J.S., Cronin, J.C., Sibenaller, Z., Ryken, T., Rosenberg, S.A., Ransom, H., Jean, W., Bigner, D., Yan, H., Samuels, Y., Waldman, T., 2008. Mutational inactivation of PTPRD in glioblastoma multiforme and malignant melanoma. *Cancer Res* 68, 10300-10306.
254. Souza, V., Escobar Mdel, C., Bucio, L., Hernandez, E., Gomez-Quiroz, L.E., Gutierrez Ruiz, M.C., 2009. NADPH oxidase and ERK1/2 are involved in cadmium induced-STAT3 activation in HepG2 cells. *Toxicol Lett* 187, 180-186.
255. Stechishin, O.D., Luchman, H.A., Ruan, Y., Blough, M.D., Nguyen, S.A., Kelly, J.J., Cairncross, J.G., Weiss, S., 2013. On-target JAK2/STAT3 inhibition slows disease progression in orthotopic xenografts of human glioblastoma brain tumor stem cells. *Neuro-Oncology* 15, 198-207.
256. Stefano Monti, P.T., Jill Mesirov, Todd Golub, 2003. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. *Machine Learning* 52, 91-118.
257. Steiner, H.H., Karcher, S., Mueller, M.M., Nalbantis, E., Kunze, S., Herold-Mende, C., 2004. Autocrine pathways of the vascular endothelial growth factor (VEGF) in glioblastoma multiforme: clinical relevance of radiation-induced increase of VEGF levels. *J Neurooncol* 66, 129-138.
258. Stone, M., 1977. An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion. *Journal of the Royal Statistical Society: Series B (Methodological)* 39, 44-47.

259. Stone, M., 1979. Comments on Model Selection Criteria of Akaike and Schwarz. *Journal of the Royal Statistical Society. Series B (Methodological)* 41, 276-278.
260. Sturm, D., Witt, H., Hovestadt, V., Khuong-Quang, D.A., Jones, D.T., Konermann, C., Pfaff, E., Tonjes, M., Sill, M., Bender, S., Kool, M., Zapatka, M., Becker, N., Zucknick, M., Hielscher, T., Liu, X.Y., Fontebasso, A.M., Ryzhova, M., Albrecht, S., Jacob, K., Wolter, M., Ebinger, M., Schuhmann, M.U., van Meter, T., Fruhwald, M.C., Hauch, H., Pekrun, A., Radlwimmer, B., Niehues, T., von Komorowski, G., Durken, M., Kulozik, A.E., Madden, J., Donson, A., Foreman, N.K., Drissi, R., Fouladi, M., Scheurlen, W., von Deimling, A., Monoranu, C., Roggendorf, W., Herold-Mende, C., Unterberg, A., Kramm, C.M., Felsberg, J., Hartmann, C., Wiestler, B., Wick, W., Milde, T., Witt, O., Lindroth, A.M., Schwartzentruber, J., Faury, D., Fleming, A., Zakrzewska, M., Liberski, P.P., Zakrzewski, K., Hauser, P., Garami, M., Klekner, A., Bogнар, L., Morrissy, S., Cavalli, F., Taylor, M.D., van Sluis, P., Koster, J., Versteeg, R., Volckmann, R., Mikkelsen, T., Aldape, K., Reifemberger, G., Collins, V.P., Majewski, J., Korshunov, A., Lichter, P., Plass, C., Jabado, N., Pfister, S.M., 2012. Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell* 22, 425-437.
261. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102, 15545-15550.
262. Suh, K.S., Sarojini, S., Youssif, M., Nalley, K., Milinovic, N., Elloumi, F., Russell, S., Pecora, A., Schecter, E., Goy, A., 2013. Tissue Banking, Bioinformatics, and Electronic Medical Records: The Front-End Requirements for Personalized Medicine. *Journal of Oncology* 2013, 368751.
263. Suzuki, H., Aoki, K., Chiba, K., Sato, Y., Shiozawa, Y., Shiraishi, Y., Shimamura, T., Niida, A., Motomura, K., Ohka, F., Yamamoto, T., Tanahashi, K., Ranjit, M., Wakabayashi, T., Yoshizato, T., Kataoka, K., Yoshida, K., Nagata, Y., Sato-Otsubo, A., Tanaka, H., Sanada, M., Kondo, Y., Nakamura, H., Mizoguchi, M., Abe, T., Muragaki, Y., Watanabe, R., Ito, I., Miyano, S., Natsume, A., Ogawa, S., 2015. Mutational landscape and clonal architecture in grade II and III gliomas. *Nat Genet* 47, 458-468.
264. Tan, M.S.Y., Sandanaraj, E., Chong, Y.K., Lim, S.W., Koh, L.W.H., Ng, W.H., Tan, N.S., Tan, P., Ang, B.T., Tang, C., 2019. A STAT3-based gene signature stratifies glioma patients for targeted therapy. *Nat Commun* 10, 3601.
265. Tan, S.Y., Sandanaraj, E., Tang, C., Ang, B.T., 2016. Biobanking: An Important Resource for Precision Medicine in Glioblastoma. *Adv Exp Med Biol* 951, 47-56.
266. Tang, L.Y., Heller, M., Meng, Z., Yu, L.R., Tang, Y., Zhou, M., Zhang, Y.E., 2017. Transforming Growth Factor-beta (TGF-beta) Directly Activates the JAK1-STAT3 Axis to Induce Hepatic Fibrosis in Coordination with the SMAD Pathway. *J Biol Chem* 292, 4302-4312.
267. Team, R.C., 2019. R: A Language and Environment for Statistical Computing.
268. Terry TM, G.P.G., 2000. *Statistics for Biology and Health*. Springer-Verlag New York, New York, 39-77 pp.
269. The Cancer Genome Atlas Research, N., McLendon, R., Friedman, A., Bigner, D., Van Meir, E.G., Brat, D.J., M. Mastrogiannis, G., Olson, J.J., Mikkelsen, T., Lehman, N., Aldape, K., Alfred Yung, W.K., Bogler, O., VandenBerg, S., Berger, M., Prados, M., Muzny, D., Morgan, M., Scherer, S., Sabo, A., Nazareth, L., Lewis, L., Hall, O., Zhu, Y., Ren, Y., Alvi, O., Yao, J., Hawes, A., Jhangiani, S., Fowler, G., San Lucas, A., Kovar, C., Cree, A., Dinh, H., Santibanez, J., Joshi, V., Gonzalez-Garay, M.L., Miller, C.A., Milosavljevic, A.,

- Donehower, L., Wheeler, D.A., Gibbs, R.A., Cibulskis, K., Sougnez, C., Fennell, T., Mahan, S., Wilkinson, J., Ziaugra, L., Onofrio, R., Bloom, T., Nicol, R., Ardlie, K., Baldwin, J., Gabriel, S., Lander, E.S., Ding, L., Fulton, R.S., McLellan, M.D., Wallis, J., Larson, D.E., Shi, X., Abbott, R., Fulton, L., Chen, K., Koboldt, D.C., Wendl, M.C., Meyer, R., Tang, Y., Lin, L., Osborne, J.R., Dunford-Shore, B.H., Miner, T.L., Delehaunty, K., Markovic, C., Swift, G., Courtney, W., Pohl, C., Abbott, S., Hawkins, A., Leong, S., Haipok, C., Schmidt, H., Wiechert, M., Vickery, T., Scott, S., Dooling, D.J., Chinwalla, A., Weinstock, G.M., Mardis, E.R., Wilson, R.K., Getz, G., Winckler, W., Verhaak, R.G.W., Lawrence, M.S., O'Kelly, M., Robinson, J., Alexe, G., Beroukhi, R., Carter, S., Chiang, D., Gould, J., Gupta, S., Korn, J., Mermel, C., Mesirov, J., Monti, S., Nguyen, H., Parkin, M., Reich, M., Stransky, N., Weir, B.A., Garraway, L., Golub, T., Meyerson, M., Chin, L., Protopopov, A., Zhang, J., Perna, I., Aronson, S., Sathiamoorthy, N., Ren, G., Yao, J., Wiedemeyer, W.R., Kim, H., Won Kong, S., Xiao, Y., Kohane, I.S., Seidman, J., Park, P.J., Kucherlapati, R., Laird, P.W., Cope, L., Herman, J.G., Weisenberger, D.J., Pan, F., Van Den Berg, D., Van Neste, L., Mi Yi, J., Schuebel, K.E., Baylin, S.B., Absher, D.M., Li, J.Z., Southwick, A., Brady, S., Aggarwal, A., Chung, T., Sherlock, G., Brooks, J.D., Myers, R.M., Spellman, P.T., Purdom, E., Jakkula, L.R., Lapuk, A.V., Marr, H., Dorton, S., Gi Choi, Y., Han, J., Ray, A., Wang, V., Durinck, S., Robinson, M., Wang, N.J., Vranizan, K., Peng, V., Van Name, E., Fontenay, G.V., Ngai, J., Conboy, J.G., Parvin, B., Feiler, H.S., Speed, T.P., Gray, J.W., Brennan, C., Socci, N.D., Olshen, A., Taylor, B.S., Lash, A., Schultz, N., Reva, B., Antipin, Y., Stukalov, A., Gross, B., Cerami, E., Qing Wang, W., Qin, L.-X., Seshan, V.E., Villafania, L., Cavatore, M., Borsu, L., Viale, A., Gerald, W., Sander, C., Ladanyi, M., Perou, C.M., Neil Hayes, D., Topal, M.D., Hoadley, K.A., Qi, Y., Balu, S., Shi, Y., Wu, J., Penny, R., Bittner, M., Shelton, T., Lenkiewicz, E., Morris, S., Beasley, D., Sanders, S., Kahn, A., Sfeir, R., Chen, J., Nassau, D., Feng, L., Hickey, E., Zhang, J., Weinstein, J.N., Barker, A., Gerhard, D.S., Vockley, J., Compton, C., Vaught, J., Fielding, P., Ferguson, M.L., Schaefer, C., Madhavan, S., Buetow, K.H., Collins, F., Good, P., Guyer, M., Ozenberger, B., Peterson, J., Thomson, E., 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455, 1061.
270. Tibshirani, R., Hastie, T., Narasimhan, B., Chu, G., 2002. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proceedings of the National Academy of Sciences of the United States of America* 99, 6567-6572.
 271. Tusher, V.G., Tibshirani, R., Chu, G., 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 98, 5116-5121.
 272. Van Brocklyn, J.R., Wojton, J., Meisen, W.H., Kellough, D.A., Ecsedy, J.A., Kaur, B., Lehman, N.L., 2014. Aurora-A inhibition offers a novel therapy effective against intracranial glioblastoma. *Cancer Res* 74, 5364-5370.
 273. Veeriah, S., Brennan, C., Meng, S., Singh, B., Fagin, J.A., Solit, D.B., Paty, P.B., Rohle, D., Vivanco, I., Chmielecki, J., Pao, W., Ladanyi, M., Gerald, W.L., Liao, L., Cloughesy, T.C., Mischel, P.S., Sander, C., Taylor, B., Schultz, N., Major, J., Heguy, A., Fang, F., Mellinghoff, I.K., Chan, T.A., 2009. The tyrosine phosphatase PTPRD is a tumor suppressor that is frequently inactivated and mutated in glioblastoma and other human cancers. *Proc Natl Acad Sci U S A* 106, 9435-9440.
 274. Venables, W.N., Ripley, B.D., Venables, W.N., 2002. *Modern applied statistics with S, Statistics and computing*. Springer, New York, pp. 1 online resource (xi, 495 p.) ill.
 275. Verhaak, R.G., Hoadley, K.A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M.D., Miller, C.R., Ding, L., Golub, T., Mesirov, J.P., Alexe, G., Lawrence, M., O'Kelly, M., Tamayo, P., Weir, B.A., Gabriel, S., Winckler, W., Gupta, S., Jakkula, L.,

- Feiler, H.S., Hodgson, J.G., James, C.D., Sarkaria, J.N., Brennan, C., Kahn, A., Spellman, P.T., Wilson, R.K., Speed, T.P., Gray, J.W., Meyerson, M., Getz, G., Perou, C.M., Hayes, D.N., 2010. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 17, 98-110.
276. Videira, P.A., Correia, M., Malagolini, N., Crespo, H.J., Ligeiro, D., Calais, F.M., Trindade, H., Dall'Olio, F., 2009. ST3Gal.I sialyltransferase relevance in bladder cancer tissues and cell lines. *BMC Cancer* 9, 357.
 277. Vitucci, M., Hayes, D.N., Miller, C.R., 2011. Gene expression profiling of gliomas: merging genomic and histopathological classification for personalised therapy. *Br J Cancer* 104, 545-553.
 278. Vitucci, M., Karpinich, N.O., Bash, R.E., Werneke, A.M., Schmid, R.S., White, K.K., McNeill, R.S., Huff, B., Wang, S., Van Dyke, T., Miller, C.R., 2013. Cooperativity between MAPK and PI3K signaling activation is required for glioblastoma pathogenesis. *Neuro Oncol* 15, 1317-1329.
 279. Wald, N.J., Hackshaw, A.K., Frost, C.D., 1999. When can a risk factor be used as a worthwhile screening test? *BMJ* 319, 1562-1565.
 280. Wang, H., Wang, H., Zhang, W., Huang, H.J., Liao, W.S., Fuller, G.N., 2004. Analysis of the activation status of Akt, NFkappaB, and Stat3 in human diffuse gliomas. *Lab Invest* 84, 941-951.
 281. Wang, H., Wang, H., Shen, W., Huang, H., Hu, L., Ramdas, L., Zhou, Y.H., Liao, W.S., Fuller, G.N., Zhang, W., 2003. Insulin-like growth factor binding protein 2 enhances glioblastoma invasion by activating invasion-enhancing genes. *Cancer Res* 63, 4315-4321.
 282. Wang, Q., Hu, B., Hu, X., Kim, H., Squatrito, M., Scarpace, L., deCarvalho, A.C., Lyu, S., Li, P., Li, Y., Barthel, F., Cho, H.J., Lin, Y.H., Satani, N., Martinez-Ledesma, E., Zheng, S., Chang, E., Gabriel Sauve, C.E., Olar, A., Lan, Z.D., Finocchiaro, G., Phillips, J.J., Berger, M.S., Gabrusiewicz, K.R., Wang, G., Eskilsson, E., Hu, J., Mikkelsen, T., DePinho, R.A., Muller, F., Heimberger, A.B., Sulman, E.P., Nam, D.H., Verhaak, R.G.W., 2018. Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer Cell* 33, 152.
 283. Wang, Z., Zhang, S., Siu, T.L., Huang, S., 2015. Glioblastoma multiforme formation and EMT: role of FoxM1 transcription factor. *Curr Pharm Des* 21, 1268-1271.
 284. Wei, C., Li, J., Bumgarner, R.E., 2004. Sample size for detecting differentially expressed genes in microarray experiments. *BMC Genomics* 5, 87.
 285. Wesseling, P., Capper, D., 2018. WHO 2016 Classification of gliomas. *Neuropathology and Applied Neurobiology* 44, 139-150.
 286. Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
 287. Wiedemeyer, W.R., Dunn, I.F., Quayle, S.N., Zhang, J., Chheda, M.G., Dunn, G.P., Zhuang, L., Rosenbluh, J., Chen, S., Xiao, Y., Shapiro, G.I., Hahn, W.C., Chin, L., 2010. Pattern of retinoblastoma pathway inactivation dictates response to CDK4/6 inhibition in GBM. *Proc Natl Acad Sci U S A* 107, 11501-11506.
 288. Wilkerson, M.D., Hayes, D.N., 2010. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 26, 1572-1573.
 289. Wilkerson, M.D., Yin, X., Walter, V., Zhao, N., Cabanski, C.R., Hayward, M.C., Miller, C.R., Socinski, M.A., Parsons, A.M., Thorne, L.B., Haithcock, B.E., Veeramachaneni, N.K., Funkhouser, W.K., Randell, S.H., Bernard, P.S., Perou, C.M., Hayes, D.N., 2012. Differential pathogenesis of lung adenocarcinoma subtypes involving sequence mutations, copy number, chromosomal instability, and methylation. *PloS one* 7, e36530.

290. Wiranowska, M., Ladd, S., Smith, S.R., Gottschall, P.E., 2006. CD44 adhesion molecule and neuro-glial proteoglycan NG2 as invasive markers of glioma. *Brain Cell Biol* 35, 159-172.
291. Wit, E., Heuvel, E.v.d., Romeijn, J.-W., 2012. All models are wrong...': an introduction to model uncertainty. *Statistica Neerlandica* 66, 217-236.
292. Wolf, A., Agnihotri, S., Guha, A., 2010. Targeting metabolic remodeling in glioblastoma multiforme. *Oncotarget* 1, 552-562.
293. Wrensch, M., Jenkins, R.B., Chang, J.S., Yeh, R.F., Xiao, Y., Decker, P.A., Ballman, K.V., Berger, M., Buckner, J.C., Chang, S., Giannini, C., Halder, C., Kollmeyer, T.M., Kosel, M.L., LaChance, D.H., McCoy, L., O'Neill, B.P., Patoka, J., Pico, A.R., Prados, M., Quesenberry, C., Rice, T., Rynearson, A.L., Smirnov, I., Tihan, T., Wiemels, J., Yang, P., Wiencke, J.K., 2009. Variants in the CDKN2B and RTEL1 regions are associated with high-grade glioma susceptibility. *Nat Genet* 41, 905-908.
294. Wu, F.-X., Zhang, W.J., Kusalik, A.J., 2003. A Genetic K-means Clustering Algorithm Applied to Gene Expression Data. 2671, 520-526.
295. Wu, W., Punt, J.A., Granger, L., Sharrow, S.O., Kearse, K.P., 1997. Developmentally regulated expression of peanut agglutinin (PNA)-specific glycans on murine thymocytes. *Glycobiology* 7, 349-356.
296. Xiaowei, X., Ester, M., Kriegel, H.P., Sander, J., 1998. A distribution-based clustering algorithm for mining in large spatial databases. 324-331.
297. Xu, L., Chen, Y., Mayakonda, A., Koh, L., Chong, Y.K., Buckley, D.L., Sandanaraj, E., Lim, S.W., Lin, R.Y., Ke, X.Y., Huang, M.L., Chen, J., Sun, W., Wang, L.Z., Goh, B.C., Dinh, H.Q., Kappei, D., Winter, G.E., Ding, L.W., Ang, B.T., Berman, B.P., Bradner, J.E., Tang, C., Koeffler, H.P., 2018. Targetable BET proteins- and E2F1-dependent transcriptional program maintains the malignancy of glioblastoma. *Proc Natl Acad Sci U S A* 115, E5086-E5095.
298. Yamamoto, H., Saito, T., Kaneko, Y., Kersey, D., Yong, V.W., Bremer, E.G., Mkrdichian, E., Cerullo, L., Leestma, J., Moskal, J.R., 1997. alpha2,3-sialyltransferase mRNA and alpha2,3-linked glycoprotein sialylation are increased in malignant gliomas. *Brain Res* 755, 175-179.
299. Yang, P., Zhang, Y., Pang, J., Zhang, S., Yu, Q., He, L., Wagner, K.U., Zhou, Z., Wang, C.Y., 2013. Loss of Jak2 impairs endothelial function by attenuating Raf-1/MEK1/Sp-1 signaling along with altered eNOS activities. *Am J Pathol* 183, 617-625.
300. Ye, X.Z., Xu, S.L., Xin, Y.H., Yu, S.C., Ping, Y.F., Chen, L., Xiao, H.L., Wang, B., Yi, L., Wang, Q.L., Jiang, X.F., Yang, L., Zhang, P., Qian, C., Cui, Y.H., Zhang, X., Bian, X.W., 2012. Tumor-associated microglia/macrophages enhance the invasion of glioma stem-like cells via TGF-beta1 signaling pathway. *J Immunol* 189, 444-453.
301. Yeo, C.W., Ng, F.S., Chai, C., Tan, J.M., Koh, G.R., Chong, Y.K., Koh, L.W., Foong, C.S., Sandanaraj, E., Holbrook, J.D., Ang, B.T., Takahashi, R., Tang, C., Lim, K.L., 2012. Parkin pathway activation mitigates glioma cell proliferation and predicts patient survival. *Cancer Res* 72, 2543-2553.
302. Yoo, M., Shin, J., Kim, J., Ryall, K.A., Lee, K., Lee, S., Jeon, M., Kang, J., Tan, A.C., 2015. DSigDB: drug signatures database for gene set analysis. *Bioinformatics* 31, 3069-3071.
303. Yoshihara, K., Shahmoradgoli, M., Martinez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Trevino, V., Shen, H., Laird, P.W., Levine, D.A., Carter, S.L., Getz, G., Stemke-Hale, K., Mills, G.B., Verhaak, R.G., 2013. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 4, 2612.
304. Yu, G., He, Q.Y., 2016. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol Biosyst* 12, 477-479.

305. Yu, H., Jove, R., 2004. The STATs of cancer--new molecular targets come of age. *Nat Rev Cancer* 4, 97-105.
306. Yu, H., Pardoll, D., Jove, R., 2009. STATs in cancer inflammation and immunity: a leading role for STAT3. *Nat Rev Cancer* 9, 798-809.
307. Zeileis, A., Meyer, D., Hornik, K., 2007. Residual-based shadings for visualizing (Conditional) independence. *J Comput Graph Stat* 16, 507-525.
308. Zeng, H., Edwards, M.D., Liu, G., Gifford, D.K., 2016. Convolutional neural network architectures for predicting DNA-protein binding. *Bioinformatics* 32, i121-i127.
309. Zhang, M., Mathews Griner, L.A., Ju, W., Duveau, D.Y., Guha, R., Petrus, M.N., Wen, B., Maeda, M., Shinn, P., Ferrer, M., Conlon, K.D., Bamford, R.N., O'Shea, J.J., Thomas, C.J., Waldmann, T.A., 2015. Selective targeting of JAK/STAT signaling is potentiated by Bcl-xL blockade in IL-2-dependent adult T-cell leukemia. *Proc Natl Acad Sci U S A* 112, 12480-12485.
310. Zhang, S.D., Gant, T.W., 2009. sscMap: an extensible Java application for connecting small-molecule drugs using gene-expression signatures. *BMC Bioinformatics* 10, 236.
311. Zhang, W., Zong, C.S., Hermanto, U., Lopez-Bergami, P., Ronai, Z.e., Wang, L.-H., 2006. RACK1 Recruits STAT3 Specifically to Insulin and Insulin-Like Growth Factor 1 Receptors for Activation, Which Is Important for Regulating Anchorage-Independent Growth. *Molecular and Cellular Biology* 26, 413-424.
312. Zheng, H., Ying, H., Yan, H., Kimmelman, A.C., Hiller, D.J., Chen, A.J., Perry, S.R., Tonon, G., Chu, G.C., Ding, Z., Stommel, J.M., Dunn, K.L., Wiedemeyer, R., You, M.J., Brennan, C., Wang, Y.A., Ligon, K.L., Wong, W.H., Chin, L., DePinho, R.A., 2008. p53 and Pten control neural and glioma stem/progenitor cell renewal and differentiation. *Nature* 455, 1129-1133.
313. Zhu, J., Li, W., Mao, Z., 2011. Cdk5: mediator of neuronal development, death and the response to DNA damage. *Mech Ageing Dev* 132, 389-394.
314. Zhu, Y., Qiu, P., Ji, Y., 2014. TCGA-assembler: open-source software for retrieving and processing TCGA data. *Nat Methods* 11, 599-600.
315. Zinn, P.O., Singh, S.K., Colen, R.R., 2015. Shedding light on glioblastoma cellular heterogeneity. *Neuro Oncol* 17, 327-328.
316. Zou, J., Huss, M., Abid, A., Mohammadi, P., Torkamani, A., Telenti, A., 2019. A primer on deep learning in genomics. *Nat Genet* 51, 12-18.
317. Zou, Y., Wang, Q., Li, B., Xie, B., Wang, W., 2014. Temozolomide induces autophagy via ATMAMPKULK1 pathways in glioma. *Mol Med Rep* 10, 411-416.

9 AUTHOR'S PUBLICATIONS

Peer-reviewed journal articles:

1. Tan MS*, **Sandanaraj E***, Chong YK, Lim SW, Koh LWH, Ng WH, Tan NS, Tan P, Ang BT, Tang C. Signal transducer and activator of transcription 3-based glioma patient stratification and its application in precision oncology. *Nat Commun.* **2019 Aug 9;10(1):3601. doi: 10.1038/s41467-019-11614-x.** * Authors contributed equally to this work
2. Xu L, Chen Y, Mayakonda A, Koh L, Chong YK, **Sandanaraj E**, Huang ML, Chen JX, Lim SW, Dinh HQ, Kappei D, Winter GE, Ding LW, Ang BT, Berman BP, Bradner JE, Tang C, Koeffler P. A targetable BET proteins- and E2F1-dependent transcriptional program maintains the malignancy of glioblastoma. *Proc Natl Acad Sci U S A.* **2018 May 29;115(22): E5086-E5095.**
3. Chen B, Ng G, Gao Y, Low SW, **Sandanaraj E**, Ramasamy B, Sekar S, Bhakoo K, Soong TW, Nilius B, Tang C, Robins EG, Goggi J, Liao P. Non-Invasive Multimodality Imaging Directly Shows TRPM4 Inhibition Ameliorates Stroke Reperfusion Injury. *Transl Stroke Res.* **2018 Mar 22. doi: 10.1007/s12975-018-0621-3.**
4. Nina KLM, Lim JK, Leong MF, **Sandanaraj E**, Ang BT, Tang C, Wan ACA. Collaboration of 3D Context and Extracellular Matrix in the Development of Glioma Stemness in a 3D Model. *Biomaterials.* **2016 Feb;78:62-73.**

5. Chong YK*, **Sandanaraj E***, Koh LW, Thangaveloo M, Tan MS, Koh GR, Toh TB, Lim GG, Holbrook JD, Kon OL, Nadarajah M, Ng I, Ng WH, Tan NS, Lim KL, Tang C, Ang BT. ST3GAL1-Associated Transcriptomic Program in Glioblastoma Tumor Growth, Invasion, and Prognosis. *J Natl Cancer Inst.* 2015 Nov 7;108(2). * **Authors contributed equally to this work**

Conference poster presentations:

1. **Sandanaraj E**, Tan MS, Chong YK, Lim SW, Koh LW, Ng WH, Tan NS, Ang BT, Tang C. High-Content Kinome Screen identifies Molecular Mechanisms underlying STAT3-Inhibitor Resistance in Brain Tumors. Frontiers in Cancer Science (2018). Singapore.

2. Yeo J, Fadhlullah SF, Ho R, Abdul Halim N, **Sandanaraj E**, Ang BT, Tang C, Ng WH, Virshup D, Ho I. NF1 inhibits glioma cells invasion and reverts mesenchymal transition. 25th European Association for cancer research conference, Amsterdam, The Netherlands (2018)

3. Tan MS, **Sandanaraj E**, Lim SW, Chong YK, Koh LW, Ng WH, Tan NS, Tang C, Ang BT. STAT3 Signaling stratifies glioma patients and identifies cohorts most likely to receive treatment benefit from inhibition therapy. Cancer Research UK Brain tumour conference, UK (2018).

4. Chen Y, Xu L, Dutra-Clarke M, Mayakonda A, Lin DC, Koh LSW, Chong YK, **Sandanaraj E**, Madan V, Yang H, Doan N, Said JW, Yong WH, Müschen M, Ang BT, Tang C, Breunig JJ, Koeffler P. BCL6 modulates the TP53 and STAT pathways in glioma. American Association for Cancer Research (AACR), USA (2017).

5. Tay FC, Tan MS, Lim SW, Chong YK, **Sandanaraj E**, Koh LW, Chang YT, Tang C, Ang BT. Fatty Acid Binding Protein 7 (FABP7) and CDr3 – Potential Prognostic Indicator and Intracellular Imaging Probe for Anti-EGFR Inhibition Glioblastoma Therapy. *Frontiers in Cancer Science* (2017), Singapore.
6. Tan MS, **Sandanaraj E**, Chong YK, Koh LW, Lim SW, Ng WH, Tan NS, Tang C, Ang BT. STAT3 signaling stratifies glioma patients and identifies cohorts most likely to receive treatment benefit from inhibition therapy. *Frontiers in Cancer Science* (2016), Singapore.
7. Koh LW, **Sandanaraj E**, Chong YK, Lim SW, Ng WH, Tan NS, Tang C, Ang BT. Epigenomic manipulation to facilitate a ROS-based strategy for precision medicine in glioblastoma. *Frontiers in Cancer Science* (2016), Singapore.
8. Chong YK, **Sandanaraj E**, Koh LW, Thangaveloo M, Tan MS, Koh GR, Toh TB, Lim GG, Holbrook JD, Kon OL, Nadarajah M, Ng I, Ng WH, Tan NS, Lim KL, Tang C, Ang BT. ST3GAL1-associated transcriptomic program portends poor prognosis in glioblastoma. American Association for Cancer Research (AACR), New Orleans, USA (2016).
9. Tan MS, **Sandanaraj E**, Chong YK, Thangaveloo M, Koh LW, Thakkar D, Ng WH, Tan NS, Tang C, Ang BT, 2015. Patient Stratification in Brain Tumors – The JAK/STAT Story. *Frontiers in Cancer Science* (2015), Singapore.
10. Tan MS, **Sandanaraj E**, Chong YK, Thangaveloo M, Ng WH, Tan NS, Tang C, Ang BT, 2015. Patient Stratification in Brain Tumors – The

JAK/STAT Story. National Neuroscience Institute Research Day (2015), Singapore.

Oral presentations:

1. **Sandanaraj E.** High-Content Kinome Screen identifies Molecular Mechanisms underlying STAT3-Inhibitor Resistance in Brain Tumors. Lightning talk presentation in Frontiers in Cancer Science., Singapore. 12th November 2018. (*Selected as a top-quality abstract to be presented in the conference*)
2. **Sandanaraj E.** Artificial Intelligence: Understanding the development of chemoresistance in glioma patients, a case study (STAT3). NNRI's research day., 27th April 2018 (*Won Gold award for Basic science and translational category research talk*)
3. **Sandanaraj E.** Use of databases to explore a novel target. Experimental Therapeutic Centre's Drug Discovery Course., Agency for Science, Technology and Research (A*Star). 21st November 2017.
4. **Sandanaraj E.** Computational prediction of epigenetic hallmarks driving the invasive phenotype of glioma tumors. NNI-Neuroscience Academic Clinical Program's research day., 13th April 2017 (*Won Bronze award for Basic science and translational category research talk*)
5. **Sandanaraj E.** Computational analyses of patient-centric genomic databases reveal molecular heterogeneity. Singapore Neuro-Oncology Symposium at The Academia., 11 March 2016.

6. **Sandanaraj E.** Genomics-guided patient stratification approaches to address molecular heterogeneity in glioma invasion. EMBO workshop on the analysis of high-throughput sequencing data. University of Cambridge., 14th September 2016.

7. **Sandanaraj E.** Integrating patient tumor molecular information with prognosis outcome, using a systems biology approach. NNI-Neuroscience Academic Clinical Program's research day research day., 23rd January 2015. (*Won Gold award for Basic science and translational category research talk*)

Book chapters:

1. Tan MS, **Sandanaraj E**, Ang BT, Tang CS. Biobanking: An important resource for precision medicine in glioblastoma. Vol.951 (4). Biobanking in the 21st Century: Applications: Springer; 2016

10 APPENDICES

SUPPLEMENTARY FIGURE-1. Empirical Bayes model adjustment on merged GPC transcriptome databases.

SUPPLEMENTARY TABLES

Supplementary Table 1. Top 5 stem cell-related gene sets significantly enriched with high *ST3GAL1* patient groups

REMBRANDT						
Name	Description	Size	NES*	Nominal <i>p</i> -value†	FDR <i>q</i> -value‡	<i>ST3GAL1</i> class association
†DAZARD_RESPONSE_T O_UV_NHEK_UP	Activated signaling genes responsible for resistance to UVB-induced apoptotic pathway in normal human epidermal keratinocytes (N=45)	134	1.75	0.004	0.19	The transcriptome pattern of REMBRANDT patients with high <i>ST3GAL1</i> expression positively enriched with resistance pathway genes
KANG_IMMORTALIZED_B Y_TERT_DN	Down-regulated genes in the signature of adipose stromal cells immortalized by forced expression of telomerase (N=102)	88	1.72	<0.001	0.13	The transcriptome pattern of REMBRANDT patients with high <i>ST3GAL1</i> expression positively enriched with pathway related to proliferative and multilineage potential
OSWALD_HEMATOPOIET IC_STEM_CELL_IN_COLL AGEN_GEL_DN	Genes down-regulated in hematopoietic stem cells (HSC, CD34+) cultured in a three-dimensional collagen gel compared to the cells grown in suspension (N=275)	207	1.67	0.01	0.15	The transcriptome pattern of REMBRANDT patients with high <i>ST3GAL1</i> expression positively enriched with growth factors, cytokines and chemokines related pathway
OSWALD_HEMATOPOIET IC_STEM_CELL_IN_COLL AGEN_GEL_UP	Genes up-regulated in hematopoietic stem cells (HSC, CD34+) cultured in a three-dimensional collagen gel compared to the cells grown in suspension (N=233)	207	1.67	0.01	0.11	The transcriptome pattern of REMBRANDT patients with high <i>ST3GAL1</i> expression positively enriched with growth factors, cytokines and chemokines related pathway

†GAUSSMANN_MLL_AF4_FUSION_TARGETS_F_UP	Activated signaling genes responsible for pluripotency and self-renewal in embryonic stem cells (N=185)	127	1.66	<0.001	0.10	The transcriptome pattern of REMBRANDT patients with high <i>ST3GAL1</i> expression positively enriched with pluripotency and self-renewal in embryonic stem cells
Gravendeel						
Name	Description	Size	NES*	NOM <i>p</i>-value†	FDR <i>q</i>-value‡	<i>ST3GAL1</i> class association
†DAZARD_RESPONSE_TO_UV_NHEK_UP	Activated signaling genes responsible for resistance to UVB-induced apoptotic pathway in normal human epidermal keratinocytes (N=45)	146	1.72	<0.001	0.09	The transcriptome pattern of Gravendeel patients with high <i>ST3GAL1</i> expression positively enriched with resistance pathway genes
REN_ALVEOLAR_RHABDOMYOSARCOMA_DN	Genes commonly down-regulated in human alveolar rhabdomyosarcoma (ARMS) and its mouse model overexpressing PAX3-FOXO1 (N=408)	392	1.71	<0.001	0.06	The transcriptome pattern of Gravendeel patients with high <i>ST3GAL1</i> expression positively enriched with mesenchymal stem cells pattern causing human alveolar rhabdomyosarcoma
†GAUSSMANN_MLL_AF4_FUSION_TARGETS_F_UP	Activated signaling genes responsible for pluripotency and self-renewal in embryonic stem cells (N=185)	130	1.70	<0.001	0.05	The transcriptome pattern of Gravendeel patients with high <i>ST3GAL1</i> expression positively enriched with pluripotency and self-renewal in embryonic stem cells
BENPORATH_MYC_TARGETS_WITH_EBOX	Targets of c-Myc identified by ChIP on chip in cultured cell lines, focusing on E-box-containing genes; high affinity bound subset (N=230)	214	1.70	0.002	0.04	The transcriptome pattern of Gravendeel patients with high <i>ST3GAL1</i> expression positively enriched with embryonic stem cell like phenotype associated with estrogen receptor-negative tumors, often of the basal-like subtype with poor clinical outcome

HOFMANN_MYELODYSP LASTIC_SYNDROM_RISK _DN	Genes down-regulated in bone marrow hematopoietic stem cells (HSC, CD34+) from patients with low risk of myelodysplastic syndrom (N=23)	20	1.68	0.006	0.04	The transcriptome pattern of Gravendeel patients with high <i>ST3GAL1</i> expression positively enriched with genes, resulting low risk patients susceptible to cell damage
---	---	----	------	-------	------	---

*Normalized enrichment score (NES) was calculated using Kolmogorov–Smirnov-like statistics.

†Nominal *P* values were estimated using empirical phenotype-based permutation test.

‡False discovery rate (FDR) *q*-value is the estimated probability that the normalized enrichment score represents a false positive finding.

Supplementary Table 2. Top 5 Reactome pathway clusters significantly enriched for the core markers from Dazard *et al.*, resistance pathway genes

ID	Description	Back ground ratio	* <i>p</i> -value	<i>q</i> -value†	Gene ID	Count
380108	Homo sapiens: Chemokine receptors bind chemokines	59/29492	<0.001	0.005	CCL20/PPBP/CXCL3	3
162582	Homo sapiens: Signal Transduction	1627/29492	<0.001	0.005	TGFB1/HBEGF/STAT6/LAMA2/PPP1R15A/CCL20/PPBP/CXCL3/ADM/ITPR1	10
1643685	Homo sapiens: Disease	948/29492	0.001	0.01	TGFB1/HBEGF/ST3GAL1/IGFBP6/ATF3/CYBA/ITPR1	7
2173788	Homo sapiens: Downregulation of TGF-beta receptor signaling	26/29492	0.001	0.01	TGFB1/PPP1R15A	2
2173789	Homo sapiens: TGF-beta receptor signaling activates SMADs	29/29492	0.001	0.01	TGFB1/PPP1R15A	2

*Nominal *p*-values were estimated using empirical phenotype-based permutation test.

†False discovery rate (FDR) *q*-value is the estimated probability that the normalized enrichment score represents a false positive finding.

Supplementary Table 3. List of significantly different genes in *ST3GAL1* knockdown clones of GPCs

Probe ID	Entrez Gene ID	Gene symbol	Description	Log ₂ -fold change	<i>p</i> -value*	Adjusted <i>p</i> -value†
1552552_s_at	170482	CLEC4C	C-type lectin domain family 4, member C	1.68	0.001	0.007
1552617_a_at	64326	RFWD2	ring finger and WD repeat domain 2, E3 ubiquitin protein ligase	-1.62	<0.001	<0.001
1552843_at	10861	SLC26A1	solute carrier family 26 (sulfate transporter), member 1	-2.31	<0.001	<0.001
1552880_at	89866	SEC16B	SEC16 homolog B (<i>S. cerevisiae</i>)	-1.63	0.009	0.04
1553292_s_at	124923	SGK494	uncharacterized serine/threonine-protein kinase SgK494	-1.59	<0.001	<0.001
1553356_at	---	---	---	-1.79	0.005	0.02
1553357_at	158696	LOC158696	uncharacterized LOC158696	-2.22	<0.001	<0.001
1553601_a_at	259236	TMIE	transmembrane inner ear	-2.45	<0.001	<0.001
1553721_at	220441	RNF152	ring finger protein 152	-1.92	<0.001	0.001
1554140_at	79819	WDR78	WD repeat domain 78	2.52	<0.001	<0.001
1554141_s_at	79819	WDR78	WD repeat domain 78	1.92	0.002	0.01
1554168_a_at	30011	SH3KBP1	SH3-domain kinase binding protein 1	-2.48	<0.001	<0.001
1554242_a_at	1690	COCH	coagulation factor C homolog, cochlin (<i>Limulus polyphemus</i>)	-2.23	<0.001	0.002
1554307_at	644852	LOC644852	uncharacterized LOC644852	1.67	0.006	0.03

1554445_at	7639	ZNF85	zinc finger protein 85	-1.72	<0.001	0.003
1554700_at	1005	CDH7	cadherin 7, type 2	1.93	0.002	0.009
1554724_at	6538	SLC6A11	solute carrier family 6 (neurotransmitter transporter, GABA), member 11	-1.60	<0.001	0.003
1554981_at	54586	C9orf11	chromosome 9 open reading frame 11	1.59	0.002	0.01
1555003_at	5933	RBL1	retinoblastoma-like 1 (p107)	-1.60	0.003	0.02
1555169_at	3980	LIG3	ligase III, DNA, ATP-dependent	1.86	<0.001	<0.001
1555677_s_at	8243	SMC1A	structural maintenance of chromosomes 1A	-1.68	0.002	0.01
1555801_s_at	151126	ZNF385B	zinc finger protein 385B	-2.00	0.007	0.03
1555950_a_at	1604	CD55	CD55 molecule, decay accelerating factor for complement (Cromer blood group)	-2.22	0.001	0.007
1555997_s_at	3488	IGFBP5	insulin-like growth factor binding protein 5	-1.81	<0.001	0.004
1556357_s_at	157697	ERICH1	glutamate-rich 1	-2.08	0.002	0.01
1556368_at	5261	PHKG2	phosphorylase kinase, gamma 2 (testis)	2.06	0.001	0.008
1556369_a_at	5261	PHKG2	phosphorylase kinase, gamma 2 (testis)	2.04	<0.001	0.003
1556429_a_at	93594	WDR67	WD repeat domain 67	-3.07	<0.001	<0.001
1556486_at	---	---	---	-1.80	<0.001	0.003
1558152_at	1E+08	LOC100131262	uncharacterized LOC100131262	-1.64	<0.001	<0.001
1558164_s_at	5194	PEX13	peroxisomal biogenesis factor 13	-1.70	<0.001	<0.001
1558189_a_at	284570	LOC284570	uncharacterized LOC284570	-2.12	<0.001	0.003

1559147_at	---	---	---	-1.63	0.003	0.01
1559265_at	387640	SKIDA1	SKI/DACH domain containing 1	-2.24	0.001	0.009
1559521_at	---	---	---	1.69	0.007	0.03
1560648_s_at	7259	TSPYL1	TSPY-like 1	-2.03	<0.001	<0.001
1561342_at	150005	LOC150005	uncharacterized LOC150005	1.59	<0.001	0.04
1561431_at	---	---	---	-1.71	0.01	0.04
1561511_at	---	---	---	-1.69	0.007	0.03
1561882_at	94120	SYTL3	synaptotagmin-like 3	-1.71	0.007	0.03
1561998_at	---	---	---	-1.74	0.002	<0.001
1562745_at	---	---	---	1.93	<0.001	0.003
1563254_a_at	1.01E+08	LOC100506733	uncharacterized LOC100506733	-2.10	<0.001	0.002
1563379_at	---	---	---	-1.69	0.01	0.04
1563620_at	8945	BTRC	beta-transducin repeat containing E3 ubiquitin protein ligase	1.70	0.001	0.008
1563904_at	1E+08	LOC100129620	uncharacterized LOC100129620	1.80	<0.001	0.001
1565820_x_at	---	---	---	2.22	0.005	0.02
1565863_at	---	---	---	2.12	<0.001	0.002
1566162_x_at	---	---	---	-1.83	<0.001	0.005

1566269_at	253959	RALGAPA1	Ral GTPase activating protein, alpha subunit 1 (catalytic)	1.87	0.001	0.007
1567664_at	---	---	---	-1.61	0.007	0.03
1569061_at	128239	IQGAP3	IQ motif containing GTPase activating protein 3	-1.78	0.002	0.01
1569723_a_at	84501	SPIRE2	spire homolog 2 (Drosophila)	-1.66	<0.001	0.002
1569872_a_at	64788	LMF1	lipase maturation factor 1	-1.85	<0.001	<0.001
1569873_at	142910	LIPJ	lipase, family member J	1.97	0.008	0.03
1570290_at	---	---	---	-1.63	0.008	0.03
200606_at	1832	DSP	desmoplakin	-2.02	<0.001	0.006
200644_at	65108	MARCKSL1	MARCKS-like 1	-2.47	<0.001	<0.001
200660_at	6282	S100A11	S100 calcium binding protein A11	-2.11	<0.001	<0.001
200791_s_at	8826	IQGAP1	IQ motif containing GTPase activating protein 1	-2.38	<0.001	<0.001
200862_at	1718	DHCR24	24-dehydrocholesterol reductase	-1.70	<0.001	<0.001
201129_at	6432	SRSF7	serine/arginine-rich splicing factor 7	-1.98	<0.001	<0.001
201150_s_at	7078	TIMP3	TIMP metalloproteinase inhibitor 3	-1.73	<0.001	0.002
201291_s_at	7153	TOP2A	topoisomerase (DNA) II alpha 170kDa	-1.72	<0.001	<0.001
201292_at	7153	TOP2A	topoisomerase (DNA) II alpha 170kDa	-1.80	<0.001	<0.001
201489_at	10105	PPIF	peptidylprolyl isomerase F	-1.84	<0.001	<0.001
201490_s_at	10105	PPIF	peptidylprolyl isomerase F	-1.66	<0.001	<0.001
201508_at	3487	IGFBP4	insulin-like growth factor binding protein 4	-1.70	<0.001	0.002

201517_at	22916	NCBP2	nuclear cap binding protein subunit 2, 20kDa	-2.11	<0.001	<0.001
201589_at	8243	SMC1A	structural maintenance of chromosomes 1A	-1.69	<0.001	<0.001
201645_at	3371	TNC	tenascin C	-1.62	<0.001	<0.001
201710_at	4605	MYBL2	v-myb myeloblastosis viral oncogene homolog (avian)-like 2	-2.80	<0.001	<0.001
201851_at	6455	SH3GL1	SH3-domain GRB2-like 1	-2.13	<0.001	<0.001
201890_at	6241	RRM2	ribonucleotide reductase M2	-2.67	<0.001	<0.001
201897_s_at	1163	CKS1B	CDC28 protein kinase regulatory subunit 1B	-1.73	<0.001	<0.001
202094_at	332	BIRC5	baculoviral IAP repeat containing 5	-2.29	<0.001	<0.001
202095_s_at	332	BIRC5	baculoviral IAP repeat containing 5	-2.34	<0.001	<0.001
202145_at	4061	LY6E	lymphocyte antigen 6 complex, locus E	-1.90	0.002	0.01
202338_at	7083	TK1	thymidine kinase 1, soluble	-1.85	<0.001	0.001
202363_at	6695	SPOCK1	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 1	-1.59	0.005	0.02
202499_s_at	6515	SLC2A3	solute carrier family 2 (facilitated glucose transporter), member 3	1.59	<0.001	<0.001
202580_x_at	2305	FOXM1	forkhead box M1	-1.87	<0.001	<0.001
202705_at	9133	CCNB2	cyclin B2	-1.85	<0.001	<0.001
202765_s_at	2200	FBN1	fibrillin 1	-1.88	<0.001	<0.001
202766_s_at	2200	FBN1	fibrillin 1	-1.77	<0.001	<0.001

202855_s_at	9123	SLC16A3	solute carrier family 16, member 3 (monocarboxylic acid transporter 4)	1.81	0.002	0.01
202898_at	9672	SDC3	syndecan 3	-1.79	<0.001	<0.001
202954_at	11065	UBE2C	ubiquitin-conjugating enzyme E2C	-1.94	<0.001	<0.001
203154_s_at	10298	PAK4	p21 protein (Cdc42/Rac)-activated kinase 4	-2.06	<0.001	<0.001
203213_at	983	CDK1	cyclin-dependent kinase 1	-2.36	<0.001	<0.001
203214_x_at	983	CDK1	cyclin-dependent kinase 1	-1.97	<0.001	<0.001
203379_at	6195	RPS6KA1	ribosomal protein S6 kinase, 90kDa, polypeptide 1	-2.20	<0.001	<0.001
203505_at	19	ABCA1	ATP-binding cassette, sub-family A (ABC1), member 1	1.73	<0.001	<0.001
203548_s_at	4023	LPL	lipoprotein lipase	-1.93	<0.001	<0.001
203549_s_at	4023	LPL	lipoprotein lipase	-1.68	<0.001	<0.001
203755_at	701	BUB1B	budding uninhibited by benzimidazoles 1 homolog beta (yeast)	-1.89	<0.001	<0.001
203764_at	9787	DLGAP5	discs, large (Drosophila) homolog-associated protein 5	-1.89	<0.001	<0.001
203870_at	64854	USP46	ubiquitin specific peptidase 46	-1.93	<0.001	<0.001
203930_s_at	4137	MAPT	microtubule-associated protein tau	-1.62	<0.001	0.002
203963_at	771	CA12	carbonic anhydrase XII	-1.74	<0.001	<0.001
204126_s_at	8318	CDC45	cell division cycle 45 homolog (S. cerevisiae)	-1.83	0.005	0.02
204159_at	1031	CDKN2C	cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4)	-1.68	<0.001	<0.001

204162_at	10403	NDC80	NDC80 kinetochore complex component homolog (S. cerevisiae)	-1.59	<0.001	0.002
204317_at	55687	TRMU	tRNA 5-methylaminomethyl-2-thiouridylate methyltransferase	-1.84	<0.001	<0.001
204374_s_at	2584	GALK1	galactokinase 1	-1.61	<0.001	<0.001
204444_at	3832	KIF11	kinesin family member 11	-1.59	<0.001	<0.001
204492_at	9824	ARHGAP11 A	Rho GTPase activating protein 11A	-1.66	<0.001	<0.001
204508_s_at	771	CA12	carbonic anhydrase XII	-1.89	<0.001	<0.001
204709_s_at	9493	KIF23	kinesin family member 23	-2.10	<0.001	<0.001
204825_at	9833	MELK	maternal embryonic leucine zipper kinase	-1.85	<0.001	<0.001
204954_s_at	9149	DYRK1B	dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1B	1.64	0.001	0.007
204962_s_at	1058 /// 54978	C2orf18 /// CENPA	chromosome 2 open reading frame 18 /// centromere protein A	-1.69	<0.001	<0.001
205023_at	5888	RAD51	RAD51 homolog (S. cerevisiae)	-1.95	<0.001	0.001
205046_at	1062	CENPE	centromere protein E, 312kDa	-2.04	<0.001	<0.001
205167_s_at	995	CDC25C	cell division cycle 25 homolog C (S. pombe)	-1.65	<0.001	<0.001
205565_s_at	2395	FXN	frataxin	-2.10	<0.001	<0.001
205715_at	683	BST1	bone marrow stromal cell antigen 1	-1.91	<0.001	0.003
205722_s_at	2675	GFRA2	GDNF family receptor alpha 2	-1.72	0.001	0.007
205742_at	7137	TNNI3	troponin I type 3 (cardiac)	-1.60	0.004	0.02

205934_at	5334	PLCL1	phospholipase C-like 1	1.75	<0.001	<0.001
206023_at	10874	NMU	neuromedin U	-2.13	<0.001	<0.001
206074_s_at	3159	HMGA1	high mobility group AT-hook 1	-1.98	<0.001	<0.001
206198_s_at	1087	CEACAM7	carcinoembryonic antigen-related cell adhesion molecule 7	-1.95	<0.001	0.005
206355_at	2774	GNAL	guanine nucleotide binding protein (G protein), alpha activating activity polypeptide, olfactory type	-1.65	0.004	0.02
206364_at	9928	KIF14	kinesin family member 14	-1.97	<0.001	<0.001
206481_s_at	9079	LDB2	LIM domain binding 2	-1.85	0.002	0.01
206488_s_at	948	CD36	CD36 molecule (thrombospondin receptor)	1.83	0.01	0.04
206500_s_at	55320	MIS18BP1	MIS18 binding protein 1	-1.59	<0.001	<0.001
206785_s_at	3821 /// 3822	KLRC1 /// KLRC2	killer cell lectin-like receptor subfamily C, member 1 /// killer cell lectin-like receptor subfamily C, member 2	1.61	0.005	0.02
206884_s_at	8796	SCEL	sciellin	2.41	<0.001	0.001
207299_s_at	2911	GRM1	glutamate receptor, metabotropic 1	-1.62	0.01	0.04
207425_s_at	10801	Sep-09	septin 9	-1.59	<0.001	<0.001
207584_at	4018	LPA	lipoprotein, Lp(a)	1.59	0.005	0.02
207742_s_at	2649	NR6A1	nuclear receptor subfamily 6, group A, member 1	2.48	<0.001	<0.001
207751_at	---	---	---	-1.67	0.007	0.03
208025_s_at	8091	HMGA2	high mobility group AT-hook 2	1.61	<0.001	<0.001

208079_s_at	6790	AURKA	aurora kinase A	-1.69	<0.001	<0.001
208190_s_at	51599	LSR	lipolysis stimulated lipoprotein receptor	-1.82	0.001	0.009
208289_s_at	9538	EI24	etoposide induced 2.4 mRNA	-2.51	<0.001	<0.001
208368_s_at	675	BRCA2	breast cancer 2, early onset	-1.91	<0.001	0.002
208396_s_at	5136	PDE1A	phosphodiesterase 1A, calmodulin-dependent	2.22	<0.001	<0.001
208561_at	10060	ABCC9	ATP-binding cassette, sub-family C (CFTR/MRP), member 9	-2.13	0.004	0.02
208631_s_at	3030	HADHA	hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase/enoyl-CoA hydratase (trifunctional protein), alpha subunit	-1.64	<0.001	<0.001
208657_s_at	10801	Sep-09	septin 9	-2.17	<0.001	<0.001
208736_at	10094	ARPC3	actin related protein 2/3 complex, subunit 3, 21kDa	1.64	<0.001	<0.001
208998_at	7351	UCP2	uncoupling protein 2 (mitochondrial, proton carrier)	-2.48	<0.001	<0.001
209393_s_at	100507 948 /// 9470	EIF4E2 /// LOC100507 948	eukaryotic translation initiation factor 4E family member 2 /// uncharacterized LOC100507948	1.78	<0.001	<0.001
209398_at	3006	HIST1H1C	histone cluster 1, H1c	1.63	<0.001	<0.001
209408_at	11004	KIF2C	kinesin family member 2C	-2.18	<0.001	<0.001
209464_at	9212	AURKB	aurora kinase B	-1.70	<0.001	<0.001
209610_s_at	6509	SLC1A4	solute carrier family 1 (glutamate/neutral amino acid transporter), member 4	-1.87	<0.001	<0.001
209680_s_at	3833	KIFC1	kinesin family member C1	-2.59	<0.001	<0.001

209773_s_at	6241	RRM2	ribonucleotide reductase M2	-2.88	<0.001	<0.001
209959_at	8013	NR4A3	nuclear receptor subfamily 4, group A, member 3	-1.65	<0.001	0.001
210052_s_at	22974	TPX2	TPX2, microtubule-associated, homolog (<i>Xenopus laevis</i>)	-1.71	<0.001	<0.001
210121_at	8707	B3GALT2	UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 2	2.11	<0.001	<0.001
210334_x_at	332	BIRC5	baculoviral IAP repeat containing 5	-1.73	<0.001	<0.001
210448_s_at	5026	P2RX5	purinergic receptor P2X, ligand-gated ion channel, 5	-2.19	<0.001	<0.001
210559_s_at	983	CDK1	cyclin-dependent kinase 1	-1.87	<0.001	<0.001
210815_s_at	10203	CALCRL	calcitonin receptor-like	2.06	0.003	0.01
210840_s_at	8826	IQGAP1	IQ motif containing GTPase activating protein 1	-1.92	<0.001	<0.001
211122_s_at	6373	CXCL11	chemokine (C-X-C motif) ligand 11	2.15	<0.001	0.005
211162_x_at	6319	SCD	stearoyl-CoA desaturase (delta-9-desaturase)	-1.68	<0.001	<0.001
211421_s_at	5979	RET	ret proto-oncogene	-1.59	0.005	0.02
211519_s_at	11004	KIF2C	kinesin family member 2C	-1.68	<0.001	<0.001
211767_at	84296	GINS4	GINS complex subunit 4 (Sld5 homolog)	-1.61	<0.001	0.005
212002_at	26099	C1orf144	chromosome 1 open reading frame 144	-1.62	<0.001	<0.001
212005_at	26099	C1orf144	chromosome 1 open reading frame 144	-2.82	<0.001	<0.001
212023_s_at	4288	MKI67	antigen identified by monoclonal antibody Ki-67	-1.86	<0.001	<0.001
212155_at	149603	RNF187	ring finger protein 187	-2.25	<0.001	<0.001

212295_s_at	6541	SLC7A1	solute carrier family 7 (cationic amino acid transporter, y+ system), member 1	-1.65	<0.001	<0.001
212415_at	23157	Sep-06	septin 6	-1.66	<0.001	0.003
212625_at	8677	STX10	syntaxin 10	-1.69	<0.001	<0.001
212739_s_at	4833	NME4	NME/NM23 nucleoside diphosphate kinase 4	-2.00	<0.001	<0.001
212810_s_at	6509	SLC1A4	solute carrier family 1 (glutamate/neutral amino acid transporter), member 4	-1.87	<0.001	<0.001
212811_x_at	6509	SLC1A4	solute carrier family 1 (glutamate/neutral amino acid transporter), member 4	-1.78	<0.001	<0.001
213013_at	644172 /// 9479	LOC644172 /// MAPK8IP1	mitogen-activated protein kinase 8 interacting protein 1 pseudogene /// mitogen-activated protein kinase 8 interacting protein 1	-1.89	<0.001	<0.001
213045_at	23031	MAST3	microtubule associated serine/threonine kinase 3	-1.79	<0.001	<0.001
213171_s_at	10893	MMP24	matrix metalloproteinase 24 (membrane-inserted)	-1.79	0.004	0.02
213251_at	8467	SMARCA5	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 5	-1.80	<0.001	<0.001
213264_at	5094	PCBP2	poly(rC) binding protein 2	-1.67	<0.001	<0.001
213422_s_at	54587	MXRA8	matrix-remodelling associated 8	-1.60	0.002	0.01
213425_at	7474	WNT5A	wingless-type MMTV integration site family, member 5A	-1.61	0.001	0.007
213446_s_at	8826	IQGAP1	IQ motif containing GTPase activating protein 1	-1.76	<0.001	<0.001
213520_at	9401	RECQL4	RecQ protein-like 4	-1.63	<0.001	0.006

213571_s_at	100507 948 /// 9470	EIF4E2 /// LOC100507 948	eukaryotic translation initiation factor 4E family member 2 /// uncharacterized LOC100507948	1.96	<0.001	<0.001
213599_at	11339	OIP5	Opa interacting protein 5	-1.77	<0.001	<0.001
213606_s_at	396	ARHGDIA	Rho GDP dissociation inhibitor (GDI) alpha	-1.83	<0.001	<0.001
213728_at	3916	LAMP1	lysosomal-associated membrane protein 1	-1.70	<0.001	<0.001
213742_at	9295	SRSF11	serine/arginine-rich splicing factor 11	1.65	<0.001	<0.001
213763_at	28996	HIPK2	homeodomain interacting protein kinase 2	-1.64	<0.001	<0.001
213846_at	1350	COX7C	cytochrome c oxidase subunit VIIc	1.98	<0.001	<0.001
213977_s_at	25792	CIZ1	CDKN1A interacting zinc finger protein 1	-1.95	<0.001	<0.001
214290_s_at	723790 /// 8337	HIST2H2AA 3 /// HIST2H2AA 4	histone cluster 2, H2aa3 /// histone cluster 2, H2aa4	1.86	<0.001	<0.001
214455_at	8339 /// 8343 /// 8344 /// 8346 /// 8347	HIST1H2BC /// HIST1H2BE /// HIST1H2BF /// HIST1H2BG /// HIST1H2BI	histone cluster 1, H2bc /// histone cluster 1, H2be /// histone cluster 1, H2bf /// histone cluster 1, H2bg /// histone cluster 1, H2bi	1.74	<0.001	<0.001
214727_at	675	BRCA2	breast cancer 2, early onset	-1.65	<0.001	<0.001
214793_at	1849	DUSP7	dual specificity phosphatase 7	-2.35	<0.001	<0.001

215508_at	699	BUB1	budding uninhibited by benzimidazoles 1 homolog (yeast)	-1.60	<0.001	0.04
216396_s_at	9538	EI24	etoposide induced 2.4 mRNA	-2.27	<0.001	<0.001
216672_s_at	23040	MYT1L	myelin transcription factor 1-like	1.83	0.004	0.02
216880_at	5890	RAD51B	RAD51 homolog B (<i>S. cerevisiae</i>)	-1.65	<0.001	0.003
217438_at	---	---	---	-1.99	<0.001	0.002
217452_s_at	8707	B3GALT2	UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 2	1.91	<0.001	<0.001
217489_s_at	3570	IL6R	interleukin 6 receptor	-2.16	<0.001	0.004
217491_x_at	1350	COX7C	cytochrome c oxidase subunit VIIc	1.94	<0.001	<0.001
217913_at	27183	VPS4A	vacuolar protein sorting 4 homolog A (<i>S. cerevisiae</i>)	-1.86	<0.001	<0.001
218009_s_at	9055	PRC1	protein regulator of cytokinesis 1	-1.72	<0.001	<0.001
218039_at	51203	NUSAP1	nucleolar and spindle associated protein 1	-2.01	<0.001	<0.001
218280_x_at	723790 /// 8337	HIST2H2AA 3 /// HIST2H2AA 4	histone cluster 2, H2aa3 /// histone cluster 2, H2aa4	1.89	<0.001	<0.001
218403_at	51499	TRIAP1	TP53 regulated inhibitor of apoptosis 1	-1.64	<0.001	<0.001
218498_s_at	30001	ERO1L	ERO1-like (<i>S. cerevisiae</i>)	-1.75	<0.001	<0.001
218542_at	55165	CEP55	centrosomal protein 55kDa	-1.96	<0.001	<0.001

218585_s_at	51514	DTL	denticleless E3 ubiquitin protein ligase homolog (Drosophila)	-2.08	<0.001	<0.001
218662_s_at	64151	NCAPG	non-SMC condensin I complex, subunit G	-1.60	<0.001	0.002
218726_at	55355	HJURP	Holliday junction recognition protein	-1.64	<0.001	<0.001
218782_s_at	29028	ATAD2	ATPase family, AAA domain containing 2	-1.60	<0.001	<0.001
219028_at	28996	HIPK2	homeodomain interacting protein kinase 2	-1.85	<0.001	<0.001
219121_s_at	54845	ESRP1	epithelial splicing regulatory protein 1	1.90	<0.001	0.003
219148_at	55872	PBK	PDZ binding kinase	-1.71	<0.001	0.001
219475_at	29948	OSGIN1	oxidative stress induced growth inhibitor 1	-1.80	<0.001	0.004
219646_at	54849	DEF8	differentially expressed in FDCP 8 homolog (mouse)	1.84	<0.001	<0.001
219733_s_at	10998	SLC27A5	solute carrier family 27 (fatty acid transporter), member 5	-1.69	<0.001	<0.001
219804_at	79933	SYNPO2L	synaptopodin 2-like	-2.29	<0.001	<0.001
219860_at	80741	LY6G5C	lymphocyte antigen 6 complex, locus G5C	-1.77	<0.001	<0.001
219918_s_at	259266	ASPM	asp (abnormal spindle) homolog, microcephaly associated (Drosophila)	-1.82	<0.001	<0.001
219978_s_at	51203	NUSAP1	nucleolar and spindle associated protein 1	-2.18	<0.001	<0.001
219990_at	79733	E2F8	E2F transcription factor 8	-1.77	<0.001	0.006
219993_at	64321	SOX17	SRY (sex determining region Y)-box 17	-1.84	0.001	0.008
220317_at	9227	LRAT	lecithin retinol acyltransferase (phosphatidylcholine--retinol O-acyltransferase)	-1.70	0.003	0.02

220325_at	54457	TAF7L	TAF7-like RNA polymerase II, TATA box binding protein (TBP)-associated factor, 50kDa	1.92	<0.001	0.005
220483_s_at	25897	RNF19A	ring finger protein 19A, E3 ubiquitin protein ligase	1.70	<0.001	0.003
220504_at	11081	KERA	keratocan	1.76	0.002	0.01
220651_s_at	55388	MCM10	minichromosome maintenance complex component 10	-1.81	0.002	0.01
220892_s_at	29968	PSAT1	phosphoserine aminotransferase 1	-1.91	<0.001	<0.001
221406_s_at	100532 732 /// 401251 /// 4439	MSH5 /// MSH5- SAPCD1 /// SAPCD1	mutS homolog 5 (E. coli) /// MSH5-SAPCD1 readthrough (non-protein coding) /// suppressor APC domain containing 1	-1.61	<0.001	0.004
221436_s_at	83461	CDCA3	cell division cycle associated 3	-1.73	<0.001	<0.001
221476_s_at	6138	RPL15	ribosomal protein L15	-1.78	<0.001	<0.001
221520_s_at	55143	CDCA8	cell division cycle associated 8	-1.68	<0.001	<0.001
221703_at	83990	BRIP1	BRCA1 interacting protein C-terminal helicase 1	-1.59	<0.001	0.001
221795_at	4915	NTRK2	neurotrophic tyrosine kinase, receptor, type 2	-2.04	<0.001	<0.001
221796_at	4915	NTRK2	neurotrophic tyrosine kinase, receptor, type 2	-2.42	<0.001	<0.001
222088_s_at	144195 /// 6515	SLC2A14 /// SLC2A3	solute carrier family 2 (facilitated glucose transporter), member 14 /// solute carrier family 2 (facilitated glucose transporter), member 3	1.69	<0.001	<0.001
222116_s_at	125058	TBC1D16	TBC1 domain family, member 16	-1.85	<0.001	<0.001
222170_at	---	---	---	-1.68	<0.001	0.005

222248_s_at	23409	SIRT4	sirtuin 4	1.65	<0.001	<0.001
222256_s_at	100137 047 /// 100137 049 /// 8681	JMJD7 /// JMJD7- PLA2G4B /// PLA2G4B	jumonji domain containing 7 /// JMJD7-PLA2G4B readthrough /// phospholipase A2, group IVB (cytosolic)	-2.01	<0.001	0.003
222403_at	23788	MTCH2	mitochondrial carrier 2	-2.27	<0.001	<0.001
222409_at	23603	CORO1C	coronin, actin binding protein, 1C	-1.60	<0.001	<0.001
222476_at	57472	CNOT6	CCR4-NOT transcription complex, subunit 6	-1.76	<0.001	<0.001
222632_s_at	54585	LZTFL1	leucine zipper transcription factor-like 1	1.67	<0.001	<0.001
222740_at	29028	ATAD2	ATPase family, AAA domain containing 2	-1.87	<0.001	<0.001
222848_at	64105	CENPK	centromere protein K	-1.61	<0.001	<0.001
222936_s_at	51029	DESI2	desumoylating isopeptidase 2	-2.63	<0.001	0.001
223039_at	83606	C22orf13	chromosome 22 open reading frame 13	-1.79	<0.001	<0.001
223062_s_at	29968	PSAT1	phosphoserine aminotransferase 1	-1.73	<0.001	<0.001
223082_at	30011	SH3KBP1	SH3-domain kinase binding protein 1	-2.53	<0.001	<0.001
223229_at	29089	UBE2T	ubiquitin-conjugating enzyme E2T (putative)	-1.62	<0.001	<0.001
223381_at	83540	NUF2	NUF2, NDC80 kinetochore complex component, homolog (S. cerevisiae)	-1.76	<0.001	<0.001
223614_at	4325	MMP16	matrix metalloproteinase 16 (membrane-inserted)	-1.84	<0.001	<0.001
223723_at	4241	MFI2	antigen p97 (melanoma associated) identified by monoclonal antibodies 133.2 and 96.5	-1.61	0.001	0.008

223825_at	57589	KIAA1432	KIAA1432	1.61	<0.001	<0.001
224339_s_at	9068	ANGPTL1	angiopoietin-like 1	1.92	<0.001	<0.001
224383_at	56163	RNF17	ring finger protein 17	-1.82	0.001	0.009
224539_s_at	56134	PCDHAC2	protocadherin alpha subfamily C, 2	1.73	<0.001	0.004
224576_at	57222	ERGIC1	endoplasmic reticulum-golgi intermediate compartment (ERGIC) 1	-1.77	<0.001	<0.001
224659_at	57190	SEPN1	selenoprotein N, 1	-1.89	<0.001	<0.001
224846_at	92799	SHKBP1	SH3KBP1 binding protein 1	-2.10	<0.001	0.002
224901_at	79966	SCD5	stearoyl-CoA desaturase 5	-1.85	<0.001	<0.001
224929_at	340061	TMEM173	transmembrane protein 173	-1.59	<0.001	0.001
224972_at	140823	ROMO1	reactive oxygen species modulator 1	-1.69	<0.001	<0.001
225033_at	6482	ST3GAL1	ST3 beta-galactoside alpha-2,3-sialyltransferase 1	-2.00	<0.001	<0.001
225097_at	28996	HIPK2	homeodomain interacting protein kinase 2	-2.14	<0.001	<0.001
225108_at	8540	AGPS	alkylglycerone phosphate synthase	-1.67	<0.001	<0.001
225113_at	8540	AGPS	alkylglycerone phosphate synthase	-2.10	<0.001	<0.001
225115_at	28996	HIPK2	homeodomain interacting protein kinase 2	-2.06	<0.001	<0.001
225116_at	28996	HIPK2	homeodomain interacting protein kinase 2	-1.89	<0.001	<0.001
225142_at	80853	JHDM1D	jumonji C domain containing histone demethylase 1 homolog D (S. cerevisiae)	1.65	<0.001	<0.001
225368_at	28996	HIPK2	homeodomain interacting protein kinase 2	-2.19	<0.001	<0.001

225612_s_at	100505 668 /// 84002	B3GNT5 /// LOC100505 668	UDP-GlcNAc:betaGal beta-1,3-N- acetylglucosaminyltransferase 5 /// uncharacterized LOC100505668	-1.64	<0.001	<0.001
225687_at	81610	FAM83D	family with sequence similarity 83, member D	-1.63	<0.001	<0.001
225835_at	6558	SLC12A2	solute carrier family 12 (sodium/potassium/chloride transporters), member 2	-1.90	<0.001	<0.001
225836_s_at	83695	C12orf32	chromosome 12 open reading frame 32	-2.14	<0.001	<0.001
225837_at	83695	C12orf32	chromosome 12 open reading frame 32	-2.00	<0.001	<0.001
225847_at	57552	NCEH1	neutral cholesterol ester hydrolase 1	1.82	<0.001	<0.001
225972_at	169200	TMEM64	transmembrane protein 64	-2.12	<0.001	<0.001
226096_at	252995	FNDC5	fibronectin type III domain containing 5	2.00	<0.001	0.005
226408_at	8463	TEAD2	TEA domain family member 2	-2.08	<0.001	<0.001
226834_at	---	---	---	-2.26	<0.001	<0.001
227137_at	143384	C10orf46	Chromosome 10 open reading frame 46	2.79	<0.001	<0.001
227203_at	64839	FBXL17	F-box and leucine-rich repeat protein 17	-1.80	<0.001	<0.001
227296_at	113655	MFSD3	major facilitator superfamily domain containing 3	-2.40	<0.001	<0.001
227519_at	191585	PLAC4	placenta-specific 4	-1.71	<0.001	0.005
227578_at	1E+08	TMPO-AS1	TMPO antisense RNA 1 (non-protein coding)	-1.66	<0.001	<0.001
227828_s_at	84141	FAM176A	family with sequence similarity 176, member A	-1.77	<0.001	0.003
227940_at	339803	LOC339803	uncharacterized LOC339803	-1.86	0.002	0.01
227984_at	64788	LMF1	lipase maturation factor 1	-1.66	<0.001	<0.001

228014_at	138428	PTRH1	peptidyl-tRNA hydrolase 1 homolog (S. cerevisiae)	-1.81	<0.001	<0.001
228205_at	7086	TKT	transketolase	-3.17	<0.001	<0.001
228281_at	220042	C11orf82	chromosome 11 open reading frame 82	-1.75	<0.001	<0.001
228401_at	29028	ATAD2	ATPase family, AAA domain containing 2	-1.94	<0.001	<0.001
228403_at	375704	ENHO	energy homeostasis associated	-2.11	<0.001	<0.001
228522_at	57666	FBRSL1	fibrosin-like 1	-1.80	<0.001	<0.001
228540_at	9444	QKI	QKI, KH domain containing, RNA binding	-1.66	<0.001	<0.001
228846_at	4084	MXD1	MAX dimerization protein 1	2.13	<0.001	<0.001
228868_x_at	81620	CDT1	Chromatin licensing and DNA replication factor 1	-1.90	<0.001	0.002
229032_at	9671	WSCD2	WSC domain containing 2	-1.61	0.006	0.03
229296_at	1.01E+08	LOC100506119	uncharacterized LOC100506119	-1.94	0.002	0.01
229419_at	55294	FBXW7	F-box and WD repeat domain containing 7, E3 ubiquitin protein ligase	1.68	<0.001	<0.001
229442_at	162681	C18orf54	chromosome 18 open reading frame 54	-1.66	<0.001	<0.001
229490_s_at	---	---	---	-2.18	<0.001	<0.001
229538_s_at	128239	IQGAP3	IQ motif containing GTPase activating protein 3	-1.76	<0.001	<0.001
229607_at	1.01E+08	LOC100652912	uncharacterized LOC100652912	-1.72	0.001	0.007
229653_at	55275	VPS53	vacuolar protein sorting 53 homolog (S. cerevisiae)	1.63	<0.001	<0.001
230021_at	90381	C15orf42	chromosome 15 open reading frame 42	-1.68	<0.001	<0.001

230045_at	6900	CNTN2	contactin 2 (axonal)	-1.89	0.002	0.01
230418_s_at	57452	GALNTL1	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-like 1	-1.90	<0.001	0.001
230425_at	2047	EPHB1	EPH receptor B1	-1.72	<0.001	<0.001
230521_at	84904	C9orf100	chromosome 9 open reading frame 100	-2.12	<0.001	0.003
230778_at	---	---	---	2.51	<0.001	0.003
230830_at	123264	OSTBETA	organic solute transporter beta	2.14	<0.001	0.002
230974_at	11269	DDX19B	DEAD (Asp-Glu-Ala-Asp) box polypeptide 19B	-2.41	<0.001	<0.001
231120_x_at	5570	PKIB	protein kinase (cAMP-dependent, catalytic) inhibitor beta	-2.26	<0.001	0.004
231353_at	---	---	---	1.80	<0.001	<0.001
231358_at	83876	MRO	maestro	-2.77	<0.001	<0.001
231515_at	---	---	---	-1.93	<0.001	<0.001
231608_at	---	---	---	-1.71	0.003	0.02
231741_at	1903	S1PR3	sphingosine-1-phosphate receptor 3	1.85	<0.001	0.004
231835_at	127281	FAM213B	family with sequence similarity 213, member B	-1.64	<0.001	<0.001
232124_at	729085	FAM198A	family with sequence similarity 198, member A	-2.11	<0.001	0.006
232238_at	259266	ASPM	asp (abnormal spindle) homolog, microcephaly associated (Drosophila)	-1.80	<0.001	<0.001
232475_at	90381	C15orf42	chromosome 15 open reading frame 42	-1.97	<0.001	<0.001
232544_at	---	---	---	-1.96	<0.001	0.003

232574_at	64131	XYLT1	xylosyltransferase I	-2.12	<0.001	0.005
232651_at	1E+08	LOC100287813	uncharacterized LOC100287813	-1.60	0.004	0.02
232714_at	---	---	---	1.70	<0.001	0.004
232812_at	401052	LOC401052	uncharacterized LOC401052	-1.76	0.005	0.02
232976_at	100033 429 /// 100033 431 /// 100033 432 /// 692236 /// 727708	SNORD116-17 /// SNORD116-19 /// SNORD116-20 /// SNORD116-21 /// SNORD116 @	small nucleolar RNA, C/D box 116-17 /// small nucleolar RNA, C/D box 116-19 /// small nucleolar RNA, C/D box 116-20 /// small nucleolar RNA, C/D box 116-21 /// small nucleolar RNA, C/D box 116 cluster	2.14	<0.001	<0.001
233064_at	23217	ZFR2	zinc finger RNA binding protein 2	-1.71	0.003	0.02
233115_at	---	---	---	-1.66	0.006	0.03
233142_at	---	---	---	-1.89	<0.001	0.002
233174_at	1E+08	LOC100287015	uncharacterized LOC100287015	-1.62	<0.001	0.04
233549_at	5136	PDE1A	phosphodiesterase 1A, calmodulin-dependent	-1.63	0.006	0.03
234126_at	---	---	---	-1.84	0.003	0.01
234150_at	---	---	---	2.30	<0.001	<0.001
234739_at	---	---	---	1.98	0.002	0.01

234932_s_at	64866	CDCP1	CUB domain containing protein 1	-1.60	0.001	0.008
235201_at	93986	FOXP2	forkhead box P2	1.73	0.001	0.008
235266_at	29028	ATAD2	ATPase family, AAA domain containing 2	-1.63	<0.001	<0.001
235404_at	84159	ARID5B	AT rich interactive domain 5B (MRF1-like)	1.63	<0.001	<0.001
235456_at	---	---	---	1.75	<0.001	<0.001
235572_at	147841	SPC24	SPC24, NDC80 kinetochore complex component, homolog (S. cerevisiae)	-1.86	<0.001	<0.001
235609_at	83990	BRIP1	BRCA1 interacting protein C-terminal helicase 1	-1.66	<0.001	<0.001
235681_at	---	---	---	1.86	<0.001	<0.001
235843_at	---	---	---	-1.70	<0.001	<0.001
236120_at	---	---	---	-1.69	<0.001	0.002
236289_at	1.01E+08	LOC100506563	uncharacterized LOC100506563	-1.66	0.007	0.03
236422_at	---	---	---	-1.75	0.005	0.02
236448_at	90249	UNC5A	unc-5 homolog A (C. elegans)	-1.61	0.004	0.02
236481_at	2184	FAH	fumarylacetoacetate hydrolase (fumarylacetoacetase)	-2.27	<0.001	<0.001
236641_at	9928	KIF14	kinesin family member 14	-1.81	<0.001	<0.001
237031_at	---	---	---	-2.23	<0.001	<0.001
237137_at	677766	SCARNA2	small Cajal body-specific RNA 2	2.45	<0.001	<0.001
237169_at	3371	TNC	tenascin C	-2.29	<0.001	<0.001

237177_at	152330	CNTN4	contactin 4	-1.59	0.01	0.05
237692_at	---	---	---	2.21	<0.001	<0.001
238477_at	10749	KIF1C	kinesin family member 1C	-1.66	<0.001	<0.001
238957_at	283970	PDXDC2P	Pyridoxal-dependent decarboxylase domain containing 2, pseudogene	-1.83	0.006	0.03
239002_at	259266	ASPM	asp (abnormal spindle) homolog, microcephaly associated (Drosophila)	-2.13	<0.001	<0.001
239253_at	---	---	---	-1.69	<0.001	<0.001
239521_at	1.01E+08	MFI2-AS1	MFI2 antisense RNA 1 (non-protein coding)	-1.62	<0.001	0.005
239688_at	8243	SMC1A	structural maintenance of chromosomes 1A	-2.00	<0.001	<0.001
239790_s_at	---	---	---	-1.88	<0.001	0.001
239852_at	166785	MMAA	methylmalonic aciduria (cobalamin deficiency) cblA type	1.69	<0.001	0.001
240462_at	171425	CLYBL	citrate lyase beta like	-1.91	0.004	0.02
240466_at	---	---	---	-1.97	<0.001	<0.001
240575_at	---	---	---	-1.89	0.004	0.02
240785_at	---	---	---	1.62	0.001	0.009
241041_at	---	---	---	-1.73	0.003	0.02
241088_at	---	---	---	1.68	<0.001	<0.001
241436_at	6340	SCNN1G	sodium channel, non-voltage-gated 1, gamma subunit	-2.08	0.004	0.02

241569_at	---	---	---	-1.81	<0.001	<0.001
241653_x_at	---	---	---	2.27	<0.001	0.003
241850_at	---	---	---	1.64	0.006	0.03
242328_at	115827	RAB3C	RAB3C, member RAS oncogene family	-1.74	0.007	0.03
242527_at	---	---	---	-1.78	<0.001	<0.001
242560_at	2177	FANCD2	Fanconi anemia, complementation group D2	-1.87	<0.001	<0.001
243032_at	---	---	---	-1.62	<0.001	0.002
243229_at	---	---	---	-1.68	<0.001	<0.001
243272_at	---	---	---	-2.40	<0.001	<0.001
244083_at	---	---	---	2.06	0.003	0.01
244324_at	162681	C18orf54	chromosome 18 open reading frame 54	-1.64	<0.001	<0.001
244623_at	56479	KCNQ5	potassium voltage-gated channel, KQT-like subfamily, member 5	-1.86	0.004	0.02
244647_at	---	---	---	-1.77	<0.001	<0.001
244721_at	94241	TP53INP1	tumor protein p53 inducible nuclear protein 1	1.77	<0.001	0.006

*Two-sided *p*-values were calculated in moderated t-statistics in linear models for microarray data.

†Adjusted *p*-values were calculated using Benjamini and Hochberg's method.

GPC, glioma-propagating cell.

Supplementary Table 4. List of REMBRANDT samples predicted using *ST3GAL1*-associated signature

Samples	Activation score	Normalized score	<i>p</i>-value*	<i>ST3GAL1</i> status	Survival	Status	Histology	Grade	Age
E10462	0.60	1.00	0.1	Low	35.9	1	GBM	IV	45
HF1511	0.59	0.98	0.03	Low	56.6	1	ASTROCYTOMA	II	25
HF1235	0.59	0.97	<0.001	Low	98.4	0	ASTROCYTOMA	II	35
HF1381	0.59	0.97	<0.001	Low	136.2	0	ASTROCYTOMA	II	25
HF0048	0.58	0.95	0.01	Low	59	1	GBM	IV	40
E09278B	0.57	0.95	<0.001	Low	36.6	0	GBM	IV	55
HF0802	0.56	0.93	<0.001	Low	88	0	ASTROCYTOMA	II	30
HF1677	0.55	0.91	0.01	Low	63.5	0	ASTROCYTOMA	II	50
HF0152	0.54	0.90	0.003	Low	131.5	0	ASTROCYTOMA	III	30
E09430	0.53	0.89	0.05	Low	32	1	GBM	IV	45
HF1606	0.53	0.88	0.03	Low	73.6	0	ASTROCYTOMA	II	60
HF1551	0.53	0.88	0.005	Low	70.9	0	ASTROCYTOMA	II	30
E09867	0.53	0.88	0.04	Low	38.7	0	OLIGODENDROGLIOMA	II	75
E09860	0.53	0.87	0.005	Low	36.8	0	OLIGODENDROGLIOMA	II	40
HF0962	0.52	0.87	0.003	Low	116.5	0	OLIGODENDROGLIOMA	II	45
HF0087	0.52	0.87	0.004	Low	78.8	1	ASTROCYTOMA	III	60
HF0305	0.51	0.85	0.1	Low	49.2	1	ASTROCYTOMA	II	30

E09394	0.51	0.84	0.07	Low	51.7	1	MIXED	III	30
HF1489	0.51	0.84	<0.001	Low	68.3	1	ASTROCYTOMA	II	50
HF0726	0.50	0.84	0.006	Low	89.4	0	ASTROCYTOMA	II	35
HF0899	0.50	0.84	0.005	Low	66.7	1	ASTROCYTOMA	II	55
HF1568	0.50	0.82	<0.001	Low	42.6	1	ASTROCYTOMA	II	30
HF1409	0.49	0.82	<0.001	Low	12.7	1	ASTROCYTOMA	III	50
E10097	0.49	0.81	0.02	Low	36.6	0	OLIGODENDROGLIOMA	II/III	60
HF0223	0.48	0.80	<0.001	Low	43.1	1	ASTROCYTOMA	II	45
E09262	0.48	0.80	0.1	Low	13.6	1	MIXED	II/III	50
HF1156	0.48	0.79	0.002	Low	105.3	0	ASTROCYTOMA	II	30
E09804	0.48	0.79	0.009	Low	42.4	0	OLIGODENDROGLIOMA	II	70
HF0434	0.48	0.79	<0.001	Low	6.1	1	ASTROCYTOMA	II	60
HF0022	0.47	0.77	0.006	Low	133.9	0	ASTROCYTOMA	II	20
HF0835	0.47	0.77	0.03	Low	45.3	1	OLIGODENDROGLIOMA	II	20
HF0017	0.47	0.77	0.003	Low	133.2	0	ASTROCYTOMA	III	45
HF0914	0.46	0.76	0.004	Low	146.9	0	ASTROCYTOMA	II	40
HF1587	0.46	0.76	0.002	Low	75.3	0	ASTROCYTOMA	III	30
HF0931	0.46	0.76	0.004	Low	119	0	ASTROCYTOMA	II	35
E10433	0.45	0.75	0.03	Low	10.7	1	GBM	IV	45
HF1613	0.45	0.75	0.11	Low	66.8	1	ASTROCYTOMA	III	35

HF0975	0.45	0.75	<0.001	Low	36.5	1	ASTROCYTOMA	III	60
E09956	0.45	0.74	0.04	Low	21	1	GBM	IV	70
E10138	0.44	0.73	0.004	Low	46.9	0	ASTROCYTOMA	II/III	25
E09670	0.44	0.73	0.003	Low	14.2	1	GBM	IV	75
HF0960	0.44	0.73	<0.001	Low	88.7	1	OLIGODENDROGLIOMA	II	45
E10144	0.44	0.72	0.04	Low	25.9	1	GBM	IV	50
HF0435	0.43	0.72	0.1	Low	110.5	0	ASTROCYTOMA	III	40
E09688	0.43	0.72	0.04	Low	50.8	1	MIXED	II	55
E09997	0.43	0.71	0.003	Low	46.9	0	ASTROCYTOMA	II/III	35
E09673	0.43	0.71	0.02	Low	213.8	0	ASTROCYTOMA	II	45
E10258	0.42	0.70	0.14	Low	20.6	1	GBM	IV	50
E10026	0.42	0.70	0.18	Low	15.4	1	GBM	IV	65
HF1442	0.42	0.70	0.02	Low	47.2	1	ASTROCYTOMA	II	35
E09358	0.42	0.70	<0.001	Low	127.1	0	MIXED	II	55
HF1640	0.42	0.70	<0.001	Low	5.5	1	GBM	IV	50
HF1325	0.42	0.69	0.02	Low	84.5	0	ASTROCYTOMA	II	35
HF1357	0.42	0.69	0.007	Low	30.8	1	OLIGODENDROGLIOMA	III	40
E09278	0.42	0.69	0.19	Low	36.6	0	GBM	IV	40
E10102	0.42	0.69	0.01	Low	38.8	0	GBM	IV	45
HF0757	0.41	0.69	0.06	Low	74.9	1	ASTROCYTOMA	II	40

HF0654	0.41	0.68	0.76	Low	14.6	1	GBM	IV	20
E09818	0.41	0.68	0.08	Low	38.1	0	ASTROCYTOMA	II/III	30
HF1588	0.41	0.68	0.002	Low	75.2	0	ASTROCYTOMA	II	40
E09890	0.41	0.67	0.07	Low	1.8	NA	OLIGODENDROGLIOMA	II	35
HF0026	0.41	0.67	0.001	Low	57.1	1	ASTROCYTOMA	II	60
HF0252	0.41	0.67	<0.001	Low	123.1	0	ASTROCYTOMA	II	35
HF1058	0.40	0.67	0.003	Low	18	1	GBM	IV	40
E10158	0.40	0.67	0.02	Low	40.6	0	GBM	IV	70
E10193	0.40	0.67	0.23	Low	34.2	NA	GBM	IV	50
E09661	0.40	0.67	0.007	Low	48.5	0	OLIGODENDROGLIOMA	II	65
HF1345	0.40	0.66	0.21	Low	83.5	0	ASTROCYTOMA	II	15
E10483	0.40	0.66	0.5	Low	66.4	1	ASTROCYTOMA	II	25
HF1492	0.39	0.65	0.02	Low	2.2	1	GBM	IV	30
HF0505	0.39	0.65	<0.001	Low	3.2	1	GBM	IV	35
E09858	0.39	0.65	0.01	Low	129.5	0	OLIGODENDROGLIOMA	II/III	65
E09192	0.39	0.64	0.02	Low	13.4	1	GBM	IV	75
HF1708	0.39	0.64	<0.001	Low	67.8	0	ASTROCYTOMA	II	20
E09801	0.38	0.64	0.005	Low	42.5	1	ASTROCYTOMA	III	30
E09531	0.38	0.63	0.01	Low	19.3	1	ASTROCYTOMA	II/III	55
E09786	0.38	0.63	0.007	Low	46.3	1	GBM	IV	50

HF0285	0.37	0.62	0.03	Low	14.4	1	OLIGODENDROGLIOMA	II	80
HF0608	0.37	0.62	0.07	Low	10.6	1	ASTROCYTOMA	II	50
E09454	0.37	0.62	<0.001	Low	18	1	GBM	IV	55
HF1509	0.37	0.62	0.02	Low	2.7	1	GBM	IV	40
HF1097	0.37	0.61	0.03	Low	9	1	GBM	IV	45
HF1338	0.37	0.61	0.26	Low	5.9	1	GBM	IV	35
HF0050	0.36	0.60	0.04	Low	26.2	1	GBM	IV	35
E10072	0.36	0.60	0.05	Low	37.2	0	OLIGODENDROGLIOMA	II/III	65
HF1186	0.36	0.60	0.22	Low	20.1	1	ASTROCYTOMA	III	30
E10077	0.36	0.60	0.02	Low	25.8	1	GBM	IV	55
E10318	0.36	0.59	0.09	Low	26.4	NA	GBM	IV	70
HF0268	0.36	0.59	0.4	Low	21.9	1	GBM	IV	45
E09606	0.35	0.58	0.004	Low	13.3	1	GBM	IV	30
E10211	0.35	0.58	0.05	Low	28.4	1	GBM	IV	85
E09513	0.34	0.57	0.72	Low	13.4	1	OLIGODENDROGLIOMA	II/III	55
E09471	0.34	0.56	0.13	Low	27.2	1	ASTROCYTOMA	II	70
HF0450	0.34	0.56	0.1	Low	29.5	1	ASTROCYTOMA	II	30
E09759	0.33	0.55	0.15	Low	20.1	1	GBM	IV	45
HF1581	0.33	0.55	0.02	Low	0.2	1	ASTROCYTOMA	III	65
HF1000	0.33	0.54	0.63	Low	40.3	1	ASTROCYTOMA	II	30

HF0291	0.33	0.54	0.15	Low	94.5	1	OLIGODENDROGLIOMA	II	35
E10041	0.32	0.54	0.14	Low	8.2	1	GBM	IV	35
HF1502	0.32	0.53	0.05	Low	6.4	1	OLIGODENDROGLIOMA	III	70
HF1220	0.32	0.53	0.36	Low	10.4	1	GBM	IV	35
E09852	0.32	0.52	0.28	Low	48.3	1	GBM	IV	50
E09690	0.30	0.50	0.13	Low	62.3	0	GBM	IV	75
HF1469	0.30	0.50	0.42	Low	22.2	1	GBM	IV	25
E09664	0.28	0.47	0.51	Low	27.1	1	OLIGODENDROGLIOMA	II/III	40
E09602	0.25	0.42	0.2	Low	4.1	1	GBM	IV	60
E09832	-0.24	-0.38	0.78	High	5.6	1	GBM	IV	50
HF0583	-0.28	-0.45	0.15	High	15.8	1	GBM	IV	25
E09334	-0.29	-0.47	0.37	High	85.2	1	MIXED	III	50
HF0855	-0.31	-0.49	0.03	High	13.7	1	ASTROCYTOMA	II	55
E09802	-0.31	-0.50	0.03	High	32.2	1	GBM	IV	80
HF0963	-0.31	-0.51	0.06	High	10.6	1	GBM	IV	10
HF1517	-0.32	-0.52	0.07	High	8.3	1	GBM	IV	55
E10300	-0.32	-0.52	0.33	High	9.4	1	GBM	IV	60
E09647	-0.32	-0.52	0.03	High	19.3	1	GBM	IV	55
E10250	-0.33	-0.53	0.06	High	6	NA	GBM	IV	35
HF1286	-0.34	-0.54	0.07	High	13.2	1	ASTROCYTOMA	III	75

HF1334	-0.34	-0.55	0.41	High	115.7	1	ASTROCYTOMA	II	20
E10226	-0.34	-0.56	0.52	High	18.7	1	GBM	IV	65
HF0180	-0.35	-0.56	0.01	High	0.3	1	GBM	IV	35
HF0996	-0.35	-0.56	0.007	High	120.5	0	GBM	IV	50
E10444	-0.35	-0.56	0.01	High	17	1	GBM	IV	60
HF0894	-0.35	-0.57	0.05	High	14.1	1	GBM	IV	40
HF1585	-0.36	-0.58	0.01	High	19.9	1	GBM	IV	25
HF1292	-0.36	-0.58	0.01	High	57.7	1	GBM	IV	35
E09730	-0.36	-0.58	0.48	High	61.7	1	GBM	IV	40
HF0790	-0.36	-0.59	0.42	High	7.5	1	GBM	IV	45
E09782	-0.37	-0.59	0.004	High	17.3	0	GBM	IV	50
E09698	-0.37	-0.59	0.37	High	4.1	NA	GBM	IV	50
HF0543	-0.37	-0.60	0.003	High	67.6	0	GBM	IV	30
HF1493	-0.38	-0.61	0.78	High	41.9	1	OLIGODENDROGLIOMA	III	65
HF1057	-0.38	-0.61	0.1	High	24.5	1	OLIGODENDROGLIOMA	III	55
E10488	-0.38	-0.62	<0.001	High	21.9	1	GBM	IV	55
E09139	-0.38	-0.62	0.01	High	36.5	1	GBM	IV	45
HF0510	-0.38	-0.62	0.32	High	19.6	1	OLIGODENDROGLIOMA	II	45
HF1397	-0.39	-0.62	0.01	High	30	1	GBM	IV	20
HF0460	-0.39	-0.63	0.13	High	10.7	1	OLIGODENDROGLIOMA	III	45

HF0891	-0.40	-0.64	<0.001	High	28.8	1	GBM	IV	35
E10105	-0.40	-0.64	0.004	High	1.1	NA	ASTROCYTOMA	III	40
E10001	-0.40	-0.65	<0.001	High	8.7	1	OLIGODENDROGLIOMA	III	50
HF0445	-0.41	-0.66	0.02	High	47.2	1	GBM	IV	40
HF0142	-0.41	-0.66	0.02	High	0.3	1	GBM	IV	25
E09740	-0.41	-0.66	<0.001	High	16.8	1	GBM	IV	50
HF1139	-0.41	-0.66	0.27	High	15.8	1	GBM	IV	40
E09833B	-0.41	-0.67	0.04	High	20.8	1	GBM	IV	45
E50123	-0.41	-0.67	0.01	High	17.5	1	GBM	IV	30
E09938	-0.41	-0.67	0.02	High	25.2	1	GBM	IV	45
E09744	-0.42	-0.68	0.05	High	46.8	1	GBM	IV	70
E10110	-0.42	-0.68	0.05	High	23.1	1	GBM	IV	50
HF1618	-0.42	-0.69	0.09	High	2.4	1	GBM	IV	50
HF1540	-0.42	-0.69	0.04	High	10.2	1	GBM	IV	55
E10284	-0.43	-0.70	0.12	High	4.8	1	GBM	IV	55
HF1382	-0.43	-0.70	<0.001	High	48.3	1	GBM	IV	35
E09988	-0.44	-0.70	0.02	High	64.6	1	OLIGODENDROGLIOMA	III	70
E10031	-0.44	-0.71	0.005	High	27.5	1	GBM	IV	55
E10290	-0.44	-0.71	0.005	High	6.9	1	GBM	IV	30
E10292	-0.44	-0.71	<0.001	High	5.5	1	GBM	IV	60

HF1137	-0.44	-0.72	0.16	High	18.3	1	GBM	IV	35
E10002	-0.45	-0.72	0.02	High	8.4	1	GBM	IV	55
E09451	-0.45	-0.73	0.001	High	5.3	1	GBM	IV	60
HF1078	-0.46	-0.74	0.005	High	22.8	1	GBM	IV	50
E10013	-0.46	-0.74	<0.001	High	4.9	1	GBM	IV	55
E10262	-0.46	-0.75	<0.001	High	18	1	GBM	IV	50
HF1608	-0.46	-0.75	0.002	High	7.9	1	GBM	IV	60
HF0066	-0.47	-0.76	0.02	High	9.1	1	GBM	IV	50
HF1136	-0.48	-0.77	<0.001	High	45.3	1	ASTROCYTOMA	III	35
E10162	-0.48	-0.77	0.007	High	15.4	NA	ASTROCYTOMA	III	55
HF0108	-0.48	-0.77	0.04	High	132	0	ASTROCYTOMA	III	35
E08021	-0.48	-0.77	0.007	High	81.5	1	OLIGODENDROGLIOMA	III	40
E10313	-0.48	-0.78	<0.001	High	19.6	1	GBM	IV	70
HF0702	-0.48	-0.78	0.001	High	8.5	1	ASTROCYTOMA	III	50
HF0408	-0.49	-0.79	0.11	High	15.8	1	GBM	IV	30
HF1458	-0.49	-0.79	0.006	High	8.8	1	OLIGODENDROGLIOMA	III	75
HF0327	-0.50	-0.81	0.13	High	19.6	1	OLIGODENDROGLIOMA	II	75
E10551	-0.50	-0.81	0.006	High	12.3	1	GBM	IV	65
HF0244	-0.51	-0.82	0.05	High	7.5	1	ASTROCYTOMA	III	60
HF0992	-0.51	-0.82	0.02	High	20	1	GBM	IV	30

E10267	-0.51	-0.83	0.08	High	12	1	GBM	IV	75
HF1191	-0.51	-0.83	0.02	High	0.3	1	GBM	IV	25
E09917	-0.52	-0.84	0.004	High	6.1	1	GBM	IV	50
HF1297	-0.52	-0.85	0.01	High	17.2	1	GBM	IV	60
E09951	-0.52	-0.85	0.001	High	22.6	1	GBM	IV	45
HF1628	-0.56	-0.90	<0.001	High	14.6	1	GBM	IV	70
HF0251	-0.56	-0.91	<0.001	High	22.7	1	OLIGODENDROGLIOMA	III	65
E09448	-0.61	-0.99	0.002	High	229.1	1	OLIGODENDROGLIOMA	III	60
HF0627	-0.62	-1.00	<0.001	High	10.7	1	GBM	IV	35

**p*-values were calculated using Kolmogorov-Smirnov test in the connectivity map approach.

GBM, glioblastoma.

Supplementary Table 5. List of Gravendeel samples predicted using *ST3GAL1*-associated signature

Samples	Activation score	Normalized score	<i>p</i>-value*	<i>ST3GAL1</i> status	Survival	Status	Histology	Grade	Age
GSM405429	0.67	1.00	<0.001	Low	18.72	1	MIXED	III	43.93
GSM405258	0.64	0.97	<0.001	Low	1.85	1	ASTROCYTOMA	II	59.35
GSM405211	0.55	0.82	0.02	Low	1.83	1	OLIGODENDROGLIOMA	III	34.93
GSM405340	0.53	0.80	<0.001	Low	5.56	0	GBM	IV	52.88
GSM405311	0.53	0.79	<0.001	Low	7.48	0	ASTROCYTOMA	II	43
GSM405356	0.52	0.78	<0.001	Low	0.18	1	GBM	IV	43.21
GSM405319	0.51	0.77	<0.001	Low	5.62	1	OLIGODENDROGLIOMA	III	53.65
GSM405325	0.50	0.76	<0.001	Low	3.65	1	MIXED	III	42.98
GSM405331	0.50	0.75	<0.001	Low	8.23	0	OLIGODENDROGLIOMA	III	59.56
GSM405327	0.50	0.75	0.02	Low	3.3	1	OLIGODENDROGLIOMA	III	66.86
GSM405217	0.49	0.74	0.003	Low	6.77	0	GBM	IV	33.2
GSM405318	0.48	0.72	0.03	Low	6.21	0	OLIGODENDROGLIOMA	III	62.46
GSM405365	0.48	0.71	0.08	Low	2.66	1	GBM	IV	64.29
GSM405287	0.47	0.71	0.002	Low	6.87	1	OLIGODENDROGLIOMA	III	43.57
GSM405433	0.47	0.71	0.02	Low	4.91	0	ASTROCYTOMA	II	43.85
GSM405343	0.47	0.70	0.19	Low	NA	0	GBM	IV	66.8
GSM405381	0.46	0.70	0.02	Low	15.82	1	OLIGODENDROGLIOMA	III	38.37

GSM405248	0.46	0.69	0.04	Low	0.84	1	GBM	IV	55.2
GSM405338	0.45	0.67	<0.001	Low	7.96	1	OLIGODENDROGLIOMA	II	50.23
GSM405394	0.45	0.67	0.005	Low	0.86	0	ASTROCYTOMA	II	31.77
GSM405252	0.45	0.67	0.09	Low	0.37	1	GBM	IV	68.04
GSM405286	0.44	0.67	0.72	Low	3.16	1	MIXED	III	62.2
GSM405393	0.44	0.67	0.08	Low	0.08	1	GBM	IV	70.67
GSM405222	0.44	0.66	0.05	Low	1.3	1	GBM	IV	54.06
GSM405408	0.44	0.66	<0.001	Low	1.78	1	OLIGODENDROGLIOMA	II	63.1
GSM405313	0.44	0.66	0.009	Low	3.31	1	GBM	IV	32.5
GSM405249	0.44	0.66	0.62	Low	0.04	1	GBM	IV	23.02
GSM405202	0.43	0.65	0.05	Low	11.67	1	OLIGODENDROGLIOMA	III	28.69
GSM405244	0.43	0.65	0.02	Low	2.05	1	GBM	IV	29.55
GSM405295	0.43	0.64	0.003	Low	1.99	1	ASTROCYTOMA	III	41.77
GSM405281	0.42	0.63	0.16	Low	3.32	1	ASTROCYTOMA	III	37.12
GSM405227	0.42	0.63	0.006	Low	4.77	1	OLIGODENDROGLIOMA	III	48.1
GSM405469	0.42	0.62	<0.001	Low	5.56	0	OLIGODENDROGLIOMA	III	46.86
GSM405275	0.42	0.62	0.13	Low	3.46	1	GBM	IV	60.39
GSM405203	0.42	0.62	0.1	Low	8.92	1	OLIGODENDROGLIOMA	III	38.58
GSM405395	0.42	0.62	<0.001	Low	3.76	1	MIXED	II	55.42
GSM405406	0.41	0.61	<0.001	Low	1.69	1	OLIGODENDROGLIOMA	III	78.82

GSM405439	0.41	0.61	<0.001	Low	3.7	0	MIXED	II	32.92
GSM405299	0.41	0.61	<0.001	Low	1.75	1	GBM	IV	54.94
GSM405251	0.40	0.61	0.23	Low	3.21	1	GBM	IV	26.17
GSM405335	0.40	0.60	<0.001	Low	7.44	1	OLIGODENDROGLIOMA	III	65.59
GSM405437	0.40	0.60	0.01	Low	3.44	0	OLIGODENDROGLIOMA	II	51.63
GSM405334	0.40	0.60	0.002	Low	1.47	1	OLIGODENDROGLIOMA	III	57.01
GSM405425	0.40	0.60	<0.001	Low	16.05	1	OLIGODENDROGLIOMA	II	32.98
GSM405212	0.39	0.58	<0.001	Low	17.49	1	OLIGODENDROGLIOMA	III	23.33
GSM405350	0.38	0.58	0.18	Low	0.61	1	GBM	IV	40.74
GSM405483	0.38	0.57	0.006	Low	0.19	0	MIXED	I	32.35
GSM405256	0.38	0.56	0.03	Low	4.79	1	ASTROCYTOMA	II	38.05
GSM405250	0.37	0.56	0.11	Low	1.48	1	ASTROCYTOMA	II	30.69
GSM405310	0.37	0.56	<0.001	Low	6.8	0	MIXED	III	35.92
GSM405447	0.37	0.55	0.15	Low	2.79	1	GBM	IV	59.03
GSM405387	0.37	0.55	<0.001	Low	11.41	1	MIXED	III	37.11
GSM405344	0.37	0.55	0.002	Low	1.19	1	OLIGODENDROGLIOMA	II	34.71
GSM405265	0.36	0.54	0.01	Low	1.81	1	ASTROCYTOMA	III	32.14
GSM405255	0.36	0.54	0.1	Low	1.08	1	GBM	IV	47.93
GSM405320	0.35	0.53	0.12	Low	0.6	1	GBM	IV	70.28
GSM405462	0.35	0.53	0.002	Low	7.52	1	ASTROCYTOMA	II	49.83

GSM405423	0.35	0.52	0.01	Low	6.12	1	ASTROCYTOMA	III	35.67
GSM405245	0.34	0.51	0.15	Low	1.4	1	GBM	IV	37.98
GSM405207	0.34	0.51	0.04	Low	8.12	1	OLIGODENDROGLIOMA	III	44.41
GSM405278	0.34	0.51	0.003	Low	0.73	1	GBM	IV	58.23
GSM405478	0.34	0.51	0.01	Low	6.06	1	OLIGODENDROGLIOMA	III	26.66
GSM405298	0.34	0.51	0.28	Low	1.96	1	MIXED	III	42.56
GSM405442	0.34	0.51	0.13	Low	0.3	1	GBM	IV	63.61
GSM405473	0.34	0.50	0.43	Low	1.12	1	GBM	IV	47.29
GSM405277	0.33	0.50	0.46	Low	2.76	1	ASTROCYTOMA	III	43.89
GSM405420	0.33	0.50	0.01	Low	3	1	OLIGODENDROGLIOMA	III	49.94
GSM405283	0.33	0.50	0.23	Low	4.07	1	MIXED	III	38.07
GSM405261	0.33	0.50	0.45	Low	0.98	1	OLIGODENDROGLIOMA	III	60.46
GSM405404	0.32	0.48	0.06	Low	0.04	1	OLIGODENDROGLIOMA	III	54.35
GSM405450	0.32	0.48	0.04	Low	13.3	0	MIXED	III	36.66
GSM405451	0.32	0.48	<0.001	Low	3.69	1	MIXED	II	28.55
GSM405329	0.32	0.47	0.26	Low	5.02	1	OLIGODENDROGLIOMA	III	60.33
GSM405427	0.31	0.47	0.31	Low	1.53	1	GBM	IV	50.83
GSM405378	0.31	0.46	0.35	Low	0	1	OLIGODENDROGLIOMA	II	57.22
GSM405259	0.31	0.46	0.52	Low	0.41	1	ASTROCYTOMA	III	64.01
GSM405424	0.30	0.45	0.001	Low	3.2	1	ASTROCYTOMA	II	33.25

GSM405418	0.30	0.45	0.23	Low	5.17	1	GBM	IV	46.61
GSM405307	0.30	0.45	0.55	Low	0.33	1	GBM	IV	72.22
GSM405413	0.30	0.45	0.03	Low	1.2	1	ASTROCYTOMA	III	46.47
GSM405412	0.29	0.44	0.14	Low	0.65	1	GBM	IV	55.71
GSM405333	0.29	0.44	0.35	Low	9.11	1	OLIGODENDROGLIOMA	III	57.64
GSM405445	0.28	0.43	0.12	Low	3.33	0	OLIGODENDROGLIOMA	III	56.15
GSM405468	0.28	0.42	0.31	Low	7.04	0	MIXED	I	33.48
GSM405257	0.26	0.39	0.05	Low	10.86	0	OLIGODENDROGLIOMA	III	46.04
GSM405471	0.26	0.38	0.41	Low	0.15	1	GBM	IV	78.12
GSM405376	0.25	0.38	0.51	Low	1.85	1	GBM	IV	52.98
GSM405225	0.23	0.35	0.13	Low	3.47	1	MIXED	III	31.56
GSM405389	-0.24	-0.39	0.1	High	5.56	0	GBM	IV	52.88
GSM405339	-0.25	-0.41	0.44	High	NA	0	GBM	IV	78.08
GSM405456	-0.26	-0.42	0.4	High	NA	0	GBM	IV	52.5
GSM405436	-0.27	-0.43	0.3	High	0.48	1	GBM	IV	79.06
GSM405330	-0.27	-0.43	0.03	High	0.71	1	GBM	IV	33.12
GSM405377	-0.27	-0.44	0.08	High	0.6	1	OLIGODENDROGLIOMA	III	41.98
GSM405464	-0.27	-0.44	0.08	High	0.56	1	GBM	IV	54.72
GSM405454	-0.28	-0.45	0.26	High	3.32	1	GBM	IV	22.52
GSM405305	-0.30	-0.49	0.12	High	1.06	1	GBM	IV	52.69

GSM405229	-0.31	-0.50	0.02	High	0.66	1	GBM	IV	54.18
GSM405241	-0.31	-0.50	0.02	High	0.16	1	GBM	IV	55.98
GSM405363	-0.32	-0.51	0.27	High	9.79	1	GBM	IV	48.84
GSM405271	-0.32	-0.52	0.13	High	0.3	1	GBM	IV	69.89
GSM405366	-0.33	-0.53	0.07	High	2.21	1	OLIGODENDROGLIOMA	III	75.13
GSM405214	-0.33	-0.53	0.01	High	1.5	1	GBM	IV	37.84
GSM405237	-0.33	-0.53	0.34	High	0.41	1	GBM	IV	51.17
GSM405361	-0.33	-0.54	0.19	High	1.64	1	OLIGODENDROGLIOMA	III	79.19
GSM405470	-0.33	-0.54	0.14	High	1.92	1	GBM	IV	31.72
GSM405453	-0.33	-0.54	0.03	High	0.21	1	GBM	IV	70.23
GSM405215	-0.33	-0.54	0.07	High	2.3	1	GBM	IV	51.44
GSM405303	-0.35	-0.56	0.05	High	0.55	1	GBM	IV	56.64
GSM405236	-0.35	-0.56	0.009	High	1.03	1	GBM	IV	52.2
GSM405475	-0.35	-0.57	0.14	High	1.05	1	GBM	IV	33.74
GSM405273	-0.35	-0.57	0.03	High	0.54	1	MIXED	III	56.42
GSM405382	-0.35	-0.57	0.07	High	4.86	1	MIXED	III	44.15
GSM405414	-0.35	-0.57	0.04	High	5.33	1	OLIGODENDROGLIOMA	III	35.23
GSM405449	-0.36	-0.58	<0.001	High	2.02	1	OLIGODENDROGLIOMA	III	45.39
GSM405374	-0.36	-0.58	0.02	High	1.21	1	GBM	IV	58.58
GSM405337	-0.36	-0.59	0.04	High	0.28	1	GBM	IV	15.02

GSM405232	-0.37	-0.59	<0.001	High	0.98	1	GBM	IV	35.7
GSM405234	-0.37	-0.60	0.008	High	0.62	1	GBM	IV	57.68
GSM405383	-0.38	-0.61	0.24	High	1.32	1	ASTROCYTOMA	II	37.6
GSM405292	-0.38	-0.62	0.1	High	1.11	1	GBM	IV	65.52
GSM405289	-0.38	-0.62	0.07	High	0.19	1	ASTROCYTOMA	III	37.44
GSM405388	-0.39	-0.62	0.01	High	0.53	1	MIXED	III	78.52
GSM405262	-0.39	-0.62	0.02	High	2.89	1	GBM	IV	43.26
GSM405448	-0.39	-0.62	0.09	High	0.48	1	GBM	IV	58.71
GSM405426	-0.39	-0.62	0.32	High	0.05	1	GBM	IV	67.1
GSM405384	-0.39	-0.63	0.03	High	0.02	0	GBM	IV	70.3
GSM405239	-0.39	-0.63	<0.001	High	0.59	1	GBM	IV	73.65
GSM405391	-0.40	-0.64	0.12	High	1.05	1	GBM	IV	55.55
GSM405440	-0.40	-0.64	0.03	High	0.53	1	GBM	IV	69.88
GSM405317	-0.41	-0.66	<0.001	High	0.45	1	GBM	IV	62.36
GSM405392	-0.41	-0.67	0.004	High	0.47	1	GBM	IV	48.35
GSM405226	-0.41	-0.67	0.01	High	0.28	1	MIXED	III	48.6
GSM405463	-0.41	-0.67	<0.001	High	2.22	1	GBM	IV	65.53
GSM405285	-0.42	-0.67	0.08	High	1.27	1	MIXED	III	36.61
GSM405280	-0.43	-0.69	0.01	High	2.46	1	GBM	IV	30.43
GSM405312	-0.43	-0.69	0.06	High	1.02	1	GBM	IV	61.31

GSM405247	-0.43	-0.69	0.001	High	1.59	1	GBM	IV	39.36
GSM405349	-0.43	-0.69	0.002	High	2.41	1	GBM	IV	24.42
GSM405399	-0.43	-0.70	0.003	High	1.18	1	MIXED	III	52.07
GSM405467	-0.43	-0.70	<0.001	High	1.34	1	OLIGODENDROGLIOMA	III	43.9
GSM405352	-0.43	-0.70	0.02	High	0.4	1	GBM	IV	69.95
GSM405254	-0.44	-0.72	<0.001	High	0.59	1	GBM	IV	40.58
GSM405347	-0.45	-0.72	0.02	High	0.19	1	MIXED	III	43.11
GSM405460	-0.45	-0.73	0.002	High	0.48	1	OLIGODENDROGLIOMA	III	52.52
GSM405375	-0.47	-0.76	0.09	High	0.06	1	GBM	IV	67.03
GSM405351	-0.48	-0.77	<0.001	High	1.22	1	GBM	IV	42.7
GSM405474	-0.48	-0.77	0.13	High	0.29	1	GBM	IV	61.28
GSM405263	-0.49	-0.79	0.06	High	1.55	1	GBM	IV	61.74
GSM405221	-0.52	-0.85	0.06	High	0.69	1	GBM	IV	55.52
GSM405431	-0.53	-0.86	0.009	High	0.92	1	GBM	IV	80.65
GSM405279	-0.53	-0.86	<0.001	High	0.71	1	MIXED	III	46.15
GSM405465	-0.54	-0.88	<0.001	High	0.63	1	GBM	IV	69.38
GSM405417	-0.60	-0.97	<0.001	High	0.02	1	GBM	IV	77.31
GSM405220	-0.61	-0.98	<0.001	High	1.27	1	GBM	IV	54.12
GSM405309	-0.62	-1.00	<0.001	High	0.26	1	GBM	IV	54.6

* p -values were calculated using Kolmogorov-Smirnov test in the connectivity map approach.

GBM, glioblastoma. Mixed histological type includes anaplastic oligodendroglioma and pilocytic astrocytoma, categories that are consistent with REMBRANDT.

Supplementary Table 6. Enrichment of *ST3GAL1*-signature stratified patients with cellular invasion marked by T2-zone volume calculated from MRI scans

Sample	Activation score	Normalized score	p-value*	<i>ST3GAL1</i> status	Survival	Status	Histology	Grade	Age	MRI ID	T2 zone	Volume group
HF0962	0.52	0.87	0.003	Low	116.5	0	OD	II	45	114894	ND	ND
HF0608	0.37	0.62	0.07	Low	10.6	1	A	II	50	334787	ND	ND
HF1139	-0.41	-0.66	0.27	High	15.8	1	GBM	IV	40	132301	250014.2	BIG
HF0960	0.44	0.73	<0.001	Low	88.7	1	OD	II	45	248713	198642.9	BIG
HF1409	0.49	0.82	<0.001	Low	12.7	1	A	III	50	306817	197273.2	BIG
HF1458	-0.49	-0.79	0.006	High	8.8	1	OD	III	75	137257	163977	BIG
HF1334	-0.34	-0.55	0.41	High	115.7	1	A	II	20	314253	161892.6	BIG
HF0931	0.46	0.76	0.004	Low	119	0	A	II	35	661885	154333.9	BIG
HF1397	-0.39	-0.62	0.01	High	30	1	GBM	IV	20	151599	134251.2	BIG
HF1078	-0.46	-0.74	0.005	High	22.8	1	GBM	IV	50	164631	125179.9	BIG
HF1292	-0.36	-0.58	0.01	High	57.7	1	GBM	IV	35	125307	117424.9	BIG
HF1540	-0.42	-0.69	0.04	High	10.2	1	GBM	IV	55	305041	110427	BIG
HF1058	0.40	0.67	0.003	Low	18	1	GBM	IV	40	289730	107147.2	BIG
HF1357	0.42	0.69	0.007	Low	30.8	1	OD	III	40	155551	102987.7	BIG
HF1568	0.50	0.82	<0.001	Low	42.6	1	A	II	30	916779	91184.73	BIG
HF1297	-0.52	-0.85	0.01	High	17.2	1	GBM	IV	60	271880	86996.98	BIG

HF1097	0.37	0.61	0.03	Low	9	1	GBM	IV	45	828527	78241.37	BIG
HF0855	-0.31	-0.49	0.03	High	13.7	1	A	II	55	316211	73268.22	BIG
HF1493	-0.38	-0.61	0.78	High	41.9	1	OD	III	65	335452	70824.06	BIG
HF1137	-0.44	-0.72	0.16	High	18.3	1	GBM	IV	35	155669	56166.96	BIG
HF1442	0.42	0.70	0.02	Low	47.2	1	A	II	35	327759	54200.33	BIG
HF1628	-0.56	-0.90	<0.001	High	14.6	1	GBM	IV	70	140204	53339.53	BIG
HF1708	0.39	0.64	<0.001	Low	67.8	0	A	II	20	331577	52067.19	BIG
HF1489	0.51	0.84	<0.001	Low	68.3	1	A	II	50	164886	51214.24	BIG
HF1057	-0.38	-0.61	0.1	High	24.5	1	OD	III	55	110620	46390.85	SMALL
HF1156	0.48	0.79	0.002	Low	105.3	0	A	II	30	682545	46002.34	SMALL
HF1587	0.46	0.76	0.002	Low	75.3	0	A	III	30	104772	44861.94	SMALL
HF1551	0.53	0.88	0.005	Low	70.9	0	A	II	30	224028	39929.64	SMALL
HF1613	0.45	0.75	0.11	Low	66.8	1	A	III	35	403475	29951.94	SMALL
HF1325	0.42	0.69	0.02	Low	84.5	0	A	II	35	158134	28916.79	SMALL
HF1517	-0.32	-0.52	0.07	High	8.3	1	GBM	IV	55	310308	25117.03	SMALL
HF0899	0.50	0.84	0.005	Low	66.7	1	A	II	55	164851	21375.4	SMALL
HF0996	-0.35	-0.56	0.007	High	120.5	0	GBM	IV	50	303253	21300	SMALL
HF1502	0.32	0.53	0.05	Low	6.4	1	OD	III	70	238119	20970.13	SMALL
HF1511	0.59	0.98	0.03	Low	56.6	1	A	II	25	311798	20809.91	SMALL
HF1000	0.33	0.54	0.63	Low	40.3	1	A	II	30	170746	18299.78	SMALL

HF1136	-0.48	-0.77	<0.001	High	45.3	1	A	III	35	181119	16524.78	SMALL
HF1191	-0.51	-0.83	0.02	High	0.3	1	GBM	IV	25	693301	15692.26	SMALL
HF0835	0.47	0.77	0.03	Low	45.3	1	OD	II	20	230183	15098.49	SMALL
HF1588	0.41	0.68	0.002	Low	75.2	0	A	II	40	845655	14294.25	SMALL
HF1345	0.40	0.66	0.21	Low	83.5	0	A	II	15	218864	10780.38	SMALL
HF1606	0.53	0.88	0.03	Low	73.6	0	A	II	60	178427	10549.47	SMALL
HF1235	0.59	0.97	<0.001	Low	98.4	0	A	II	35	111076	8922.123	SMALL
HF1677	0.55	0.91	0.01	Low	63.5	0	A	II	50	720946	4046.371	SMALL
HF1381	0.59	0.97	<0.001	Low	136.2	0	A	II	25	867625	2304.358	SMALL

**p*-values were calculated using Kolmogorov-Smirnov test in the connectivity map approach.

MRI, magnetic resonance imaging; GBM, glioblastoma; A, Astrocytoma; OD, Oligodendroglioma; ND, not determined.

Supplementary Table 7. List of Gravendeel samples predicted using functionally-tuned *STAT3* signature

Samples	Up score	Down score	Raw score	Activation score	p-value	STAT3	SURVIVAL	STATUS	WHO	Age
GSM405256	0.603589	-0.56063	1.16422	1	0	Low	4.79	1	IDHmut-NCDL	38.05
GSM405378	0.65369	-0.49259	1.146283	0.984593	0	Low	0	1	IDH wt	57.22
GSM405342	0.487331	-0.60353	1.090861	0.936989	0	Low	2.99	1	IDHmut-NCDL	46.51
GSM405433	0.493485	-0.56168	1.055168	0.906331	0	Low	4.91	0	NA	43.85
GSM405216	0.429048	-0.61331	1.042354	0.895324	0	Low	3.28	1	IDH wt	52.06
GSM405435	0.421416	-0.61031	1.031722	0.886192	0	Low	0.26	1	IDHmut-NCDL	54.98
GSM405346	0.563373	-0.46496	1.028329	0.883278	0	Low	6.5	0	NA	22.28
GSM405203	0.511248	-0.51039	1.021637	0.877529	0	Low	8.92	1	IDH wt	38.58
GSM405251	0.508214	-0.51309	1.021301	0.877241	0	Low	3.21	1	NA	26.17
GSM405444	0.450016	-0.56211	1.012128	0.869362	0	Low	2.84	0	IDH wt	39.92
GSM405315	0.462477	-0.54367	1.006144	0.864222	0	Low	0.35	1	IDH wt	64.61
GSM405287	0.444803	-0.54289	0.987693	0.848373	0	Low	6.87	1	NA	43.57
GSM405424	0.433325	-0.54872	0.98204	0.843518	0	Low	3.2	1	IDHmut-NCDL	33.25
GSM405394	0.412009	-0.5524	0.96441	0.828375	0	Low	0.86	0	IDH wt	31.77
GSM405222	0.499545	-0.45963	0.95917	0.823874	0	Low	1.3	1	NA	54.06
GSM405462	0.343587	-0.60403	0.94762	0.813953	0	Low	7.52	1	IDH wt	49.83
GSM405310	0.386144	-0.53791	0.92405	0.793707	0	Low	6.8	0	IDHmut-CDL	35.92
GSM405478	0.352374	-0.56057	0.91294	0.784165	0	Low	6.06	1	NA	26.66
GSM405451	0.307971	-0.59889	0.906865	0.778946	0	Low	3.69	1	IDHmut-NCDL	28.55
GSM405338	0.37795	-0.52658	0.904535	0.776945	0	Low	7.96	1	IDHmut-CDL	50.23

GSM405441	0.386521	-0.5079	0.894422	0.768259	0	Low	3.27	0	IDH wt	44.74
GSM405298	0.384721	-0.50284	0.887565	0.762369	0	Low	1.96	1	IDHmut-NCDL	42.56
GSM405425	0.337414	-0.54668	0.88409	0.759384	0	Low	16.05	1	IDHmut-CDL	32.98
GSM405311	0.360586	-0.51448	0.875062	0.75163	0	Low	7.48	0	IDH wt	43
GSM405206	0.384441	-0.48615	0.870587	0.747785	0	Low	6.81	1	IDH wt	31.53
GSM405469	0.341548	-0.52378	0.865325	0.743266	0	Low	5.56	0	IDHmut-CDL	46.86
GSM405439	0.385173	-0.47216	0.857336	0.736404	0	Low	3.7	0	IDH wt	32.92
GSM405244	0.358323	-0.49642	0.854738	0.734173	0	Low	2.05	1	IDHmut-NCDL	29.55
GSM405409	0.259136	-0.59224	0.851371	0.731281	0	Low	10.36	1	NA	54
GSM405207	0.33361	-0.51461	0.848224	0.728577	0	Low	8.12	1	IDH wt	44.41
GSM405395	0.232683	-0.61212	0.8448	0.725636	0	Low	3.76	1	IDHmut-CDL	55.42
GSM405437	0.338618	-0.49993	0.838545	0.720263	0	Low	3.44	0	IDHmut-NCDL	51.63
GSM405429	0.346162	-0.47552	0.821681	0.705779	0	Low	18.72	1	IDHmut-NCDL	43.93
GSM405408	0.187787	-0.63266	0.820443	0.704715	0	Low	1.78	1	NA	63.1
GSM405333	0.241656	-0.56377	0.805424	0.691814	0	Low	9.11	1	NA	57.64
GSM405319	0.225594	-0.57829	0.803884	0.690491	0	Low	5.62	1	IDHmut-CDL	53.65
GSM405387	0.290825	-0.50604	0.79686	0.684459	0	Low	11.41	1	IDHmut-NCDL	37.11
GSM405381	0.381292	-0.41191	0.793198	0.681313	0	Low	15.82	1	IDHmut-NCDL	38.37
GSM405335	0.308207	-0.47261	0.780818	0.670679	0	Low	7.44	1	IDHmut-CDL	65.59
GSM405318	0.325038	-0.45371	0.778746	0.668899	0	Low	6.21	0	IDHmut-NCDL	62.46

GSM405227	0.304897	-0.46906	0.773959	0.664788	0	Low	4.77	1	NA	48.1
GSM405480	0.475421	-0.2974	0.772823	0.663812	0	Low	6.31	0	IDH wt	17.55
GSM405356	0.28816	-0.46308	0.751241	0.645275	0	Low	0.18	1	IDHmut- NCDL	43.21
GSM405210	0.274183	-0.47674	0.750927	0.645005	0	Low	10.28	1	IDHmut- CDL	38.53
GSM405331	0.213424	-0.53693	0.750351	0.64451	0	Low	8.23	0	IDHmut- CDL	59.56
GSM405275	0.312935	-0.42224	0.735176	0.631475	0	Low	3.46	1	NA	60.39
GSM405401	0.229537	-0.50362	0.733158	0.629742	0	Low	1.97	1	NA	37.18
GSM405350	0.341815	-0.38828	0.730091	0.627108	0	Low	0.61	1	IDHmut- NCDL	40.74
GSM405327	0.265702	-0.46178	0.727485	0.624869	0	Low	3.3	1	IDHmut- CDL	66.86
GSM405212	0.332805	-0.39353	0.726331	0.623878	0	Low	17.49	1	IDH wt	23.33
GSM405291	0.184288	-0.53636	0.720648	0.618997	0	Low	6.27	1	IDHmut- NCDL	23.88
GSM405201	0.25291	-0.44857	0.701478	0.602531	0	Low	9.82	1	IDHmut- CDL	44.57
GSM405246	0.254956	-0.44262	0.697581	0.599183	0	Low	6.31	1	IDHmut- CDL	32.59
GSM405258	0.253625	-0.43892	0.692546	0.594858	0	Low	1.85	1	IDH wt	59.35
GSM405382	0.344772	-0.34742	0.692189	0.594552	0	Low	4.86	1	NA	44.15
GSM405420	0.262905	-0.42645	0.689352	0.592115	0	Low	3	1	IDHmut- NCDL	49.94
GSM405386	0.183276	-0.49423	0.677506	0.58194	0	Low	3.76	1	IDHmut- CDL	53.85
GSM405250	0.269151	-0.40679	0.675937	0.580592	0	Low	1.48	1	IDHmut- NCDL	30.69
GSM405406	0.161689	-0.50893	0.670624	0.576028	0	Low	1.69	1	NA	78.82
GSM405249	0.332387	-0.33513	0.667512	0.573356	0	Low	0.04	1	IDHmut- NCDL	23.02

GSM405413	0.202912	-0.46406	0.666973	0.572892	0	Low	1.2	1	IDHmut-NCDL	46.47
GSM405483	0.383834	-0.28303	0.666861	0.572797	0	Low	0.19	0	IDH wt	32.35
GSM405482	0.343889	-0.32079	0.664675	0.570919	0	Low	11.21	0	IDH wt	37.02
GSM405340	0.213905	-0.42767	0.641577	0.551079	0	Low	5.56	0	IDHmut-NCDL	52.88
GSM405348	0.380443	-0.2597	0.640141	0.549846	0	Low	0.03	0	NA	11.72
GSM405225	0.439454	-0.19224	0.63169	0.542586	0	Low	3.47	1	NA	31.56
GSM405418	0.263434	-0.35505	0.618486	0.531245	0	Low	5.17	1	IDHmut-CDL	46.61
GSM405368	0.397996	-0.21951	0.617502	0.5304	0	Low	1.81	1	NA	32.14
GSM405211	0.182759	-0.43297	0.615726	0.528875	0	Low	1.83	1	IDHmut-CDL	34.93
GSM405257	0.227501	-0.38488	0.612385	0.526004	0	Low	10.86	0	IDHmut-CDL	46.04
GSM405410	0.259022	-0.35089	0.609914	0.523882	0	Low	20.68	1	IDH wt	37.08
GSM405306	0.385493	-0.22186	0.607349	0.521679	4.00E-04	Low	8.16	0	IDHmut-NCDL	15.67
GSM405402	0.357656	-0.2463	0.603957	0.518765	0	Low	4.55	0	IDH wt	23.72
GSM405369	0.211024	-0.38636	0.59738	0.513116	0	Low	4.13	1	IDHmut-CDL	50.34
GSM405344	0.217365	-0.36791	0.585276	0.502719	0	Low	1.19	1	IDH wt	34.71
GSM405300	0.304346	-0.27536	0.57971	0.497938	0	Low	2.84	1	IDHmut-NCDL	33.26
GSM405303	0.353904	-0.21466	0.568564	0.488365	2.00E-04	Low	0.55	1	IDH wt	56.64
GSM405296	0.193885	-0.37022	0.564109	0.484538	0	Low	0.29	1	IDH wt	41.09
GSM405461	0.209963	-0.3445	0.554463	0.476253	0	Low	0.76	1	IDH wt	49.14
GSM405208	0.252697	-0.29722	0.549917	0.472348	0	Low	3.04	1	IDHmut-CDL	51.4
GSM405370	0.172712	-0.37122	0.543933	0.467208	0	Low	1.61	1	IDH wt	46.52

GSM405329	0.154966	-0.38733	0.542291	0.465798	0	Low	5.02	1	IDHmut-CDL	60.33
GSM405481	0.358186	-0.16827	0.526453	0.452194	1.00E-04	Low	10.37	0	NA	33.88
GSM405283	0.177331	-0.3445	0.52183	0.448223	0	Low	4.07	1	NA	38.07
GSM405325	0.165422	-0.35483	0.520255	0.44687	0	Low	3.65	1	IDHmut-NCDL	42.98
GSM405421	0.275236	-0.24476	0.519991	0.446644	0	Low	6.08	0	IDH wt	38.4
GSM405450	0.222623	-0.2912	0.513826	0.441348	0	Low	13.3	0	NA	36.66
GSM405380	0.240012	-0.26977	0.509778	0.437871	0	Low	6.04	1	IDHmut-CDL	39.99
GSM405467	0.229877	-0.27777	0.507646	0.436039	0	Low	1.34	1	IDHmut-NCDL	43.9
GSM405295	0.245466	-0.26118	0.506645	0.43518	0	Low	1.99	1	NA	41.77
GSM405464	0.357486	-0.14723	0.504719	0.433526	0	Low	0.56	1	IDH wt	54.72
GSM405202	0.319551	-0.17934	0.498892	0.428521	2.00E-04	Low	11.67	1	IDHmut-CDL	28.69
GSM405288	0.198525	-0.28298	0.481506	0.413587	0	Low	1.53	1	NA	41.93
GSM405224	0.191253	-0.28003	0.47128	0.404803	0	Low	0.48	1	IDH wt	47.74
GSM405366	0.169106	-0.27623	0.445338	0.382521	0	Low	2.21	1	IDHmut-CDL	75.13
GSM405475	0.237263	-0.20647	0.443737	0.381146	0	Low	1.05	1	IDH wt	33.74
GSM405423	0.221316	-0.22081	0.44213	0.379765	0	Low	6.12	1	IDHmut-NCDL	35.67
GSM405400	0.291114	-0.14762	0.438731	0.376846	0	Low	0.45	1	IDH wt	49.64
GSM405419	0.242327	-0.18061	0.42294	0.363282	1.00E-04	Low	2.93	1	IDHmut-NCDL	36.27
GSM405376	0.159006	-0.25851	0.417515	0.358622	0	Low	1.85	1	IDH wt	52.98
GSM405286	0.16705	-0.24999	0.417036	0.358211	0	Low	3.16	1	NA	62.2
GSM405330	0.192787	-0.21524	0.408023	0.350469	0	Low	0.71	1	IDHmut-NCDL	33.12
GSM405271	0.21023	-0.19743	0.407661	0.350158	2.00E-04	Low	0.3	1	NA	69.89

GSM405449	0.173033	-0.20807	0.381098	0.327342	0	Low	2.02	1	IDH wt	45.39
GSM405470	0.188604	-0.1855	0.3741	0.321331	1.00E-04	Low	1.92	1	IDHmut-NCDL	31.72
GSM405452	0.140765	-0.21753	0.358296	0.307756	0	Low	0.35	1	IDH wt	71.02
GSM405278	0.21879	-0.13727	0.356065	0.30584	6.00E-04	Low	0.73	1	IDH wt	58.23
GSM405361	0.200855	-0.15086	0.35171	0.302099	0	Low	1.64	1	IDH wt	79.19
GSM405277	0.173908	-0.15104	0.324946	0.279111	8.00E-04	Low	2.76	1	IDHmut-NCDL	43.89
GSM405313	0.155057	-0.16879	0.323845	0.278165	3.00E-04	Low	3.31	1	IDHmut-NCDL	32.5
GSM405477	-0.1455	0.181006	-0.32651	-0.2999	0	High	0.11	1	NA	73.64
GSM405265	-0.16045	0.167316	-0.32777	-0.30106	2.00E-04	High	1.81	1	IDH wt	32.14
GSM405367	-0.21203	0.162917	-0.37495	-0.3444	8.00E-04	High	2.62	1	IDH wt	46.83
GSM405472	-0.17417	0.208358	-0.38252	-0.35136	0	High	0.34	1	IDHmut-NCDL	62.11
GSM405459	-0.1756	0.208999	-0.3846	-0.35326	0	High	1.14	1	IDH wt	64.28
GSM405245	-0.16432	0.229995	-0.39432	-0.36219	0	High	1.4	1	IDH wt	37.98
GSM405412	-0.16303	0.233022	-0.39605	-0.36378	0	High	0.65	1	IDH wt	55.71
GSM405217	-0.24941	0.148496	-0.3979	-0.36548	1.00E-04	High	6.77	0	IDHmut-NCDL	33.2
GSM405411	-0.20935	0.189803	-0.39916	-0.36663	1.00E-04	High	0.05	1	IDH wt	37.55
GSM405302	-0.22005	0.180439	-0.40049	-0.36786	0	High	0.74	1	IDH wt	41.39
GSM405312	-0.2486	0.167892	-0.41649	-0.38256	5.00E-04	High	1.02	1	IDH wt	61.31
GSM405228	-0.20753	0.211733	-0.41927	-0.3851	0	High	0.71	1	IDH wt	45.71
GSM405414	-0.28055	0.144454	-0.42501	-0.39038	1.00E-04	High	5.33	1	IDH wt	35.23
GSM405314	-0.15936	0.285716	-0.44508	-0.40881	0	High	0.65	1	IDH wt	54.37
GSM405388	-0.25134	0.195656	-0.44699	-0.41057	0	High	0.53	1	IDH wt	78.52
GSM405377	-0.22097	0.229511	-0.45048	-0.41377	0	High	0.6	1	IDH wt	41.98
GSM405205	-0.28802	0.163786	-0.45181	-0.415	4.00E-04	High	3.24	1	IDHmut-CDL	48.03

GSM405383	-0.17266	0.289355	-0.46202	-0.42437	0	High	1.32	1	IDH wt	37.6
GSM405264	-0.18123	0.284508	-0.46574	-0.42779	0	High	0.32	1	IDH wt	36.5
GSM405407	-0.20186	0.268194	-0.47005	-0.43175	0	High	0.82	1	IDH wt	81.18
GSM405466	-0.20554	0.268322	-0.47387	-0.43526	0	High	0.28	1	IDH wt	67.11
GSM405458	-0.32235	0.15636	-0.47871	-0.43971	4.00E-04	High	1.58	1	IDH wt	68.46
GSM405428	-0.29376	0.186228	-0.47999	-0.44088	2.00E-04	High	0.79	1	IDH wt	70.51
GSM405289	-0.25981	0.223731	-0.48354	-0.44414	0	High	0.19	1	IDH wt	37.44
GSM405229	-0.22975	0.25839	-0.48814	-0.44836	0	High	0.66	1	IDH wt	54.18
GSM405268	-0.23351	0.255949	-0.48946	-0.44958	0	High	0.64	1	IDH wt	48.04
GSM405323	-0.24055	0.251075	-0.49163	-0.45157	0	High	0.62	1	IDH wt	58.78
GSM405417	-0.18787	0.306694	-0.49456	-0.45427	0	High	0.02	1	NA	77.31
GSM405309	-0.20846	0.289538	-0.498	-0.45742	0	High	0.26	1	IDHmut- NCDL	54.6
GSM405285	-0.34018	0.160348	-0.50053	-0.45975	0	High	1.27	1	NA	36.61
GSM405223	-0.26515	0.23968	-0.50483	-0.46369	0	High	1.92	1	IDH wt	53.26
GSM405255	-0.17114	0.334979	-0.50612	-0.46488	0	High	1.08	1	IDH wt	47.93
GSM405304	-0.18866	0.319534	-0.50819	-0.46678	0	High	0.56	1	IDH wt	66.39
GSM405266	-0.23003	0.286045	-0.51608	-0.47403	0	High	0.25	1	IDH wt	51.62
GSM405422	-0.31616	0.208029	-0.52419	-0.48148	0	High	0.91	1	IDH wt	70.67
GSM405403	-0.22868	0.297714	-0.5264	-0.48351	0	High	0.04	1	NA	38.42
GSM405345	-0.17558	0.353214	-0.5288	-0.48571	0	High	0.89	1	IDH wt	45.63
GSM405219	-0.31222	0.217202	-0.52942	-0.48628	1.00E-04	High	0.58	1	NA	57.25
GSM405476	-0.33249	0.201481	-0.53397	-0.49046	6.00E-04	High	1.31	1	IDH wt	64.97
GSM405233	-0.3061	0.23315	-0.53925	-0.49531	0	High	0.86	1	IDH wt	51.64
GSM405274	-0.20379	0.337019	-0.54081	-0.49675	0	High	1.29	1	IDH wt	63.08
GSM405455	-0.18143	0.361043	-0.54248	-0.49827	0	High	0.23	1	IDHmut- NCDL	55.49
GSM405292	-0.35401	0.200439	-0.55445	-0.50927	5.00E-04	High	1.11	1	IDH wt	65.52

GSM405405	-0.26457	0.293196	-0.55777	-0.51232	0	High	0.61	1	IDH wt	71.09
GSM405426	-0.42711	0.131285	-0.5584	-0.5129	0	High	0.05	1	IDHmut-NCDL	67.1
GSM405226	-0.29547	0.264874	-0.56035	-0.51469	0	High	0.28	1	NA	48.6
GSM405248	-0.30325	0.258052	-0.5613	-0.51556	0	High	0.84	1	IDHmut-NCDL	55.2
GSM405474	-0.34111	0.227663	-0.56877	-0.52243	0	High	0.29	1	IDH wt	61.28
GSM405362	-0.33425	0.235711	-0.56996	-0.52352	0	High	1.06	1	IDH wt	38.11
GSM405434	-0.31015	0.265249	-0.5754	-0.52852	0	High	0.24	1	NA	67.01
GSM405320	-0.31908	0.256589	-0.57567	-0.52876	0	High	0.6	1	IDH wt	70.28
GSM405371	-0.40299	0.177796	-0.58079	-0.53346	1.00E-04	High	1.92	1	NA	53.26
GSM405294	-0.30537	0.277019	-0.58239	-0.53493	0	High	0.8	1	IDH wt	56.41
GSM405347	-0.37031	0.213105	-0.58342	-0.53588	1.00E-04	High	0.19	1	IDHmut-NCDL	43.11
GSM405393	-0.29431	0.293324	-0.58764	-0.53976	0	High	0.08	1	IDH wt	70.67
GSM405213	-0.39842	0.191797	-0.59022	-0.54213	1.00E-04	High	1.34	1	IDH wt	44.09
GSM405399	-0.3998	0.192455	-0.59226	-0.544	1.00E-04	High	1.18	1	IDH wt	52.07
GSM405415	-0.28149	0.316196	-0.59769	-0.54899	0	High	0.5	1	IDH wt	67.48
GSM405299	-0.35928	0.239753	-0.59904	-0.55023	1.00E-04	High	1.75	1	NA	54.94
GSM405364	-0.25205	0.350005	-0.60206	-0.553	0	High	9.85	1	IDHmut-CDL	37.61
GSM405442	-0.25536	0.348578	-0.60394	-0.55473	0	High	0.3	1	IDH wt	63.61
GSM405349	-0.41346	0.203887	-0.61735	-0.56705	7.00E-04	High	2.41	1	IDH wt	24.42
GSM405392	-0.33759	0.283658	-0.62125	-0.57063	0	High	0.47	1	NA	48.35
GSM405215	-0.34184	0.289236	-0.63107	-0.57965	0	High	2.3	1	IDH wt	51.44
GSM405479	-0.29896	0.333571	-0.63253	-0.58099	0	High	0.38	1	IDHmut-NCDL	63.3
GSM405473	-0.38661	0.249227	-0.63584	-0.58403	0	High	1.12	1	IDH wt	47.29
GSM405273	-0.3644	0.273571	-0.63797	-0.58599	0	High	0.54	1	IDH wt	56.42

GSM405290	-0.42901	0.213041	-0.64205	-0.58974	2.00E-04	High	1.16	1	NA	45.5
GSM405322	-0.37365	0.269913	-0.64356	-0.59112	1.00E-04	High	1.14	1	IDH wt	62.28
GSM405351	-0.20499	0.441692	-0.64668	-0.59399	0	High	1.22	1	IDHmut-NCDL	42.7
GSM405440	-0.31734	0.333653	-0.65099	-0.59795	0	High	0.53	1	NA	69.88
GSM405237	-0.25407	0.400604	-0.65467	-0.60133	0	High	0.41	1	IDH wt	51.17
GSM405432	-0.33731	0.32396	-0.66127	-0.60739	0	High	0.55	1	NA	60.78
GSM405276	-0.26	0.403841	-0.66384	-0.60975	0	High	0.28	1	IDH wt	51.85
GSM405343	-0.44643	0.217522	-0.66395	-0.60985	6.00E-04	High	NA	0	IDH wt	66.8
GSM405247	-0.32351	0.341866	-0.66538	-0.61116	0	High	1.59	1	IDH wt	39.36
GSM405373	-0.37424	0.295391	-0.66963	-0.61507	0	High	1.21	1	NA	58.58
GSM405317	-0.2326	0.437037	-0.66964	-0.61508	0	High	0.45	1	IDH wt	62.36
GSM405293	-0.3216	0.349291	-0.67089	-0.61622	0	High	0.88	1	IDHmut-NCDL	63.73
GSM405260	-0.32367	0.353242	-0.67692	-0.62176	0	High	0.82	1	IDH wt	57.58
GSM405334	-0.33807	0.344124	-0.6822	-0.62661	0	High	1.47	1	NA	57.01
GSM405471	-0.30962	0.375473	-0.68509	-0.62927	0	High	0.15	1	IDHmut-NCDL	78.12
GSM405214	-0.38612	0.302414	-0.68853	-0.63243	0	High	1.5	1	IDH wt	37.84
GSM405454	-0.29988	0.392446	-0.69233	-0.63592	0	High	3.32	1	IDHmut-NCDL	22.52
GSM405339	-0.38406	0.309511	-0.69358	-0.63706	0	High	NA	0	IDH wt	78.08
GSM405252	-0.36733	0.330197	-0.69753	-0.64069	0	High	0.37	1	IDH wt	68.04
GSM405270	-0.22481	0.482606	-0.70742	-0.64978	0	High	0.58	1	IDH wt	73.11
GSM405430	-0.46151	0.256891	-0.7184	-0.65986	4.00E-04	High	0.35	1	IDH wt	61.1
GSM405448	-0.33828	0.384582	-0.72286	-0.66396	0	High	0.48	1	NA	58.71
GSM405465	-0.37197	0.361966	-0.73394	-0.67414	0	High	0.63	1	IDH wt	69.38
GSM405328	-0.41732	0.317147	-0.73447	-0.67462	0	High	2.26	1	IDH wt	67.6
GSM405416	-0.48692	0.248221	-0.73514	-0.67524	4.00E-04	High	0.34	1	NA	64.26

GSM405446	-0.44491	0.290709	-0.73562	-0.67568	0	High	0.73	1	IDHmut-NCDL	68.18
GSM405280	-0.34238	0.394577	-0.73696	-0.67691	0	High	2.46	1	IDH wt	30.43
GSM405307	-0.30686	0.434092	-0.74095	-0.68058	0	High	0.33	1	IDH wt	72.22
GSM405231	-0.40392	0.341152	-0.74507	-0.68436	0	High	1.26	1	NA	62.96
GSM405453	-0.38941	0.363137	-0.75255	-0.69123	0	High	0.21	1	IDH wt	70.23
GSM405282	-0.33787	0.416626	-0.75449	-0.69302	0	High	0.12	1	IDH wt	51.92
GSM405220	-0.37777	0.378034	-0.75581	-0.69423	0	High	1.27	1	IDH wt	54.12
GSM405438	-0.50626	0.252684	-0.75895	-0.69711	4.00E-04	High	2.3	1	IDH wt	42.67
GSM405384	-0.38057	0.396699	-0.77727	-0.71393	0	High	0.02	0	IDH wt	70.3
GSM405263	-0.23767	0.540357	-0.77802	-0.71463	0	High	1.55	1	IDH wt	61.74
GSM405447	-0.48607	0.315071	-0.80114	-0.73586	0	High	2.79	1	NA	59.03
GSM405396	-0.37405	0.428304	-0.80235	-0.73698	0	High	0.02	1	IDHmut-NCDL	77.31
GSM405456	-0.3885	0.41668	-0.80518	-0.73958	0	High	NA	0	IDH wt	52.5
GSM405460	-0.5122	0.294915	-0.80712	-0.74135	1.00E-04	High	0.48	1	IDH wt	52.52
GSM405253	-0.3772	0.435373	-0.81257	-0.74636	0	High	12.56	0	IDH wt	34.84
GSM405391	-0.5074	0.323356	-0.83076	-0.76307	0	High	1.05	1	NA	55.55
GSM405254	-0.39965	0.432282	-0.83193	-0.76415	0	High	0.59	1	IDH wt	40.58
GSM405272	-0.46632	0.366219	-0.83254	-0.76471	0	High	1	0	IDH wt	43.64
GSM405301	-0.44897	0.396168	-0.84514	-0.77627	0	High	0.3	1	NA	65.35
GSM405374	-0.3011	0.548038	-0.84914	-0.77995	0	High	1.21	1	IDH wt	58.58
GSM405221	-0.41607	0.435354	-0.85142	-0.78205	0	High	0.69	1	NA	55.52
GSM405341	-0.52886	0.325487	-0.85435	-0.78474	0	High	0.63	1	IDHmut-NCDL	71.11
GSM405230	-0.46032	0.401326	-0.86164	-0.79143	0	High	0.47	1	IDH wt	63.33
GSM405305	-0.44837	0.41829	-0.86666	-0.79605	0	High	1.06	1	IDHmut-NCDL	52.69
GSM405279	-0.35076	0.521829	-0.87259	-0.80149	0	High	0.71	1	IDH wt	46.15

GSM405353	-0.49387	0.39262	-0.88649	-0.81426	0	High	0.7	1	IDH wt	55.39
GSM405326	-0.61145	0.280091	-0.89154	-0.8189	0.0011	High	0.27	1	IDHmut-NCDL	74.56
GSM405398	-0.3518	0.539945	-0.89175	-0.81909	0	High	NA	1	IDH wt	56.85
GSM405324	-0.45225	0.442908	-0.89516	-0.82222	0	High	0.67	1	IDH wt	14.38
GSM405235	-0.45834	0.442241	-0.90058	-0.8272	0	High	0.21	1	IDH wt	71.09
GSM405239	-0.41266	0.500457	-0.91312	-0.83872	0	High	0.59	1	IDH wt	73.65
GSM405236	-0.55629	0.390187	-0.94648	-0.86936	0	High	1.03	1	IDH wt	52.2
GSM405241	-0.51547	0.432565	-0.94804	-0.87079	0	High	0.16	1	NA	55.98
GSM405463	-0.49872	0.475528	-0.97425	-0.89487	0	High	2.22	1	IDH wt	65.53
GSM405375	-0.52259	0.459195	-0.98179	-0.90179	0	High	0.06	1	IDH wt	67.03
GSM405218	-0.61107	0.477641	-1.08871	-1	0	High	0.64	1	IDH wt	32.36

**p*-values were calculated using Kolmogorov-Smirnov test in the connectivity map approach.

Wt, wild-type; mut, Mutant; CDL Chromosomal *1p/19q* co-deletions; NCDL, non-codeletion status for chromosomal *1p/19q*.

Supplementary Table 8. List of TCGA samples predicted using functionally-tuned *STAT3* signature

ID	Up score	Down score	Raw score	Activation score	p-value	STAT3	SURVIVAL	STATUS	WHO	Age
TCGA-06-0221	0.237452	-0.21052	0.447977	0.42756	8.00E-04	Low	19.8115	1	IDHmut-non-codel	31
TCGA-26-1442	0.231688	-0.18134	0.413032	0.394208	0.0011	Low	31.31072	0	IDHmut-non-codel	43
TCGA-CS-4938	0.31105	-0.11031	0.421358	0.402155	0.0178	Low	4.698251	0	IDHmut-non-codel	31
TCGA-CS-4942	0.299977	-0.20185	0.501827	0.478956	0.0041	Low	43.86129	1	IDHmut-non-codel	44
TCGA-CS-4944	0.175955	-0.19058	0.366532	0.349828	0.0017	Low	10.61213	0	IDHmut-non-codel	50
TCGA-CS-5390	0.43114	-0.24811	0.679252	0.648295	0.003	Low	64.59273	0	IDHmut-codel	47
TCGA-CS-5393	0.39252	-0.34251	0.735028	0.701529	0	Low	40.14869	0	IDHmut-non-codel	39
TCGA-CS-5394	0.282406	-0.21955	0.501957	0.47908	0.0016	Low	0.098565	0	IDHmut-non-codel	40
TCGA-CS-6290	0.224396	-0.17555	0.399947	0.381719	0.0107	Low	17.93878	0	IDHmut-non-codel	31
TCGA-CS-6667	0.406605	-0.30139	0.707992	0.675726	1.00E-04	Low	7.556627	0	IDHmut-non-codel	39
TCGA-CS-6668	0.284624	-0.29434	0.578965	0.552579	0	Low	8.016596	0	IDHmut-codel	57

TCGA-CS-6669	0.403313	-0.38155	0.784858	0.749088	0	Low	7.392353	0	IDHwt	26
TCGA-DB-5275	0.275417	-0.32739	0.602811	0.575338	0	Low	47.90244	0	IDHmut-non-codel	36
TCGA-DB-5276	0.33567	-0.16706	0.50273	0.479818	0.0168	Low	72.87217	0	IDHmut-non-codel	32
TCGA-DB-5277	0.158724	-0.13079	0.289509	0.276315	0.0583	Low	50.82653	1	IDHmut-non-codel	34
TCGA-DB-5278	0.331595	-0.37359	0.705187	0.673048	0	Low	50.13658	0	IDHmut-codel	17
TCGA-DB-5279	0.42625	-0.33022	0.756468	0.721992	0	Low	44.48553	0	IDHmut-codel	59
TCGA-DB-5280	0.326292	-0.30128	0.627575	0.598973	0	Low	36.53465	0	IDHmut-non-codel	43
TCGA-DB-5281	0.376752	-0.22548	0.602236	0.57479	0.0128	Low	78.22752	0	IDHmut-non-codel	61
TCGA-DB-A4X9	0.259162	-0.37482	0.633979	0.605086	0	Low	30.06223	0	IDHmut-non-codel	33
TCGA-DB-A4XA	0.274088	-0.36918	0.643272	0.613956	0	Low	10.28358	0	IDHmut-codel	30
TCGA-DB-A4XB	0.383643	-0.28405	0.667696	0.637266	0	Low	12.71485	0	IDHmut-non-codel	38
TCGA-DB-A4XC	0.409213	-0.39425	0.803462	0.766844	0	Low	27.59812	0	IDHmut-non-codel	26
TCGA-DB-A4XD	0.400374	-0.27409	0.674461	0.643723	3.00E-04	Low	27.00673	0	IDHmut-non-codel	32

TCGA-DB-A4XE	0.306927	-0.36593	0.672859	0.642193	0	Low	18.82586	0	IDHmut-non-codel	27
TCGA-DB-A4XF	0.317683	-0.22193	0.539615	0.515022	0.0019	Low	18.36589	0	IDHmut-non-codel	41
TCGA-DB-A4XG	0.332243	-0.32375	0.655997	0.6261	0	Low	72.90502	0	IDHmut-codel	34
TCGA-DB-A4XH	0.498257	-0.39446	0.892718	0.852033	0	Low	12.58343	0	IDHmut-codel	53
TCGA-DH-5141	0.202032	-0.2964	0.498429	0.475713	0	Low	31.80354	0	IDHmut-codel	32
TCGA-DH-5143	0.382923	-0.32594	0.708864	0.676558	0	Low	14.39045	0	IDHmut-non-codel	30
TCGA-DH-5144	0.302327	-0.30114	0.603463	0.57596	0	Low	8.016596	0	IDHmut-codel	56
TCGA-DU-5849	0.510342	-0.44701	0.957355	0.913724	0	Low	14.55472	0	IDHmut-codel	48
TCGA-DU-5851	0.513904	-0.41106	0.924968	0.882813	0	Low	17.44595	0	IDHmut-non-codel	40
TCGA-DU-5853	0.388154	-0.3612	0.749354	0.715202	0	Low	13.37194	0	IDHmut-non-codel	29
TCGA-DU-5855	0.234586	-0.13487	0.369453	0.352615	0.0202	Low	6.800964	0	IDHmut-non-codel	49
TCGA-DU-5870	0.296026	-0.26579	0.561811	0.536207	0	Low	182.2133	0	IDHmut-codel	34
TCGA-DU-5871	0.335131	-0.27327	0.608397	0.580669	0	Low	18.92442	0	IDHmut-non-codel	37

TCGA-DU-5872	0.238294	-0.1431	0.381389	0.364007	0.0376	Low	17.47881	0	IDHmut-non-codel	43
TCGA-DU-5874	0.268877	-0.3395	0.608379	0.580652	0	Low	15.14611	0	IDHmut-codel	62
TCGA-DU-6393	0.272243	-0.27188	0.544125	0.519327	0	Low	52.07502	1	IDHmut-codel	66
TCGA-DU-6395	0.233543	-0.3063	0.539841	0.515238	0	Low	48.98666	1	IDHmut-non-codel	31
TCGA-DU-6397	0.334048	-0.2872	0.621249	0.592936	1.00E-04	Low	46.02971	1	IDHmut-codel	45
TCGA-DU-6399	0.389619	-0.15454	0.544161	0.519362	0.0065	Low	65.7098	1	IDHmut-non-codel	54
TCGA-DU-6400	0.358997	-0.37526	0.734261	0.700797	0	Low	1.215631	1	IDHmut-codel	66
TCGA-DU-6401	0.274016	-0.12679	0.400804	0.382538	0.0507	Low	87.39403	1	IDHmut-non-codel	31
TCGA-DU-6407	0.292673	-0.16114	0.45381	0.433128	0.0058	Low	94.45784	1	IDHmut-non-codel	35
TCGA-DU-6408	0.362522	-0.20219	0.564713	0.538976	0.0169	Low	114.0065	1	IDHmut-non-codel	23
TCGA-DU-6410	0.315671	-0.23035	0.546025	0.52114	0.0012	Low	7.950886	0	IDHmut-codel	56
TCGA-DU-6542	0.298961	-0.15324	0.452202	0.431593	0.0265	Low	7.720902	0	IDHmut-non-codel	25
TCGA-DU-7008	0.380606	-0.33014	0.710749	0.678357	0	Low	156.1265	0	IDHmut-non-codel	41

TCGA-DU-7009	0.449602	-0.32711	0.776713	0.741315	0	Low	138.1877	0	IDHmut-codel	32
TCGA-DU-7014	0.201631	-0.17715	0.378783	0.36152	0.0143	Low	117.3248	1	NA	59
TCGA-DU-7015	0.330585	-0.16025	0.490834	0.468465	0.0034	Low	90.71238	0	IDHmut-non-codel	41
TCGA-DU-7018	0.270207	-0.27807	0.548279	0.523291	1.00E-04	Low	30.65362	1	IDHmut-codel	57
TCGA-DU-7019	0.403971	-0.1534	0.557366	0.531964	0.062	Low	26.28392	0	IDHmut-non-codel	39
TCGA-DU-7292	0.453858	-0.14463	0.598493	0.571217	0.0022	Low	7.950886	1	IDHwt	69
TCGA-DU-7294	0.402617	-0.34654	0.749153	0.715011	0	Low	94.26071	0	IDHmut-codel	53
TCGA-DU-7298	0.282782	-0.14042	0.423199	0.403912	0.0279	Low	18.92442	1	IDHmut-non-codel	38
TCGA-DU-7299	0.42503	-0.33009	0.755125	0.720711	0	Low	43.99271	1	IDHmut-non-codel	33
TCGA-DU-7300	0.469698	-0.35664	0.826336	0.788676	0	Low	61.96434	1	IDHmut-codel	53
TCGA-DU-7302	0.351754	-0.37856	0.730314	0.69703	0	Low	60.25589	0	IDHmut-codel	48
TCGA-DU-7304	0.457272	-0.28114	0.738415	0.704762	4.00E-04	Low	23.29412	1	IDHmut-non-codel	43
TCGA-DU-7306	0.257346	-0.13518	0.392524	0.374635	0.0086	Low	41.95571	0	IDHmut-non-codel	67
TCGA-DU-7309	0.459406	-0.41276	0.872169	0.832421	0	Low	2.759812	0	IDHmut-non-codel	41
TCGA-DU-8162	0.410458	-0.32489	0.735352	0.701838	0	Low	14.58758	1	IDHwt	61

TCGA-DU-8163	0.409031	-0.21042	0.61945	0.591219	0.0112	Low	20.66573	0	IDHmut-non-codel	29
TCGA-DU-8164	0.383435	-0.34935	0.732789	0.699393	0	Low	21.38854	0	IDHmut-codel	51
TCGA-DU-8166	0.38724	-0.27805	0.665294	0.634973	2.00E-04	Low	16.95313	0	IDHmut-non-codel	29
TCGA-DU-8167	0.167224	-0.24226	0.40948	0.390818	0	Low	15.47466	0	IDHmut-non-codel	69
TCGA-DU-8168	0.239062	-0.30608	0.54514	0.520295	0	Low	14.16046	0	IDHmut-codel	55
TCGA-DU-A5TS	0.225762	-0.2824	0.508165	0.485005	0	Low	15.11325	0	IDHmut-non-codel	42
TCGA-DU-A5TW	0.141252	-0.26715	0.408405	0.389792	0	Low	5.651043	0	IDHmut-non-codel	33
TCGA-E1-5302	0.298355	-0.19202	0.490375	0.468026	0.0131	Low	50.10372	1	IDHmut-non-codel	41
TCGA-E1-5303	0.21741	-0.14759	0.365003	0.348368	0.0127	Low	67.41825	1	IDHmut-non-codel	38
TCGA-E1-5305	0.40024	-0.25512	0.65536	0.625492	0.0012	Low	79.93597	1	IDHmut-non-codel	34
TCGA-E1-5307	0.419482	-0.34908	0.768562	0.733535	0	Low	57.89033	1	IDHmut-non-codel	62
TCGA-E1-5311	0.44926	-0.3789	0.828161	0.790418	0	Low	134.1794	1	IDHmut-codel	31
TCGA-E1-5318	0.258443	-0.21187	0.470313	0.448879	0.0027	Low	78.16181	1	IDHmut-codel	42

TCGA-E1-5319	0.304523	-0.27532	0.579848	0.553422	0	Low	95.50919	1	IDHmut-codel	48
TCGA-EZ-7264	0.279206	-0.29016	0.569367	0.543418	0	Low	14.32474	0	IDHmut-codel	47
TCGA-FG-5962	0.424726	-0.45342	0.878147	0.838126	0	Low	47.73817	0	IDHmut-codel	54
TCGA-FG-5964	0.209555	-0.3688	0.578356	0.551998	0	Low	34.33337	0	IDHmut-codel	62
TCGA-FG-6689	0.565754	-0.32093	0.886686	0.846276	1.00E-04	Low	14.91612	0	IDHmut-non-codel	30
TCGA-FG-6690	0.375802	-0.29669	0.672491	0.641843	0	Low	24.96972	0	IDHmut-non-codel	70
TCGA-FG-6691	0.270674	-0.16422	0.434893	0.415073	0.0056	Low	23.95122	0	IDHmut-non-codel	23
TCGA-FG-7636	0.278527	-0.30127	0.579799	0.553375	0	Low	17.87307	0	IDHmut-non-codel	48
TCGA-FG-7638	0.495843	-0.35668	0.852525	0.813671	1.00E-04	Low	22.53846	0	IDHmut-codel	31
TCGA-FG-7641	0.443042	-0.49225	0.93529	0.892665	0	Low	20.60002	0	IDHmut-codel	31
TCGA-FG-7643	0.469151	-0.26954	0.738686	0.705021	7.00E-04	Low	20.07434	0	IDHwt	49
TCGA-FG-8181	0.330743	-0.34744	0.678183	0.647275	0	Low	12.94483	0	IDHwt	23
TCGA-FG-8182	0.386167	-0.29121	0.677375	0.646504	1.00E-04	Low	13.66764	0	IDHmut-non-codel	35
TCGA-FG-8185	0.276807	-0.26521	0.542017	0.517315	1.00E-04	Low	14.22617	0	IDHmut-non-codel	37
TCGA-FG-8186	0.375045	-0.32265	0.697692	0.665895	0	Low	16.00034	0	IDHmut-codel	42

TCGA-FG-8187	0.355657	-0.33702	0.692678	0.661109	0	Low	20.07434	0	IDHmut-codel	31
TCGA-FG-8188	0.339438	-0.28156	0.620994	0.592693	0	Low	14.94898	0	IDHmut-non-codel	41
TCGA-FG-8189	0.37258	-0.40915	0.781731	0.746104	0	Low	11.86062	0	IDHmut-non-codel	33
TCGA-FG-8191	0.205283	-0.21687	0.422152	0.402912	1.00E-04	Low	15.50751	0	IDHmut-non-codel	30
TCGA-FG-A4MY	0.231653	-0.17815	0.409807	0.39113	0.0087	Low	9.659341	0	IDHmut-non-codel	44
TCGA-FG-A60J	0.27004	-0.35482	0.624863	0.596385	0	Low	7.030949	0	IDHmut-non-codel	47
TCGA-FN-7833	0.288514	-0.14455	0.433065	0.413328	0.0217	Low	0.13142	0	IDHmut-non-codel	25
TCGA-HT-7467	0.51889	-0.34406	0.862954	0.823625	0	Low	0.098565	0	IDHmut-codel	54
TCGA-HT-7468	0.399093	-0.18559	0.584684	0.558037	0.007	Low	6.669545	0	IDHmut-codel	30
TCGA-HT-7470	0.531621	-0.37141	0.903034	0.861878	0	Low	8.805113	0	IDHmut-non-codel	37
TCGA-HT-7471	0.23044	-0.24748	0.477923	0.456142	0	Low	0.13142	0	IDHmut-codel	37
TCGA-HT-7472	0.329488	-0.14811	0.477596	0.45583	0.0027	Low	0.032855	0	IDHmut-non-codel	38
TCGA-HT-7473	0.371387	-0.11427	0.485656	0.463522	0.0175	Low	16.52601	0	IDHmut-non-codel	28

TCGA-HT-7474	0.489088	-0.45518	0.944273	0.901238	0	Low	3.745459	0	IDHmut-non-codel	52
TCGA-HT-7475	0.441944	-0.43272	0.874665	0.834802	0	Low	17.4131	0	IDHmut-non-codel	67
TCGA-HT-7476	0.353254	-0.33305	0.686301	0.655023	0	Low	6.538125	0	IDHmut-non-codel	26
TCGA-HT-7479	0.326451	-0.25415	0.580604	0.554143	6.00E-04	Low	7.096658	0	IDHmut-non-codel	44
TCGA-HT-7480	0.360766	-0.45277	0.813535	0.776458	0	Low	75.13916	0	IDHmut-codel	33
TCGA-HT-7481	0.506708	-0.33787	0.844574	0.806083	0	Low	59.13882	0	IDHmut-codel	39
TCGA-HT-7482	0.335804	-0.12595	0.461757	0.440713	0.0331	Low	59.99305	0	IDHmut-non-codel	18
TCGA-HT-7483	0.313415	-0.23523	0.548647	0.523642	0.0031	Low	154.0895	0	IDHmut-non-codel	14
TCGA-HT-7485	0.458687	-0.2021	0.660783	0.630668	0.1136	Low	4.008298	0	IDHmut-non-codel	42
TCGA-HT-7602	0.246356	-0.18601	0.432371	0.412666	0.0111	Low	29.83225	0	IDHmut-non-codel	21
TCGA-HT-7603	0.307807	-0.35402	0.661831	0.631668	0	Low	23.1627	0	IDHmut-non-codel	29
TCGA-HT-7604	0.291929	-0.24007	0.532	0.507754	0.001	Low	107.8626	0	IDHmut-non-codel	50

TCGA-HT-7605	0.534548	-0.46475	0.999301	0.953758	0	Low	4.566831	0	IDHmut-codel	38
TCGA-HT-7607	0.475649	-0.38279	0.858437	0.819314	0	Low	3.15407	1	IDHmut-codel	61
TCGA-HT-7608	0.355743	-0.22406	0.579805	0.55338	0.0089	Low	22.04564	0	IDHmut-codel	61
TCGA-HT-7609	0.381545	-0.2558	0.637341	0.608294	0.0021	Low	45.96401	0	IDHmut-non-codel	34
TCGA-HT-7610	0.449809	-0.4097	0.859507	0.820336	0	Low	56.05046	0	IDHmut-non-codel	25
TCGA-HT-7611	0.301964	-0.18701	0.488978	0.466693	0.0299	Low	57.56178	0	IDHmut-non-codel	36
TCGA-HT-7616	0.207162	-0.12517	0.33233	0.317184	0.0383	Low	0.229984	1	IDHmut-codel	75
TCGA-HT-7620	0.337462	-0.31088	0.648346	0.618798	0	Low	14.25903	0	IDHmut-codel	40
TCGA-HT-7676	0.224933	-0.27802	0.502956	0.480034	0	Low	0.164275	0	IDHmut-non-codel	26
TCGA-HT-7677	0.264146	-0.31976	0.583908	0.557296	0	Low	16.23032	0	IDHmut-codel	53
TCGA-HT-7681	0.284113	-0.44938	0.733493	0.700064	0	Low	44.64981	0	IDHmut-codel	29
TCGA-HT-7687	0.338194	-0.32711	0.665303	0.634982	0	Low	0.098565	0	IDHmut-codel	74
TCGA-HT-7688	0.467845	-0.43391	0.901755	0.860657	0	Low	31.67212	0	IDHmut-non-codel	59
TCGA-HT-7689	0.365826	-0.1834	0.54923	0.524199	0.0565	Low	14.94898	0	IDHmut-non-codel	58

TCGA-HT-7692	0.216661	-0.29188	0.508537	0.485361	0	Low	2.956941	0	IDHmut-codel	43
TCGA-HT-7693	0.177371	-0.2043	0.381668	0.364273	1.00E-04	Low	17.51166	0	IDHmut-non-codel	51
TCGA-HT-7694	0.481174	-0.4193	0.900473	0.859435	0	Low	6.899529	0	IDHmut-codel	60
TCGA-HT-7695	0.494709	-0.42918	0.923893	0.881786	0	Low	14.52187	0	IDHmut-codel	29
TCGA-HT-7854	0.348132	-0.11789	0.466019	0.44478	0.0196	Low	39.45873	0	IDHwt	62
TCGA-HT-7855	0.359705	-0.28401	0.643712	0.614375	3.00E-04	Low	19.22012	0	IDHmut-non-codel	39
TCGA-HT-7856	0.520646	-0.48265	1.003301	0.957576	0	Low	39.06448	0	IDHmut-codel	35
TCGA-HT-7858	0.227737	-0.30025	0.527985	0.503922	0	Low	50.59655	0	IDHmut-non-codel	28
TCGA-HT-7873	0.23753	-0.27608	0.513607	0.4902	0	Low	23.58982	0	IDHmut-non-codel	29
TCGA-HT-7874	0.546255	-0.45057	0.996824	0.951394	0	Low	37.12604	0	IDHmut-codel	41
TCGA-HT-7875	0.321439	-0.39797	0.71941	0.686623	0	Low	0.328549	0	IDHmut-codel	56
TCGA-HT-7877	0.468795	-0.46956	0.938359	0.895594	0	Low	0.13142	0	IDHmut-codel	20
TCGA-HT-7879	0.209994	-0.32764	0.537632	0.51313	0	Low	3.679749	0	IDHmut-non-codel	31
TCGA-HT-7880	0.49027	-0.45068	0.940949	0.898066	0	Low	5.322494	0	IDHmut-non-codel	30
TCGA-HT-7881	0.39802	-0.43303	0.831054	0.793179	0	Low	2.924086	0	IDHmut-codel	38

TCGA-HT-7884	0.37085	-0.38097	0.751815	0.717552	0	Low	11.26923	0	IDHmut-non-codel	44
TCGA-HT-7902	0.309805	-0.23301	0.542819	0.51808	0.0027	Low	31.40928	0	IDHmut-non-codel	30
TCGA-HT-8010	0.544779	-0.41182	0.956598	0.913001	0	Low	1.642745	0	IDHmut-codel	64
TCGA-HT-8012	0.346596	-0.42712	0.773716	0.738454	0	Low	9.396501	0	IDHmut-codel	30
TCGA-HT-8013	0.196016	-0.1513	0.34732	0.331491	0.0069	Low	63.50852	1	IDHmut-non-codel	37
TCGA-HT-8018	0.244663	-0.26017	0.50483	0.481823	1.00E-04	Low	21.4871	0	IDHmut-non-codel	40
TCGA-HT-8019	0.352681	-0.40098	0.753663	0.719315	0	Low	32.98632	0	IDHwt	34
TCGA-HT-8105	0.283234	-0.39943	0.68266	0.651548	0	Low	6.242431	0	IDHmut-codel	54
TCGA-HT-8107	0.462492	-0.37124	0.833727	0.79573	0	Low	0.459969	0	IDHwt	62
TCGA-HT-8108	0.304403	-0.31918	0.623585	0.595165	0	Low	2.496972	0	IDHmut-non-codel	26
TCGA-HT-8109	0.454358	-0.38949	0.843852	0.805394	0	Low	3.942588	0	IDHmut-codel	64
TCGA-HT-8111	0.365729	-0.18331	0.549043	0.52402	0.0028	Low	0.229984	0	IDHmut-non-codel	32
TCGA-HT-8113	0.404129	-0.41726	0.821391	0.783956	0	Low	29.56941	0	IDHmut-non-codel	49
TCGA-HT-8114	0.341925	-0.16432	0.506244	0.483172	0.01	Low	3.876878	0	IDHmut-non-codel	36

TCGA-HT-8564	0.376532	-0.24202	0.618552	0.590362	0.0017	Low	15.70464	0	IDHwt	47
TCGA-HT-A4DV	0.209406	-0.35441	0.56382	0.538124	0	Low	10.48071	0	IDHmut-codel	51
TCGA-HT-A5R5	0.183691	-0.17221	0.355902	0.339682	0.0112	Low	0.13142	0	IDHmut-non-codel	33
TCGA-HT-A5R9	0.23127	-0.22766	0.458933	0.438018	1.00E-04	Low	0.06571	0	IDHmut-codel	48
TCGA-HT-A5RB	0.408995	-0.41533	0.824325	0.786757	0	Low	0.098565	0	IDHmut-non-codel	24
TCGA-HW-7486	0.399495	-0.26398	0.663478	0.63324	9.00E-04	Low	29.89796	0	IDHmut-codel	37
TCGA-HW-7487	0.561217	-0.48653	1.047751	1	0	Low	13.43765	0	IDHmut-codel	39
TCGA-HW-7489	0.48343	-0.45009	0.933524	0.890979	0	Low	8.31229	0	IDHmut-non-codel	38
TCGA-HW-7490	0.231823	-0.18133	0.413155	0.394326	0.0075	Low	19.22012	0	IDHmut-non-codel	41
TCGA-HW-7491	0.353291	-0.29257	0.645861	0.616426	0	Low	40.14869	0	IDHmut-codel	35
TCGA-HW-7493	0.519622	-0.2445	0.764124	0.7293	0.0114	Low	35.87755	0	NA	40
TCGA-HW-7495	0.526695	-0.42274	0.94944	0.906169	0	Low	4.303992	0	IDHmut-codel	45
TCGA-IK-8125	0.491026	-0.38259	0.873616	0.833801	0	Low	42.74422	0	IDHmut-codel	62
TCGA-P5-A5EW	0.382631	-0.29355	0.676183	0.645366	0	Low	6.7024	0	IDHmut-non-codel	20
TCGA-02-0047	-0.16547	0.347054	-0.51252	-0.51001	0	High	14.719	1	IDHwt	78
TCGA-02-0055	-0.38025	0.415931	-0.79618	-0.79228	0	High	2.496972	1	IDHwt	62

TCGA-02-2483	-0.42707	0.194412	-0.62148	-0.61843	1.00E-04	High	15.31038	0	IDHmut-non-codel	43
TCGA-02-2485	-0.3145	0.351282	-0.66578	-0.66252	0	High	15.4418	0	IDHwt	53
TCGA-02-2486	-0.39801	0.375946	-0.77395	-0.77016	0	High	16.19747	0	IDHwt	64
TCGA-06-0125	-0.38732	0.38485	-0.77217	-0.76839	0	High	47.5739	1	IDHwt	63
TCGA-06-0129	-0.15099	0.135128	-0.28611	-0.28471	0.0085	High	33.64342	1	IDHmut-non-codel	30
TCGA-06-0130	-0.45749	0.38829	-0.84578	-0.84163	0	High	12.94483	1	IDHwt	54
TCGA-06-0132	-0.18127	0.142811	-0.32408	-0.32249	0.0038	High	25.33113	1	IDHwt	49
TCGA-06-0138	-0.36224	0.257267	-0.61951	-0.61647	1.00E-04	High	24.21406	1	IDHwt	43
TCGA-06-0141	-0.42957	0.285884	-0.71545	-0.71194	3.00E-04	High	10.28358	1	IDHwt	62
TCGA-06-0152	-0.43729	0.436942	-0.87423	-0.86994	0	High	12.32059	1	IDHwt	68
TCGA-06-0157	-0.45092	0.339974	-0.79089	-0.78701	0	High	3.186925	1	IDHwt	63
TCGA-06-0158	-0.33916	0.367128	-0.70629	-0.70282	0	High	10.80926	1	IDHwt	73
TCGA-06-0168	-0.36548	0.252669	-0.61815	-0.61512	0.0012	High	19.64723	1	IDHwt	59
TCGA-06-0171	-0.19972	0.277256	-0.47698	-0.47464	0	High	13.10911	1	IDHwt	65
TCGA-06-0174	-0.3817	0.340996	-0.7227	-0.71916	0	High	3.21978	1	IDHwt	54
TCGA-06-0178	-0.33517	0.177475	-0.51265	-0.51013	0.0126	High	53.94775	0	IDHmut-non-codel	38
TCGA-06-0184	-0.40287	0.291052	-0.69393	-0.69052	0	High	40.34582	0	IDHwt	63
TCGA-06-0187	-0.36103	0.434041	-0.79507	-0.79118	0	High	27.20386	1	IDHwt	69
TCGA-06-0190	-0.36519	0.434123	-0.79932	-0.7954	0	High	10.415	1	IDHwt	62
TCGA-06-0210	-0.4002	0.426639	-0.82684	-0.82279	0	High	7.392353	1	IDHwt	72
TCGA-06-0211	-0.53649	0.42877	-0.96526	-0.96052	0	High	11.82776	1	IDHwt	47
TCGA-06-0219	-0.2868	0.206891	-0.4937	-0.49127	0.0049	High	0.722808	1	IDHwt	67
TCGA-06-0238	-0.33528	0.149488	-0.48477	-0.48239	0.0279	High	13.30623	1	IDHwt	46

TCGA-06-0644	-0.42054	0.471035	-0.89158	-0.88721	0	High	12.32059	0	IDHwt	71
TCGA-06-0645	-0.49605	0.39498	-0.89103	-0.88666	0	High	5.749608	1	IDHwt	55
TCGA-06-0646	-0.4976	0.387268	-0.88487	-0.88053	0	High	5.749608	1	IDHwt	60
TCGA-06-0649	-0.41143	0.266991	-0.67843	-0.6751	0.0025	High	2.102714	1	IDHwt	73
TCGA-06-0686	-0.37476	0.173337	-0.5481	-0.54541	0.0084	High	9.396501	0	IDHwt	53
TCGA-06-0743	-0.3361	0.214401	-0.5505	-0.5478	0.0204	High	22.30848	0	IDHwt	69
TCGA-06-0744	-0.45185	0.182136	-0.63398	-0.63087	0.0018	High	19.54867	0	IDHwt	66
TCGA-06-0745	-0.49916	0.299876	-0.79904	-0.79512	2.00E-04	High	7.852321	1	IDHwt	59
TCGA-06-0747	-0.34116	0.366126	-0.70728	-0.70381	0	High	2.694102	1	IDHwt	53
TCGA-06-0750	-0.45171	0.260876	-0.71259	-0.70909	0.0061	High	0.919937	1	IDHwt	43
TCGA-06-0878	-0.48267	0.522255	-1.00493	-1	0	High	7.162368	0	IDHwt	74
TCGA-06-0882	-0.33028	0.259056	-0.58934	-0.58645	1.00E-04	High	5.421059	0	IDHwt	30
TCGA-06-1804	-0.21591	0.198338	-0.41425	-0.41222	8.00E-04	High	13.60193	1	IDHwt	81
TCGA-06-2557	-0.42499	0.24785	-0.67284	-0.66954	0.0017	High	1.084212	1	IDHwt	76
TCGA-06-2558	-0.25516	0.255896	-0.51106	-0.50855	0	High	12.48486	1	IDHwt	75
TCGA-06-2559	-0.29055	0.346615	-0.63717	-0.63404	0	High	4.928235	1	IDHwt	83
TCGA-06-2561	-0.2597	0.300453	-0.56015	-0.55741	0	High	9.265082	0	IDHwt	53
TCGA-06-2562	-0.32979	0.434819	-0.76461	-0.76086	0	High	2.825521	0	IDHwt	81
TCGA-06-2563	-0.45278	0.382639	-0.83542	-0.83133	0	High	8.509419	0	IDHwt	72
TCGA-06-2564	-0.30406	0.405714	-0.70978	-0.70629	0	High	5.946737	0	IDHwt	50
TCGA-06-2565	-0.35589	0.30938	-0.66527	-0.66201	0	High	6.800964	0	IDHwt	59
TCGA-06-2567	-0.27489	0.315698	-0.59059	-0.58769	0	High	4.369702	1	IDHwt	65
TCGA-06-2570	-0.18074	0.234837	-0.41557	-0.41354	1.00E-04	High	9.363647	0	IDHmut- non- codel	21
TCGA-06-5408	-0.41969	0.3885	-0.80819	-0.80423	0	High	11.7292	1	IDHwt	54
TCGA-06-5410	-0.36939	0.346068	-0.71546	-0.71195	0	High	3.548329	1	IDHwt	72

TCGA-06-5412	-0.38303	0.385371	-0.7684	-0.76463	0	High	4.533976	1	IDHwt	78
TCGA-06-5413	-0.43013	0.335478	-0.76561	-0.76185	0	High	8.805113	0	IDHwt	67
TCGA-06-5414	-0.39985	0.485499	-0.88535	-0.88101	0	High	8.969388	0	IDHwt	61
TCGA-06-5415	-0.46922	0.315528	-0.78475	-0.7809	0	High	8.542274	0	IDHwt	60
TCGA-06-5416	-0.31021	0.245583	-0.55579	-0.55307	6.00E-04	High	6.7024	0	NA	23
TCGA-06-5418	-0.42502	0.424257	-0.84927	-0.84511	0	High	2.726957	1	IDHwt	75
TCGA-06-5856	-0.30482	0.366876	-0.67169	-0.6684	0	High	3.745459	1	IDHwt	58
TCGA-06-5858	-0.34221	0.406176	-0.74838	-0.74471	0	High	6.143866	0	IDHwt	45
TCGA-06-5859	-0.25371	0.31816	-0.57187	-0.56907	0	High	4.566831	0	IDHwt	63
TCGA-08-0386	-0.37535	0.228511	-0.60386	-0.6009	0.0112	High	18.00449	1	IDHwt	74
TCGA-12-0616	-0.46095	0.268591	-0.72955	-0.72597	0.0017	High	14.719	1	IDHwt	36
TCGA-12-0618	-0.15029	0.130833	-0.28112	-0.27974	0.0164	High	12.97769	1	IDHwt	49
TCGA-12-0619	-0.36642	0.351113	-0.71754	-0.71402	0	High	34.8919	1	IDHwt	60
TCGA-12-0821	-0.35686	0.313824	-0.67069	-0.6674	0	High	10.61213	1	IDHwt	62
TCGA-12-3650	-0.45377	0.244268	-0.69804	-0.69462	3.00E-04	High	10.94068	1	IDHwt	46
TCGA-12-3652	-0.42003	0.26493	-0.68496	-0.6816	1.00E-04	High	34.8919	1	IDHwt	60
TCGA-12-3653	-0.38771	0.309993	-0.69771	-0.69429	0	High	14.52187	1	IDHwt	34
TCGA-12-5295	-0.38931	0.333327	-0.72264	-0.7191	0	High	14.91612	1	IDHwt	60
TCGA-14-0736	-0.27095	0.160347	-0.4313	-0.42918	0.0047	High	15.11325	1	NA	49
TCGA-14-0781	-0.3456	0.393637	-0.73923	-0.73561	0	High	0.952792	1	IDHwt	49
TCGA-14-0787	-0.41921	0.406671	-0.82588	-0.82183	0	High	2.234133	1	IDHwt	69
TCGA-14-0789	-0.32783	0.358943	-0.68677	-0.6834	0	High	11.23638	1	IDHwt	54
TCGA-14-0790	-0.46236	0.326141	-0.7885	-0.78463	1.00E-04	High	13.7662	1	IDHwt	64
TCGA-14-0817	-0.40762	0.249318	-0.65693	-0.65371	0.0091	High	5.388204	1	IDHwt	69
TCGA-14-0871	-0.31535	0.205888	-0.52124	-0.51869	0.0038	High	28.91231	1	IDHwt	74
TCGA-14-1034	-0.23213	0.460409	-0.69254	-0.68914	0	High	15.93463	1	IDHwt	60
TCGA-14-1402	-0.43311	0.249339	-0.68245	-0.67911	0.0072	High	32.03353	1	NA	58

TCGA-14-1823	-0.34628	0.307126	-0.6534	-0.6502	0	High	17.84021	1	IDHwt	58
TCGA-14-1825	-0.32274	0.241394	-0.56414	-0.56137	5.00E-04	High	7.622337	1	IDHwt	70
TCGA-14-1829	-0.30561	0.252484	-0.55809	-0.55536	6.00E-04	High	7.162368	0	IDHwt	57
TCGA-14-2554	-0.29904	0.37324	-0.67228	-0.66898	0	High	17.47881	1	IDHwt	52
TCGA-15-0742	-0.40118	0.324272	-0.72545	-0.72189	0	High	13.7662	1	IDHwt	65
TCGA-16-0846	-0.22181	0.159878	-0.38169	-0.37981	0.0097	High	3.909733	1	IDHwt	85
TCGA-16-1045	-0.3655	0.383029	-0.74853	-0.74486	0	High	29.01088	1	IDHwt	49
TCGA-19-1389	-0.3568	0.459439	-0.81624	-0.81224	0	High	4.632541	1	IDHwt	51
TCGA-19-1390	-0.21459	0.158979	-0.37357	-0.37174	0	High	25.36398	1	IDHwt	63
TCGA-19-1787	-0.30381	0.30976	-0.61357	-0.61056	0	High	12.64914	1	NA	48
TCGA-19-2619	-0.25899	0.329101	-0.58809	-0.58521	0	High	9.659341	0	IDHwt	55
TCGA-19-2620	-0.25043	0.239306	-0.48974	-0.48734	0	High	4.862525	1	IDHwt	70
TCGA-19-2624	-0.30479	0.154827	-0.45962	-0.45737	0.0249	High	0.164275	1	IDHwt	51
TCGA-19-2625	-0.41379	0.295745	-0.70953	-0.70605	0	High	4.074008	1	IDHwt	76
TCGA-19-2629	-0.34373	0.310045	-0.65377	-0.65057	0	High	24.21406	1	IDHmut- non- codel	60
TCGA-19-5960	-0.17387	0.212127	-0.386	-0.3841	1.00E-04	High	5.421059	0	IDHwt	56
TCGA-26-5132	-0.52771	0.307226	-0.83493	-0.83084	6.00E-04	High	9.396501	0	IDHwt	74
TCGA-26-5133	-0.23359	0.2862	-0.51979	-0.51724	0	High	14.85041	0	IDHwt	59
TCGA-26-5134	-0.29635	0.198949	-0.4953	-0.49287	0.0168	High	5.486768	0	IDHwt	74
TCGA-26-5135	-0.40537	0.242248	-0.64762	-0.64445	0.0047	High	8.870823	1	IDHwt	72
TCGA-26-5136	-0.32391	0.248882	-0.57279	-0.56999	0.0022	High	1.215631	0	IDHwt	78
TCGA-26-5139	-0.45994	0.448736	-0.90868	-0.90422	0	High	1.577035	0	IDHwt	65
TCGA-27-1830	-0.33162	0.402843	-0.73446	-0.73086	0	High	5.059655	1	IDHwt	57
TCGA-27-1831	-0.37308	0.317931	-0.69101	-0.68762	0	High	16.59172	1	IDHwt	66
TCGA-27-1832	-0.38719	0.337096	-0.72429	-0.72073	0	High	9.85647	1	IDHwt	59

TCGA-27-1834	-0.29676	0.263372	-0.56013	-0.55738	0	High	40.51009	1	IDHwt	56
TCGA-27-1835	-0.2331	0.325882	-0.55899	-0.55624	0	High	21.28998	1	IDHwt	53
TCGA-27-1837	-0.39718	0.336426	-0.7336	-0.73001	0	High	14.02904	1	IDHwt	36
TCGA-27-2519	-0.37061	0.353257	-0.72386	-0.72031	0	High	10.38215	0	IDHwt	48
TCGA-27-2521	-0.41898	0.183533	-0.60251	-0.59956	0.0013	High	10.38215	0	IDHmut-non-codel	34
TCGA-27-2523	-0.47826	0.372183	-0.85044	-0.84627	0	High	16.06605	1	IDHwt	63
TCGA-27-2524	-0.44033	0.322565	-0.7629	-0.75916	1.00E-04	High	7.589482	1	IDHwt	56
TCGA-27-2526	-0.26745	0.257146	-0.5246	-0.52203	0	High	2.858376	1	IDHwt	79
TCGA-27-2528	-0.34306	0.415898	-0.75896	-0.75524	0	High	15.77035	1	IDHwt	62
TCGA-28-1747	-0.28318	0.304566	-0.58775	-0.58487	0	High	2.529827	1	IDHwt	44
TCGA-28-1753	-0.49024	0.347344	-0.83758	-0.83347	0	High	1.215631	0	IDHwt	53
TCGA-28-2509	-0.42033	0.420637	-0.84097	-0.83685	0	High	4.763961	0	IDHwt	77
TCGA-28-2513	-0.47236	0.413084	-0.88544	-0.8811	0	High	7.293788	0	IDHwt	69
TCGA-28-2514	-0.28614	0.338526	-0.62466	-0.6216	0	High	5.256784	0	IDHwt	45
TCGA-28-5204	-0.36352	0.414961	-0.77848	-0.77466	0	High	14.91612	1	IDHwt	72
TCGA-28-5207	-0.32213	0.243909	-0.56604	-0.56327	0.0012	High	11.26923	1	IDHwt	71
TCGA-28-5208	-0.48953	0.449396	-0.93892	-0.93432	0	High	15.57322	0	IDHwt	52
TCGA-28-5209	-0.44052	0.522165	-0.96268	-0.95796	0	High	2.529827	0	IDHwt	66
TCGA-28-5213	-0.42771	0.28459	-0.7123	-0.70881	1.00E-04	High	9.79076	0	IDHwt	72
TCGA-28-5215	-0.28266	0.359801	-0.64246	-0.63931	0	High	11.00639	1	IDHwt	62
TCGA-28-5216	-0.40285	0.385111	-0.78796	-0.7841	0	High	13.63478	0	IDHwt	52
TCGA-28-5218	-0.36952	0.352158	-0.72168	-0.71814	0	High	5.158219	1	IDHwt	63
TCGA-28-5220	-0.41702	0.263297	-0.68032	-0.67698	0	High	10.48071	0	IDHwt	67
TCGA-32-1970	-0.30487	0.272917	-0.57778	-0.57495	0	High	15.37609	1	IDHwt	59
TCGA-32-1980	-0.1691	0.165266	-0.33437	-0.33273	0.0025	High	1.182776	1	IDHwt	72

TCGA-32-1982	-0.39589	0.287396	-0.68328	-0.67993	1.00E-04	High	4.665396	1	IDHwt	76
TCGA-32-2615	-0.53357	0.367862	-0.90143	-0.89701	1.00E-04	High	15.93463	1	IDHwt	62
TCGA-32-2616	-0.37459	0.325823	-0.70042	-0.69698	0	High	7.359498	1	NA	48
TCGA-32-2632	-0.38179	0.327289	-0.70908	-0.7056	0	High	8.837968	1	IDHwt	80
TCGA-32-2634	-0.3098	0.27544	-0.58524	-0.58237	0	High	8.870823	0	IDHwt	82
TCGA-32-2638	-0.48479	0.464062	-0.94886	-0.9442	0	High	7.359498	0	IDHwt	67
TCGA-32-4213	-0.44602	0.383499	-0.82952	-0.82545	0	High	11.7292	0	IDHwt	47
TCGA-32-5222	-0.41336	0.335971	-0.74933	-0.74566	0	High	5.421059	0	IDHwt	66
TCGA-41-2571	-0.25185	0.123714	-0.37556	-0.37372	0.017	High	0.854227	1	IDHwt	89
TCGA-41-2572	-0.47206	0.301419	-0.77348	-0.76969	1.00E-04	High	13.33909	1	IDHwt	67
TCGA-41-3915	-0.43835	0.414902	-0.85325	-0.84906	0	High	6.045302	0	IDHwt	48
TCGA-41-4097	-0.29656	0.320778	-0.61734	-0.61431	0	High	0.197129	1	IDHwt	63
TCGA-41-5651	-0.16346	0.136756	-0.30021	-0.29874	0.005	High	11.53207	0	IDHwt	59
TCGA-76-4925	-0.37567	0.333969	-0.70964	-0.70616	0	High	4.796815	1	IDHwt	76
TCGA-76-4926	-0.23313	0.294461	-0.52759	-0.525	0	High	4.533976	1	IDHwt	68
TCGA-76-4928	-0.4525	0.466774	-0.91928	-0.91477	0	High	3.088361	1	IDHwt	85
TCGA-76-4929	-0.26286	0.220003	-0.48286	-0.48049	0.0031	High	3.646894	1	IDHwt	76
TCGA-76-4931	-0.41809	0.315169	-0.73326	-0.72967	0	High	9.166517	1	IDHwt	70
TCGA-CS-4941	-0.29155	0.219397	-0.51094	-0.50844	0	High	7.688047	1	IDHwt	67
TCGA-CS-6188	-0.22887	0.413806	-0.64268	-0.63953	0	High	23.8198	0	IDHwt	48
TCGA-DH-5142	-0.24328	0.167061	-0.41034	-0.40833	3.00E-04	High	30.29222	0	IDHmut- non- codel	29
TCGA-DU-5847	-0.37701	0.323723	-0.70073	-0.6973	0	High	18.00449	0	IDHwt	34
TCGA-DU-5852	-0.20587	0.249935	-0.4558	-0.45357	0	High	6.735255	1	IDHwt	61
TCGA-DU-5854	-0.17958	0.234598	-0.41418	-0.41215	0	High	8.443709	0	IDHwt	57
TCGA-DU-6402	-0.32662	0.455416	-0.78204	-0.7782	0	High	7.030949	1	IDHwt	52

TCGA-DU-6405	-0.18056	0.258371	-0.43893	-0.43678	0	High	19.87721	1	IDHwt	51
TCGA-DU-6406	-0.26447	0.450078	-0.71455	-0.71105	0	High	16.82171	1	IDHwt	59
TCGA-DU-7006	-0.2402	0.444241	-0.68444	-0.68108	0	High	11.46636	1	IDHwt	60
TCGA-DU-7012	-0.2457	0.263508	-0.5092	-0.50671	0	High	6.538125	1	IDHwt	74
TCGA-DU-7290	-0.23498	0.26672	-0.5017	-0.49924	0	High	10.34929	1	IDHwt	45
TCGA-DU-8158	-0.16974	0.183381	-0.35312	-0.35139	0.0036	High	5.09251	1	IDHwt	57
TCGA-DU-8165	-0.24248	0.27868	-0.52116	-0.5186	0	High	19.12155	0	IDHwt	60
TCGA-DU-A5TY	-0.32273	0.280413	-0.60315	-0.60019	0	High	12.18917	0	IDHwt	46
TCGA-FG-6688	-0.16911	0.195468	-0.36458	-0.36279	8.00E-04	High	18.76015	0	IDHwt	59
TCGA-FG-6692	-0.47525	0.209582	-0.68483	-0.68147	0.0027	High	18.4316	1	IDHwt	63
TCGA-FG-A4MU	-0.46754	0.500867	-0.96841	-0.96366	0	High	0.591388	0	IDHwt	58
TCGA-FG-A4MW	-0.25592	0.166844	-0.42276	-0.42069	0.0043	High	5.256784	0	IDHwt	63
TCGA-HT-7691	-0.13594	0.157067	-0.29301	-0.29157	0.0028	High	0.098565	0	IDHwt	31
TCGA-HT-7857	-0.23691	0.331019	-0.56793	-0.56514	0	High	0.229984	0	IDHwt	24
TCGA-HT-7860	-0.16398	0.183738	-0.34772	-0.34601	2.00E-04	High	0.492824	0	IDHwt	60
TCGA-HT-7882	-0.3622	0.376003	-0.73821	-0.73459	0	High	3.712604	1	IDHwt	66
TCGA-HT-8011	-0.20236	0.244949	-0.44731	-0.44512	0	High	16.23032	0	IDHwt	55
TCGA-HT-8104	-0.18732	0.310478	-0.4978	-0.49536	0	High	12.22202	0	IDHwt	51
TCGA-HT-8106	-0.18616	0.243724	-0.42989	-0.42778	0	High	0.098565	0	IDHmut-non-codel	53
TCGA-HT-8110	-0.21816	0.282844	-0.50101	-0.49855	0	High	4.698251	0	IDHwt	57
TCGA-HT-8563	-0.17189	0.341645	-0.51354	-0.51102	0	High	16.03319	0	IDHmut-non-codel	30
TCGA-HT-A4DS	-0.26275	0.123689	-0.38644	-0.38454	0.0021	High	0.229984	0	IDHwt	55
TCGA-HT-A5RA	-0.33779	0.410183	-0.74797	-0.74431	0	High	14.68614	0	IDHwt	65

Wt, wild-type; mut, Mutant; codel, Chromosomal *1p/19q* co-deletions.