

UNIVERSITA' VITA-SALUTE SAN RAFFAELE

**CORSO DI DOTTORATO DI RICERCA
INTERNAZIONALE IN MEDICINA MOLECOLARE**

CURRICULUM IN MEDICINA CLINICA E SPERIMENTALE

Validating advanced MRI features as surrogate biomarkers of the molecular subgroups of glioblastoma by exploiting patient-specific cancer stem cell (CSC)-based animal models.

DoS: Dr. Rossella Galli 

Second Supervisor: Dr. Sevin Turcan

Tesi di DOTTORATO di RICERCA di Alberto Luigi Gallotti
Matr. 015647
Ciclo di dottorato XXXV
SSD BIO/13 MED/04 MED/37

Anno Accademico 2021/2022

CONSULTAZIONE TESI DI DOTTORATO DI RICERCA

Il sottoscritto/I Alberto Luigi Gallotti
Matricola / registration number 015647

Nato a/ born at Pavia
il/on 17/12/1989

autore della tesi di Dottorato di ricerca dal titolo / author of the PhD Thesis titled
Validating advanced MRI features as surrogate biomarkers of the molecular
subgroups of glioblastoma by exploiting patient-specific cancer stem cell (CSC)-
based animal models.

AUTORIZZA la Consultazione della tesi / AUTHORIZES the public release of the
thesis
 NON AUTORIZZA la Consultazione della tesi per .48. mesi /DOES NOT AUTHORIZE
the public release of the thesis for .48. months

a partire dalla data di conseguimento del titolo e precisamente / from the PhD thesis
date,

specifically

Dal / from/...../..... Al / to/...../.....

Poiché /because:

l'intera ricerca o parti di essa sono potenzialmente soggette a brevettabilità/ The
whole project or part of it might be subject to patentability;

ci sono parti di tesi che sono già state sottoposte a un editore o sono in attesa di
pubblicazione/ Parts of the thesis have been or are being submitted to a publisher or
are in press;

la tesi è finanziata da enti esterni che vantano dei diritti su di esse e sulla loro
pubblicazione/ the thesis project is financed by external bodies that have rights over
it and on its publication.

E' fatto divieto di riprodurre, in tutto o in parte, quanto in essa contenuto / Copyright
the contents of the thesis in whole or in part is forbidden

Data /Date 24/10/2022.

Firma /Signature

DECLARATION

This thesis has been:

- composed by myself and has not been used in any previous application for a degree. Throughout the text I use both 'I' and 'We' interchangeably.
- has been written according to the editing guidelines approved by the University.

Permission to use images and other material covered by copyright has been sought. For the following images (*Figures 3.1, 3.2, 3.3, 3.4, 3.5, 3.6*), it was not possible to obtain permission and are therefore included in thesis under the "fair use" exception (Italian legislative Decree no. 68/2003).

All the results presented here were obtained by myself, except for:

1) **MRI radiomic analyses on patients and xenografts** (Methods, Chapters 7.8-7.11 and Results, paragraphs 5.1.4 and 5.1.5, tables 5.3-5.10, figures 5.3-5.6) were performed in collaboration with Dr. Antonella Castellano and Ing. Nicolo Pecco from the Department of Neuroradiology, San Raffaele Scientific Institute, Milan, Italy.

2) **MRI acquisition of xenografts** (Methods, Chapter 7.7 and Results, figure 5.25) was performed by Tamara Canu from Experimental Imaging Center, San Raffaele Scientific Institute, Milan, Italy.

3) **RNAseq of GBM tissues and GSCs** (Methods, Chapter 7.12 and Results, paragraphs 5.2.1, 5.3.1 and 5.12.2, figures 5.13 and 5.34), was processed by Dr. Giulia Maria Scotti, Center for Omics Sciences, San Raffaele Scientific Institute, Milan, Italy (but not subsequent GSEA, EnrichR, signatures analyses, and DGE and subsequent analyses on TCGA human samples, which were performed by me).

4) **IHC analysis on human and xenografts and IHC based transcriptional subgroup affiliation** (Methods, Chapter 7.17 and paragraphs 5.1.1 and 5.12.4, tables 5.2, 5.12, 5.26 and 5.27, figures 5.1, 5.24, 5.26 and 5.44-47) were performed by Prof. Pietro Luigi Poliani and Dr. Manuela Cominelli from the Pathology Unit, Department of Molecular and Translational Medicine, University of Brescia, Italy, and by Ilaria Pagano, Neural Stem Cell Biology Unit (NSCBU), San Raffaele Scientific Institute, Milan, Italy.

5) **WES of GSC lines** (Methods, Chapter 7.22 and Results, Chapters 5.11 and 5.12, table 5.39, figure 5.34) was performed and analyzed by Dr. Alessandra Gasperini, Basic and Translational Oncology Unit, Veneto Institute of Oncology IOV, Padova, Italy, in the context of GBM taskforce – Alliance Against Cancer (Alleanza contro il Cancro, ACC).

All sources of information are acknowledged by means of reference.

Ad Ale, Claudia, Gaia, Giulia, Greta, Marco, Omar

Difficile fare dei ringraziamenti dopo 14 anni insieme, al termine di un percorso così importante. Ci sono già la laurea e la specializzazione per i discorsi più lunghi e i ringraziamenti strappalacrime. Ora è diverso. Ora è veramente finita un'epoca. Ti giri, ti guardi indietro e guardi te stesso e tutto quello che hai fatto ed è successo per quasi la metà della tua vita. E rivedi tutto, ma soprattutto tutti quelli che ti hanno accompagnato, colpito, segnato, deviato, rimesso in piedi.

Vorrei ringraziare Rossella per tutto il supporto e gli insegnamenti di questi anni, sia in ambito di ricerca che da un punto di vista umano. Sicuramente un mentore ed un'amica che di sicuro non perderò nella vita futura.

Ringrazio le compagne del laboratorio NSCBU che si sono susseguite in questi anni, affiancandomi prima da specializzando e poi da dottorando, Valentina (Vevs), Valentina (Conti), Ilaria e Paola, con le quali ho condiviso gioie e dolori, dubbi e traguardi e dalle quali ho avuto modo di imparare come essere un buon ricercatore.

Ringrazio Antonella per la collaborazione e il supporto fin dal 2016, fin dalla maglietta di Zerocalcare "Solo tute acetate!".

Ringrazio tutte quelle persone che, durante questi 14 anni di San Raffaele, fin da quel lontano Ottobre 2008, sono entrate a far parte della mia vita. Amici, colleghi medici e ricercatori, infermieri, fisioterapisti, amministrativi... chi stabilmente e chi solo per un periodo, chi è ancora al San Raffaele e chi se n'è andato, chi mi ha insegnato qualcosa, chi mi ha supportato e sopportato e anche chi non lo ha fatto. Sarebbero troppe persone da nominare e, in fondo, non ce n'è nemmeno bisogno, perché tutte le parole che potrei scrivere sarebbero insieme superflue e insufficienti. Bastano gli sguardi, i sorrisi e gli abbracci che ci scambiamo, anche dopo tanto tempo, a dire tutto.

Insomma, grazie San Raffaele, per tutto quello che mi hai dato e tutto quello che mi hai preso, perché se non l'avessi fatto, ora non sarei la persona che sono.

Ed infine grazie a voi, Ale, Claudia, Gaia, Giulia, Greta, Marco, Omar, i primi compagni di questo viaggio, fin da quel 6 Ottobre 2008, quando ci trovammo in aula per la prima volta (se non addirittura prima, alla stazione di Lambrate, cercando di arrivare in università durante uno sciopero dei mezzi). Abbiamo tutti preso strade differenti, non è più semplice trovarsi come una volta, com'è normale che accada, ma quando ci vediamo sembra non sia passato neanche un momento dall'ultima volta. Volevo dirvi che sono fiero ed orgoglioso di voi, delle vostre strade, delle persone che siete. Siete sempre stati e sempre sarete fonte di ispirazione e motivazione per me, vi voglio bene.

Ora è veramente finita un'epoca. E rivedi tutto, ma soprattutto tutti. E rivedi te stesso, cerchi le tue radici, le hai cercate per tanto tempo, ma non serve più perché non ti servono radici, ma un terreno solido e quello ce l'hai. È sempre stato lì. Le radici ti ancorano, il terreno solido ti permette di andare avanti. Ed è arrivata l'ora di andare.

PS. Grazie a Federica, perché nonostante il disagio a gogo, sei sempre fantastica ed instancabile nello stare vicina alle sventure ed i dolori della gente. Grazie ad Ambra che fin da piccolo mi hai insegnato che se prometti, poi mantieni.

ABSTRACT

Introduction: Glioblastoma (GBM) is the most malignant brain tumor in adults. Transcriptional subgroups (proneural, PN, classical, CL, and mesenchymal, MES) have been identified, with MES being the most aggressive. Radiomics has been recently applied to neurooncology, but generally not to transcriptional affiliation. GBM derived glioma stem cells (GSCs) are used to model GBM, but culturing conditions may affect effective modeling.

Materials and Methods: 36 IDHwt GBMs have been studied with advanced MRI protocols including diffusion sequences. 14 GSC lines were established from 48 GBMs and were transplanted to generate xenografts. Radiomic features were extracted from humans and xenografts to train a model predicting MES affiliation. Human GBMs and their GSC lines were profiled at the transcriptional and protein level.

Results: In the first track of our study, we exploited radiomic features extracted from DTI and NODDI that indicate that MES tumors are more locally infiltrative and have more heterogeneous signal than non-MES tumors, probably due to more proliferative, less migrating cells and deposition of extracellular matrix. Models based on such features can predict MES affiliation. In the second study track, we demonstrated a progressive *in vitro* drift in transcriptional affiliation of GBM-derived GSCs, with some diverging to a PN profile, while other to a MES. CL component was generally downregulated *in vitro*. Still, PN lines efficiently model PN GBMs, as do MES GSCs for MES GBMs. We also demonstrated that protein-based categorizations effectively approximate transcriptional classification. In the third track, we demonstrated an increasing transcriptional distance of PN, CL and MES GBMs from healthy brain tissue, suggesting a likely progression from PN to MES. MES GBMs are more hypoxic and angiogenic and more dependent on extracellular matrix. On the contrary, PN tumors exploit neuronal ontologies, likely to establish synapses with neurons to guide infiltration along white matter tracts. In the last track, we identified IL7R as a candidate MES-specific mediator. Of note, tumor expression of IL7R in immunocompetent but not in immunocompromised murine recipients suggests a crosstalk between the immune microenvironment and tumor cells.

Conclusions: We identified novel diffusion MRI radiomic features that correlate and predict MES affiliation of GBMs. We confirmed GSCs as powerful tools to model GBM heterogeneity, especially at early and intermediate passages, notwithstanding the progressive *in vitro* drift in transcriptional affiliation. We postulated a transcriptional-based evolution of GBMs, suggesting also different mechanisms of infiltration. We also propose a role of IL7R in MES GBM as potential biomarker.

TABLE OF CONTENTS

TABLE OF CONTENTS.....	1
1 ACRONYMS AND ABBREVIATIONS	5
2 LIST OF FIGURES AND TABLES	10
2.1 Figures.....	10
2.2 Tables.....	13
3 INTRODUCTION.....	16
3.1 Glioblastoma	16
3.1.1 <i>Epidemiology</i>	16
3.1.2 <i>Symptoms and diagnosis</i>	17
3.1.3 <i>Biology and molecular pathogenesis</i>	18
3.1.3.1 <i>Proneural (PN) subtype</i>	19
3.1.3.2 <i>Classical (CL) subtype</i>	20
3.1.3.3 <i>Mesenchymal (MES) subtype</i>	22
3.1.3.4 <i>Neural (NEU) subtype</i>	23
3.1.3.5 <i>Glioblastoma heterogeneity</i>	24
3.1.3.6 <i>Subtype and clinical correlation</i>	28
3.1.3.7 <i>Subtype and treatment</i>	28
3.1.3.8 <i>Proneural to mesenchymal transition</i>	29
3.1.4 <i>Pathology and classification</i>	30
3.1.4.1 <i>Histological variants of GBMs</i>	32
3.1.4.2 <i>Transcriptional variants of GBMs</i>	33
3.1.4.3 <i>Immunohistochemical panel for GBM transcriptional affiliation</i> . 33	
3.1.5 <i>Management</i>	35
3.1.5.1 <i>Surgery</i>	35
3.1.5.2 <i>Adjuvant treatment</i>	38
3.1.5.3 <i>Elderly patients</i>	39
3.1.5.4 <i>Recurrent Glioblastoma</i>	40
3.1.5.5 <i>Supportive Care</i>	41
3.1.5.6 <i>Targeted therapies</i>	41

3.1.5.7	<i>Immunotherapies</i>	42
3.1.6	<i>Challenges in glioblastoma management</i>	44
3.2	Magnetic Resonance Imaging (MRI)	46
3.2.1	<i>Conventional MRI</i>	46
3.2.2	<i>Advanced MRI</i>	46
3.2.2.1	<i>Diffusion MRI (DWI)</i>	46
3.2.2.2	<i>Diffusion Tensor Imaging (DTI)</i>	47
3.2.2.3	<i>Neurite Orientation Dispersion and Density Imaging (NODDI)</i> ..	48
3.2.2.4	<i>Perfusion MRI (PWI)</i>	50
3.2.2.5	<i>MR Spectroscopy (MRS)</i>	50
3.2.3	<i>Glioblastoma Imaging</i>	50
3.2.4	<i>Radiomics</i>	52
3.3	Glioma Stem Cells (GSCs)	57
3.3.1	<i>Neurosphere Assay</i>	60
3.3.2	<i>Culture media and supplements</i>	63
3.3.3	<i>Other strategies to isolate glioma stem cells</i>	65
3.3.4	<i>Glioma stem cells as a tool to investigate and model GBM</i>	66
3.3.5	<i>Controversies on glioma stem cells</i>	68
4	AIM OF THE WORK	70
5	RESULTS	73
5.1	Identification of radiogenomic features discriminating GBM subtypes. ...	73
5.1.1	<i>GBM patient population enrollment and subgroup affiliation</i>	73
5.1.2	<i>Characteristics of GBM patients utilized for radiogenomic analyses</i> ..	79
5.1.3	<i>Generation of xenografts for radiogenomic feature analysis</i>	80
5.1.4	<i>Radiomic feature selection and validation</i>	81
5.1.5	<i>Refinement of radiomic feature selection and prediction model</i>	86
5.2	Dynamics of transcriptional subgroup affiliation plasticity of GSCs and their derived xenografts.	93
5.2.1	<i>Investigation of transcriptional subtype drift from the original tissue to the in vitro stem cell line by GSEA</i>	93
5.2.2	<i>GBM tissues and GSCs subgroup affiliation can be assessed by specific gene classifiers at the protein level</i>	94
5.2.3	<i>Comparison between IHC-based GBM subgroup definition and WB- based minimal signature affiliation</i>	99
5.3	Identification and testing of GSC-derived gene signatures for GBM transcriptional subgroup determination.	101

5.3.1	<i>RNA Sequencing, unsupervised clustering of GSCs and differential gene expression.</i>	101
5.3.2	<i>Identification and validation of GSC-derived transcriptional signatures.</i>	105
5.4	Identification of subgroup-restricted altered pathways and phenotypes in MES and PN GSCs.	118
5.5	Assessment of the in vivo abilities of MES and PN GSCs to model GBM subgroups.	123
5.6	Summary of subgroup evolution from original human GBM to <i>in vitro</i> GSC line, to <i>in vivo</i> xenograft.	129
5.7	Analysis of metabolic pathways differentiating MES and PN GSCs.	130
5.8	Dynamics of transcriptional subgroup evolution of GBMs with respect to healthy tissue.	132
5.8.1	<i>Inferring a possible evolutionary model of GBM transcriptional subgroup composition.</i>	132
5.9	Identification of molecular pathways and phenotypes characterizing each GBM subgroup in contrast to healthy brain tissue.	142
5.10	Exploiting transcriptional analyses for pinpointing subgroup-specific drug vulnerability.	149
5.11	Identification of GBM-related mutations in GSCs by whole exome sequencing (WES) analysis.	151
5.12	Identification and characterization of IL7R as a putative marker for GBMs.	153
5.12.1	<i>In silico preliminary analysis of IL7R correlation with GBMs and transcriptional subgroups.</i>	154
5.12.2	<i>Transcriptional analysis of interleukin-receptors and IL7 pathway.</i>	156
5.12.3	<i>Protein expression analysis of IL7R in vitro.</i>	162
5.12.4	<i>Protein expression analysis of IL7R in vivo.</i>	165
6	DISCUSSION	170
6.1	Radiogenomic analysis of GBM transcriptional subtypes.	170
6.1.1	<i>Limitations and conclusions</i>	176
6.2	GSC evolution and model reliability.	177
6.2.1	<i>Limitations and conclusions</i>	184
6.3	GBM transcriptional subtype evolution model.	185
6.3.1	<i>Limitations and conclusions</i>	191

6.4	Identification of novel putative players in determining GBM transcriptional subtypes.	192
6.4.1	<i>Limitations and conclusions</i>	196
7	MATERIALS AND METHODS	197
7.1	Patient enrollment	197
7.2	Patients' MRI acquisition	198
7.3	Surgical sample collection	198
7.4	GBM specimen processing	199
7.5	GSCs culture propagation	199
7.6	Orthotopic implantation of GSCs	200
7.7	Xenograft MRI acquisition	200
7.8	MRI image preprocessing	201
7.9	Radiomic features extraction	202
7.10	Radiomic analysis and feature selection.....	203
7.11	Radiomic analysis and feature selection with data augmentation	204
7.12	RNA Sequencing (RNAseq) analysis of GBM tissues and GSCs.....	205
7.13	Differentially Expressed Genes (DEG) in TCGA dataset	205
7.14	Gene Set Enrichment Analysis (GSEA).....	206
7.15	Western blot analysis (WB)	206
7.16	k-Means Clustering analysis based on transcriptional profile.....	207
7.17	Immunohistochemical (IHC) analysis and subgroup affiliation.....	208
7.18	Determination of GSC-derived subtype signatures.....	209
7.19	Analysis of enriched pathways in GSCs clusters.....	210
7.20	Metabolomic-transcriptomic integrated analyses	210
7.21	Transcriptional identification of drug sensitivity	211
7.22	Whole Exome Sequencing (WES) analysis of GSCs	211
7.23	In silico IL7R analysis.....	212
7.24	Quantitative real time PCR (qRT-PCR)	212
7.25	Flow cytometry	213
7.26	Statistical analyses	213
8	REFERENCES	214
9	APPENDIX: R SCRIPTS	246

1 ACRONYMS AND ABBREVIATIONS

2HG: 2-hydroxyglutarate

5-ALA: 5-aminolevulinic acid

AC: astrocyte

ADC: apparent diffusion coefficient

AEDs: anti-epileptic drugs

AIF1: allograft inflammatory factor 1

aMRI: advanced magnetic resonance imaging

AMPA: alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid

AMPA-R: AMPA receptors

ASCL1: achaete-scute homolog 1

ATRX: alpha thalassemia/mental retardation syndrome X-linked

AUC: area under the curve

bFGF2: basic fibroblast growth factor 2

BSA: bovine serum albumin

CAR: chimeric antigen receptor

CD: cluster of differentiation

CDK4/6: cyclin-dependent kinase 4/6

CDKN1A: cyclin-dependent kinase inhibitor 1A

CDKN2A/B: cyclin-dependent kinase inhibitor 2A/B

CHI3L1: chitinase 3 like 1

CL: classical

cMRI: conventional magnetic resonance imaging

CNS: central nervous system

CSCs: cancer stem cells

CT: computed tomography

DAB: diaminobenzydine

DCE: dynamic contrast-enhancement

DEGs: differentially expressed genes

DGE: differential gene expression

DLL3: delta-like ligand 3

DMEM: dulbecco's modified eagle's medium

DMSO: dimethyl sulfoxide
DSC: dynamic susceptibility contrast
DTI: diffusion tensor imaging
DWI: diffusion-weighted imaging
EBSS: earle's balanced salts
ECM: extracellular matrix
EDTA: ethylenediaminetetraacetic acid
EGF: epidermal growth factor
EGFR: epidermal growth factor receptor
EGFRvI/II/III/IV/V: EGFR variant I/II/III/IV/V
EMT: epithelial to mesenchymal transition
EOR: extent of resection
EORTC: european organization for research and treatment of cancer
F: female
FA: fractional anisotropy
FBS: fetal bovine serum
FGFR3: fibroblast growth factor receptor 3
FDR: false discovery rate
FECV: extra-neurite component of diffusion
FICV: intra-neurite component of diffusion
FISO: isotropic component of diffusion
FLAIR: Fluid-attenuated inversion recovery
FWER: family-wise error rate
GBM: glioblastoma
GCLs: glioma cell lines
Gd: gadolinium
GFAP: glial fibrillary acidic protein
GLCM: grey level co-occurrence matrix
GLDM: grey level dependence matrix
GLRLM: grey level run length matrix
GLSZM: grey level size zone matrix
GO: gene ontology
GPM: glycolytic/plurimetabolic
GSCs: GBM stem cells
GSEA: gene set enrichment analysis
GTR: gross total resection
HARDI: high angular resolution DWI

H&E: hematoxylin and eosin
HEPES: N-2-hydroxyethylpiperazine-N-2 ethane sulfonic acid
HLA: human leukocyte antigens
HPLC: high performance liquid chromatography
IBA1: ionized calcium binding adaptor molecule 1/allograft inflammatory factor 1
IDH1/2: isocitrate dehydrogenase 1/2
IDHwt: IDH wild-type
IFNg: interferon gamma
IHC: immunohistochemistry
IL: interleukin
IL7R: IL7 receptor
IMC1/2: informational measure of correlation 1/2
IRF8: interferon regulatory factor 8
ITGAM: integrin subunit alpha M
KNN: k-nearest neighbor
KPS: karnofsky performance status
LIF: leukemia inhibitory factor
LMWH: low molecular weight heparin
LOH: loss of heterozygosity
LOO: leave-one-out cross validation
LR: linear regression
LSD1: lysine-specific demethylase 1
M: male
MD: mean diffusivity
MES: mesenchymal
MGMT: O6-methylguanine-DNA methyltransferase
MHC: major histocompatibility complex
M-MDSCs: monocytic-myeloid derived suppressor cells
MMR: DNA mismatch repair
MMP2/9: matrix metalloproteinase 2/9
MRI: magnetic resonance imaging
MRS: magnetic resonance spectroscopy
MTC: mitochondrial
mTOR: mammalian target of rapamycin
(p)NDRG1: (phosphorylated) N-Myc downstream-regulated gene 1
NES: Normalized enrichment score
NEU: neural

NF1: neurofibromatosis type 1
NF-kB: nuclear factor kappa B
NGTDM: neighboring grey tone difference matrix
NK: natural killer
NODDI: neurite orientation dispersion and density imaging
NOD-SCID: non-obese diabetic severe combined immunodeficiency
NOG or NSG: NOD-SCID IL2Rgamma-null
NPC: neural-progenitor cell
NSA: neurosphere assay
nu/nu: nude
ODI: orientation dispersion index
OLIG1/2: oligodendrocyte transcription factor 1/2
OPC: oligodendrocyte-progenitor cell
OR: odds ratio
OS: overall survival
PBS: phosphate-buffered saline
PD1: programmed cell death protein 1
PDGFRA: platelet derived growth factor alpha
PDL1: programmed death-ligand 1
PDX: patient-derived xenografts
PFA: paraformaldehyde
PFS: progression-free survival
PN: proneural
PPR: proliferative/progenitor
PPV: positive predictive value
PTEN: phosphatase and tensin homolog
PWI: perfusion magnetic resonance imaging
qRT-PCR: quantitative real time polymerase chain reaction
RB: retinoblastoma gene
rCBV: relative cerebral blood volume
RF: random forest
RIPA: radio-immunoprecipitation assay
RNAseq: RNA sequencing
ROC: receiver operating curve
ROI: region of interest
ROS: reactive oxygen species
RT: radiotherapy

RTK: receptor tyrosine kinase
RTOG: radiation therapy oncology group
SCID: severe combined immunodeficiency
SD: standard deviation
SMOTE: synthetic minority oversampling technique
SNP: single nucleotide polymorphism
SVM: support vector machine
SVZ: subventricular zone
TACC3: transforming acidic coiled-coil-containing protein 3
TAMs: tumor-associated macrophages
TBS(-T): tris buffer saline (-Tween-20)
TCGA: the cancer genome atlas
TE: time to echo
TERT: reverse transcriptase
TGFb: transforming growth factor beta
TGFBI: transforming growth factor-beta-induced
TI: inversion time
TMZ: temozolomide
TP53: tumor protein 53
TR: repetition time
TRADD: tumor necrosis factor receptor type 1-associated death domain
Tuj1: neuron-specific class III beta-tubulin
UC: unsupervised cluster
VAF: variant allele frequency
VASARI: visually accessible rembrandt images
VEGF: vascular endothelial growth factor
VIF: variance inflation factor
WB: western blot
WES: whole exome sequencing
WHO: world health organization
YKL40: *another name for* CHI3L

2 LIST OF FIGURES AND TABLES

2.1 Figures

Figure 3.1. GBM heterogeneity and correspondence between GBM cell state and bulk tumor affiliation. (pag...25)

Figure 3.2. GBM transcriptional affiliation according to IHC panel classification and validated by RNAseq analysis. (pag...33)

Figure 3.3. NODDI algorithm. (pag...48)

Figure 3.4. Radiomic and radiogenomic pipeline. (pag...55)

Figure 3.5. Stochastic and hierarchical model for cancer stem cells. (pag...56)

Figure 3.6. Neurosphere Assay (NSA) to isolate glioma stem cells (GSCs). (pag...60)

Figure 4.1. Schematic representation of study design. (pag...71)

Figure 5.1. Transcriptional subgroup affiliation of GBMs based on immunohistochemical panel tested on a representative MES, CL, and PN tumor. (pag...76)

Figure 5.2. Kaplan Meier curves for overall survival. (pag...78)

Figure 5.3. Significant radiomic features discriminating MES (red) from non-MES (green) tumors, both in patients (blue) and xenografts (orange). (pag...84)

Figure 5.4. K-nearest neighbor prediction model performance in differentiating MES from non-MES GBMs. (pag...85)

Figure 5.5. Significant radiomic features discriminating MES (red) from non-MES (green) tumors, both in patients (blue) and xenografts (orange) as isolated from augmented datasets. (pag...90)

Figure 5.6. ROC curves for the 4 prediction models based on the 6 radiomic features identified on augmented datasets. (pag...91)

Figure 5.7. Gene Set Enrichment Analysis of Verhaak's transcriptional signatures in GBM tissues and their derivative GSCs. (pag...92)

Figure 5.8. K-means clustering of TCGA GBM samples for which transcriptional subgroup was known, based on expression levels of minimal gene signature. (pag...93)

Figure 5.9. Minimal gene signature western blot analysis of GBM tissues and derived GSCs showing upregulation of the PN marker. (pag...94)

Figure 5.10. Minimal gene signature western blot analysis of GBM tissues and derived GSCs showing maintenance or upregulation of the MES marker. (pag...95)

Figure 5.11. Minimal gene signature western blot analysis of GBM tissues and derived GSCs showing regulation of CL marker. (pag...96)

Figure 5.12. Minimal gene signature western blot analysis of GBM tissues and derived GSCs showing regulation of CL marker in EGFRvIII positive samples. (pag...97)

Figure 5.13. Scheme of GSC lines studied with RNA Sequencing. (pag...101)

Figure 5.14. Gene set enrichment analysis on differential gene expression between WB-based MES and PN GSC lines. (pag...102)

Figure 5.15. Gene set enrichment analysis on differential gene expression between RNA Sequencing unsupervised clusters A and B. (pag...103)

Figure 5.16. K-means clustering analysis of unrelated human GBMs, based on transcriptional signatures identified from the WB-MES vs WB-PN comparison. (pag...106)

Figure 5.17. K-means clustering analysis of unrelated human GBMs, based on transcriptional signatures identified from the RNA Sequencing Unsupervised Clusters A and B. (pag...110)

Figure 5.18. K-means clustering analysis of unrelated human GBMs, based on transcriptional signatures identified from the RNA Sequencing Unsupervised Clusters A and B. (pag...111)

Figure 5.19. Enrichment plots and statistics of our transcriptional signatures in the TCGA CL vs PN GBMs. (pag...112)

Figure 5.20. Enrichment plots and statistics of our transcriptional signatures in the TCGA MES vs CL GBMs. (pag...112)

Figure 5.21. Enrichment plots and statistics of our transcriptional signatures in the TCGA MES vs PN GBMs. (pag...113)

Figure 5.22. Venn diagrams showing the intersections between the MES (A) and PN (B) transcriptional signatures published by Verhaak and those identified from the RNA Sequencing of our GSCs. (pag...115)

Figure 5.23. Venn diagrams showing the intersections between the MES (A) and PN (B) transcriptional signatures published by Wang and those identified from the

RNA Sequencing of our GSCs. MES: mesenchymal, PN: proneural, WB: western blot. (pag...116)

Figure 5.24. In vivo growth pattern of established MES and PN GSCs. (pag...122)

Figure 5.25. Radiological analysis of growth pattern of GSCs. (pag...125)

Figure 5.26. IHC panel comparing the expression of subgroup-specific classifiers in the original human GBM samples and their corresponding GSC-derived xenografts. (pag...126)

Figure 5.27. Enrichment plots and statistics of Verhaak's transcriptional signatures in the TCGA PN GBMs vs Healthy controls. (pag...133)

Figure 5.28. Enrichment plots and statistics of Verhaak's transcriptional signatures in the TCGA CL GBMs vs Healthy controls. (pag...134)

Figure 5.29. Enrichment plots and statistics of Verhaak's transcriptional signatures in the TCGA MES GBMs vs Healthy controls. (pag...135)

Figure 5.30. Enrichment plots and statistics of Verhaak's transcriptional signatures in the TCGA CL vs PN GBMs. (pag...136)

Figure 5.31. Enrichment plots and statistics of Verhaak's transcriptional signatures in the TCGA MES vs CL GBMs. (pag...137)

Figure 5.32. Enrichment plots and statistics of Verhaak's transcriptional signatures in the TCGA MES vs PN GBMs. (pag...138)

Figure 5.33. Enrichment plots and statistics of our transcriptional signatures in the TCGA MES GBMs vs Healthy controls. (pag...141)

Figure 5.34. Comparison between unsupervised clustering of GSC lines by means of whole exome sequencing (WES, A) and RNA Sequencing analysis (RNAseq, B). (pag...151)

Figure 5.35. In silico analysis of IL7R expression in TCGA GBMs. (pag...154)

Figure 5.36. In silico analysis of IL7R expression in TCGA cancers and in cancer cell lines. (pag...155)

Figure 5.37. Comprehensive Interleukin receptors gene set enrichment analysis on TCGA GBM samples and healthy tissues. (pag...157)

Figure 5.38. IL7 pathway gene set enrichment analysis on TCGA GBM samples and healthy tissues. (pag...158)

Figure 5.39. Comprehensive Interleukin receptors gene set enrichment analysis on GSCs. (pag...160)

Figure 5.40. RT-qPCR analysis of IL7R in GSCs (A and B) and their parental human tumors (C and D). (pag...161)

Figure 5.41. Western blot quantification of IL7R in MES and PN/non-MES cell lines. (pag...162)

Figure 5.42. IL7R protein expression modulation at progressive culturing passages in each GSC line. (pag...163)

Figure 5.43. IL7R protein expression in tumorigenic and non-tumorigenic MES GSC lines as measured by flow cytometry. (pag...164)

Figure 5.44. IL7R protein expression in GSC and GCL-derived orthotopic xenografts in immunocompromised mice. (pag...165)

Figure 5.45. IL7R IHC on human GBMs of different transcriptional subgroup: PN (top), CL (middle), and MES (bottom). (pag...166)

Figure 5.46. IL7R IHC on a human PN GBM with a MES component. (pag...166)

Figure 5.47. IL7R expression in the GL261 murine glioma cell line and in tumors derived from its implantation. (pag...168)

2.2 Tables

Table 5.1. Location of 56 CNS grade 4 gliomas. (pag...73)

Table 5.2. Transcriptional subgroup affiliation of CNS WHO Grade 4 gliomas. (pag...74-75)

Table 5.3. Significant radiomic features discriminating human MES vs non-MES GBMs (3D-FLAIR mask). (pag...81)

Table 5.4. Significant radiomic features discriminating human MES vs non-MES GBMs (post-Gd T1 mask). (pag...81)

Table 5.5. Significant radiomic features discriminating MES vs non-MES GSC-derived xenografts. (pag...82-83)

Table 5.6. Significant radiomic features discriminating MES vs non-MES GSC-derived xenografts after data augmentation and VIF. (pag...87)

Table 5.7. Significant radiomic features discriminating human MES vs non-MES GBMs after data augmentation and VIF (3D-FLAIR mask). (pag...88)

Table 5.8. Significant radiomic features discriminating human MES vs non-MES GBMs after data augmentation and VIF (post-Gd T1 mask). (pag...89)

Table 5.9. Confusion matrices for the 4 prediction models based on the 6 radiomic features identified. (pag...91)

Table 5.10. Performances of the 4 prediction models based on the 6 radiomic features identified. (pag...91)

Table 5.11. Subgroup affiliation of original GBM tissue and derived GSC lines at early and intermediate/late in vitro passages under NSA conditions, based on minimal gene signature classification. (pag...98)

Table 5.12. Subgroup affiliation of original GBM tissue based on IHC panel and WB minimal gene signature classification. (pag...99)

Table 5.13. 81-gene MES signature derived from WB-based MES vs PN GSC differential gene expression. (pag...104)

Table 5.14. 43-gene PN signature derived from WB-based MES vs PN GSC differential gene expression. (pag...105)

Table 5.15. 49 gene - Cluster A (MES) signature derived from RNA Sequencing unsupervised cluster A vs B GSC differential gene expression. (pag...107)

Table 5.16. 98 gene Cluster B (non-MES) signature derived from RNA Sequencing unsupervised cluster A vs B GSC differential gene expression. (pag...108)

Table 5.17. Genes common to the three MES Signatures. (pag...114)

Table 5.18. Genes common to the three PN Signatures. (pag...114)

Table 5.19. Genes common to the three MES Signatures. (pag...116)

Table 5.20. Genes common to the three PN Signatures. (pag...116)

Table 5.21. Characterizing features of GSC Unsupervised cluster A with respect to B. (pag...117-118)

Table 5.22. Characterizing features of GSC Unsupervised cluster B with respect to A. (pag...118)

Table 5.23. Characterizing features of WB-defined MES GSCs with respect to WB-defined PN GSCs. (pag...119)

Table 5.24. Characterizing features of WB-defined PN GSCs with respect to WB-defined MES GSCs. (pag...120)

Table 5.25. Radiological growth pattern of GSC-derived xenografts compared to WB classification of the same lines. (pag...124)

Table 5.26. Transcriptional subgroup affiliation of parental human tumors and their GSC-derived xenografts as determined by IHC. (pag...127)

Table 5.27. Percentage of transcriptional subgroup-specific components in parental human tumors and their GSC-derived xenografts as assessed by IHC. (pag...127)

Table 5.28. Table summarizing the subgroup affiliation of the original GBM samples, their derived GSC lines at early and intermediate/late in vitro passages under NSA conditions, and of their corresponding xenografts. (pag...129)

Table 5.29. TCGA GBM samples for which both mRNA sequencing data and transcriptional affiliation were retrievable. (pag...132)

Table 5.30. Number of total and significant differentially expressed genes for each comparison between pathologic subgroups and healthy state or within subgroups. (pag...132)

Table 5.31. Enrichment statistics of Verhaak's transcriptional signatures in pairwise contrasts between the three pathological subtypes of TCGA GBMs. (pag...139)

Table 5.32. Enrichment statistics of GSC-derived transcriptional signatures in contrasts between the three pathological subtypes of TCGA GBMs and healthy controls. (pag...140)

Table 5.33. Characterizing features of TCGA PN GBMs with respect to normal brain tissue. (pag...142)

Table 5.34. Characterizing features of TCGA CL GBMs with respect to normal brain tissue. (pag...143)

Table 5.35. Characterizing features of TCGA MES GBMs with respect to normal brain tissue. (pag...144)

Table 5.36. Characterizing pathway enrichment features of TCGA PN vs MES GBMs. (pag...146)

Table 5.37. Characterizing pathway enrichment features of TCGA MES vs PN GBMs. (pag...147)

Table 5.38. Summary of drugs putatively effective against each GBM subgroup based on cMap analyses. (pag...149)

Table 5.39. Known GBM-related mutations for each WES subcluster of GSC lines. (pag...152)

Table 5.40. IL7R ranking and core enrichment status for gene set enrichment analysis on differentially expressed genes of TCGA GBMs and healthy controls with comprehensive interleukin receptor gene set. (pag...156)

Table 5.41. IL7 ranking and core enrichment status for gene set enrichment analysis on differentially expressed genes of TCGA GBMs and healthy controls with IL7 pathway gene set. (pag...158)

3 INTRODUCTION

3.1 Glioblastoma

Glioblastoma (GBM) is the most common malignant tumor of the central nervous system affecting adults, second only to brain metastases and first among primary lesions. It accounts for the majority of deaths in patients with primary CNS neoplasia (Wen *et al*, 2020; Ostrom *et al*, 2019a). It accounts for 48.3% of primary malignant brain lesions, reaching up to 57.3% of all gliomas (Conti Nibali *et al*, 2021).

GBMs cause disproportionate morbidity and mortality (Ostrom *et al*, 2019a). Despite progresses in deciphering their biology and despite surgical and medical advances, a significant amelioration in treatment schemes and outcomes have not yet been attained. With a median overall survival (OS) of less than 18 months (Conti Nibali *et al*, 2021) and 5-year OS of 6.8% (Wen *et al*, 2020) it represents a pathology with very poor prognosis, even if a recent investigation on almost 90,000 patients treated between 2004 and 2013 reports a minor increase in 3 years overall survival after the diagnosis (Conti Nibali *et al*, 2021).

The focus of our work is on the IDH wild-type (IDHwt) GBMs, which represent 90-95% of CNS WHO Grade 4 astrocytomas and have the worst prognosis (Wen *et al*, 2020; Yan *et al*, 2009; Louis *et al*, 2016).

3.1.1 Epidemiology

The overall incidence of GBM in the western world is 3-5/100,000 persons after age-adjustment, being more frequent in males (M/F ratio of 1.6) (McKinnon *et al*, 2021; Brodbelt *et al*, 2015; Ostrom *et al*, 2019a). Incidence rises with age, reaching its highest in the range between 65 and 84 years old, with a 2-fold increase with respect to people aged 55-64 (Conti Nibali *et al*, 2021). A regional variation in incidence has also been reported worldwide (Leece *et al*, 2017).

As reported in the literature, there is no increasing incidence in North America (Davis *et al*, 2020), as opposed to what reported for the UK (Philips *et al*, 2018). Such disproportions might be due to dissimilar surveillance protocols in the various countries besides changes in GBM classifications (Wen *et al*, 2020).

Known risk factors explain GBMs to a very low extent (Ostrom *et al*, 2019b). Ionizing radiation to head and neck has been correlated to increased risk of developing GBM, while atopic diseases may play a protective role (Ostrom *et al*, 2019b). On the contrary, there is no evidence for non-ionizing radiations, such as those associated with phone use, to be dangerous in this sense (Wen *et al*, 2020).

Most GBM patients are sporadic. Still, roughly 5% of all gliomas show a familial predisposition, and few rare Mendelian syndromes have been implicated in GBM development (Ranger *et al*, 2014; Wen *et al*, 2020). Genome-wide association studies confirmed 25 single nucleotide polymorphisms to confer an increased risk to develop glial tumors, 11 of which are specific to GBMs (Melin *et al*, 2017). In most cases, the identified loci contain critical genes for gliomas, such as telomerase reverse transcriptase (TERT), epidermal growth factor receptor (EGFR), and cyclin-dependent kinase inhibitor 2B (CDKN2B) (Melin *et al*, 2017; Labreche *et al*, 2018).

3.1.2 Symptoms and diagnosis

Rapid tumor growth and displacement or infiltration of eloquent structures brings about a variety of symptoms leading GBM patients to seek for medical attention and, eventually, diagnosis. Often, the initial disturbances are non-specific, being similar to those experienced with other primary or secondary brain tumors, or more commonly, with benign neurological conditions. Circa 50% of patients initially seek for emergency hospital care (McKinnon *et al*, 2021).

Among the various symptoms, new onset epilepsy, progressive headaches, focal neurologic deficits, signs of increased intracranial pressure, and mental status alterations are the most indicative (Weller *et al*, 2017). However, the earliest and most common presenting symptom is headache. Still, only 0.1-0.2% of people with unusual, new onset headache bear an underlying brain tumor (McKinnon *et al*, 2021; Ozawa *et al*, 2019). Headache stigmata depend on tumor location, size, and growth rate, being either tension- or migraine-like (Kirby & Purdy, 2014). Exacerbated by supine position, it mostly presents on waking, is accentuated by the Valsalva maneuver, and frequently increases in occurrence and severity. In case it associates

with other neurological focuses, the likelihood of an underlying expanding lesion significantly increases (McKinnon *et al*, 2021; Ozawa *et al*, 2019).

Seizures present at diagnosis in about 20% of GBM patients, even though they are more common as presenting symptoms in lower-grade gliomas. An additional 20% of GBM patients develops epilepsy during the course of the disease (McKinnon *et al*, 2021). Notably, new onset seizures in adults have the greatest positive predictive value (PPV, 1.6%), followed by motor weakness (1.5%) and confusion (1.4%) (Ozawa *et al*, 2019). Thus, a combinations of such complaints strongly suggests an underlying intracranial mass, especially if progressive (Posti *et al*, 2015).

Cognitive deficits, personality changes, and mood disturbances often complicate the clinical presentation (Pace *et al*, 2017). Interestingly, up to 91% of patients present cognitive deficits, though only a minority complains about them (Gehring *et al*, 2015; Wen *et al*, 2020). Cognitive impairment is frequently heightened by fatigue and sleep disturbances (Armstrong *et al*, 2017; Armstrong & Gilbert, 2012).

Contrast-enhanced MRI is the gold standard diagnostic tool for GBM detection.

3.1.3 Biology and molecular pathogenesis

GBMs have been suggested to originate from neuroglial stem or progenitor cells and are molecularly heterogeneous. Genes and pathways typically mutated in sporadic GBMs have been identified (McLendon *et al*, 2008; Parsons *et al*, 2008; Barthel *et al*, 2018) and integrated with gene expression profiling and methylomic studies, allowing to identify molecular GBM subgroups enriched for peculiar alterations (Wen *et al*, 2020).

As a matter of fact, The Cancer Genome Atlas (TCGA) Research Network provided a list of the genomic changes in 206 GBM patient samples providing a catalog of tumorigenic genomic abnormalities. For instance, previously reported TP53 and RB mutations were confirmed and new GBM-associated mutations such as in PIK3R1, neurofibromatosis type 1 (NF1), and ERBB2 were described. Further copy number and mutational analyses showed that most GBMs harbor abnormalities in TP53, RB, and receptor tyrosine kinase (RTK) pathways, suggesting their importance in glioblastomagenesis (Verhaak *et al*, 2010). Another typical feature, common in sporadic adult GBMs, is oncogene amplification on extrachromosomal DNA, which allows tumor cells to thrive even in case of scarce microenvironmental resources (Turner *et al*, 2017; Decarvalho *et al*, 2018).

Studies on 200 GBMs and 2 normal brain samples allowed identification of 1740 genes with consistent but highly variable expression across the samples. Based on these, average linkage hierarchical clustering identified 4 groups which were named as Proneural, Neural, Classical, and Mesenchymal. Characterizing genes for each group were then deposited as signatures (Verhaak *et al*, 2010).

The Proneural/RTK I group, more frequent in younger adults, was marked by cyclin-dependent kinase 4 (CDK4) and platelet derived growth factor alpha (PDGFRA) amplifications. Hallmark of the Classical/RTK II group are aberrations of EGFR and homozygous loss of CDKN2A/B, whereas the Mesenchymal group was characterized by NF1 loss and increased immune infiltrate. Transversal to these groups are the TERT promoter mutations (Wen *et al*, 2020; Brennan *et al*, 2013; Ceccarelli *et al*, 2016; Wang *et al*, 2017).

Mutually exclusive with TERT mutations, aberration of alpha thalassemia/mental retardation syndrome X-linked (ATRX) is associated with TP53 mutation and usually affects rarer GBMs harboring alterations in H3K27M or H3G34R (Wen *et al*, 2020). Another rare subset of GBMs is hallmarked by FGFR3-TACC3 fusion protein and shows stigmata of oxidative phosphorylation, rendering it metabolically divergent from the more common glycolytic GBMs (Frattini *et al*, 2018).

Notably, treatment like chemo-radiation, may induce or select cell subclones with alterations in the DNA mismatch repair (MMR) machinery (Kim *et al*, 2015; Körber *et al*, 2019). As a matter of fact, up to 10% of recurrent, post-temozolomide (TMZ) GBMs show a markedly high mutational load (Draaisma *et al*, 2020). The hypermutational phenotype may be coupled with germline MMR defects, but most of the times it is acquired following alkylating agents administration (Touat *et al*, 2020; Hunter *et al*, 2006; Johnson *et al*, 2014) and usually occurs in O6-methylguanine-DNA methyltransferase (MGMT) methylated gliomas (Wen *et al*, 2020). Other frequent mutations developed in recurrent tumors affect TP53, EGFR, and phosphatase and tensin homolog (PTEN) (Wen *et al*, 2020).

3.1.3.1 Proneural (PN) subtype

The Proneural class is characterized by alterations in PDGFRA, such as focal amplifications at 4q12, which are detected in all subtypes, yet at a much higher rate in PN samples. The PDGFRA signature leads to accumulation of many copies of PDGFRA gene product, which is almost exclusive of this subtype. Occasionally, besides amplification, PN samples also harbor point mutations in this gene, generally

observed in the immunoglobulin-domain, that might lead to disruption of ligand interaction (Verhaak *et al*, 2010).

Mutations and loss of heterozygosity (LOH) of TP53 are frequent events in PN subtype. High expression of genes related to oligodendrocyte development, such as NKX2-2 and OLIG2 (Noble *et al*, 2004), are quite common. Increased OLIG2 expression proved to inhibit the tumor suppressor function of CDKN1A, thus promoting proliferation (Ligon *et al*, 2007). Notably, PN tumors with no PDGFRA abnormalities harbor most of the PIK3CA/PIK3R1 mutations detected in GBMs (Verhaak *et al*, 2010). On the contrary, the typical chromosome 7 amplification and 10 loss is strikingly less common as opposed to other subtypes (only 54% of PN samples) (Verhaak *et al*, 2010).

Remarkably, the PN signature consists of many crucial genes for neural differentiation, such as SOX genes, DCX, DLL3, ASCL1, and TCF4 (Phillips *et al*, 2006). PN tumors are characteristically enriched for gene ontologies (GO) involving neurodevelopmental processes and cell cycle/proliferation signatures (Whitfield *et al*, 2002).

As far as the tumor microenvironment is concerned, PN GBMs present a reduced immune infiltrate and a less immunosuppressive tumor milieu (Gangoso *et al*, 2021).

Before the new CNS WHO tumor classification (Louis *et al*, 2021), secondary IDH1/2 mutant lesions were mainly included in the PN subclass. As a matter of fact, 11/12 IDH1/2 mutations in Verhaak's study were detected in PN tumors, mostly in those not harboring PDGFRA abnormalities (Verhaak *et al*, 2010).

3.1.3.2 Classical (CL) subtype

The Classical subtype is distinguished by chromosome 7 amplification and chromosome 10 loss, which is present in 100% of CL samples. Although chromosome 7 amplification is detected in lesions of other subtypes, amplification of epidermal growth factor receptor (EGFR) is generally spotted in CL tumors (97%) and very infrequently in other subgroups. The neural precursor and stem cell marker NES is typically expressed (Verhaak *et al*, 2010).

EGFR/ERBB1/HER1 is a transmembrane glycoprotein belonging to tyrosine kinase receptor family, which includes also ERBB2/HER2/Neu, ERBB3/HER3, and ERBB4/HER4. They are activated upon ligand binding and receptor dimerization. Activation of EGFR induce downstream signaling through Ras/Raf/MAPK, PI3K/AKT, JAK/STAT, or PLC/PKC pathways, with an impact on cell proliferation, metabolism, apoptosis, survival, and differentiation (Rutkowska *et al*, 2019).

The gene encoding for EGFR consists of 28 exons. It comprises an extracellular domain for ligand binding and receptor dimerization (exons 1-16), an hydrophobic transmembrane domain (exon 17), and an intracellular domain endowed with tyrosine kinase activity plus a C-terminal linker region (exons 18-28) (Rutkowska *et al*, 2019).

A majority of mutations affect the extracellular and tyrosine kinase domains and induce an increased or prolonged EGFR signaling (Su Huang *et al*, 1997), with resulting enhanced proliferation, apoptosis inhibition, and angiogenesis. (Rutkowska *et al*, 2019). In GBMs, EGFR amplification is mostly associated to rearrangements and deletions involving different exons and are designated as EGFRvI (deletion of N-terminal part), EGFRvII (exons 14 and 15), EGFRvIII (exons 2-7), EGFRvIV (exons 25-27), and EGFRvV (exons 25-28) (Francis *et al*, 2014; Wong *et al*, 1992; Cho *et al*, 2011). Most frequently, EGFRvIII is the pathogenic variant detected in GBMs (Rutkowska *et al*, 2019), with more than 50% of CL GBMs in Verhaak's study displaying a point mutation in EGFR or EGFRvIII (Verhaak *et al*, 2010).

As for the mechanism of action, EGFRvIII exhibits a constitutive activity (Nishikawa *et al*, 1994), since it can dimerize independently of the ligand, in a conformation which might resemble either active or inactive EGFR. Apparently, EGFRvIII constitutive activity is not extremely stark; however its stability might allow longer kinase activity translating into significant biological effects (Rutkowska *et al*, 2019). Another possible mechanism of action could be linked to stable heterodimer formation with triggered wild-type EGFR (Fan *et al*, 2013). Additionally, EGFRvIII is likely to heterodimerize with other inactive receptors, such as MET and PDGFR, affecting their action (Chakravarty *et al*, 2017; Garnett *et al*, 2013). Moreover, EGFRvIII-positive cells show both autocrine and paracrine effects by secreting factors like leukemia inhibitory factor (LIF) and IL6, which in turn activate receptors on wild-type EGFR-positive cells. Through NF- κ B activation and induction of survivin, IL6 may amplify cell resistance to apoptosis (Rutkowska *et al*, 2019).

Additionally, CL GBMs are distinctly lacking TP53 mutations, even if such gene is habitually mutated in GBMs (Verhaak *et al*, 2010). CDKN2A homozygous deletion is frequently associated to CL lesions, appearing together with EGFR amplification in 94% of cases, and is mutually exclusive with other aberrations such those affecting RB1, CDK4, and CCDN2, all impairing RB pathway. Other pathways hyperactivated in CL lesions are those of Notch (NOTCH3, JAG1, and LFNG) and Sonic hedgehog (SMO, GAS1, and GLI2) (Verhaak *et al*, 2010).

3.1.3.3 Mesenchymal (MES) subtype

Mesenchymal tumors express CHI3L1/YKL40, CD44 and MET (Phillips *et al*, 2006). Remarkably, increased expression of both mesenchymal and astrocytic markers resembles the epithelial-to-mesenchymal transition, observed in dedifferentiated or transdifferentiated epithelial tumors (Thiery, 2002). Genes pertaining the tumor necrosis factor superfamily and NF- κ B pathway, such as TRADD, RELB, and TNFRSF1A, are highly expressed, possibly due to inflammatory cells infiltrating the extensive necrosis seen in MES tumors (Verhaak *et al*, 2010). Additionally, genes regulating extracellular matrix (ECM) turnover and wound healing correlate with MES tumors (Richards *et al*, 2021).

MES GBMs frequently deactivate NF1 through genomic copy loss or somatic mutations. As a matter of fact, in the pivotal work by Verhaak *et al.*, focal hemizygous deletions of NF1 locus predominantly occurred in the MES cluster, and 53% of MES samples expressed less NF1. Conversely, NF1 mutations were detected in 20 specimens, 14 of which were MES (Verhaak *et al*, 2010). Moreover, most co-mutations of NF1 and PTEN intersecting with the AKT pathway were observed in MES lesions (Verhaak *et al*, 2010). Notably, in an experimental model optimized by Gangoso *et al.*, NF1 loss was enough to induce enrichment of various gene signatures relevant to glioblastomagenesis, including angiogenesis and cell migration, which suggests that NF1 plays a role in priming cells for malignant transformation. (Gangoso *et al*, 2021)

Even more interestingly, NF1-deleted/mutated GBMs show reduced tumor purity due to a higher stromal and immune infiltration, when compared to NF1 wild-type GBMs. This association between reduced tumor purity and MES affiliation has been commonly identified across different cancers (Wang *et al*, 2017). Indeed, the interaction between tumor cells and microenvironment favors tumor adaptability and progression, particularly effective in MES neoplasia (Olar & Aldape, 2014).

Among GBM-associated stromal cells, tumor-associated macrophages (TAMs) may derive from resident microglia or peripheral blood monocytes and are marked by integrin subunit alpha M (ITGAM or CD11B) and allograft inflammatory factor 1 (AIF1 or IBA1). They account for up to 30-50% of the cells in the GBM microenvironment and through bidirectional communications with tumor cells are associated with tumor aggressiveness and resistance to therapies (Hambardzumyan *et al*, 2015; Lu-Emerson *et al*, 2013). Accordingly, Wang *et al.* proved that tumor-promoting M2 macrophages were associated to a higher extent with the MES subtype (13%) rather than to the PN (5%) and CL (6%) subgroups (Wang *et al*, 2017). Moreover, MES samples were significantly enriched in proinflammatory M1 macrophage and

neutrophil gene signatures. Contrarily, activated natural killer (NK) cells were significantly underrepresented in MES GBMs (Wang *et al*, 2017). Also, evidence suggests how radiotherapy could support TAM recruitment, thus selecting radioresistant GBM cells and inducing their MES differentiation (Bhat *et al*, 2013; Kim *et al*, 2021).

The high burden of infiltrating non-tumoral cells might play a confounding role in defining the stigmata of MES tumors. However, Wang *et al*. demonstrated that MES signatures are actually intrinsic to tumor cells and not simply due to contaminating immune or endothelial cells. Nevertheless, it remains unclear how such molecular program is acquired (Wang *et al*, 2017).

The immune infiltrate was further studied by Gangoso *et al*. in a murine model of MES GBMs. They reported major changes in the immune cell repertoire with an increase in CD8/CD4 T-cells, M1 and M2 macrophages, and a decrease in microglia. They also documented dysfunctional lymphoid populations marked by PD1 and TIM3. Interestingly, they reported an increased immune infiltrate in immune evasive MES tumors relative to non-immune evasive ones, with predominant monocytic-myeloid derived suppressor cells (M-MDSCs) and higher levels of PDL1-positive macrophages (Gangoso *et al*, 2021).

Parallely, the authors described an upregulation of several pro-tumorigenic chemokines, as well as immune-associated genes and gene ontologies (GO) in the immune evasive tumors. Among them, upregulation of interferon regulatory factor 8 (IRF8), which is a myeloid-specific transcription factor expressed in hematopoietic cells (Driggers *et al*, 1990) to induce myeloid specification and macrophage differentiation, was strikingly detected in GBM cells (Gangoso *et al*, 2021). As a matter of fact, IRF8 responds to interferon gamma (IFN γ) and is normally inactive in neural and glial cells. However, chronic immune attack and sustained IFN γ signal from the microenvironment, might lead GBM cells to hijack its expression. Subsequent changes in DNA methylation then would be fundamental to stabilize the transcriptional changes imposed by immune attack (Gangoso *et al*, 2021). Such findings described in murine models would suggest a similar mechanism in human GBMs, undergoing similar transcriptional reconfiguration and epigenetic stabilization (Gangoso *et al*, 2021).

3.1.3.4 Neural (NEU) subtype

Neural tumors were defined based on neuronal markers, such as NEFL, GABRA1, SYT1, and SLC12A5. Related gene ontologies included neuron projection and axonal

transmission through synapses. Notably, the healthy brain samples included in the TCGA were catalogued as NEU (Verhaak *et al*, 2010).

In the following work by Wang *et al.*, the authors adopted a different strategy to investigate the transcriptional heterogeneity of GBMs subgroups, concentrating on tumor-intrinsic and microenvironment-independent factors. Therefore, they identified genes uniquely expressed by glioma cells by filtering the gene list obtained from the comparison between GBMs and their derived cultured neurospheres and by removing genes overexpressed in GBMs but not in their derivative neurospheres, thus eliminating microenvironmental contaminations. A further refinement was carried out eliminating genes pertaining to infiltrative tumoral margins with no more than 10% tumoral cells as opposed to core lesions (100% tumoral cells). The resulting filtered list (Affymetrix U133A microarray) was adopted to re-classify the TCGA GBMs, identifying 3 subtypes out of 369 IDHwt GBMs. Upon comparison of the resulting clusters with the preceding classification (Brennan *et al*, 2013; Verhaak *et al*, 2010), the newly identified clusters fitted with PN, CL, and MES GBMs. None of them was designated as NEU, advocating this last phenotype as non-tumoral. Therefore, the previous NEU class probably identified mostly tumoral margins, where mostly healthy white matter was contaminated by few infiltrating GBM cells (Wang *et al*, 2017). Contamination by normal healthy tissue might provide the underlying reason for the lack of characteristic gene aberrations in the NEU subgroup (Brennan *et al*, 2013).

3.1.3.5 Glioblastoma heterogeneity

Genetic, epigenetic, and microenvironmental signals influence biological programs and induce GBM heterogeneity. First, the above-described transcriptional subtypes can be differentially enriched in different regions of the same lesion, as proven by their multiple sampling. In addition, longitudinal analyses revealed treatment strategies can induce subtype switch over time, with different cells from the very same GBM activating pathways pertaining to different subtypes (Patel *et al*, 2014; Wang *et al*, 2017; Sottoriva *et al*, 2013).

Secondly, GBMs hijack neural developmental pathways and contain subsets of GBM stem cells (GSCs) endowed with tumorigenic ability and resistance to chemo- and radiotherapy (Bao *et al*, 2006a; Chen *et al*, 2012; Parada *et al*, 2017; Neftel *et al*, 2019). Yet, it is still debated whether such cells exist in univocal or diverging cellular states and if distinct GSC subpopulations generate tumors of similar or varied cellular composition (Neftel *et al*, 2019), just as it still remains unknown whether

unidirectional hierarchies or reversible transitions govern GBM and GSC biology (Nefitel *et al*, 2019; Suvà *et al*, 2014).

In a recent work by Nefitel *et al.*, single cell RNA-Sequencing of 20 adult and 8 pediatric GBMs allowed to define how GBM cells exist in a limited set of cellular states. Each state recapitulates one out of six meta-modules and is defined by a set of highly recurring genes, which can be clustered into 4 sets, 2 of which can be further divided into two subgroups (Nefitel *et al*, 2019).

In particular, two meta-modules enriched for mesenchymal genes were identified, the former being strongly associated with response to hypoxia, stress, and glycolysis, as opposed to the latter. As a consequence, the authors defined them as mesenchymal-like (MES-like) states, and subdivided them into hypoxia-independent (MES1) and dependent (MES2) meta-modules (Nefitel *et al*, 2019).

Four additional meta-modules presented genes implicated in neurodevelopment, reminding of neuronal or glial progenitors. Accordingly, they were enriched for astrocytic markers (S100B, GFAP, SLC1A3, GLAST, and MLC1), oligodendroglial markers (OLIG1, OMG, PLP1, PLLP, TNFR, and ALCAM), and stem cell signatures, including neural progenitor markers (SOX4, SOX11, and DCX) (Tirosh *et al*, 2016; Venteicher *et al*, 2017; Darmanis *et al*, 2015, 2017; Nowakowski *et al*, 2017). Therefore, the states mimicking these developmental states were named as astrocyte-like (AC-like), oligodendrocyte-progenitor-like (OPC-like), and neural-progenitor-like (NPC-like). Intuitively, being hijacked by tumor cells, all of these 3 developmental-related modules presented important aberrations that distinguished them from their normal counterparts. Remarkably, NPC-like module was in turn split into NPC1 and NPC2 based on the inclusion of genes related to oligodendrocyte progenitors (NPC1) or to neuronal lineage (NPC2), likely reflecting the NPC ability to differentiate into either OPCs or neurons (Nefitel *et al*, 2019).

The authors reported that each tumor sample contains cells in multiple states. However, their percentual make-up varies from lesion to lesion and depends on aberrations in CDK4, PDGFRA, EGFR, and NF1, each favoring a specific state. As a matter of fact, EGFR is maximally upregulated in AC-high tumors, while high amplifications of PDGFRA and CDK4 induce OPC-like and NPC-like states specification, respectively (Nefitel *et al*, 2019). On the contrary, NF1 alterations are correlated with MES-high tumors. Quite remarkably, deletions in chromosome 5q are negatively associated with MES-like state, suggesting that 5q loss could impair MES affiliation. In fact, this chromosome arm contains controllers of mesenchymal differentiation as well as numerous chemokines and cytokines that could interact with

microglia/macrophages and other immune cells (Neftel *et al*, 2019; Wang *et al*, 2017).

Lastly, even though most GBM cells correspond primarily to one of the four states, up to 15% of cells express two modules, configuring hybrid states, with some combinations being common (AC-like/MES-like, NPC-like/OPC-like, and AC-like/OPC-like), while others almost totally undetected. On the other hand, from a whole-tumor point of view, each GBM contains cells pertaining to at least two of the four cellular states, mostly including all four of them to varying extent. As a matter of fact, most tumors consist either of NPC-like and OPC-like cells, or of AC-like and MES-like cells, though different combinations might be detected (Neftel *et al*, 2019). In turn, the predominance of a specific module or combination consistently concurs in defining the global affiliation of each tumor. In fact, the CL and MES subtypes represent tumors showing mainly the AC-like and MES-like states, respectively. On the contrary, PN GBMs entail the combination of OPC-like and NPC-like, which typically co-occur (Neftel *et al*, 2019).

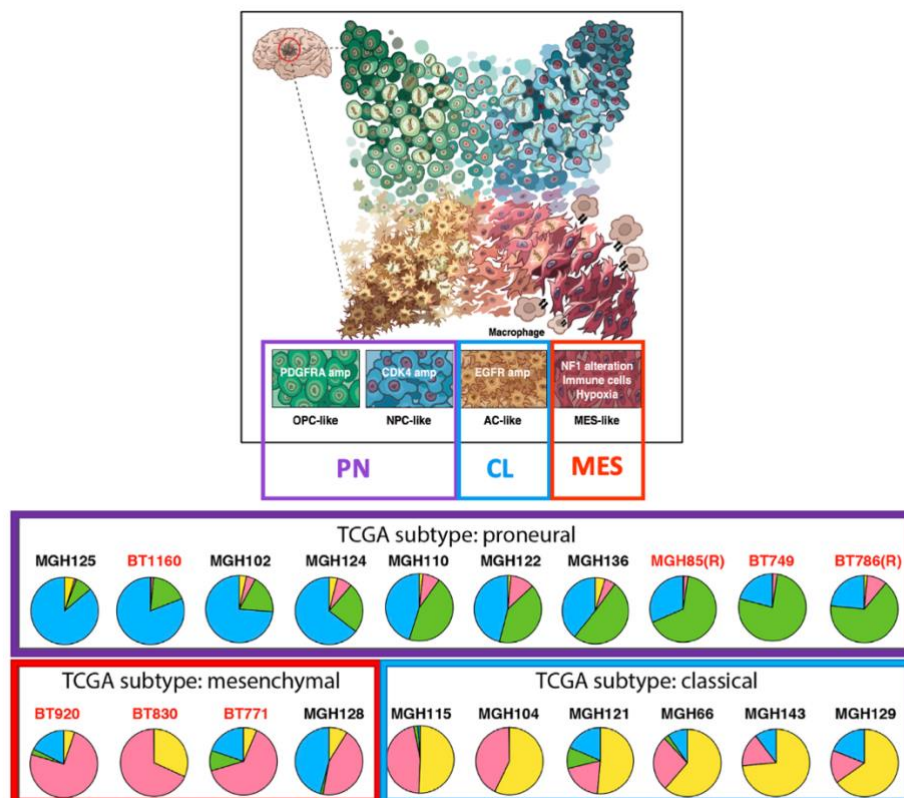


Figure 3.1. GBM heterogeneity and correspondence between GBM cell state and bulk tumor affiliation. Top panel, depiction of segregation of the 4 cell states according to single-cell RNAseq analysis. Bottom panel, composition of single tumors with each cell state adding up to final transcriptional affiliation of the bulk lesion. AC: astrocyte-like, CL: Classical, MES: mesenchymal, NPC: neural-progenitor-like, OPC: oligodendrocyte-progenitor-like, PN: proneural. Adapted from (Neftel *et al*, 2019)

Therefore, intra-lesional GBM heterogeneity globally matches cellular states shifting over a continuum of hybrid states, each with proliferative potential. Still, the highest proliferation potential has been linked to NPC-like and OPC-like states (Neftel *et al*, 2019). The immune microenvironment might then act on such gradient by eroding the epigenetic setting and altering transcription factors, thus conferring a versatile nature to GBMs (Gangoso *et al*, 2021).

Another level of heterogeneity was recently identified in a work by Garofano *et al*., who submitted 36 adult high-grade gliomas to single-cell transcriptional analysis and subsequent pathway deconvolution. The authors proposed a transcriptional classification of IDHwt GBMs based on the core biological pathways exploited by individual cells. This classification identified 4 modules described by traits of either development (neuronal, NEU and proliferative/progenitor, PPR) or metabolism (glycolytic/plurimetabolic, GPM and mitochondrial, MTC) (Garofano *et al*, 2021).

On the metabolic side, GPM cells exploited pathways such as lipidic and aminoacidic metabolism; however, they did not present features of mitochondrial activity and oxidative phosphorylation and were enriched in mesenchymal and immune-related functions. On the contrary, MTC cells depended on mitochondrial metabolism, oxidative phosphorylation, and fatty acid oxidation. Notably, these two states stand in a metabolic spectrum imposed by opposing transcriptomic programs and mutations, inducing definite metabolic needs. In fact, each state is associated with stable inactivation of genetic programs necessary to induce the opposed metabolic phenotype. Therefore, MTC GBMs are susceptible to inhibition of oxidative phosphorylation, and portend increased sensitivity to radiotherapy through ROS produced by mitochondria. Conversely, activation of several pathways in GPM GBMs induces a redundant metabolic profile that might explain overcoming of glycolysis inhibition and resistance to multiple treatment strategies (Garofano *et al*, 2021).

On the development-recapitulating side, the NEU cluster harbored stigmata of specialized neuronal functions like axon genesis and synaptic transmission, whereas the PPR cluster activated pathways associated with cell cycle progression, DNA replication and damage response, mitosis, and expressed neural stem/progenitor cell markers. Confronting this classification with the 6 abovementioned cellular states, PPR and NEU were enriched in NPC1, NPC2 and OPC signatures, while GPM and MTC preferentially correlated with the MES and AC cell states (Garofano *et al*, 2021).

Interestingly, survival analysis exhibited significantly better survival for MTC GBMs independent of age, gender or MGMT methylation. The MTC state was evenly distributed across the CL, PN and MES subgroups, without restrictions of oxidative

phosphorylation to specific transcriptional affiliations, while MES cells were positively associated with the GPM state, suggesting them as codependent features in GBMs. On the other hand, NEU and PPR states were mostly undetected in MES GBMs (Garofano *et al*, 2021).

3.1.3.6 Subtype and clinical correlation

As reported by Verhaak *et al.*, the most consistent clinical association for tumor subtypes is age, with younger patients overrepresented in the PN subtype. Additionally, a longer survival for PN GBM-bearing patients was reported, although it did not reach statistical significance (Verhaak *et al*, 2010). However, these findings could have been confounded by analyzing conjointly IDH-mutant and IDHwt GBMs, now considered two very distinct entities, as only IDHwt tumors are proper GBMs (Louis *et al*, 2021; Verhaak *et al*, 2010; Noushmehr *et al*, 2010).

Notably, in the following study by Wang *et al.*, the authors compared outcomes according to transcriptional affiliation restricted to IDHwt GBMs only, and observed a median overall survival of 11.5, 14.7, and 17.0 months in MES, CL, and PN cases, respectively. Patients with MES tumors, being them primary or recurrent, showed a worse overall survival. Consequently, patients bearing both primary and recurrent MES lesions showed the worst outcome, suggesting a likely additive or synergic effect of transcriptional affiliation at different times in the same pathology (Wang *et al*, 2017). This difference in survival between MES and non-MES tumors (p-value = 0.03) is consistent with what described by Verhaak, in addition to a slightly milder progression for the PN subtype (Wang *et al*, 2017).

3.1.3.7 Subtype and treatment

The PN and MES subtypes are the most constantly addressed subtypes in the literature and have been described to bear different responses to therapeutic schemes (Huse *et al*, 2011; Phillips *et al*, 2006; Wang *et al*, 2017). Analyses of the Murat and TCGA datasets showed that patients treated with concurrent chemo-radiotherapy or > 3 cycles of chemotherapy had reduced mortality if affected by CL (hazard ratio 0.45; p-value = 0.02) or MES tumors (hazard ratio 0.54; p-value = 0.02), while survival was not affected in PN lesions (hazard ratio 0.8; p-value = 0.4) (Murat *et al*, 2008).

Methylation of the DNA repair gene MGMT, a known prognostic factor for response to alkylating agents, was not associated with GBM subtype (Hegi *et al*, 2005). On the contrary, resistance to antiangiogenic therapy with bevacizumab, recently introduced in GBM treatment, is closely linked to MES affiliation (Piao *et al*, 2013). Sandmann *et al*. in a retrospective analysis of a trial investigating the use of bevacizumab in GBMs, found that only IDHwt PN GBM patients derived an overall and progression-free survival benefit compared to placebo (Sandmann *et al*, 2015). Contrarily, although MES GBMs presented a longer PFS with bevacizumab, their OS was not affected. Other evidence showed that radiations induced a MES drift which drive to resistance to alkylating agents (Kim *et al*, 2021). Remarkably, temozolomide can in its turn induce a hypermutation state in recurrent tumors (Hunter *et al*, 2006; Kim & Verhaak, 2015; Felsberg *et al*, 2011; McFaline-Figueroa *et al*, 2015; Indraccolo *et al*, 2019), possibly generating neoantigens to be exploited and targeted by CD8+ T cells (Schumacher & Schreiber, 2015).

Considering the immune scars of each transcriptional subgroup, CL tumors are enriched for the activated dendritic cell gene signature, and therefore may benefit from dendritic cell vaccines (Palucka & Banchereau, 2012). Similarly, MES GBMs show increased levels of M2 macrophages and might be suitable for therapies attacking tumor-associated macrophages (Wang *et al*, 2017).

3.1.3.8 Proneural to mesenchymal transition

GBMs have been theorized to evolve from PN to MES. Cytoreductive surgery and adjuvant treatments provide survival benefits but induce tumor evolution by selecting therapy-resistant tumor cells. This often leads to a switch upon recurrence that has been linked to treatment resistance, although its frequency and relevance remains controversial (Bao *et al*, 2006a; Bhat *et al*, 2013; Ozawa *et al*, 2014; Phillips *et al*, 2006). To determine the significance of this transition in IDHwt GBMs, Wang *et al*. performed a longitudinal transcriptional analysis on paired samples obtained at diagnosis and recurrence. The MES subtype was the most stable between primary and recurrent tumors (65%), as opposed to CL (51%) and PN (41%) classes. At recurrence, 10, 13, and 18 tumors switched affiliation to become CL, MES, and PN, respectively. Therefore, at recurrence, PN and MES were increased, while CL became less frequent. In fact, as the CL cells are the most sensitive to intensive therapy, there may be an actual advantage for non-CL cells (Verhaak *et al*, 2010), leading to underrepresentation of CL subtype after treatment (Van Den Bent *et al*, 2015). Furthermore, subtype stability at recurrence was linked to lower transcriptional

heterogeneity, with same affiliation at recurrence in 30 out of 45 (67%) recurrent cases, compared to only 26 out of 46 (57%) cases of tumors with higher transcriptional heterogeneity (Wang *et al*, 2017).

The PN-to-MES transition is also affected by tumor microenvironment. Tumor-associated macrophages, deriving from peripheral blood or from resident microglia (Gabrusiewicz *et al*, 2016; Hambardzumyan *et al*, 2015), could perturb this process through NF- κ B activation (Bhat *et al*, 2013) or other proliferative signals (Patel *et al*, 2014; Wang *et al*, 2017).

In the work by Wang *et al.*, a comparison between 91 primary and recurrent IDHwt tumors pinpointed a relative depletion of peripheral blood-derived monocytes upon recurrence. However, primary non-MES tumors switching to MES showed a boosted immune infiltrate, as opposed to recurrent non-MES lesions. In particular, M2 macrophage were significantly more frequent in cases transitioning to MES, just as a higher M2 fraction was observed in primary MES relative to primary non-MES GBMs. Hence, the relationship between MES GBMs and macrophages is further strengthened, extending their codependence to disease recurrence (Wang *et al*, 2017).

3.1.4 Pathology and classification

From a histopathologic point of view, GBMs are diffuse, infiltrative astroglial neoplasms with microvascular proliferation and pseudo palisading necrosis. Nuclei are angulated with irregular chromatin. Mitoses are frequent (Wen *et al*, 2020).

The 2016 CNS WHO classification of brain tumors divided them into IDHwt GBMs (about 90% of cases), which correspond to primary GBMs and predominate in patients older than 55 years, and IDH-mutant GBMs (10% of cases), corresponding to secondary tumors deriving from dedifferentiation of prior lower-grade gliomas and affecting younger people (Louis *et al*, 2016).

Sometimes GBMs may lack their typical histologic hallmarks. Prior to the integration of molecular analyses, they used to be classified as lower-grade gliomas. However, several studies demonstrated that if these tumors harbor specific molecular scars, they act like proper GBMs and should be treated as such (Brat *et al*, 2018). In particular, when IDHwt tumors harbor either TERT promoter mutation, EGFR amplification, or combined +7/-10 chromosome alterations, they are rightfully assigned the highest WHO grade, even when histology meets only WHO grade 2 or 3 criteria (Brat *et al*, 2018; Tesileanu *et al*, 2020; Louis *et al*, 2021).

Conversely, mutations in IDH1/2 portends extended patient survival (Yan *et al*, 2009). Usually IDH status is initially screened by mutation-specific immunohistochemical search for IDH1-R132H, which makes up to >90% of IDH mutations in GBMs (Wen *et al*, 2020). Sequencing analyses to detect other non-canonical mutations, such as those in IDH2 or non-R132H mutations in IDH1, is optional in patients older than 55, since they are very unusual at this age (Wen *et al*, 2020). It is also very unlikely to diagnose an IDH-mutant lesion when microthrombi or pseudo palisading necrosis are spotted. Remarkably, most IDH-mutant GBMs display concurrent loss of ATRX, easily detected at immunohistochemistry (Wen *et al*, 2020).

Given this difference in clinical behavior as a function of IDH mutation, the new 2021 CNS WHO classification of brain tumors clustered adult diffuse gliomas into IDH-mutant and IDHwt astrocytomas, IDH-mutant and 1p/19q-codeleted oligodendrogliomas, and IDHwt GBMs. Therefore, all IDH-mutant diffuse astrocytic tumors are considered a single entity graded as CNS WHO grade 2, 3, or 4. Additionally, grading no longer relies on pure histology, as the presence of at least one of the three aforementioned genetic aberrations or of homozygous deletion of CDKN2A/B results in a CNS WHO grade 4 astrocytic tumor, even if microvascular proliferation or necrosis are not detected (Louis *et al*, 2021).

Consequently, IDHwt GBM are diagnosed in case of a IDHwt diffuse astrocytic tumor with microvascular proliferation or necrosis, or TERT promoter mutation, or EGFR amplification, or +7/–10 chromosome aberrations (Louis *et al*, 2021).

Another important molecular parameter to predict treatment response is the status of MGMT promoter. Multiple trials have shown a significant survival advantage (an increase of circa 50%) with TMZ treatment in case of MGMT methylation (Hegi *et al*, 2005; Stupp *et al*, 2005). On the contrary, unmethylated MGMT GBMs derive little to no benefit from TMZ (Hegi *et al*, 2005; Weller, 2018). Perhaps one of the most considerable limitations to this prognostic factor is the absence of standardized cutoff values for MGMT status, generally set at 8-10% (Butler *et al*, 2020). However, not all studies use the same cutoffs, potentially confounding the clinical correlation of MGMT status (Brigliadori *et al*, 2016; Dunn *et al*, 2009). Even MGMT protein expression cutoffs vary, ranging from 10% to 25%, 30%, or 50% (Butler *et al*, 2020). Furthermore, approximately 10% of patients bear tumors that cannot be definitely classified as methylated or unmethylated, but derive a significant advantage from TMZ compared to those truly unmethylated (Hegi *et al*, 2019). This likely confounds the clarification between MGMT methylation thresholds and patient outcome (Mansouri *et al*, 2019; Taylor & Schiff, 2015). Therefore, it is vital to standardize

relevant cutoffs for diagnostic assays, possibly introducing a third intermediate class containing patients with intermediate methylation, who, may still profit from TMZ treatment (Butler *et al*, 2020).

3.1.4.1 Histological variants of GBMs

A few histological variants of GBMs have been described. Gliosarcomas account for a small subset of GBMs. They present a biphasic tissue pattern with a sarcomatous component mixed with more canonical cells (Dardis *et al*, 2021). Expression of SNAI2, TWIST, MMP2 and MMP9 is characteristic of mesenchymal areas, suggesting a more pronounced epithelial to mesenchymal transition (EMT). They typically have lower EGFR copy number alteration with respect to GBMs (8% vs up to 50%). Canonical genetic and chromosomal alterations detected both in the glial and mesenchymal components indicate a common clonal origin. Gliosarcomas are described to be more immuno-evasive via PD1/PDL1, which, however, may be exploited as therapeutic target (Dardis *et al*, 2021).

A second variant is giant cell GBM, which is composed of large cells with polymorphic nuclei, eosinophilic cytoplasm, and increased reticular fibers. It accounts for 2-5% of all GBMs, primarily affecting the temporal lobes in younger men. There is rare endothelium proliferation (Xue *et al*, 2021). From a molecular point of view, giant cell GBMs often harbor TP53 mutations but only rare EGFR amplifications. Although with similar clinical presentation, the prognosis of giant cell GBMs is better than that of typical GBMs, with >10% experiencing prolonged survival (12.3% 5-year OS in giant cell GBMs vs 3.4% in typical GBMs) (Xue *et al*, 2021). This might be explained by the fact that giant cell GBMs have clearer, well demarcated boundaries compared to usual GBMs, rendering them easier for surgeons to remove. Hence, maximal surgical resection followed by adjuvant chemoradiation improves the prognosis of these patients (Xue *et al*, 2021).

One recently added histologic variant of GBM is the epithelioid GBM. It features large cells with abundant cytoplasm, vesicular chromatin, and prominent nucleoli, often resembling melanoma or poorly differentiated carcinoma (Louis *et al*, 2016). It generally affects children or young adults and forms superficial or diencephalic masses, frequently marked by BRAF V600E mutation. They often lack other canonical molecular features (Louis *et al*, 2016).

GBMs with primitive neuronal component consists in diffuse astrocytomas of any grade (or even oligodendrogliomas) with well-demarcated nodules displaying neuronal differentiation (e.g., Homer Wright rosettes, synaptophysin-positive, GFAP-

negative cells). MYC or MYCN amplification is occasionally reported, and lesions tend to disseminate in cerebrospinal fluid. Around 25% of cases affect people with a prior history of lower-grade glioma. Very remarkably, this pattern should suggest the clinician to investigate the craniospinal axis in its entirety to exclude tumor seeding (Louis *et al*, 2016).

Small cell GBM and granular cell GBM are tumor variants, characterized either by uniform, small, EGFR-amplified cells reminding of oligodendroglioma, or by granular, macrophage-like, lysosome-rich tumor cells. In both cases, the prognosis is extremely poor even without other histological stigmata of grade 4 tumors (Louis *et al*, 2016).

3.1.4.2 Transcriptional variants of GBMs

As previously described, GBMs can be clustered into 3 transcriptional subtypes expressing peculiar genesets and with diverging therapeutic implications (Wang *et al*, 2017; Phillips *et al*, 2006). Genes typically expressed in neurons (e.g., ASCL1, OLIG2) and associated to CpG island methylator phenotype (Noussmeh *et al*, 2010) define the PN subgroup, frequently associated to mutations of TP53. CL GBMs portend EGFR alterations, such as amplification or EGFRvIII. Inactivation of CDKN2A, absence of TP53 mutations, and alterations in RB pathway are also common (Verhaak *et al*, 2010). The MES class is marked by genes such as YKL40, MET, and CD44, besides frequent NF1 deletions, anomalies in CDK6, CDKN2A and RB1, and activation of NF- κ B (Wang *et al*, 2017).

3.1.4.3 Immunohistochemical panel for GBM transcriptional affiliation

Transcriptional signatures are associated with different molecular alterations (Verhaak *et al*, 2010), key signaling pathways (Brennan *et al*, 2009), and DNA methylation (Noussmeh *et al*, 2010). However, GBM transcriptional classification has not yet converted into routine practice primarily because of the absence of reliable and cost-effective clinical assays. Further confounding factors include analysis on homogenized samples, which cannot discern microenvironmental contamination, reaching up to 70% of tumor bulk (Orzan *et al*, 2020). This is especially true for MES tumors, which reach the highest percentage of immune/inflammatory component. Likewise, the previously described GBM heterogeneity further complicates transcriptional classification, which may not be

illustrative of the whole neoplasm. This conundrum may be partially overcome by immunohistochemistry (IHC), which allows to discern different compartments within the entire sample, scoring the biomarkers only in neoplastic cells and excluding stroma and immune components (Orzan *et al*, 2020).

Orzan *et al*. developed a novel algorithm to predict GBM molecular profile by combining transcriptional and IHC data that can be easily applied to the routine clinical diagnostic practice (<http://fisher.med.unibs.it:3838/GBMscore>) (Orzan *et al*, 2020). The authors validated a simple and reproducible, restricted panel of 8 subgroup-specific gene classifiers (EGFR for CL, ASCL1, OLIG2, PDGFRA for PN, MET, YKL40, pNDRG1 for MES, and TP53). These markers are mutually exclusive or inversely correlated, allowing for accurate clustering analysis discerning transcriptional subtypes (Orzan *et al*, 2020). This assay attested high concordance with transcriptional analysis. Notably, concordance rates reached 81.3% and 90% for CL and MES respectively, while it reached only 69.2% in case of PN tumors. Moreover, the authors observed that CL and MES lesions reliably clustered separately, as opposed to PN lesions, which often overlapped with other subtypes, reflecting a heterogeneous phenotype (Orzan *et al*, 2020).

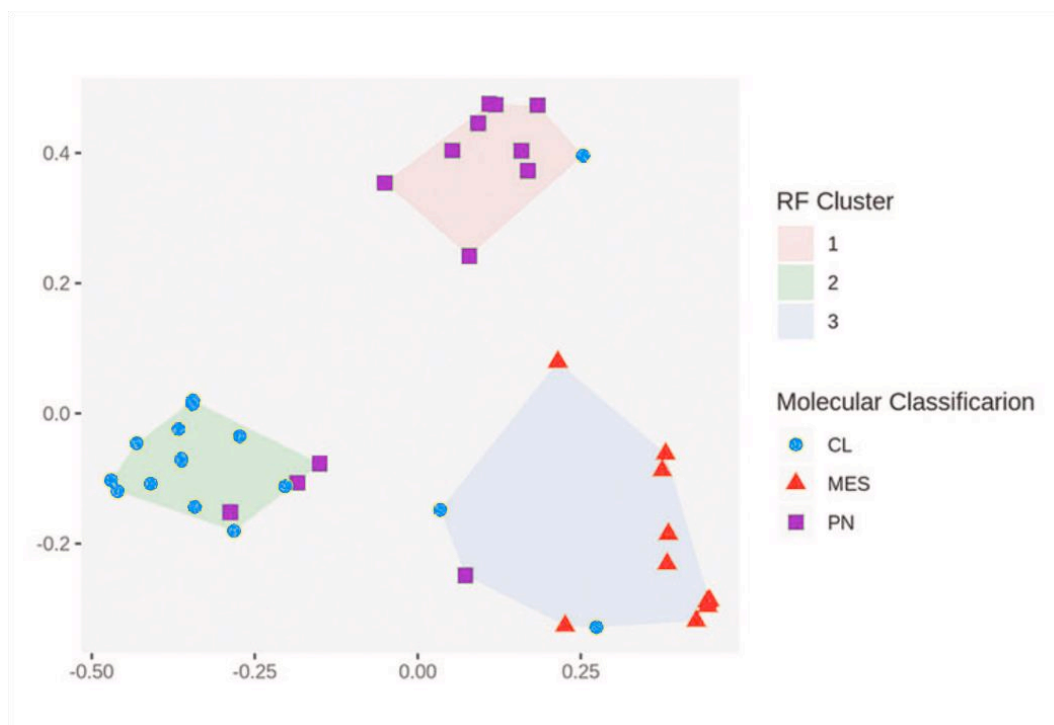


Figure 3.2. GBM transcriptional affiliation according to IHC panel classification and validated by RNAseq analysis. Random Forest performed based on IHC signal clustered RNAseq-verified CL and MES subtypes consistently in separate groups, while PN often shared overlapping features with other subtypes. CL: Classical, IHC: immunohistochemistry, MES: mesenchymal, PN: proneural, RF: random forest. Adapted from (Orzan *et al*, 2020)

3.1.5 Management

Despite advances in deciphering GBM pathogenesis, prognosis remains poor. With gold standard management, consisting of surgery, radiotherapy, and TMZ chemotherapy, median OS still attests at 15-18 months (Stupp *et al*, 2005; Gilbert *et al*, 2014), and 5-year OS hardly reaches 10% (Stupp *et al*, 2009). Upon recurrence, median OS is generally 24-44 weeks (Lamborn *et al*, 2008; Clarke *et al*, 2011). According to recent trends in medicine, treatment should be adapted to the patient, based on their performance status and age, even in case of GBMs. Additionally, palliative and best supportive care is of utmost importance in many patients (Weller *et al*, 2017).

3.1.5.1 Surgery

Surgery is crucial to obtain the histopathologic and molecular characterization of the lesion and may benefit patients also by cytoreduction and alleviating mass effect (Youngblood *et al*, 2021). Surgical management ought to be adapted to individuals, balancing risks and benefits, and weighing the effects on prognosis in each single case (Wen *et al*, 2020).

Current typical surgical aides involve stereotactic navigation, intraoperative functional monitoring, intraoperative MRI, ultrasound, and 5-aminolevulinic acid (5-ALA), which are utilized to increase the extent of resection (EOR) while containing the hazard of treatment-induced neurologic deficits (Stummer *et al*, 2006; De Witt Hamer *et al*, 2012; Gerritsen & Vincent, 2020). In particular, intraoperative brain mapping is crucial when eloquent circuitries are affected, as FLAIR-positive, non-enhancing areas often maintain neurologic function despite tumor infiltration (Youngblood *et al*, 2021). In general, avoiding new persistent deficits is more imperative than maximizing the EOR, since surgery is not sufficient for GBM eradication and postoperative deficits are negative prognostic factors (Wen *et al*, 2020).

The procedural success is commonly expressed in terms of EOR. As reported in the literature, greater EOR is associated with prolonged survival. Accordingly, surgeons aim to resect all the tumoral tissue that can be safely separated from critical neurovascular structures. Usually, EOR is defined solely on post-contrast T1-weighted MRI sequences; however, hyperintense T2/FLAIR tumor volume should be included into the resection target when feasible to complete the excision. Consequently, asportation of the enhancing nodule and additional T2/FLAIR abnormalities is the ultimate surgical goal (Wen *et al*, 2020). Although radical

microsurgical resection is unattainable due to GBM capillary invasion along white matter tracts (Sahm *et al*, 2012), maximal excision can be accomplished in many cases and only rarely surgeons perform secondary interventions on residues (Youngblood *et al*, 2021). Tumor EOR is assessed at an early postoperative contrast enhanced MRI, typically within the first 48 hours following surgery (Wen *et al*, 2020).

Gross total resection (GTR) is associated with better outcome in both newly diagnosed (Sanai *et al*, 2011; Grabowski *et al*, 2014; Stummer *et al*, 2006; Kreth *et al*, 2013; Marko *et al*, 2014; Brown *et al*, 2016) and recurrent GBMs (Ringel *et al*, 2016; Suchorska *et al*, 2016). GTR is defined based on varying cutoffs of volume reduction of the initial enhancing nodule, ranging from 90 to 100%. Notably, small changes in EOR in the range of 90-100% may bring about clinical differences (Karschnia *et al*, 2021). Beside EOR, also the residual enhancing volume independently affects survival (Youngblood *et al*, 2021).

Consequently, extensive excisions should represent the primary intent. Furthermore, resection of peripheral FLAIR anomalies and even surrounding apparently normal tissue may prove beneficial in removing a great deal of infiltrating malignant cells (Youngblood *et al*, 2021). This is corroborated by evidence comparing MRI and PET, which exposed significant tumor volumes beyond contrast enhancement (Karschnia *et al*, 2021). Therefore, supratotal resection does have a rationale in GBM treatment and has been associated with longer survival. In younger patients this approach may even result in comparable survival benefits between IDH-mutant and IDHwt neoplasms. It also delays malignant transformation of diffuse lower-grade gliomas (Youngblood *et al*, 2021; Rossi *et al*, 2021). Basically, by identifying functional boundaries with the aid of intraoperative neurophysiological monitoring, the tumor resection can be safely extended beyond the preoperative MRI contours (Youngblood *et al*, 2021). Li *et al*. identified 53% of the surrounding T2/FLAIR-abnormality as a threshold to attain an additional survival benefit (Li *et al*, 2016). In a similar study, Pessina *et al*. set this threshold to >45% of T2/FLAIR-abnormality (Pessina *et al*, 2017). As far as absolute volumes of non-enhancing T2/FLAIR residues are concerned, Molinaro *et al*. identified a limit of 5.4 cm³ under which patients had improved outcome compared to those above it (Molinaro *et al*, 2020).

Remarkably, when GTR or supratotal resection cannot be attained, patients may still derive survival benefits from near-total resection. This scenario includes those patients in which >80% of tumor volume (with <5 cm³ residue) has been resected. However, patients who undergo a >95% resection achieve better outcomes than those with a <95% resection. Therefore, patients with 95-99% EOR have an

intermediate survival curve between patients with 80-94% and those with 100% resection (Karschnia *et al*, 2021). Similar findings were reported concerning contrast enhancing tumor residues. As a matter of fact, GBM patients with 0-1 cm³ residue had more favorable outcome than patients with 1-5 cm³ or >5 cm³ residual enhancing lesion (Incekara *et al*, 2020). Interestingly, this evidence proved true independent of MGMT status. Hence, only patients with 95-99% EOR and 1 cm³ contrast-enhancing residue should properly be denoted as near-total resection cases (Karschnia *et al*, 2021).

Subtotal resection appeals to cases where a substantial proportion of the tumor is not excised. Most studies reported a minimum of 78-80% EOR to achieve a significant survival benefit. Accordingly, subtotal resection should be kept for cases with >80% resection of enhancing nodule and <5 cm³ residue, with remaining cases defined as partial resections (Karschnia *et al*, 2021).

Lastly, whenever microsurgical resection is deemed too risky due to patient's comorbidities and/or the eloquence of the areas involved, the lesion should undergo at least needle or open biopsy for histological and molecular diagnosis, by targeting the viable, contrast-enhancing areas in solid tumor, and avoiding necrotic areas or normal parenchyma (Weller *et al*, 2017). The most frequently used markers (IDH1/2 and MGMT status) are consistently represented throughout the tumor, with a low risk of false-negatives or misclassification. Still, additional markers may gain clinical importance, demanding larger samples for advanced genomic analyses (Wen *et al*, 2020).

The correlation between EOR and survival still represents a debated topic. Due to its infiltrative nature, even GTR is a temporary measure for GBM containment. As a matter of fact, surgery alone does not eradicate the tumor, which, after GTR, recurs after a median of 13 months. Nevertheless, lesion removal portends increasing survival with greater EOR, even when correcting for significant covariates (Youngblood *et al*, 2021). First, reduction of the mass effect relieves incapacitating neurologic symptoms, such as weakness, nausea, visual changes, headache, and seizures, or death following critical parenchymal herniation. However, besides parenchymal decompression, GTR seems to exert its action also through cytoreduction. In fact, extensive resections alter tumor architecture, inducing changes in the microenvironment that can freeze tumor progression. For instance, surgery primarily targets nodules containing hypoxic areas, which are associated with greater proliferation and behave more aggressively. Therefore, long-term outcome is also influenced by the biological profile of the residue, which may differ from that

of the initial lesion. Lastly, cytoreduction also helps the effects of radio-chemotherapy (Youngblood *et al*, 2021).

In conclusion, surgery plays a central role in handling GBMs and constitutes the foundation of the gold standard multimodal approach. Surgical goals are set on the benefits foreseen for the patients and limited by the preventable postoperative deficits. The subtle balance between tumor excision and operative morbidity must therefore be carefully pondered (Youngblood *et al*, 2021).

3.1.5.2 Adjuvant treatment

After maximal safe resection, GBM treatment consists of radiotherapy (RT) with concurrent radiosensitizing TMZ (75 mg/m²/day per 6 weeks) and following maintenance TMZ (150–200 mg/m²/day per 5 days for six 28-day cycles) (Stupp *et al*, 2005, 2009). Inhibition of DNA repair through MGMT methylation and loss of chromosome 10 confer TMZ sensitivity and predict treatment response (Hegi *et al*, 2005). Thus, TMZ might be withheld in case of MGMT unmethylated tumors, where it confers too limited benefit. In such cases, neither dose-dense TMZ regimens, nor prolonging treatment further benefits survival (Wen *et al*, 2020).

Standard RT approaches reckon on 60 Gy in 30 fractions. The European Organization for Research and Treatment of Cancer (EORTC) advise a single-phase treatment (60 Gy, 2 Gy/fraction) whereas the Radiation Therapy Oncology Group (RTOG) suggests an initial larger target exceeding the FLAIR abnormality by a 2-cm margin (46 Gy, 2 Gy/fraction), with a surplus 14 Gy dose (2 Gy/fraction) to the surgical bed and residue. It is paramount to spare structures sensitive to radiations, such as optic pathways, brainstem, cervical cord, cochlea, and, when possible, temporal lobes and hippocampi (Wen *et al*, 2020).

Approximately 10-30% of patients develop a transitory increase in contrast uptake at MRI for several months after radiochemotherapy, occasionally coupled with symptomatic mass effect (Ellingson *et al*, 2017; Strauss *et al*, 2019). Such “pseudoprogression” poses quite a few challenges to be discerned from actual progressive disease; perfusion MRI (Thust *et al*, 2018; Gharzeddine *et al*, 2019) and amino acid-PET may aid the differential diagnosis (Albert *et al*, 2016).

Studies with adjunct of bevacizumab reported contrasting results, with no survival benefits unless stratifying for transcriptional affiliation (Gilbert *et al*, 2014; Wick *et al*, 2017; Chinot *et al*, 2014; Sandmann *et al*, 2015). Nevertheless, bevacizumab is helpful in diminishing peritumoral edema and associated symptoms (Friedman *et al*, 2009), in elongating PFS and in reducing the needs for corticosteroids. Thus, it has

been approved in the US, but not in the EU, for recurrent GBMs (Wick *et al*, 2017). Additionally, bevacizumab is effective in radionecrosis at inferior dosages than those used for recurrent GBMs (7.5 mg/kg every 3 weeks for a maximum of 4 treatments) (Wen *et al*, 2020).

3.1.5.3 Elderly patients

GBM typically affects older people, with a median onset at 65 years. Therefore, a substantial portion of patients are actual elders (Ostrom *et al*, 2019a). Treating older people may be challenging, as they are more prone to a worse prognosis and treatment toxicities (Wick *et al*, 2018). A very thorough and systematic review of GBM treatment in elders has been carried out by Conti Nibali *et al.*, whose main points are followingly summarized. As highlighted by the authors, the elderly age starts at 65 according to the WHO; however, the National Comprehensive Cancer Network moved this threshold to 70 years in GBM patients. Nonetheless, general amelioration of life conditions and medical assistance in western countries translated into expanding the active lifespan, and present-day 70-year-old people may still have a vigorous social and intellectual life. Therefore, physiologic age is more important than its chronological counterpart (Conti Nibali *et al*, 2021).

As a matter of fact, age does not imply frailty, which, instead, reflects a physical status. Frailty consists in an unintentional weight loss, self-reported exhaustion, weakness, slow walking speed, and low physical activity. Despite age and frailty not being perfectly superimposable, age still remains an independent negative prognostic factor (Conti Nibali *et al*, 2021).

Just as in younger patients, EOR affects survival in elders as well. People undergoing GTR experience a mean gain in OS of 7.05 months, a better functional recovery, a longer PFS, and comparable mortality and morbidity to those undergoing biopsy (Conti Nibali *et al*, 2021).

As for molecular predictors, though methylation physiologically decreases in aging cells, this does not hold true for MGMT. On the contrary, VEGF expression increases in recurrent GBMs affecting people older than 55. Lastly, contrarily to younger subjects, TP53 and CDKN1A/p16 aberrations are prognostically unfavorable in people >70 years old (Conti Nibali *et al*, 2021).

As for adjuvant therapies, RT extends survival in case of unmethylated MGMT, while TMZ yields to better PFS in MGMT methylated patients. Globally, RT (50 Gy, 1.8 Gy/fraction) prolongs OS and is superior to the best supportive care in elders with a Karnofsky Performance Status of ≥ 70 . Hypofractionated regimens are

preferred and have become the standards in elders with unmethylated MGMT (Conti Nibali *et al*, 2021).

TMZ proves efficacious and is well-tolerated in elders too, with <15% of severe side effects. Post-RT TMZ should be administered in case of MGMT methylation. Also, depending on the MGMT status, exclusive TMZ represents an opportunity for elders with an extremely dismal prognosis and poor functional status (Conti Nibali *et al*, 2021).

3.1.5.4 Recurrent Glioblastoma

GBM regularly recurs after a median PFS <7 months (Stupp *et al*, 2005). Surgery may be helpful in symptomatic or large recurrences. However, parallelly to what seen with initial lesions, only recurrent patients undergoing new GTR derive survival benefit (Suchorska *et al*, 2016).

Rechallenge with TMZ could be considered in cases of relapsing MGMT methylated tumors. Yet, recurrence must be delayed, as rechallenge can be adopted only after a reasonable timelapse since first-line TMZ (Perry *et al*, 2010; Weller *et al*, 2015). However, there is no conclusive proof that TMZ rechallenge is more effective than nitrosoureas, which in turn can be started earlier than TMZ rechallenge (Wen *et al*, 2020).

Nitrosoureas (lomustine, carmustine, and fotemustine) can easily penetrate the blood-brain barrier (Brandes *et al*, 2016a). Lomustine is preferred because it can be administered orally and has a preferable administration schedule and better safety. Monotherapy lomustine (6-week cycles of 100-130 mg/m² for up to 6 cycles) is associated with median OS of 7.1-8.6 months and PFS of 1.5-3 months (Wick *et al*, 2017). Similarly to TMZ, MGMT-methylated GBMs respond better to nitrosoureas (Wick *et al*, 2017; Brandes *et al*, 2016b; Wen *et al*, 2020).

Repeated RT, such as hypofractionated radiotherapy (30-35 Gy in 5-15 fractions) or single-boost radiosurgery, is a further option, but no definitive data regarding benefit has been reported (Straube *et al*, 2019; Scocciati *et al*, 2018). Other options include bevacizumab or, when there is no space for further strategies, palliative care (Wen *et al*, 2020). Preliminary results of a phase II trial in recurrent GBMs showed that addition of re-irradiation to bevacizumab improved PFS but not OS (Wen *et al*, 2020).

3.1.5.5 Supportive Care

Corticosteroids, usually in the form of dexamethasone, are administered to reduce symptomatic vasogenic edema (Pace *et al*, 2017). Low-dose dexamethasone (4 mg/day in 1-2 doses) is efficacious in the majority of patients (Wen *et al*, 2020). Notably, increasing evidence suggests that corticosteroids may affect negatively patients' outcome, so it is advisable to avoid them in asymptomatic patients (Pitter *et al*, 2016).

Seizures affect 23% of GBM patients at presentation and an additional 20% later in the disease course (Chang *et al*, 2005). Treatment with anti-epileptic drugs (AEDs) is recommended only for patients who developed seizures, while there is no evidence supporting the prolonged use of AED in primary prophylaxis for those who never experienced seizures or in a perioperative setting. Current guidelines recommend avoiding long-term AED prophylaxis (Wen *et al*, 2020). Among AEDs, levetiracetam and lacosamide are preferentially utilized due to their fewer side effects compared to other molecules, reduced requirements to monitor blood routine, and absence of drug interactions (Schiff *et al*, 2015). Interestingly, emerging data demonstrated the establishment of synapses between neurons and glioma cells via alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA-R) receptors. Synaptic interactions are exploited by tumor cells to derive proliferation inputs and to drive infiltration along white matter fibers. Hence, AEDs inhibiting AMPA-R (e.g., perampanel) may have a doubly favorable effect, both for seizure control and for potential antitumoral action (Venkatesh *et al*, 2019; Venkataramani *et al*, 2019).

Venous thromboembolism risk is perioperatively high and persists well beyond, with 20% incidence at 1 year (Wen *et al*, 2020). Most studies balancing thrombotic and hemorrhagic hazards suggest that the bleeding risk of the lesion under anticoagulants is tolerably contained, even in case of concomitant bevacizumab treatment (Norden *et al*, 2012; Wen *et al*, 2020). In this context, low molecular weight heparin (LMWH) is usually adopted, but also direct oral anticoagulants have proven safe (Wen *et al*, 2020).

3.1.5.6 Targeted therapies

Notwithstanding progress in elucidating the pathogenetic mechanisms driving GBM, there has not been a parallel advance in discovering efficient targeted therapies (Le Rhun *et al*, 2019). Challenges include the scarcity of drugs effectively crossing the blood-brain barrier (Arvanitis *et al*, 2020), the lack of easy targets, redundant signaling pathways, tumor heterogeneity (Neftel *et al*, 2019; Draaisma *et al*, 2020),

targeting subclonal alterations instead of oncogenic drivers, and development of resistance (Wen *et al*, 2020).

Some putative treatment-exploitable biomarkers include EGFR aberrations, which have been targeted with EGFR inhibitors such as erlotinib, or addressed for vaccination with rindopepimut, or for specific drug delivery with depatuxizumab mafodotin, still with no significant outcome benefits (Wen *et al*, 2020). Raf inhibitors or combined BRAF/MEK inhibitors proved to be clinically efficacious in targeting BRAF V600E mutations. However, these aberrations are rare in GBMs, except for epithelioid variant, hence their clinical impact remains limited (Wen *et al*, 2020). Similarly, other druggable mutations, such as TRK fusions, H3K27M mutations, FGFR mutations and FGFR3-TACC3 fusions, are all quite sporadic in GBMs, hampering their strategic contribute. Microsatellite instability may be addressed with pembrolizumab (Wen *et al*, 2020).

The mammalian target of rapamycin (mTOR) Ser2448 phosphorylation, as well as PTEN loss (Cloughesy *et al*, 2008), might predict effective response to temsirolimus + RT in unmethylated GBMs (Wick *et al*, 2016).

Multiple recent studies identified metabolism as a major element in glioma progression. Oncogenes and microenvironmental cues control GBM metabolism to endorse survival, replication, and drug resistance (Venneti & Thompson, 2017; Bi *et al*, 2020). Controllers of GBM metabolic activity might provide useful prognostic and diagnostic indicators (Bi *et al*, 2020; Pieri *et al*, 2022). Moreover, tumor genotype and molecular affiliation influence GBM metabolism, inducing weaknesses that could be therapeutically exploited (Bi *et al*, 2020; Garofano *et al*, 2021; Pieri *et al*, 2022).

A classic biochemical adaptation in GBMs is the switch to anaerobic glycolysis, regardless of oxygen availability (Bi *et al*, 2020). In this sense, GBMs relying on oxidative phosphorylation provide a unique weakness profile which can be specifically addressed (Bi *et al*, 2020; Garofano *et al*, 2021). Likewise, cholesterol metabolism might be therapeutically interesting for CL tumors, which depend on cholesterol uptake and may suffer from reduced cholesterol uptake by liver receptor agonists (Bi *et al*, 2020). Therefore, targeting genes tuning metabolism might open new avenues for efficacious treatment.

3.1.5.7 Immunotherapies

Just like for other forms of targeted therapy, no considerable advance has been observed for immunotherapy in GBMs (Sampson *et al*, 2020; Lim *et al*, 2018). Although relieving glioma-associated immunosuppression to allow immune-mediated

antitumor response has been sustained by positive preclinical experiments, clinical trials involving chemokines, vaccines, or immune checkpoint blockade were unsuccessful (Sampson *et al*, 2020; Lim *et al*, 2018).

Sporadic positive results with immune checkpoint blockade in the presence of hypermutator phenotype indicate the presence of relevant neoantigens to be recognized and attacked (Wen *et al*, 2020). Notably, hypermutator phenotype is observed in circa 10% of relapsing GBMs after TMZ challenge (Wang *et al*, 2016; Draaisma *et al*, 2020). However, whereas the number of mutations may be insufficient to sustain an immune response, their intrinsic immunogenicity and clonality could drive a positive reaction to immunotherapy (Keenan *et al*, 2019).

GBMs are immunologically cold tumors characterized by scant infiltration of effector lymphocytes. Accumulating evidence suggests that microenvironment plays a role in affecting treatment response and may allow for selecting patient subsets differentially responding to specific regimens (Jackson *et al*, 2019; Sampson *et al*, 2020). Other factors driving intrinsic GBM resistance include a scarcity of neoantigens due to low mutational burden (except for lesions presenting the aforementioned hypermutator phenotype), and active immune inhibition by soluble immunosuppressors such as TGF β , IL10, and prostaglandin E2 (Jackson *et al*, 2019; Sampson *et al*, 2020; Platten *et al*, 2019). Thus, GBMs promote adaptive resistance by inducing exhaustion of infiltrating T-cells (Woroniciecka *et al*, 2018) and engaging suppressive myeloid and regulatory T-cells (Jackson *et al*, 2019; Lim *et al*, 2018). In addition, medications regularly administered, such as corticosteroids, bring about immunosuppressive effects, thus acting against immunotherapies (Wen *et al*, 2020).

Myeloid cells are instructed by GBMs to be immunosuppressive (Jackson *et al*, 2019). However, therapies with oncolytic viruses can switch macrophages from immunosuppressive M2 to proinflammatory M1 phenotype, inducing antigen presentation and promoting immune attack against the tumor (Lawler *et al*, 2017).

Over the last years, cellular therapies such as dendritic cells and chimeric antigen receptor (CAR) T-cells and CAR-transduced NK cells have earned increasing interest. Dendritic cells present tumor antigens to effector T cells. Some groups have exploited them to induce T, B and NK antitumor response, treating more than 500 gliomas with dendritic cells, with an increased survival in 15.6% of cases (Finocchiaro & Pellegatta, 2014). Remarkably, dendritic cell immunotherapy led to increased frequency and IFN γ -dependent activation of NK cells, whose levels correlated with prolonged PFS and OS. Fittingly, immunosuppressive cytokines showed an inverse correlation with patient survival. Mature rather than immature dendritic cells specifically favors interactions with NK cells and consequent NK response potentiation (Pellegatta *et al*,

2013). However, an immunosuppressive environment negatively affects NK and T cells and impairs dendritic cell action. On the contrary, radio-chemotherapy enhances immune response (Finocchiaro & Pellegatta, 2014).

CAR T-cells bear chimeric receptors which let them bind specific antigens on tumor cells with subsequent target killing independently of HLA (Bagley *et al*, 2018). As a matter of fact, normal antigen-specific cytotoxic T cells recognize antigens presented by MHC-I. Yet, GBMs frequently harbor alterations in genes associated with the very mechanisms of antigen processing and presentation. CAR T-cells have been manipulated to bear an artificial protein, binding an antigen-specific extracellular fragment to an intracellular signaling domain and a costimulatory molecule. This allows to combine the effector T cell functions with the ability of antibodies to recognize surface antigens independent of MHC restriction (Finocchiaro & Pellegatta, 2014). Thus, CAR T-cells bypass many tumor-inherent obstacles and improve T-cell specificity and tumoricidal effect. However, shortcomings regarding this strategy include the risks of unexpected on-target or off-target effects. Therefore, meticulous selection of antigens is essential to design an efficacious and safe treatment. Additionally, the costimulatory signals may induce severe cytokine storms (Finocchiaro & Pellegatta, 2014).

3.1.6 Challenges in glioblastoma management

Despite decades of research, GBMs are still among the most frightening diagnoses, with disproportionate mortality. As a matter of fact, even though they represent only 1.4% of all malignant tumors, they cause up to 2.9% of tumor-related deaths (Wen *et al*, 2020).

Several challenges must still be tackled. First, their location prompts consideration of treatment-related neurologic toxicities (Murphy *et al*, 2015), which severely affect the patients' quality of life. A second aspect is the blood-brain barrier, which prevents the diffusion of large, electrically charged, water-soluble compounds into the CNS. Unfortunately, most drugs (especially conjugated antibodies and the new target-designed molecules) do possess these features and, thus, are not able to cross this barrier to a significant degree (Heffron, 2018). Additionally, the presence of transporter proteins and active pumps on the blood-brain barrier helps extruding drugs from the brain (Heffron, 2018; Arvanitis *et al*, 2020). Although GBMs have a leaky, compromised blood-brain barrier in the central enhancing core, significant

tumor regions at the margins have a healthy barrier preventing antitumoral drugs from reaching useful concentrations (Sarkaria *et al*, 2018; Arvanitis *et al*, 2020).

A third challenging aspect is the inter- and intratumoral heterogeneity as described previously. Distinct regions within the same lesion may contain cells with different genetic compositions, transcriptional affiliations, metabolic activity and proliferation kinetics (Lan *et al*, 2017). Preclinical studies suggest that these distinct cell subpopulations can have differential responses to TMZ or RT, explaining resistance to conventional treatments (Lan *et al*, 2017; Bao *et al*, 2006a).

Lastly, GBMs also present a noteworthy plasticity to elude treatment toxicities. For instance, they can elude tyrosine kinase inhibition through regulation of DNA accessibility through epigenetic chromatin reconfiguration, new inhibition of tumor suppressors or reactivation of oncogenes (Wen *et al*, 2020).

3.2 Magnetic Resonance Imaging (MRI)

3.2.1 Conventional MRI

Conventional MRI (cMRI) is the gold standard technique to study GBMs (Wen *et al*, 2020). It assesses both the T2/fluid-attenuated inversion recovery (FLAIR) anomalies and the gadolinium uptake on postcontrast T1 images, indicating sites of abnormal blood-brain barrier, where contrast agents can leak.

cMRI provides essential morphologic information about the neoplasm and its vasculature, in addition to assessing the disease burden. These sequences are also useful in showing the deformation of healthy structures consequent to the presence of the lesion. Evaluation of these sequences is usually enough to consistently suggest a diagnosis of GBM in most cases (Riva *et al*, 2021).

3.2.2 Advanced MRI

Advanced MRI techniques (aMRI) include diffusion MRI, perfusion-weighted imaging, and proton magnetic resonance spectroscopy. They can provide quantitative measures of pathophysiologic features of GBMs (Henriksen *et al*, 2022).

3.2.2.1 Diffusion MRI (DWI)

Diffusion MRI sequences to study GBMs range from standard diffusion-weighted imaging (DWI) and apparent diffusion coefficient (ADC) to more complex elaborations, such as diffusion tensor imaging (DTI) or microstructural models.

The basic DWI sequence relies on diffusion of water molecules. Henriksen *et al*. provide a detailed description of the DWI and ADC specifics, which are here summarized. Gradient strength and duration are expressed in b-values (s/mm^2). The b-value is defined as $b = \gamma^2 G^2 \delta^2 (\Delta - \delta/3)$, where δ is the diffusion gradient length, Δ the separation, γ the gyromagnetic ratio and G strength (Henriksen *et al*, 2022). B-values for brain imaging are up to 1000 s/mm^2 and signal decay may be approximated by a normal distribution. Higher b-values correlate with more restricted diffusion employed in non-Gaussian, microstructural models (Henriksen *et al*, 2022).

ADC is the most frequently adopted diffusion parameter. By scaling the diffusion image with a non-weighted image, quantitative and qualitative maps are obtained (Henriksen *et al*, 2022). Supposing a pure Gaussian water diffusion and a log-linear signal decay with increasing b-values, S_b equals $S_0e^{-b \cdot ADC}$. Thus, ADC can be derived after acquisition of both a diffusion weighted (S_b) and an unweighted (S_0) image (Henriksen *et al*, 2022). To improve sensitivity of S_b , the gradient is set along each of the 3 gradient axes; then the images are averaged acquiring four separate volumes (the unweighted S_0 and 3 S_b) (Henriksen *et al*, 2022).

The DWI ADC has been correlated with tumor cellularity, because the motion of extracellular water is aberrantly hindered by proliferating neoplastic cells which alter the healthy histoarchitecture (Riva *et al*, 2021). Also, it has been correlated with MGMT methylation and patients' survival (Romano *et al*, 2013; Moon *et al*, 2012; Sunwoo *et al*, 2013). In postoperative setting, DWI proves extremely useful in highlighting ischemic areas in the periphery of resection cavity (Riva *et al*, 2021).

DWI is easily handled by clinicians with good expertise in interpreting images. However, it is affected by geometric distortions at the periphery of the image, or due to air and hematic byproducts, possibly impairing diagnostic ability in brain areas adjacent to bone and surgical cavity (Henriksen *et al*, 2022).

3.2.2.2 Diffusion Tensor Imaging (DTI)

Diffusion imaging allows a quantitative assessment of microstructural changes in the brain. DTI is the most frequently used technique worldwide (Caverzasi *et al*, 2016). By sampling the diffusion signal in 6 directions, it allows for creation of a tensor, a multi-dimensional vector determining quantitative parameters such as mean diffusivity (MD) and diffusion fractional anisotropy (FA) (Basser *et al*, 1994). In particular, FA ranges from 0 to 1 to indicate the degree of asymmetrical diffusion in a voxel. Given this, FA maps can reconstruct white matter tracts' routes and show where tumoral invasion destroys their organization (Henriksen *et al*, 2022).

However, DTI metrics are not specific in differentiating pathological conditions inducing microstructural changes (Caverzasi *et al*, 2016). Moreover, the ground assumption in DTI is that water diffusion is strictly Gaussian (Niendorf *et al*, 1996), which holds true in a singular, homogeneous environment, but rarely stands in highly organized human tissue (Henriksen *et al*, 2022). Thus, the tensor model is increasingly proving to be flawed for modeling biological systems (Caverzasi *et al*, 2016). Nevertheless, DTI-derived maps successfully predicted GBM recurrence (Price *et al*, 2007).

Several software exist that utilize both deterministic and probabilistic algorithms and allow for fiber track reconstructions based on DTI data (Price *et al*, 2006). Notably, newer models based on high angular resolution DWI (HARDI) have recently been elaborated to increase tractography reliability (Riva *et al*, 2021; Sarkaria *et al*, 2018).

3.2.2.3 Neurite Orientation Dispersion and Density Imaging (NODDI)

Recently, an advanced, open source diffusion MRI model called NODDI (neurite orientation dispersion and density imaging) has allowed to outdo DTI limitations, still being compatible with routine clinical activity (Zhang *et al*, 2012).

NODDI quantitatively subdivides the total voxel diffusion into 3 distinct compartments. These partitions are characterized by isotropic, anisotropic Gaussian, and anisotropic non-Gaussian diffusion, which respectively approximate cerebrospinal-fluid-like, extraneurite and intraneurite compartments (Zhang *et al*, 2012).

The signal elaboration consists of two main steps: first, discerning the isotropic diffusion (fiso) component from the total diffusion signal and subsequently identifying the intraneurite fraction (ficv) component from the residual anisotropic diffusion. The remainder represents the extraneurite component (fecv). Also, the computation returns the coherence of fiber direction through the orientation dispersion index (odi) (Zhang *et al*, 2012).

In a fundamental work by Caverzasi *et al.*, the authors elaborated a color map to visualize the actual variety of lesions looking homogeneous on cMRI in a single map that summarizes the compartmentalization of diffusion signals. As a matter of fact, the NODDI color map allows to identify vasogenic edema, characterized by a stark fiso signal (blue), from tumor-infiltrated edema, depicted by a strong fecv signal (red). On the other hand, hindered diffusion, for instance due to acute ischemia, shows increased ficv signal (green) (Caverzasi *et al*, 2016).

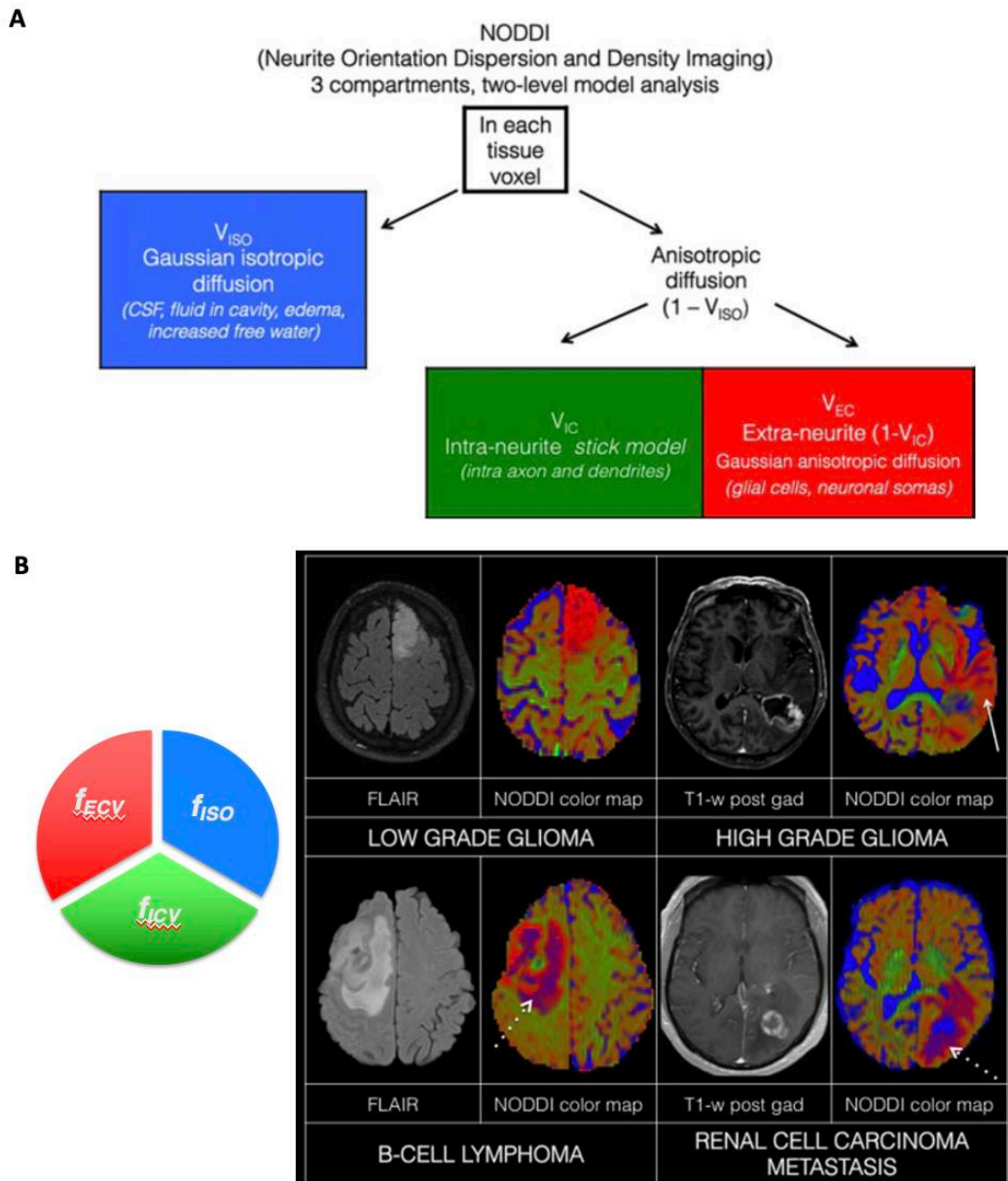


Figure 3.3. NODDI algorithm. *A*, diagram of isolation of single diffusion components per voxel. *B*, elaboration of NODDI color map for different brain lesions (red for fecv, green for ficv, blue for fiso). Post-contrast T1 sequences for each lesion are paired with corresponding NODDI color maps. Perinodular edema, which looks alike in high-grade glioma, B-cell lymphoma, and metastasis, appears different at NODDI elaboration, with higher fiso signal for the latter two (suggesting higher vasogenic edema) and higher fecv signal for high-grade glioma, suggesting more of an infiltrative nature of the edema. fecv/VEC: extra-neurite component of diffusion, ficv/VIC: intra-neurite component of diffusion, fiso/VISO: isotropic component of diffusion. Adapted from (Caverzasi et al, 2016)

3.2.2.4 Perfusion MRI (PWI)

PWI measures changes associated with tumoral neoangiogenesis. The dynamic susceptibility contrast (DSC)-derived relative cerebral blood volume (rCBV) is the most reliable feature. It reliably correlates with changes like increased microvessel density. Dynamic contrast-enhancement (DCE) is an additional PWI acquisition describing intravascular volume, microvessel permeability, and extravascular-extracellular space. It provides a multiparametric measure of microvascular texture. Remarkably, perfusion has been proven to correlate with fundamental molecular features, such as EGFRvIII positivity. Moreover, at follow-up, PWI accurately separate true viable tumor remnants from treatment-induced changes, with sensitivities and specificities of 90% and 88% for DSC, and 89% and 85% for DCE (Riva *et al*, 2021). Thus, it represents a valuable tool in the differential diagnosis between radionecrosis and GBM relapse.

3.2.2.5 MR Spectroscopy (MRS)

Hydrogen-MRS supplements MRI-derived anatomic and physiologic information by providing metabolite peaks that differ across pathologies. Besides more common, generic peaks (i.e., N-acetyl aspartate, choline), MRS also detects accumulation of 2-hydroxyglutarate (2HG), which is the hallmark of IDH-mutant gliomas. Though still technically difficult to implement, MRS based detection of 2HG peak showed outstanding performance, with sensitivity and specificity of 95% and 91%, respectively (Riva *et al*, 2021). Further pivotal implementations of 1- and 2-dimensional MRS sequences might be utilized to detect molecular subgroup-restricted metabolites identified by joint metabolomic and transcriptomic analyses.

3.2.3 Glioblastoma Imaging

Often GBMs are first investigated through CT imaging under emergency settings, due to new-onset neurological symptoms. General appearance at basal CT consists in a hypodense mass delimited by a rim which variably show iso- or hyperdense signal. It may present stigmata of fresh or dated hemorrhage, while calcifications seldom appear and suggest more of an oligodendrocytic diagnosis. Mass effect is quite common and is explained by a conjunct effect of the nodule itself and significant peritumoral edema. Upon contrast administration, strong, heterogeneous,

sometimes irregular ring enhancement is demonstrated. Hypertrophic vessels might be detected as linear enhancing streams in the proximity of the main mass (Osborn *et al*, 2017).

Diagnostic work-up is then completed with contrast-enhanced MRI, showing most frequently a poorly marginated mass with mixed signal intensity on basal T1 sequences, encased by a thick ring enhancement, which encloses a necrotic core. Sometimes further minute nodular or patchy enhancing foci are observed in the proximity of the main mass, signifying development of additional macroscopic tumor nodules into adjacent structures. T2/FLAIR sequences show heterogeneous hyperintensity characterized by faint margins and broad vasogenic edema. Other usual findings consist of necrotic or fluid-filled collections, with fluid/debris levels, hemorrhage at various evolutionary stages, and flow voids in case of deranged, extensive neoangiogenesis. Rarely, a proper dominant lesion cannot be detected, and the tumor extensively infiltrates through the white matter fibers (Osborn *et al*, 2017). Occasionally liquorol metastasization is demonstrated, with enhancing foci painting sulci and cisterns and adhering to cranial nerves, generating a picture strongly resembling that of pyogenic meningitis. Parallely, when ependymal or subependymal spread takes place, a picture reminiscent of pyogenic ventriculitis is demonstrated at MRI (Osborn *et al*, 2017).

As a rule, GBMs do not cause signal restriction on DWI sequences. At DTI, FA may be reduced in favor of a more isotropic, disorganized architecture due to disrupted white matter tracts directionality by tumor invasion. ADC inversely correlates with tumor cell density and shows an intermediate signal between lower-grade gliomas and lymphomas, suggesting progressively increasing local cellularity from lower-grade to higher-grade gliomas, to lymphoproliferative disease (Wen *et al*, 2020; Hayashida *et al*, 2006; Lu *et al*, 2019; Higano *et al*, 2006).

Microvascular proliferation due to neoangiogenesis and cooptation and hypertrophy of existing blood vessels is a hallmark of GBMs. Therefore, PWI-rCBV has been proposed to aid in differentiating GBMs from other tumor types and grades. As a matter of fact, measures of vessel density and permeability show increased values in GBMs (Wen *et al*, 2020; Kickingeder *et al*, 2014; Lee *et al*, 2018a; Suh *et al*, 2019; Law *et al*, 2004). MRS typically shows significantly higher choline peak due to augmented cell turnover, lower N-acetyl aspartate due to neuronal loss, and a lipid/lactate peak. These findings are sensitive but not specific, as analogous alterations can be detected in the setting of other tumors or inflammatory diseases (Wen *et al*, 2020; Gharzeddine *et al*, 2019).

Different from GBMs, gliosarcomas tend to appear as peripheral, heterogeneously enhancing solid masses, usually inducing marked vesogenic edema. Frequently they impinge the meninges but lack dural attachment or invasion. In case of deep-seated gliosarcomas, the radiologic findings are perfectly superimposable to those seen in other GBMs (Osborn *et al*, 2017).

The main differential diagnoses of GBMs are with IDH-mutant WHO grade 3 and 4 astrocytomas. These astrocytomas affect predominantly the frontal lobes and have significant representations of non-enhancing areas, generally lacking the classic enhancing rim around the necrotic core, though exceptions are common. MRS may reveal the scar of IDH1/2 mutation, i.e., a significant accumulation of 2HG, signaled by peak at 2.25 ppm (Osborn *et al*, 2017). Other major neoplasms in differential diagnosis are metastases. They are more frequently multiple and locate at the interface between gray and white matter. They can reach very large dimensions but usually maintain a more regular, round shape and do not infiltrate like GBMs. Notably, they cause extensive reactive edema, sometimes even more pronounced than what seen in GBMs. Finally, primary CNS lymphoma usually affects the corpus callosum and strongly enhances, often in a homogeneous manner as necrosis is seldom observed unless in HIV/AIDS patients (Osborn *et al*, 2017).

The principal non-tumoral lesion to differentiate from GBM is abscess. Abscesses classically present thinner, regular circumferential borders. The main aid in differentiating them from GBMs are DWI and MRS. In fact, infective collections restrict on DWI and display succinate and cytosolic amino acids peaks, which are uncommon in GBMs, and lack other typical peaks of glial tumors (Osborn *et al*, 2017).

3.2.4 Radiomics

In addition to qualitative assessment by neuroradiologists, MRI also provide a lot of quantitative information, which escape simple visual inspection and cannot be grasped by single measure assessment (Gillies *et al*, 2016). Advances in computational informatic techniques eased the extraction and elaboration of quantitative information on the texture and morphology of a given tumor. These factors led to the recent development of radiomics, an emerging neuroimaging branch that relies on identification of hand-crafted features from clinical images that are too complex for human eyes to be discerned (Chaddad *et al*, 2019; Lambin *et al*, 2017). Such radiomic features describe pixel-wise the image region containing GBMs and provide information regarding shape, regular or irregular tumor edges (Ismail *et*

al, 2018), intratumoral heterogeneity through gray levels (Prasanna *et al*, 2017; Rathore *et al*, 2018), as well as tumor-induced deformation of the surrounding normal brain structures (Beig *et al*, 2020). In neurooncology, radiomics has been recently applied as an emerging, non-invasive prognostic and predictive tool, also to distinguish treatment effects from tumor recurrence (Zhou *et al*, 2018; Kickingereder *et al*, 2019; Beig *et al*, 2020).

As a further subclass of radiomics, radiogenomics focuses on the relationships between radiomic phenotypes and gene expression and mutations, or other molecular profiles (Aftab *et al*, 2022; Verduin *et al*, 2018; Liu *et al*, 2021), without affecting the clinical workflow (Beig *et al*, 2020). It has already been reported in the literature to estimate GBM subgroup, mutation status and heterogeneity, and to predict progression, survival and response to targeted therapies (Verduin *et al*, 2018; Chaddad *et al*, 2019). Notably, by sampling the entire tumor image, radiogenomics offers more information as opposed to surgical biopsy which can offer only a limited view of spatial tumor heterogeneity (Chaddad *et al*, 2019), especially for recurrent GBMs (Aftab *et al*, 2022; Xi *et al*, 2018; Louis *et al*, 2016; Park *et al*, 2018).

Just as for other -omics analyses, radiogenomic studies are either exploratory or hypothesis driven. In the former, different features are tested against several genomic alterations to blindly scavenge significant associations. On the contrary, hypothesis-driven studies explore the relevant radiophenotypes that best fit an a-priori defined genetic alteration. Up to now, most studies have been designed to profile gene expression characterizing tumors discernible by radiomic features, thus being exploratory in their nature (Fathi Kazerooni *et al*, 2020).

The radiogenomic pipeline typically includes the following steps: 1. image acquisition and pre-processing (i.e., image registration, noise reduction, MRI field intensity/orientation normalization and corrections, and spatial resampling); 2. manual or semiautomatic segmentation of region of interest (ROI); 3. feature extraction; 4. feature selection through supervised or unsupervised methods, to avoid redundancy and overfitting; 5. model building and evaluation. Usually, a training dataset is used to learn the classifier parameters and a validation dataset is used to tune them. After training and validation, the final model should be tested on an unrelated, independent test dataset to evaluate the performance and show its value for clinical promotion and application (Fathi Kazerooni *et al*, 2020; Liu *et al*, 2021).

Preprocessing is a key step to correct for patient movement and image variations across different MRI machineries. The ROI is then segmented. Usually, T2/FLAIR scans are used to define edema and infiltrative volume, while postcontrast T1

sequences are adopted for necrosis and enhancing volumes (Beig *et al*, 2020). The 3 segmented tumor compartments are then superimposed and coregistered to other acquisitions, from which radiomic features pertaining only to the tumor subcompartments can be extracted. Further preprocessing includes skull stripping and signal intensity standardization and are used to homogenize magnetic field strengths and slice depths, making the studies comparable across machineries. (Beig *et al*, 2020)

Quantitative features extracted from ROIs are grouped into semantic, shape, texture/gradient-based, deformation, and wavelet features (Beig *et al*, 2020).

Semantic features describe GBM visual phenotype. They include lesion location, morphology, major and minor axis lengths, margins, and lesion vicinity. These features are known as Visually AcceSAbLe Rembrandt Images (VASARI) (Beig *et al*, 2020). The literature reports studies proving that GBM patients' OS is significantly correlated with VASARI features, providing further insight into prognosis in addition to clinical variables (Gutman *et al*, 2013). Also, several radiogenomic studies have demonstrated a correlation between GBM genetic landscape and the VASARI features (Colen *et al*, 2014; Jamshidi *et al*, 2014; Diehn *et al*, 2008).

Shape features describe qualities, such as irregular tumor infiltration, which affects tumor surface and its relationship with surrounding areas. A greater surface/volume ratio denotes a more spiculated tumor, potentially more malignant than a round mass with a smaller ratio. Shape features are divided into local and global. Local features include curvature and its sharpness, as well as shape (Beig *et al*, 2020). Global features comprise major and minor axes, and the ratio between them. Several studies demonstrated shape-based features efficacy in predicting OS in GBMs (Rathore *et al*, 2018; Chaddad *et al*, 2016; Henker *et al*, 2017; Sanghani *et al*, 2019; Czarnek *et al*, 2017). For instance, highly irregular enhancement heralds a lower OS (Gevaert *et al*, 2014).

The most common texture/gradient-based features include gray-level co-occurrence matrix (GLCM), capturing the variations in gray levels via second-order intensity statistics (e.g., entropy, angular second moment, contrast, and differential entropy); gray-level run length matrix (GLRLM), analyzing pixel runs instead of pairs of pixels; laws features, defining texture parameters (e.g., spot, edge, ripple, and level surfaces), and co-occurrence of local anisotropic gradient orientations (CoLIAGe), capturing confined anisotropic disparities within a restricted area. These features proved to predict survival in neurooncology (Prasanna *et al*, 2017; Beig *et al*, 2018; Kickingereder *et al*, 2016; Bae *et al*, 2018; Beig *et al*, 2020).

Deformation-based features quantify brain parenchyma deformation due to the presence of tumor. MRI scans are non-rigidly registered to equivalent healthy imaging atlases. Then, forward and inverse mappings allow to compute a deformation field and the deformation measurements are used for radiomic analysis. The literature reports how high deformation affecting language, visual, cognitive, and motor-control areas, particularly in the memory areas, are associated with decreased OS (Beig *et al*, 2020).

Wavelet-based features use different wavelengths, amplitudes, and frequencies to describe the lesion texture. Highly negative skewness of such features portends a meaningful increase in median OS in GBM patients (Tixier *et al*, 2019).

Since there is a plethora of radiomic features, computational costs of predictive models including even just a major part of them would be unjustifiable high. Therefore, to build classifier models, the radiomic features need to be selected to include only the most discriminative ones. Feature selection can be carried out either by univariate or multivariate statistics. Univariate methods (filter methods) depend solely on feature associations to the outcome of interest, ignoring feature redundancy in explaining overlapping pieces of information, while multivariate methods (wrapper methods) examine codependency of different features, accounting for association to outcome and redundancy. Among the filter methods, the most frequently used are Fisher, Chi-squared, and Wilcoxon tests. Wrapper methods include forward or backward feature selection, exhaustive greedy algorithms, or bidirectional search. However, wrapper models are computationally expensive and therefore less frequently used than filter methods (Beig *et al*, 2020; Lohmann *et al*, 2021).

Prediction models are divided into supervised and unsupervised approaches. The former use a-priori defined sets of labels to select features best fitting the outcome to be predicted. Then, the model is built on the selected features and their thresholds tuned on training/validation sets. On the contrary, unsupervised approaches, such as clustering, are utilized when labels are unknown. Labels are defined solely on the Euclidean, Manhattan, Minkowski, Hamming, or Cosine distances between samples' features and the so-obtained classification is correlated to molecular or clinical outcomes. Many classifiers have been adopted to model prediction algorithms, like random forest, support vector machines, or generalized linear models (Ismail *et al*, 2018; Beig *et al*, 2018; Lohmann *et al*, 2021).

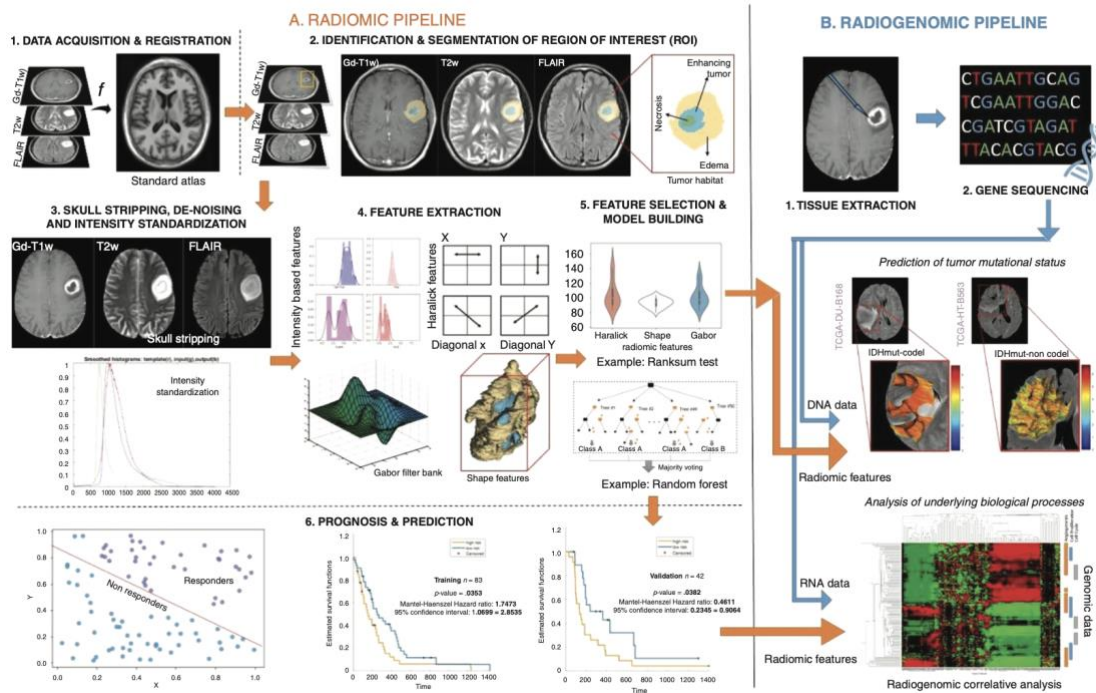


Figure 3.4. Radiomic and radiogenomic pipeline. Conventional or advanced MR image datasets are acquired and coregistered, then signal from parenchyma is stripped from skull signal and standardized. Radiomic feature are extracted and utilized to correlate with clinical or survival data in a pure radiomic approach, or with genomic/transcriptomic data in a radiogenomic approach. From (Beig et al, 2020)

3.3 Glioma Stem Cells (GSCs)

As extensively reported in the literature, tumors are not uniform masses of identical cells; rather, they are complex networks including various cellular populations and supportive stroma (Gimple *et al*, 2019). Many cancers rely on developmental and differentiation programs typical of normal tissue stem cells. Often neoplastic cells show a hierarchical organization in many neoplasms (Shibue & Weinberg, 2017; Suvà & Tirosh, 2020). In fact, according to the hierarchical cancer stem cell hypothesis, tumors behave like normal healthy tissues, where aberrantly differentiated cell subpopulations are dynamically regulated and derive from rare stem-like cells. These cancer stem cells show self-renewing capacity, tumor-propagating potential, and share markers of stemness (Gimple *et al*, 2019). As such, cancer stem cells (CSCs) are a small, apical population from which terminally differentiated cells arise to make up the bulk of the neoplasia (Kreso & Dick, 2014).

On the other hand, according to the stochastic model, tumor cells are in fact heterogeneous but virtually all of them can function as tumor-initiating cells in the proper setting (Vescovi *et al*, 2006). Notably, different cancers can behave differently, and this last model appears to best fit the gliomagenesis process, where each GBM cell can dynamically assume a stem cell profile. Anyhow, independent of the model, these cells, with the ability to regenerate the functional diversity present within the original tumor, have been identified in many neoplasms and labelled CSCs.

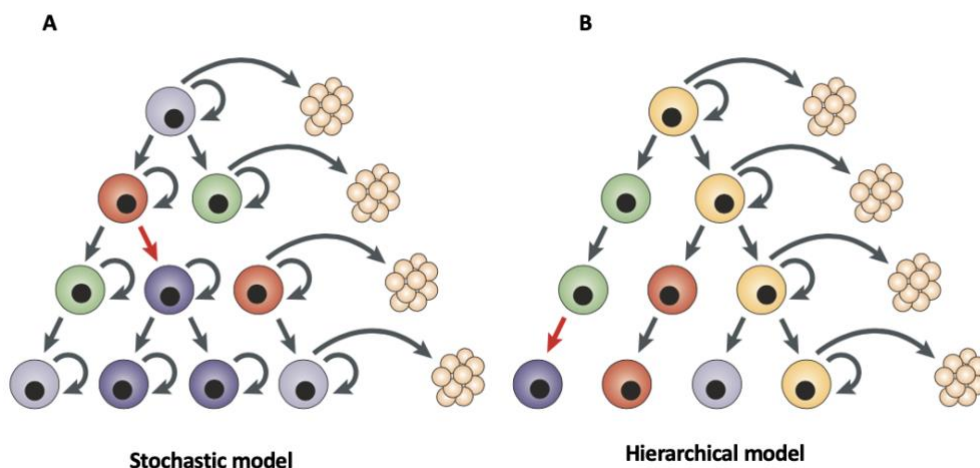


Figure 3.5. Stochastic and hierarchical model for cancer stem cells. A, the stochastic model suggests that cancer cells are heterogeneous, but all of them may act as a tumor-founding cell, although only rarely. B, the hierarchical model implies that only a small subpopulation of cancer stem cells extensively proliferates and sustain tumor growth and progression in the long run. Adapted from (Vescovi *et al*, 2006)

The initial steps of CSC theory date back to early functional studies demonstrating that injection of a single leukemic cell into rodents would regenerate the whole tumor. This theory was later corroborated by demonstration that a subset of leukemic cells could home to the bone marrow in lab rodents, where they started to proliferate and differentiate to recapitulate the initial tumoral phenotype (Gimple *et al*, 2019). Most notably, CSCs showed higher resistance to anticancer treatments than their more differentiated progeny, coherently implicating them in recurrences (Bao *et al*, 2006a; Chen *et al*, 2012). Also, the literature reports that transcriptional signatures of CSCs predict overall outcome, supporting their clinical relevance (Shibue & Weinberg, 2017; Suvà & Tirosh, 2020).

Following precursor studies in hematologic malignancies, a hunt for CSCs took place in other neoplasms, including brain cancer (Hemmati *et al*, 2003; Singh *et al*, 2004). Consequently, glioma stem cells (GSCs) were detected in IDHwt GBMs through expression of surface proteins such as CD133, CD44, SSEA1, L1CAM, CD49f, A2B5, PDGFRA, and EGFR (Anido *et al*, 2010; Lathia *et al*, 2010; Piccirillo *et al*, 2006; Singh *et al*, 2004; Son *et al*, 2009; Gimple *et al*, 2019; Mazzoleni *et al*, 2010). Like for other CSCs, GSCs have been described to long-term persist through self-renewal and to give rise to full blown tumors in animal recipients (Gimple *et al*, 2019; Lathia *et al*, 2015; Singh *et al*, 2004; Lee *et al*, 2006). Evidence in the literature supports their role in radioresistance (Bao *et al*, 2006a), chemoresistance (Liu *et al*, 2006; Chen *et al*, 2012), angiogenesis (Bao *et al*, 2006b; Cheng *et al*, 2013), local invasion of white matter tracts (Wakimoto *et al*, 2009), and recurrence (Chen *et al*, 2012). In addition, GSCs have been reported to differentiate, generally aberrantly, into multiple cellular lineages of the central nervous system (Suvà & Tirosh, 2020).

As discussed previously, GBMs consist of four states: 3 related to neurodevelopmental processes (NPC-like, OPC-like, and AC-like), and a fourth mesenchymal (MES-like) totally disconnected from neurodevelopment. These states are differently enriched in diverse GSC subpopulations (Neftel *et al*, 2019). Interestingly, most GSC surface markers are significantly associated with one of the aforementioned states, though not exclusively. As a matter of fact, CD24 is highest in NPC-like cells, CD133 in OPC-like cells, EGFR in AC-like cells, and CD44 in MES-like cells (Suvà & Tirosh, 2020).

Multiplicity of GSC subpopulations also explains GBM state plasticity. In fact, upon transplantation in animal models, cells with a specific profile do not generate xenocopies of only that very state, but re-establish the entire spectrum of cellular states spotted in the original lesion (Neftel *et al*, 2019; Mazzoleni *et al*, 2010). Thus, GSCs can switch from state to state, as pointed out by lineage-tracing experiments

(Neftel *et al*, 2019). Given this, the cellular states highest represented in a single tumor define the predominant signal detected at bulk RNA Sequencing. This in turn determines the main transcriptional affiliation of the GBM, being it PN, CL or MES (Suvà & Tirosh, 2020).

It must be noted that it is not yet proven whether GSCs are the actual GBM cells of origin, arising through mutations in normal neural stem cells. In fact, there is still much debate in the literature, with only a few reports suggesting that GBMs actually derive from normal stem cells populating the subventricular zone (SVZ) or subgranular zone. The SVZ raised particular attention due to the similarities between GSCs and neural stem cells normally residing in the vicinity of cerebral ventricles (Lee *et al*, 2018b). Transcriptional profile of immature outer SVZ radial glial cells outstandingly overlaps with that of GBM cells, suggesting them as possible early glioma precursors (Pollen *et al*, 2015; Lee *et al*, 2018b). Also, studies demonstrated the presence of GBM cells in normal SVZ, by detecting 5-ALA positive cells in its context (Piccirillo *et al*, 2015). Experimental evidence proved that induction of mutant IDH1 in adult murine SVZ causes hyperproliferation of progenitor cells and formation of aberrant nodules, as if recapitulating initial stages of gliomagenesis (Bardella *et al*, 2016). Additionally, knock-out of NF1, TP53, and PTEN in murine healthy precursors generated GBM-like lesions *in vivo*. Most notably, mutations in PTEN are able to transform neural stem cells into neoplastic counterparts but are not effective in doing the same in mesenchymal stem cells (Duan *et al*, 2015). Therefore, different neural progenitors harbor diverging molecular profiles which poise specific populations for tumorigenesis, and generate tumors with distinct properties (Gimple *et al*, 2019).

Exomic and transcriptional studies on GBMs and their matched normal SVZ showed common mutations in the tumor and SVZ in about 50% of cases. Still, most genetic and copy number aberrations were restricted to tumors. Clones from SVZ with shared mutations lacked the tumor-private aberrations (Gimple *et al*, 2019). This may suggest an initial event in a progenitor cell of SVZ or subgranular zone and subsequent development of full-blown GBM in distant regions after initiated cells migration. However, initiation of GBM from a differentiated cell and possible subsequent migration toward the SVZ due to a particular tropism after reacquisition of stem-like properties cannot be excluded.

Remarkably, recent *in vitro* and *in vivo* data showed that CD133-high cells display increased migratory and invasive potential when compared with matched CD133-low/negative cells. Coherently, infiltrating cells at the tumor margins express other stem cell markers such as L1CAM, nucleostemin, and nestin. Taken together, these

findings hint at a possible role for GSCs in GBM invasiveness (Ortensi *et al*, 2013). Therefore, both the tumor margins and the core lesion may contain GSCs that are characterized by different markers and tumorigenic potential, are enriched in different cellular states, and may deploy diverging molecular and metabolic programs. In light of this, the invasive edge could represent a new niche where GSCs switch from a stationary, proliferative phenotype to a migratory one (Ortensi *et al*, 2013).

3.3.1 Neurosphere Assay

One of the most commonly used technique to isolate and enrich GSCs *in vitro* is based on the assay used to isolate normal neural stem cells. The neurosphere assay (NSA) relies on self-renewal ability to select and enrich GSCs by growing them in nonadherent conditions, while progressively diluting and eliminating transiently amplifying progenitors and differentiated cells (Gimple *et al*, 2019).

Several studies applied this technique to GBM samples (Galli *et al*, 2004; Singh *et al*, 2003; Hemmati *et al*, 2003). In the work by Galli *et al.*, the authors applied NSA to a variety of human brain tumors, ranging from lower-grade gliomas to GBMs and medulloblastomas. After dissociation, cells were plated at clonal low-density in serum-free medium and exposed to mitogens, providing a stringent environment which specifically enriched for stem cells to proliferate exponentially (Galli *et al*, 2004). Spheroids reminiscent of healthy neurospheres were isolated from all GBMs with a clonal frequency of 0.5-31% of the total cells. Conversely, no clonal spheres was established from either grade 1 or grade 2 gliomas (Galli *et al*, 2004).

A fundamental aspect regarding NSA is that it fails in the short term to specifically enrich for proper stem cells. In fact, clone generation from disaggregated primary cancers provides only a surrogate index of their clonogenicity. However, even transiently amplifying precursors may yield neurospheres that are propagated for a limited number of passages. Therefore, only long-term propagation of primary cultures under NSA conditions safely detects and enrich for proper stem cells. Also, spheroids mimicking neurospheres may result from cell aggregation following excessive density. Besides the evidence of long-term proliferation, to further confirm the identity of these cultures as true GSCs, the authors assessed their capacity of self-renewal, multipotency and tumorigenicity (Galli *et al*, 2004).

Specifically, cultured lines were expanded well beyond 80 passages, even though they could be considered soundly established at earlier passages, with a mean

doubling time of 3-4 days. Like normal human fetal neural stem cells and glioma cell lines (GCLs), expansion of these lines followed an exponential kinetics. Notably, their proliferative activity increased at increasing culturing stages, suggesting a progressive enrichment in cells endowed with aberrant proliferative advantage, or their adaptation to culture conditions (Galli *et al*, 2004).

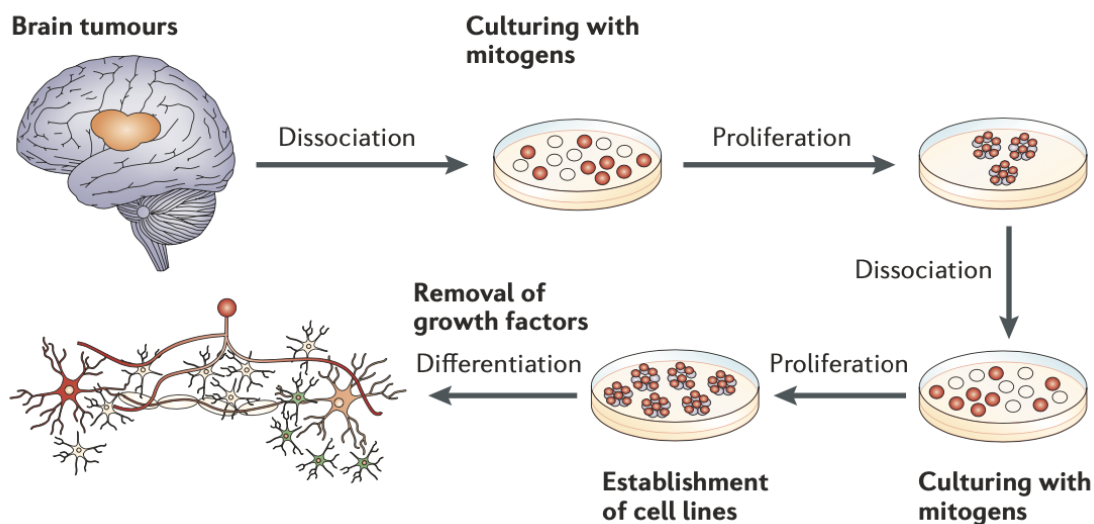


Figure 3.6. Neurosphere Assay (NSA) to isolate glioma stem cells (GSCs). The neurosphere assay is a serum-free culture system for isolation and propagation of neural stem cells or GSCs. Tumor cells are dissociated and plated in the presence of mitogens (EGF and bFGF2). In the absence of serum and at low density, most cells die, except those dividing because of mitogen stimuli. These cells proliferate to form clusters named neurospheres. Neurospheres can be subsequently dissociated and re-plated to generate daughter neurospheres, resulting in geometric expansion. Following mitogen removal, cells can differentiate into glial or neuronal cells, or aberrant hybrids. Under these conditions GSCs can be expanded indefinitely. From (Vescovi *et al*, 2006)

To determine the differentiative ability of GBM-derived cell lines, they were terminally differentiated upon mitogen withdrawal. Like normal neural stem cells, these clonal lines differentiated either into astrocytic progeny (as indicated by GFAP-positivity) or into neuron-like cells, expressing beta-tubulin, MAP5, MAP2, NF20, glutamate, and GABA, or even into GalC-immunoreactive oligodendrocyte-like cells. Multipotency was observed throughout culturing stages, from early to late passages. Notably, upon differentiation of the cultures, a few cells colabeled promiscuously with

neuronal and glial markers, indicating aberrant differentiation, typical of tumor and not of normal stem cells (Galli *et al*, 2004).

Subcloning experiments assessed self-renewal capacity: as a matter of fact, all the lines generated multiple secondary clones. Notably, lines growing slower had a significantly lower frequency of symmetric divisions than those growing faster, suggesting a constitutional difference in the self-renewal capacity (Galli *et al*, 2004).

Lastly, upon transplantation into immunosuppressed animal recipients, all the established lines generated tumors, with a take efficiency ranging from 50% for subcutaneous to up to 100% for orthotopic injections. Remarkably, orthotopic models showed high proliferation and marked infiltration, distinctive of human glial lesions. All these features were also confirmed after serial orthotopic transplantations (Galli *et al*, 2004).

In a comparison between the GSC propagation *in vitro* through NSA and *in vivo* through serial direct patient-derived xenografts, Richichi *et al*. demonstrated that 86% of the specimens was tumorigenic *in vivo*, and 54% also generated neurospheres under NSA conditions, while the remaining 46% did not. Thus, the lack of neurosphere generation under NSA does not necessarily imply a tumorigenic inability. By transplanting cells in a limiting dilution setting, the same group demonstrated that GSCs are rare and show highly variable frequencies between patients (Richichi *et al*, 2016). However, GSC content was 10-fold higher in NSA-propagated neurospheres than in freshly dissociated GBMs (Richichi *et al*, 2016). Nonetheless, GSCs obtained from either setting demonstrated the same tumorigenic potential at *in vivo* limiting dilution assay. Interestingly, when stratifying xenografts according to their GSC content, fewer was the amount of intracerebrally engrafted GSCs, greater was the mice survival, a finding that showed a parallel trend, though not significant, for PFS and GSC quantification in human lesions (Richichi *et al*, 2016).

Therefore, NSA has the advantage to high throughput functionally identify bona fide GSCs and make available long-term expanding GSC lines. However, caveats regarding this method include the inability to identify quiescent stem populations, and model cell-microenvironment interactions. Also, relying on unphysiological conditions, it induces a progressive loss of tumor heterogeneity since the very initial steps of thriving cell selection. Moreover, it fails to determine tumor formation capacity, even though such limitation can be overcome with *in vivo* xenotransplantation assays (Lee *et al*, 2006; Wan *et al*, 2010; Pastrana *et al*, 2011).

Leveraging all these findings, NSA stands as a valid method to isolate cell lines from human GBMs that properly satisfy all of the criteria deemed necessary to define GSCs (Galli *et al*, 2004). As a consequence, subsequent works from our group relied

on NSA to generate a collection of GSC lines utilized to better discern GBM pathophysiology and identify subgroup-specific theranostic candidates for the disease (Mazzoleni *et al*, 2010; Narayanan *et al*, 2019; Gagliardi *et al*, 2020; Pieri *et al*, 2022).

3.3.2 Culture media and supplements

As reported in the literature, GSCs isolated by means of NSA more faithfully replicate GBM phenotype than serum-cultured GCLs (Lee *et al*, 2006). However, a plethora of culturing media exists, possibly leading to discrepancies in the results between laboratories, as the lack of standardized media may affect experimental outputs (Zhang *et al*, 2020).

In serum-free media serum is replaced by selected hormones, promoting growth. In particular, the medium-hormone mix recipe is one of the most widely used. It consists of basal medium added with hormones and salts. The basal medium comprises Dulbecco's modified Eagle's medium (DMEM) and Ham's F-12-based medium in equal parts. DMEM has a high concentration of amino acids, and Ham's F-12 contains high amounts of trace elements. The DMEM/F-12 mixture is supplemented with 0.6% glucose, 2 mM glutamine, 3 mM sodium bicarbonate, and 5mM N-2-hydroxyethylpiperazine-N-2 ethane sulfonic acid (HEPES), to which the hormone and salt mixture is then added. Further supplementation with growth factors such as bFGF, EGF, leukemia inhibitory factor, N-acetylcysteine, and neuronal survival factor-1 has been widely used to isolate GSCs (Singh *et al*, 2004; Bao *et al*, 2006a, 2006b; Piccirillo *et al*, 2006).

Alternatively, Neurobasal medium is broadly used. It replicates DMEM/F-12, except for decreased sodium chloride to 3.0 g/liter, cysteine to 10 pM, and glutamine to 0.5 mM. The neurobasal medium is generally adopted with addition of supplements such as N2 and/or B27 and growth factors (Lee *et al*, 2006).

Three main supplements are used in GSC culture: N2 contains 5 µg/ml insulin, 100 µg/ml transferrin, 20 nM progesterone, 30 nM Na selenite, and 100 µM putrescine dihydrochloride (Zhang *et al*, 2020). B27 supplement contains 20 factors adjusted for insulin, transferrin, progesterone, putrescine, and selenium, in addition to T3, fatty acids, alpha tocopherol, and additional antioxidants. However, retinyl acetate and T3 may induce progenitor cells to differentiate. Consistently, B27 without such components slows differentiation and is more frequently adopted than its original version (Zhang *et al*, 2020). Lastly, hormone and salt mixture (hormone mix) is a

modification of N2 with 25 µg/ml insulin, 100 µg/ml transferrin, 20 nM progesterone, 60 µM putrescine, and 30 nM selenium (Zhang *et al*, 2020).

As previously stated, these media are further supplemented with growth factors. In fact, EGF and bFGF serve as mitogens essential to isolate neural and glioma stem cells (Singh *et al*, 2004; Gritti *et al*, 1996; Reynolds *et al*, 1992). In particular, EGF is fundamental for neuronal development and for propagating GSCs in suspension, and bFGF contributes to neuron production, enhances precursor proliferation, and promotes the long-term growth of GSCs (Svendsen *et al*, 1998; Reynolds *et al*, 1992). However, both bFGF receptor binding affinity and its subsequent mitogenic activity depend on heparin. Hence, either heparin or heparan sulphate is generally co-supplemented (Zhang *et al*, 2020).

Leukemia inhibitory factor is an interleukin 6-related cytokine that inhibits embryonic stem cells from differentiation and regulates GSC biology. Recently, Neuronal survival factor-1 and N-acetylcysteine were adopted as supplements to generate GSCs. Neuronal survival factor-1 endorses neural stem cells survival. N-acetylcysteine prevents oxidative DNA damage by glutathione synthesis (Zhang *et al*, 2020).

Notably, each research group uses different combinations and concentrations of growth factors. For instance, Dirks' and Rich's mix is a combination of EGF 20 ng/ml, bFGF 20 ng/ml, and leukemia inhibitory factor 10 ng/ml. Weiss' group uses EGF 20 ng/ml and bFGF 20 ng/ml with heparan sulfate 2 µg/ml, while Fine's mix consists of only EGF 50 ng/ml and bFGF 50 ng/ml. Remarkably, a recent study demonstrated a differential fate for neural stem cells cultured in the presence of low (0.1 ng/ml) or high (1-10 ng/ml) concentrations of bFGF, in the former case shifting towards neurons and in the latter towards both neurons and oligodendrocytes. Hence, varying combinations and concentrations of growth factors could produce discrepancies among various groups' results (Zhang *et al*, 2020). This further proves the necessity to standardize medium recipes to introduce fewer possible confounding factors in experimental settings.

Lastly, antibiotics are added to media at standard concentrations to prevent microbial contaminants. Generally, antibiotics do not prove toxic for GSCs, but still might affect cell function (Zhang *et al*, 2020).

3.3.3 Other strategies to isolate glioma stem cells

Beside NSA, other strategies to isolate GSCs may involve specific cellular markers or growing cells under conditions obtaining the same effects as NSA. Strategies relying on markers depend on their accuracy in selecting stem-like populations and rely on proteins used to identify normal neural stem cells, such as nuclear and cytosolic proteins (SOX2, OLIG2, MYC, and NESTIN) or surface markers (CD133, L1CAM, CD44, and A2B5) (Gimple *et al*, 2019; Brescia *et al*, 2013, 2012). Though prospectively identifying putative GSCs, this technique does not test any functional parameter. Also, the very limited set of markers may be too selective, thus missing some GSC subpopulations and depleting heterogeneity. Additionally, surface markers could be disrupted at cell dissociation steps (Singh *et al*, 2004; Son *et al*, 2009; Wan *et al*, 2010).

An alternative to NSA is adhesive culture onto poly-L-lysine/laminin-coated plates. As reported in the literature, this technique reduced cellular differentiation compared with floating neurosphere cultures (Lee *et al*, 2006). It shares with NSA the advantage of being high throughput. However, like NSA, it is limited by the inability to determine tumorigenicity, to identify quiescent stem populations, and to model interactions with the microenvironment. Besides, the artificial conditions to which cells are exposed might lead to losing tumor heterogeneity (Pollard *et al*, 2009).

Additionally, three-dimensional GSC cultures were established in the form of organoids. They recap cellular heterogeneity, hypoxic gradients, and to some extent *in vivo* tumor growth architecture (Hubert *et al*, 2016). Their main advantage is to model the *in vivo* settings with greater reliability than other *in vitro* methods. Organoids partially recapitulate tumor cells interactions with the microenvironment, still missing interactions with microglia or vasculature, but show a reduced throughput compared with NSA or adherent cultures. Moreover, the technical procedure to generate and maintain organoids is more complex than other *in vitro* methodologies (Hubert *et al*, 2016; Ogawa *et al*, 2018; Gimple *et al*, 2019).

A last technique consists in patient-derived xenografts (PDXs), which allows to simultaneously isolate GSCs and generate tumor models mimicking cellular interactions in a somehow physiologic setting. They differ from GSC-derived xenografts in that tumor cells are directly transplanted upon dissociation from specimens to the mouse striatum and propagated from primary PDXs to advanced order PDXs with no intervening *in vitro* passages. The major shortcomings of this technique are that it is very expensive by requiring a sufficient number of animals to propagate the line, and it is labor-intensive, while still providing neither mechanistic insights of GBM initiation nor interactions with adaptive immune system. Additionally,

sequential passaging from PDX to PDX may select away heterogeneity and induce private genomic aberration not shared with original tumor, just like with less expensive and less laborious *in vitro* methods (Singh *et al*, 2004; Hidalgo *et al*, 2014; Ben-David *et al*, 2017; Gimple *et al*, 2019).

3.3.4 Glioma stem cells as a tool to investigate and model GBM

A valuable tool in tumor research is the use of cancer stem cell cultures, which already allowed important steps forward in understanding other cancers. For instance, in the field of breast cancer, CSCs allowed the discovery of gene signatures predicting the risk of distant metastases, with a clear potential impact on clinical decision-making (Pece *et al*, 2019).

Parallely, GCLs and GSCs have been used to study GBMs. However, one major issue in this setting is the difficulty in effectively modeling this heterogenous disease. Therefore, the appropriate choice of the cell line to use, based on the features it replicates, is paramount since none of the available cell models reproduces the entire spectrum of GBM characteristics (Ledur *et al*, 2017).

The first studies with cell models adopted cell lines derived from induced rodent tumors of the central and peripheral nervous system. Subsequently, long-term cultures of GCLs were established from human GBMs. They provided an unlimited cell supply and reproducible results. The most common human GBM immortalized cell lines are U87MG, U252, T98G, and LN229. (Gómez-Oliva *et al*, 2021) They represent the fastest way to obtain preliminary results both *in vitro* and *in vivo*, with very rapid doubling time and short latency in generating full blown tumors in animal recipients. Still, GCLs not only fail to recreate a trustworthy microenvironment, but are also affected by decreased heterogeneity, private chromosomal aberrations and phenotypic alterations, and even higher proliferative potential than original tissues due to progressive selection by long-term serum culturing. Therefore, they are not a very reliable model, at least for experiments beyond preliminary investigations. Additionally, U87MG have been proven to have an unknown origin, with almost certain contamination by other non-glioma cell lines (Allen *et al*, 2016). Thus, experiments only relying on GCLs almost certainly fail to recapitulate the patients' disease (Gimple *et al*, 2019).

With the isolation of GSCs from patients, GBM primary cultures became the gold standard to model and investigate these tumors (Lee *et al*, 2006). They are more

challenging to maintain than GCLs, as they cannot be grown in serum but require complex media, as described previously (Lee *et al*, 2006).

Notably, there are claims that the useful experimental life of patient derived GSCs seems limited. In fact, after about 30 culturing passages, these cells accumulate genomic and transcriptional alterations in metabolic and signaling pathways which may resemble the same critical drift described for GCLs, rendering them too different from the parental GBMs (Baskaran *et al*, 2018).

The versatility of GSCs provides further modeling advantages. As a matter of fact, their ability to be propagated either as neurospheres or as adherent cultures, might combine benefits of the two culturing methods to study GBMs (Rahman *et al*, 2015). In fact, though both techniques share common pros and cons profiles, neurospheres perform better in preserving heterogeneity, while adherent cultures allow higher efficiency in generating GSC lines (close to 100%) from the primary tissue. Additionally, since GSCs are isolated from many individual patients, they may offer insights facilitating the study of inter-individual differences (Gómez-Oliva *et al*, 2021).

As a next step in modeling GBMs, *in vivo* transplantation in animal recipients allows to study neoplasm development and progression after engraftment of GSCs into immunodeficient models. The so-obtained xenografts may be either patient derived (PDX), in which dissociated tumor cells are directly implanted into mice without intervening *in vitro* culturing, or more frequently GSC derived, in which GSC line is first established and later transplanted in mice (Engebraaten *et al*, 1999; Gómez-Oliva *et al*, 2021). Notably, GSC derived engraftment is more efficient than PDX engraftment, with a more potent tumor induction still preserving a heterogeneous phenotype (Lee *et al*, 2018b). Mainly, three murine strains are utilized: nude (nu/nu) mice, which lack T cells; non-obese diabetic severe combined immunodeficiency (NOD-SCID) and SCID-beige mice, missing both T and B cells; and NOD-SCID IL2Rgamma-null (NSG/NOG) mice, completely devoid of T, B, and NK cells (Yoshida, 2020). Implantation is generally achieved subcutaneously (heterotopic) aiding the visual control over tumor formation, or in the brain (orthotopic), which provides a more physiological microenvironment (Shu *et al*, 2008). Compared to GCL xenografts, GSC derived tumors and PDX take much longer to grow but reproduce more heterogeneous lesions and preserve the original tumor histoarchitecture (Gómez-Oliva *et al*, 2021).

Many groups exploited combinations of *in vitro* and *in vivo* GSC-based models of GBMs to test new drugs or innovative drug combinations/formulations (Valtorta *et al*, 2017; Di Mascolo *et al*, 2021) and to identify novel candidates with pathogenetic and

theranostic implications (e.g., EGFR, MET, ASCL1, LSD1), some of which are already routinely evaluated (Faletti *et al*, 2021; De Bacco *et al*, 2016; Boccaccio & Comoglio, 2013; Narayanan *et al*, 2019; Mazzoleni *et al*, 2010; Park *et al*, 2017).

3.3.5 Controversies on glioma stem cells

Controversies on GSCs mostly derive from their vague and uncertain delineation. First, the term itself does not propose any assumption regarding the GBM cell of origin and are frequently used to define cell populations with the ability to long-term generate neurospheres *in vitro* and neoplasms *in vivo*. While GSCs may derive from the malignant conversion of a healthy resident stem cell, they may also result from differentiated tumoral cells reactivating stem-like developmental and survival programs (Gimple *et al*, 2019). Therefore, some authors suggest defining these cells as tumor initiating cells (TICs) or gliomasphere-forming cells (GSFCs), especially because, as previously reported, GBM stemness seems to be approximated more by the stochastic model than by a hierarchical one.

Second, GSCs are not unequivocally defined by surface markers. So far, the prognostic value of cancer stem cells has been connected in other neoplasms to well-defined features, such as markers of stemness. Similarly, these have been correlated with clinical outcomes in GBMs. Nevertheless, evidence is still controversial, as some authors reported that stem-cell markers do not invariably have a prognostic effect in GBMs (Richichi *et al*, 2016). Additionally, although some markers such as CD133, CD44, and CD15, have been reported to consistently enrich for GSCs, recent studies demonstrated that most are neither fully sensitive nor specific (Gimple *et al*, 2019). Moreover, varying marker combinations might identify cells in diverse states and with diverse genotypes. Consequently, different authors might indicate as GSCs entirely different cell populations (Suvà & Tirosh, 2020).

Third, the main GSC functional properties (i.e., self-renewal and tumorigenicity) are common to several other cancerous cells, and are tested in factitious *in vitro* settings and in xenogeneic animal models that do not perfectly replicate the native human milieu (Suvà & Tirosh, 2020). Additionally, the stem state is not static, but undergoes a plastic reorganization also by means of epigenetic changes, allowing for interconversion between distinct cellular states. Microenvironmental exposures, nutrient deprivation, hypoxia, or radiation all affect the dynamic balance between the GSC and non-GSC pools (Gimple *et al*, 2019). Thus, partially or completely conflicting results might be documented under the same experimental conditions when

heterogeneous cell populations in different states are indistinctly addressed as GSCs (Suvà & Tirosh, 2020).

Traditionally, GSCs have been associated to quiescence, implying a lower proliferation than other malignant cells and advocating the generally higher resistance of dormant cells to treatments that inherently target proliferating cells. On the contrary, recent evidence showed that GSCs could replicate more than other quiescent cell populations making up the tumor bulk (Suvà & Tirosh, 2020; Lathia *et al*, 2015).

Another criticism to the cancer stem cell theory for GBMs is that it does not address the frequency of GSCs within GBMs and diminishes the role of more differentiated progeny in maintaining complex tumor tissue systems. CSCs are relatively rare in most cancers, ranging from 0.0001% to 0.1% of the tumor bulk. However, recent studies report higher frequencies of 25-40% in GBMs (Richichi *et al*, 2016; Mazzoleni *et al*, 2010), consistent with a stochastic model, where each cell can assume a stem behavior. Various aspects may influence GSC frequency estimate (GSC isolation technique, *in vitro* cell manipulation, agents promoting cell engraftment, recipient immunodeficiency, and timelapses following GSCs injection). Additionally, the most reliable assay to determine GSC frequency is the limiting dilution cell transplantation. Yet, only rarely such experiments are performed (Richichi *et al*, 2016).

Lastly, clonal selection during *in vitro* culture for cell line propagation rapidly depletes cellular heterogeneity. This affects modeling tumor complexity and explains why some efficacious therapies in experimental settings, with homogenous tumors, fail in heterogenous human neoplasms (Gimple *et al*, 2019). Also, *in vitro* studies rely on hyperoxic, hyperglycemic, non-physiologic conditions, and are devoid of normal cell-cell interactions (Gimple *et al*, 2019). An additional reason for failures of clinical translations of drugs positively selected in preclinical studies is the fact that human brain microenvironment is not as effectively reproduced in rodents, limiting the study effectiveness. Furthermore, since these murine models have either a very weak or completely absent inflammatory and immune response, they are not suitable for testing immunomodulatory therapies (Gómez-Oliva *et al*, 2021).

4 AIM OF THE WORK

GBMs are highly malignant brain tumors, that rapidly grow and invade adjacent white matter tracts and are very resistant to standard treatment, despite multimodal therapies. Recent revisions of the pathology and progressive access to molecular information led the WHO to update the GBM classification by introducing molecular prognosticators. In this regard, accumulating evidence led to identification of molecular subtypes of GBMs (*i.e.*, proneural, classical and mesenchymal) that, although not yet included in the WHO classification, are worthy to be elucidated as for their clinical implications.

The identification of glioma stem cells (GSCs) which are not a restricted tumor component, but rather comprise most of the cells within the tumor bulk, with the ability to propagate the pathology, allowed for their exploitation in the modeling of GBMs and their molecular subtypes both *in vitro* and *in vivo*. However, primary cultures of patient-derived GSCs may suffer from the factitious conditions of the neurosphere assay (NSA) used to isolate and culture them, possibly inducing a drift as happened with the canonical high-passage glioma cell lines (GCLs), which nowadays appear inadequate. Nonetheless, xenografts derived from subgroup-specific GSC lines tend to show more homogenous subgroup-specific phenotypes than the original heterogenous human tumors, which may be an aid in identifying subgroup-restricted features.

Advancement in MRI technology and computational analysis allowed to develop new study sequences from which deriving radiomic features to relate with molecular information and, possibly, pathophysiological behavior of GBM subtypes.

Given this, the aim of my PhD project was to exploit new, advanced 3-compartment diffusion-weighted MRI sequences to study human GBMs of different subtypes (MES, the most aggressive and treatment-resistant, vs non-MES GBMs), in order to identify radiomics markers that might predict non-invasively the transcriptional subgroup of GBM.

GBM patients underwent advanced MRI and then were subjected to surgery. The tumors from these patients were processed for GSC isolation by NSA and were transplanted to generate murine xenografts to be studied with the same advanced

diffusion MRI protocol. The discriminative radiomic features between MES and non-MES tumors were selected by matching common significant ones between the human cohort and the more extremized xenograft cohort. A subsequent simplified model for non-invasive molecular diagnosis was elaborated on the selected features.

Parallely, aiming at understanding the dynamics underlying GSC molecular profile evolution observed under progressive serial culturing passages by the NSA, we subjected the original patient specimens and the early/intermediate GSC lines to Western blot analysis for protein markers of the different molecular subgroups. Then, IHC molecular diagnosis was performed on the original human tumor and on the matched GSC-derived xenografts, to assess whether and which modifications in the molecular affiliation of GBM were taking place during the establishment of GSC lines.

Finally, GSC lines were subjected to transcriptomic and whole exome sequencing analyses to identify possible new candidate genes responsible for the different biological behavior of the different GBM subtypes. To further substantiate GSCs as a reliable modeling tool for GBMs, transcriptomic data were utilized to derive gene signatures to determine subtype affiliation of unmatched human GBMs. As a further step in confirming the reliability of GSCs as GBM preclinical model, *in silico* transcriptional analyses were also carried out on publicly available datasets of MES and non-MES GBMs vs healthy brain tissue and compared to results obtained from GSCs.

Overall, all these different experimental tasks converged on the definition of a composite picture of the molecular features that GSCs are endowed with, thus helping us better understand their plasticity and their exploitability as valuable preclinical models for the different molecular subgroups of GBM, to be used also in the context of clinically relevant projects, e.g., radiogenomics.

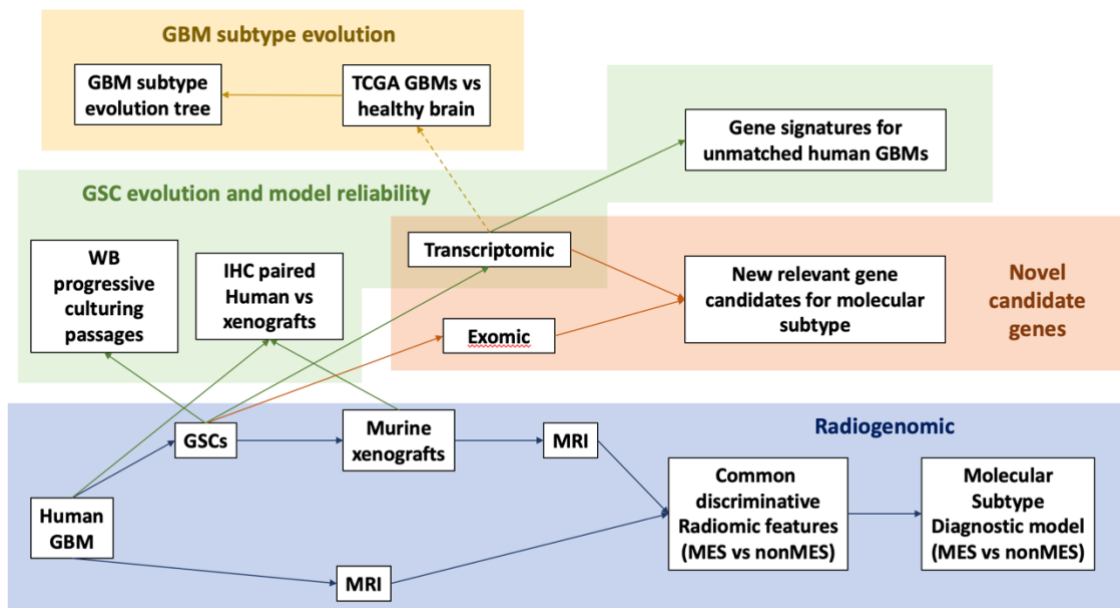


Figure 4.1. Schematic representation of study design. Global study design can be divided into 4 sub-projects (Radiogenomic, GSC evolution and model reliability, Novel candidate genes, and GBM subtype evolution) intertwined with each other as represented by the arrows depicting the project flow. Dashed arrow from "Transcriptomic" to "TCGA GBMs vs healthy brain" indicates that only the analysis and R scripts were transposed and applied to separate, independent data.

5 RESULTS

5.1 Identification of radiogenomic features discriminating GBM subtypes.

5.1.1 GBM patient population enrollment and subgroup affiliation.

Over the course of 5 years, we prospectively screened and enrolled 56 CNS WHO Grade 4 gliomas that presented to the Neurosurgery Department of San Raffaele Hospital. We first selected patients privileging those who were first diagnosed with high-grade glioma and had not yet undergone surgical resections and/or chemo- or radiotherapy treatment. However, as progressive analysis of tumors subgroup affiliation showed that MES samples were scarcely represented in our cohort, we also enrolled 5 recurrences, as it is reported in the literature a tendency to a MES shift in recurrent tumors (Bhat *et al*, 2013). We also privileged open surgeries, as biopsies would provide us with scant material for further processing.

At the end of patient enrollment, we recruited 16 females (28.6%) and 40 males (71.4%), with a mean age of 61.3 ± 11.3 years (median 61.0, range 35-80 years). Mean preoperative Karnofsky performance status (KPS) was 90.0 ± 13.6 (median 90, range 20-100). The lowest KPS was recorded in a patient who rapidly deteriorated neurologically, requiring emergent surgery. Twenty-three patients were diagnosed following clinically evident seizures, none of them due to hemorrhages; the remaining 33 patients were diagnosed following non-specific symptoms, such as intense headache, not responsive to medications, or following development of neurological deficits. In total, 41 out of 56 patients had preoperative neurological deficits, ranging from mild (e.g., seldom anomias or paraphasias, limited quadrantanopia) to more severe defects such as complete aphasia or severe hemiparesis.

Three patients had previous stereotactic biopsy for diagnosis (in all cases GBMs). Five patients had previous open surgeries with diagnosis of GBM in 2 cases, anaplastic astrocytoma in 2, and lower-grade glioma in 1. Out of these latter 5, 1 patient had no subsequent treatments, 1 underwent chemotherapy, and 3 patients both chemo- and radiotherapy after the former intervention and before the surgery for which we

enrolled them in our study. Table 5.1 reports the location of the tumors in these 56 patients. Twenty-eight lesions (50.0%) were right sided, 27 (48.2%) left sided, and 1 (1.8%) diffuse on both sides. Fifty-four patients underwent open surgery, with 40 gross-total resections of the enhancing nodule (74.1%), 12 subtotal resections (<10% residue, 22.2%), and 2 partial resections (>10% residue, 3.7%). Two patients underwent only stereotactic biopsy.

Location		%
Frontal	17	30.4
Fronto-parietal	2	3.6
Fronto-temporal	1	1.8
Fronto-temporo-insular	1	1.8
Temporal	18	32.1
Temporo-parietal	3	5.4
Temporo-occipital	1	1.8
Temporo-parietal-occipital	2	3.6
Parietal	3	5.4
Parieto-occipital	6	10.7
Occipital	1	1.8
Diffuse	1	1.8

Table 5.1. Location of 56 CNS grade 4 gliomas. Absolute and percentual (%) numbers.

Transcriptional subgroup affiliation of these tumors was determined by using an immunohistochemical panel developed by our collaborators, as described in the Method section (Figure 5.1). Based on this panel, we enrolled 5 (8.9%) MES WHO grade 4 gliomas, 19 (33.9%) CL, 20 (35.7%) PN, 8 (14.3%) mixed PN/CL, and 4 (7.2%) balanced mixes of all subtypes. Notably, PN and CL subgroups were equally represented in our cohort, while MES tumors were unexpectedly scarce.

In 2021 the WHO published an updated version of brain tumor classification, separating IDH1/2 mutant grade 4 tumors (now called grade 4 astrocytomas) from IDHwt GBMs (Louis *et al*, 2021). In this view, our cohort consisted of 50 IDHwt GBMs and 6 CNS WHO grade 4 astrocytomas. Consistently, 5/6 Grade 4 astrocytomas were PN and 1/6 CL, in line with the observation that most IDH mutant lesions are PN. Table 5.2 reports the percentual make up of each of these samples and the overall

affiliation. Out of these 56 tumors, we subjected 48 specimens to Glioma Stem Cell (GSC) derivation by the Neurosphere Assay (NSA) and established 14 GSC lines.

Sample	% PN	% CL	% MES	Affiliation	aMRI
GBM 160315	9.2	1.4	89.4	MES	Yes
AG4 160404	86.4	9.2	4.4	PN	No
GBM 160407	75.0	20.0	5.0	PN	Yes
GBM 160411	36.6	62.4	1.0	CL	Yes
GBM 160415	21.2	78.4	0.4	CL	No
GBM 160511	84.4	11.6	4.0	PN	No
AG4 160428	20.4	62.2	17.4	CL	Yes
GBM 160503	28.2	69.4	2.4	CL	Yes
GBM 160525	63.4	34.2	2.4	PN	Yes
GBM 160526	18.2	81.8	0.0	CL	No
GBM 160622	21.6	77.2	1.2	CL	No
GBM 160610	72.6	20.2	7.2	PN	No
GBM 160704 #1	21.8	77.8	0.4	CL	No
AG4 160704 #2	81.0	9.4	9.6	PN	No
GBM 160628	7.8	92.2	0.0	CL	No
GBM 160808	56.6	42.0	1.4	PN/CL	No
GBM 160817	38.6	31.2	30.2	MIX	No
GBM 161007	61.8	37.2	1.0	PN	No
GBM 161019	32.4	46.8	20.8	MIX	No
GBM 161103	13.8	26.0	60.2	MES	Yes
GBM 161116	85.8	7.0	7.2	PN	Yes
GBM 161205	53.0	44.6	2.4	PN/CL	Yes
GBM 161130	15.8	9.6	74.6	MES	No
GBM 161128	17.8	82.0	0.2	CL	Yes
AG4 170126 #1	64.6	27.2	8.2	PN	Yes
GBM 170126 #2	72.8	24.0	3.2	PN	No
GBM 170413	3.8	96.2	0.0	CL	Yes
GBM 170421	9.0	91.0	0.0	CL	Yes
GBM 170426	59.2	38.0	2.8	PN	Yes
GBM 171025	11.8	88.2	0.0	CL	Yes
GBM 171027	24.4	74.6	1.0	CL	Yes

GBM 171031	14.0	85.6	0.4	CL	Yes
GBM 171211	33.6	65.4	1.0	CL	Yes
GBM 180111	21.8	77.8	0.4	CL	Yes
GBM 180131	61.8	32.8	5.4	PN	No
GBM 180201	47.4	19.4	33.2	MIX	No
GBM 180510	55.4	39.4	5.2	PN/CL	Yes
GBM 191119	60.2	37.2	2.6	PN	Yes
GBM 190920	33.6	65.4	1.0	CL	Yes
GBM 180420	20.2	5.2	74.6	MES	Yes
GBM 180522	70.2	27.8	2.0	PN	Yes
GBM 180515	54.8	42.0	3.2	PN/CL	Yes
GBM 180507	46.0	26.2	27.8	MIX	Yes
AG4 180330	80.6	12.0	7.4	PN	Yes
GBM 190125	56.8	41.6	1.6	PN/CL	No
GBM 171214	51.0	43.2	5.8	PN/CL	Yes
AG4 180504	62.8	30.0	7.2	PN	Yes
GBM 180611	77.2	19.4	3.4	PN	Yes
GBM 190724	54.6	40.6	4.8	PN/CL	Yes
GBM 181224	80.4	17.6	2.0	PN	Yes
GBM 190222	67.0	27.8	5.2	PN	Yes
GBM 190207	34.8	64.0	1.2	CL	Yes
GBM 181012	43.4	30.2	26.4	MIX	Yes
GBM 190515	57.4	38.2	4.4	PN	Yes
GBM 201113	67.0	27.6	5.4	PN	Yes
GBM 190628	35.6	63.2	1.2	CL	Yes
GBM 181123	56.0	41.6	2.4	PN/CL	Yes
GBM 200310	11.4	20.4	68.2	MES	Yes

Table 5.2. Transcriptional subgroup affiliation of CNS WHO Grade 4 gliomas. Percentual representation of each component is reported for each sample, as well as main affiliation according to protocol established by our collaborators in the Pathology department of University of Brescia (Orzan et al, 2020). Samples in bold are those with IDH1 mutation and thus classified as CNS Grade 4 Astrocytomas by the new WHO brain tumor classification and were therefore eliminated from MRI analysis. The last column indicates samples for which also advanced MRI studies were available for further radiogenomic analyses. AG4: grade 4 astrocytoma, aMRI: advanced MRI sequences, CL: classical, GBM: glioblastoma, MES: mesenchymal, MIX: balanced mix of the three subgroups, PN: proneural.

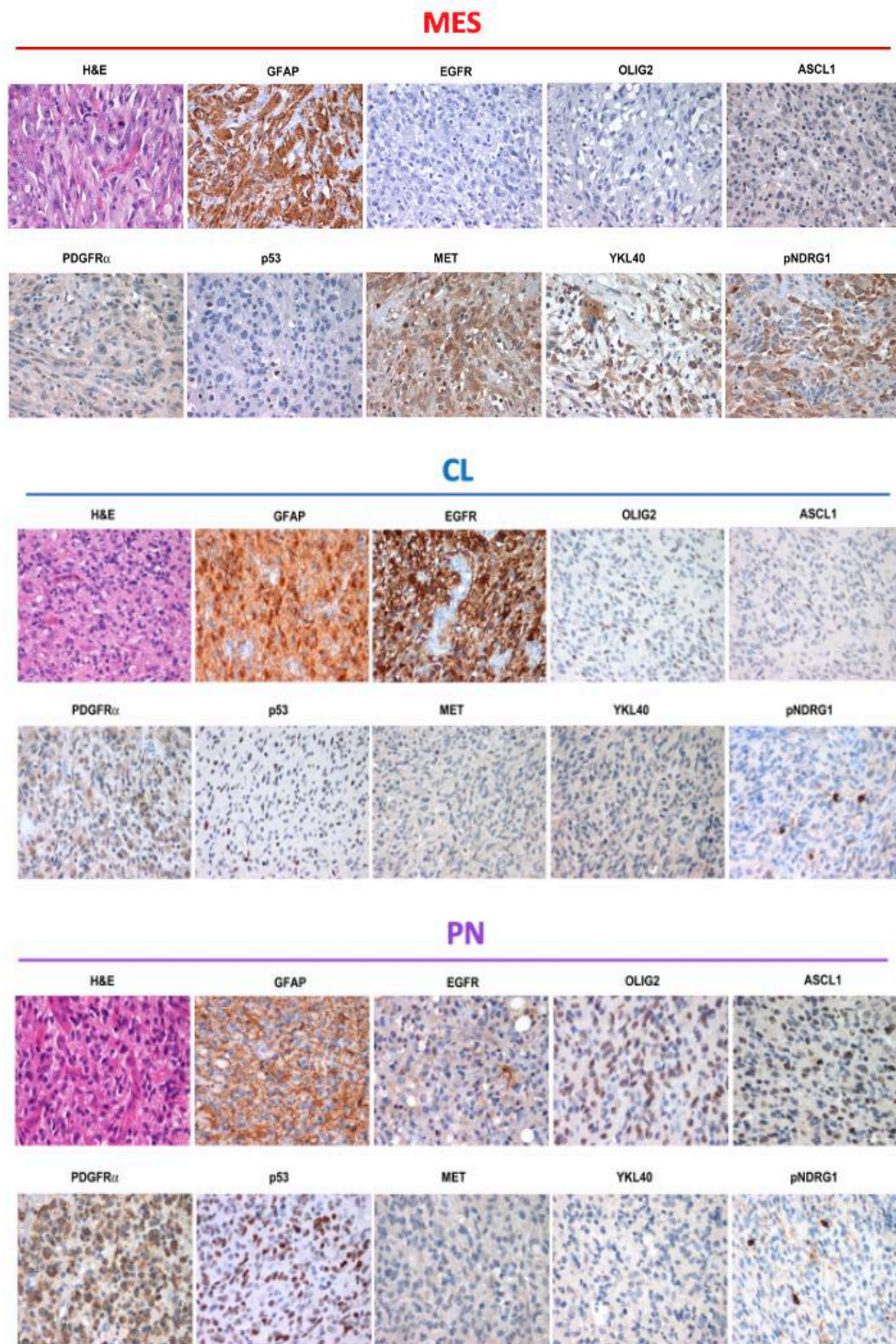


Figure 5.1. Transcriptional subgroup affiliation of GBMs based on immunohistochemical panel tested on a representative MES, CL, and PN tumor. Each panel reports a hematoxylin and eosin (H&E) staining of the specimen and IHC for GFAP, which shows consistently strong signal in all the subgroups. Signal intensity of the staining for EGFR, Olig2, ASCL1, PDGFR α , p53, MET, YKL40, and pNDRG1 is quantified and elaborated to give a percentage for each component and thus determine tumor affiliation, as described in Method section. CL: Classical, MES: mesenchymal, PN: proneural.

Considering all 56 grade 4 gliomas, 2 patients died during the postoperative hospital stay, 12 were alive at last follow-up and 44 dead, with no statistical difference between subgroups (mean follow up time 16.18 ± 15.44 months). For uncensored, dead patients at last follow up, the mean overall survival was 15.72 ± 15.14 months, corresponding to 9.35 ± 2.26 months for MES, 13.43 ± 15.66 months for PN, 19.43 ± 16.37 months for CL, 19.97 ± 19.70 months for PN/CL, and 13.46 ± 8.01 months for mixed affiliation tumors. Figure 5.2A shows Kaplan Meier curves for Grade 4 gliomas subtypes, while Figure 5.2B shows Kaplan Meier curves for Grade 4 gliomas divided in MES vs non-MES tumors. Statistical significance was not reached in either case as reported by log-rank test, even if a trend for MES tumors as being associated with a worse prognosis was evident. After removal of the 6 IDH1-mutant Grade 4 astrocytomas, thus retaining only the 50 IDHwt GBMs, 10 patients were alive and 40 dead at last follow-up, one of which died during the postoperative hospital stay. The mean overall survival of patients who died was 13.64 ± 11.68 months, corresponding to 9.35 ± 2.26 months for MES, 9.36 ± 4.04 months for PN, 16.95 ± 14.33 months for CL, 19.97 ± 19.70 months for PN/CL, and 13.46 ± 8.01 months for mixed affiliation tumors. Figure 5.2C shows Kaplan Meier curves for GBM subtypes, while Figure 5.2D shows Kaplan Meier curves for GBMs divided in MES vs non-MES tumors. Again, statistical significance was approached but not reached in either case as reported by log-rank test, suggesting however that MES GBMs were more aggressive.

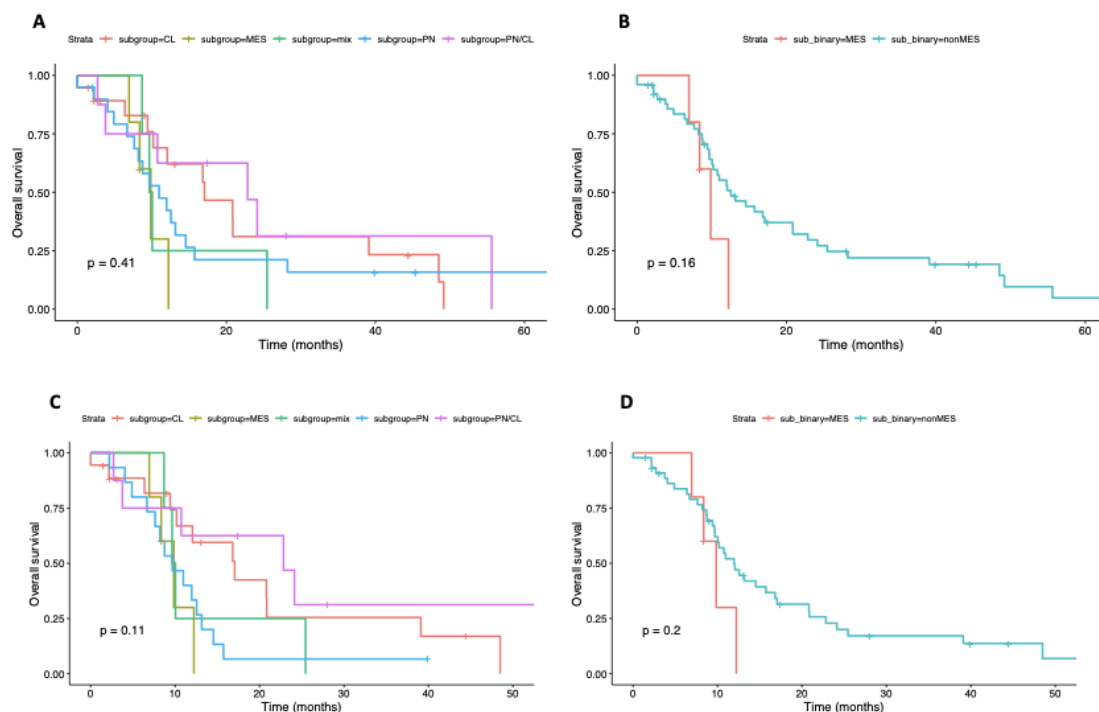


Figure 5.2. Kaplan Meier curves for overall survival. A, B, Survival analyses carried on all 56 grade 4 gliomas; A shows singular curves for each subtype (orange CL, olive green MES, cyan PN, purple PN/CL, and emerald green for balanced mixed affiliation); B shows MES vs non-MES samples (orange MES, light blue non-MES). Log rank statistics is reported in each panel. C, D, Survival analyses carried only on the 50 IDHwt GBMs; C shows singular curves for each subtype (colors as in A); B shows MES vs non-MES samples (colors as in B). Log rank statistics is reported in each panel. CL: classical, MES: mesenchymal, mix: balanced mix of the 3 subgroups, non-MES: non mesenchymal, PN: proneural, PN/CL: mixed proneural and mesenchymal.

5.1.2 Characteristics of GBM patients utilized for radiogenomic analyses.

Out of the above described 56 patients, we downsized our cohort for subsequent radiogenomic analyses. The main inclusion criterium was the presence of preoperative advanced diffusion MRI sequences (see Table 5.2 above), which were available for 40 total patients. Further reduction of patients was carried out excluding those cases that would be no more considered pure GBMs by the new 2021 WHO brain tumor classification. We therefore reduced our study sample to 36 total cases, of whom 13 were females (36.1%) and 23 males (63.9%), with a mean age of 62.6 ± 10.4 years (median 63.5, range 41-80 years). Mean preoperative KPS was $90.8 \pm$

9.7 (median 90, range 70-100). Sixteen patients (44.4%) were diagnosed due to seizures; the remaining 20 following non-specific symptoms or development of stable neurological deficits. In total, 28 (77.8%) patients had preoperative neurological deficits, either mild or severe.

Two patients had previous stereotactic biopsy for diagnosis. Three more patients had previous open surgeries with diagnosis of GBM in 2 cases and anaplastic astrocytoma in the remaining. Out of these three, 1 patient had no subsequent treatments, and 2 patients both chemo- and radiotherapy before undergoing surgery for which we enrolled them in our study. Fourteen lesions (38.9%) were right sided, 21 (58.3%) left sided, and 1 (2.8%) diffuse on both sides. Thirty-four patients underwent open surgery, with 27 gross-total resections (79.4%), 6 subtotal resections (<10% residue, 17.6%), and 1 partial resections (>10% residue, 2.9%). Two patients underwent only stereotactic biopsy.

As for transcriptional subgroup affiliation, 4 (11.1%) were MES GBMs, 13 (36.1%) CL, 11 (30.6%) PN, 6 (16.6%) mixed PN/CL, and 2 (5.6%) balanced mixes of all subtypes. Remarkably, our final patient population shows similar distribution to our initial cohort of 56 cases.

The above selected eligible patients were subjected to conventional cMRI (T2-weighted, T1-weighted, post-contrast T1-weighted and 3D-FLAIR) and aMRI by a standardized protocol implemented on a 3T scanner, including DTI and multicompartamental dMRI (35 directions at $b=711$ s/mm² and 60 directions at $b=3000$ s/mm²) such as NODDI that allows the characterization of brain tissue microstructure, neurite density and fiber orientation.

5.1.3 Generation of xenografts for radiogenomic feature analysis.

As reported earlier, we sampled 48 GBMs to generate GSC lines by the NSA. We were able to derive 14 cell lines, for an efficiency of 29.2% of our conditions in stabilizing GSCs. In vitro culturing generally selects cells with extremized features, with a trend towards purer subgroup affiliation, as opposed to human tumors, which consist of mixtures of PN, MES and CL subgroup to varying extent. Consequently, also GSC-derived xenografts tend to show more extreme characteristics of the different subgroup, and thus are helpful tools for identifying features discriminating the various subtypes. As a matter of fact, PN lines tend to generate more infiltrating tumors that spread along the white matter, but generally lack the ability to give rise to a full blown and contrast-enhancing nodule that is typical of GBMs. On the

contrary, pure MES lines, which are reported to be more difficult to establish under NSA conditions, usually grow as compact, less infiltrating masses, which model to some extent proper GBM nodules. Thorough dissertation on the features and subgroup dynamics of our cell lines are thoroughly reported in the next section.

For the radiogenomic analysis, tumors affiliated with either CL or PN subgroups were grouped together under the definition 'non-MES'.

We then transplanted MES and non-MES GSC lines in the striatum of immunocompromised mice to generate intracranial xenografts, which underwent the same advanced diffusion MRI study protocol as our patients. Transcriptional subgroup affiliation of the xenografts was confirmed by means of the same IHC panel used to assess parental human tumor affiliation. This *in vivo* modeling was very helpful, as it allowed us to generate balanced study groups of MES (9, 42.9%) and non-MES (12, 57.1%) xenografts, overcoming in this case the inherent limitation of our unbalanced patient cohort. In particular, the non-MES fraction of xenografts consisted of 8 CL (38.1%), 3 PN (14.3%) and 1 PN/CL (4.7%) tumors.

5.1.4 Radiomic feature selection and validation.

After masking the tumors' region of interest (ROI) on FLAIR and post-contrast T1-weighted sequences in humans and on diffusion B0 sequence in mice, they were imported in the diffusion maps to extract radiomic features pertaining to tumoral ROIs. Ninety-one features were extracted from each map (MD, FA for DTI, and fecv, ficv, fiso and odi for NODDI), thus isolating 1092 features (546 x2 masks) for each human study and 546 for each murine study. Discriminative features were selected independently in the patient and mice datasets by evaluating their capability in discriminating MES from non-MES tumors in a univariate manner. Thirteen features were identified for patients on 3D-FLAIR masks (Table 5.3) and 9 on post-contrast T1 masks (Table 5.4), while 45 for xenografts on the B0 mask (Table 5.5). Such disparity is not surprising as xenografts are more divergent and extreme than parental human tumors, as explained above. Notably, only features isolated on 3D-FLAIR masks in human and B0 masks in mice were comparable, since most non-MES xenografts did not uptake contrast and, thus, an equivalent to human post-Gd T1 mask was not retrievable. Therefore, we concentrated only on features isolated on 3D-FLAIR and B0 masks.

Radiomic feature	p-Value
'fecvf_FL_original_glcm_Imc2'	0.0048
'FA_FL_original_glcm_Imc2'	0.0127
'fecvf_FL_original_glcm_Imc1'	0.0127
'FA_FL_original_glcm_Imc1'	0.0286
'fecvf_FL_original_firstorder_90Percentile'	0.0296
'fecvf_FL_original_glcm_JointEnergy'	0.0324
'FA_FL_original_gldm_SmallDependenceLowGrayLevelEmphasis'	0.0367
'fecvf_FL_original_glcm_JointEntropy'	0.0367
'fecvf_FL_original_glcm_MaximumProbability'	0.0367
'fiso_FL_original_glszm_SmallAreaLowGrayLevelEmphasis'	0.0367
'FA_FL_original_glszm_LowGrayLevelZoneEmphasis'	0.0468
'FA_FL_original_glszm_SmallAreaLowGrayLevelEmphasis'	0.0468
'fiso_FL_original_glszm_LowGrayLevelZoneEmphasis'	0.0468

Table 5.3. Significant radiomic features discriminating human MES vs non-MES GBMs (3D-FLAIR mask). The first component of the feature indicates the diffusion map in which it was significant (FA, MD, fiso, fecv, ficv, odi), the second part which sequence mask was applied to the map (FL), the last part the actual radiomic feature. In bold are reported the features in common to xenografts (see Table 5.5). FA: fractional anisotropy (DTI), fecv: fraction of extraneurite volume (NODDI), ficv: fraction of intraneurite volume (NODDI), fiso: fraction of isotropic diffusion (NODDI), FL: FLAIR, MD: mean diffusivity (DTI), odi: orientation dispersion index (NODDI).

Radiomic feature	p-Value
'FA_T1_original_glcm_ClusterTendency'	0.0383
'FA_T1_original_glcm_SumSquares'	0.0194
'FA_T1_original_glrlm_GrayLevelVariance'	0.0484
'FA_T1_original_gldm_GrayLevelVariance'	0.0480
'fiso_T1_original_glcm_InverseVariance'	0.0041
'fiso_T1_original_glrlm_GrayLevelVariance'	0.0432
'fiso_T1_original_glszm_GrayLevelVariance'	0.0323
'fiso_T1_original_glszm_LargeAreaHighGrayLevelEmphasis'	0.0027
'fiso_T1_original_gldm_LargeDependenceHighGrayLevelEmphasis'	0.0040

Table 5.4. Significant radiomic features discriminating human MES vs non-MES GBMs (post-Gd T1 mask). The first component of the feature indicates the diffusion map in which it was significant (FA, MD, fiso, fecv, ficv, odi), the second part which sequence mask was applied to the map (T1), the last part the actual radiomic feature. FA: fractional anisotropy (DTI), fecv: fraction of extraneurite volume (NODDI), ficv: fraction of intraneurite volume (NODDI), fiso: fraction of isotropic diffusion (NODDI), MD: mean diffusivity (DTI), odi: orientation dispersion index (NODDI), T1: post-contrast T1.

Radiomic feature	p-Value
'odi_B0_original_glcm_Imc2'	0.0002
'odi_B0_original_glcm_Imc1'	0.0005
'fecvf_B0_original_firstorder_RootMeanSquared'	0.0018
'fecvf_B0_original_firstorder_Median'	0.0019
'fecvf_B0_original_firstorder_Mean'	0.0024
'FA_B0_original_glcm_Imc2'	0.0043
'odi_B0_original_glcm_JointEntropy'	0.0048
'fecvf_B0_original_firstorder_90Percentile'	0.0050
'fecvf_B0_original_firstorder_Maximum'	0.0062
'odi_B0_original_glcm_ClusterTendency'	0.0065
'ficvf_B0_original_glszm_SmallAreaHighGrayLevelEmphasis'	0.0071
'odi_B0_original_gldm_GrayLevelVariance'	0.0078
'odi_B0_original_glrIm_GrayLevelVariance'	0.0082
'odi_B0_original_glcm_SumSquares'	0.0083
'FA_B0_original_glcm_Imc1'	0.0086
'odi_B0_original_glrIm_GrayLevelNonUniformityNormalized'	0.0093
'odi_B0_original_firstorder_Uniformity'	0.0099
'odi_B0_original_glcm_SumEntropy'	0.0121
'odi_B0_original_firstorder_Entropy'	0.0123
'FA_B0_original_glszm_SmallAreaHighGrayLevelEmphasis'	0.0142
'odi_B0_original_glcm_ClusterProminence'	0.0156
'odi_B0_original_glcm_Correlation'	0.0169
'fecvf_B0_original_firstorder_Range'	0.0209
'odi_B0_original_firstorder_Kurtosis'	0.0255
'odi_B0_original_ngtdm_Complexity'	0.0266
'ficvf_B0_original_firstorder_10Percentile'	0.0302
'ficvf_B0_original_firstorder_Minimum'	0.0302
'ficvf_B0_original_firstorder_Range'	0.0302
'odi_B0_original_glcm_DifferenceVariance'	0.0343
'fecvf_B0_original_firstorder_10Percentile'	0.0352
'MD_B0_original_firstorder_Mean'	0.0360
'MD_B0_original_firstorder_Median'	0.0360
'MD_B0_original_firstorder_RootMeanSquared'	0.0360
'FA_B0_original_firstorder_Minimum'	0.0365

'odi_B0_original_glcm_DifferenceEntropy'	0.0371
'FA_B0_original_firstorder_10Percentile'	0.0428
'fecvf_B0_original_firstorder_Energy'	0.0428
'fecvf_B0_original_firstorder_MeanAbsoluteDeviation'	0.0428
'fecvf_B0_original_firstorder_TotalEnergy'	0.0428
'fecvf_B0_original_firstorder_Variance'	0.0428
'odi_B0_original_glcm_Contrast'	0.0429
'FA_B0_original_glcm_Correlation'	0.0430
'odi_B0_original_glcm_Idmn'	0.0447
'odi_B0_original_glrIm_RunEntropy'	0.0466
'odi_B0_original_glrIm_LongRunLowGrayLevelEmphasis'	0.0473

Table 5.5. Significant radiomic features discriminating MES vs non-MES GSC-derived xenografts. The first component of the feature indicates the diffusion map in which it was significant (FA, MD, fiso, fecv, ficv, odi), the second part which sequence mask was applied to the map (B0), the last part the actual radiomic feature. In bold are reported the features in common to parental human tumors (see Table 5.3). FA: fractional anisotropy (DTI), fecv: fraction of extraneurite volume (NODDI), ficv: fraction of intraneurite volume (NODDI), fiso: fraction of isotropic diffusion (NODDI), MD: mean diffusivity (DTI), odi: orientation dispersion index (NODDI).

By comparing the significant features retrieved in the two datasets, three appear in GBMs from both patients and mice: informational measure of correlation (IMC)1 and IMC2 of the GLCM class (grey-level-co-occurrence matrix, based on the probabilities of co-occurrence of pixel pairs with a given grey level) in the FA map of DTI, and the 90th percentile of the first-order class (i.e., based on the count of pixels in the ROI that possess a given grey-level value) in the fecv NODDI map (see features in bold in the above two tables). Figure 5.3 shows boxplots quantifying these radiomic features in MES vs non-MES tumors both in patients and mice.

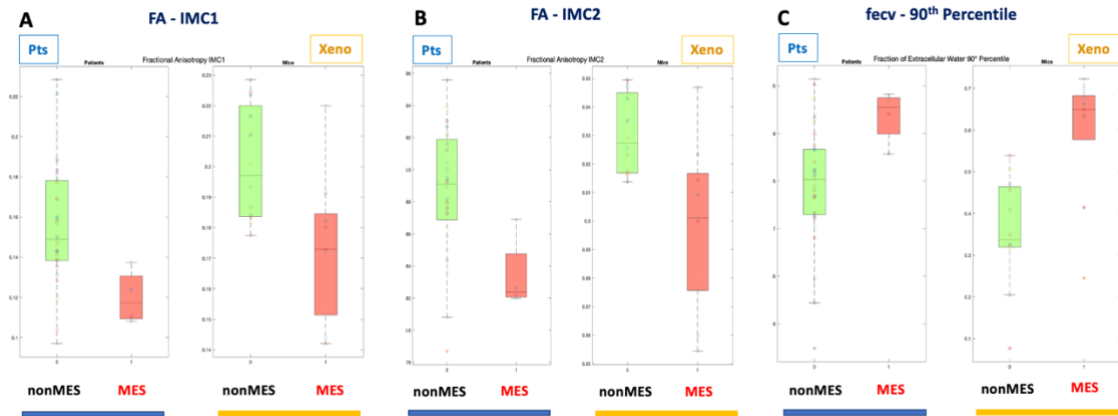


Figure 5.3. Significant radiomic features discriminating MES (red) from non-MES (green) tumors, both in patients (blue) and xenografts (orange). A, GLCM-IMC1 on FA DTI map. B, GLCM-IMC2 on FA DTI map. C, first-order 90th percentile on fecv NODDI map. FA: fractional anisotropy, fecv: fraction of extraneurite volume, GLCM: grey level co-occurrence matrix, IMC1/2: informational measure of correlation, MES: mesenchymal, non-MES: non mesenchymal, Pts: patients, Xeno: xenografts.

IMC1 and IMC2 assess the correlation between the probability distributions of grey levels in adjacent voxels, thus quantifying the complexity of the texture. In cases where the distributions are independent and there is no mutual information, these features will approach 0, indicating more heterogeneous textures. In case of uniform distribution with complete dependence and homogeneous textures, the same features will reach increasingly higher positive or negative values. Since IMC1 and IMC2 are significantly closer to 0 in MES than non-MES tumors, MES GBMs might be endowed with a more heterogeneous textural pattern than their CL or PN counterparts.

The 90th percentile feature collects the upper decile grey values in each map. MES tumors have significantly higher upper-decile values in the NODDI fecv map than non-MES GBMs. As explained in the introduction, fecv map provides a measurement of anisotropic gaussian diffusion which, under pathological conditions, indicates infiltrative edema (*i.e.*, edema due to local invasion of tumor cells, which separate axonal fibers, thus increasing the extraneurite component). This phenomenon differentiates from vasogenic, inflammatory edema, in which extreme derangement of vessel permeability leads to accumulation of water and inflammatory cells from the bloodstream, thus distorting brain texture to a much greater extent and increasing the isotropic component of the diffusion (fiso). Therefore, higher upper decile values in the fecv map in MES GBMs may indicate higher local tumoral infiltration in this subgroup, as opposed to PN and CL.

As a next step, we implemented a prediction model based on these three common radiomic features, using values detected in the xenograft dataset as reference for a k-nearest neighbor (KNN) model and for hyperparameter tuning. We then tried to predict transcriptional affiliation of patients' dataset and measured model performance, as reported in Figure 5.4. Notably, all MES human GBMs were classified correctly, with a sensitivity of 1, and 24/32 non-MES tumors were classified accordingly (specificity of 0.75). ROC statistic was considerably high, scoring 0.88. Therefore, this model represents an interesting tool in differentiating MES from non-MES GBMs, even though it may still be improved.

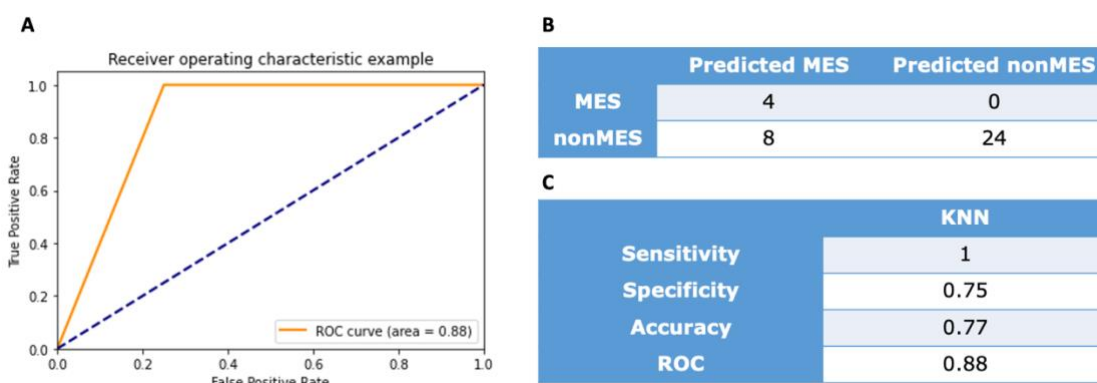


Figure 5.4. K-nearest neighbor prediction model performance in differentiating MES from non-MES GBMs. Radiomic features were selected considering the significant ones common to xenograft and patient dataset. The model was tuned on feature values in the xenograft dataset and tested on patients. A, ROC curve of the model. B, confusion matrix of the model. C, performance of the model. KNN: k-nearest neighbor, MES: mesenchymal, non-MES: non mesenchymal, ROC: receiver operating curve.

5.1.5 Refinement of radiomic feature selection and prediction model.

Given the disparity of MES and non-MES samples in our patients' dataset, we used computational strategies to augment and balance our data. Hence, we applied Python SMOTE algorithm obtaining additional random-generated human samples from the same distribution of the original unbalanced data, so not to introduce any bias. We reached a total of 64 patients equally allocated between MES and non-MES subsets. We performed the same process also on xenograft data, to equally balance this dataset with 12 MES and 12 non-MES samples.

Additionally, since radiomic features may reflect the same tumor characteristics, some of them may be redundant and might impair prediction model both in terms of overfitting to the specific study sample and by increasing computational costs with

no authentic prediction benefit. To address this question, we performed variance inflation factor (VIF) analysis and excluded radiomic features that would associate with other included features over a correlation coefficient threshold of 0.7. Consequently, we dropped 489 out of the 546 initial features for each sample mask.

Next, we repeated the univariate feature selection in the VIF-reduced xenograft and patient datasets (see Tables 5.6, 5.7 and 5.8) and identified common significant discriminative radiomic features. Notably, features were reduced in xenograft dataset (11) and slightly different from the previous analysis. On the contrary, human features were increased (19 on 3D-FLAIR masks and 21 on post-Gd T1 masks). Once again, we focused only on human 3D-FLAIR sequences, which are comparable to mice B0, as described earlier. When compared to our previous analysis, GLCM-IMC2 was ignored, as it was strongly correlated to GLCM-IMC1. When we considered significant features common to both human and murine datasets, we identified again the other 2 previously found features: GLCM-IMC1 on the FA map and the 90th percentile feature on the NODDI fecv map. Remarkably, both showed increased significance with respect to the previous analysis in humans (IMC1 p-value from 0.0286 to 4.5E-08; 90th percentile p-value from 0.0296 to 5.0E-08) as well as in xenografts (IMC1 p-value from 0.0086 to 0.0012; 90th percentile p-value from 0.0050 to 0.0019). Additionally, we selected 4 more common features, 3 pertaining to NODDI (fecvf_firstorder_10Percentile, odi_firstorder_Kurtosis and ficvf_glszm_SizeZoneNonUniformityNormalized) and 1 to DTI (GLCM_Correlation). Figure 5.5 shows boxplots for these 6 common features.

Just as the 90th Percentile for the upper decile, the 10th Percentile indicates intensities in the lower decile. MES tumors show higher values than non-MES in the lower end of their distribution in NODDI fecv map, thus suggesting that MES GBMs are characterized by more local infiltrative edema than their PN or CL counterparts.

Like IMC1, correlation ranges from 0 (uncorrelated) to 1 (perfectly correlated), representing the linear dependency of grey level values to their respective voxels in the grey level co-occurrence matrix (GLCM). In this sense, lower values in MES tumors substantiate MES ROI heterogeneity as opposed to non-MES GBMs.

A Grey Level Size Zone (GLSZM) quantifies grey level zones in an image. A grey level zone is defined as the number of connected voxels that share the same grey level intensity. In this context, the Size Zone Non-Uniformity Normalized feature measures the variability of size zone volumes throughout the image, with a lower value indicating more homogeneity. Again, MES tumors result more heterogeneous in texture than non-MES, as they present higher values for this feature in the fraction of non-Gaussian anisotropic diffusion (ficv).

Kurtosis is a measure of the “peakedness” of the ROI value distribution. Higher kurtosis indicates a flatter distribution with more consistent representation of the tails, rather than a spike-like distribution around the mean. Higher values of NODDI orientation dispersion index kurtosis in MES GBMs once more imply more heterogeneous tumoral ROI when compared to non-MES lesions.

Radiomic feature	p-Value
'FA_B0_original_glcm_Correlation'	0.0131
'FA_B0_original_glcm_Imc1'	0.0012
'fecvf_B0_original_firstorder_10Percentile'	0.0395
'fecvf_B0_original_firstorder_90Percentile'	0.0019
'ficvf_B0_original_glcm_Contrast'	0.0471
'ficvf_B0_original_glszm_SizeZoneNonUniformityNormalized'	0.0208
'odi_B0_original_firstorder_Entropy'	0.0046
'odi_B0_original_firstorder_Kurtosis'	0.0104
'odi_B0_original_glcm_ClusterProminence'	0.0077
'odi_B0_original_glcm_Imc1'	0.0002
'odi_B0_original_glrIm_LongRunLowGrayLevelEmphasis'	0.0195

Table 5.6. Significant radiomic features discriminating MES vs non-MES GSC-derived xenografts after data augmentation and VIF. The first component of the feature indicates the diffusion map in which it was significant (FA, MD, fiso, fecv, ficv, odi), the second part which sequence mask was applied to the map (B0), the last part the actual radiomic feature. In bold are reported the features in common to parental human tumors (see Table 5.7). FA: fractional anisotropy (DTI), fecv: fraction of extraneurite volume (NODDI), ficv: fraction of intraneurite volume (NODDI), fiso: fraction of isotropic diffusion (NODDI), MD: mean diffusivity (DTI), odi: orientation dispersion index (NODDI).

Radiomic feature	p-Value
'FA_FL_original_firstorder_Entropy'	0.0135
'FA_FL_original_glcmm_Autocorrelation'	0.0276
'FA_FL_original_glcmm_Correlation'	0.0467
'FA_FL_original_glcmm_Imc1'	4.5E-08
'MD_FL_original_firstorder_Skewness'	0.0003
'MD_FL_original_glcmm_ClusterProminence'	0.0287
'MD_FL_original_glcmm_Imc1'	0.0021
'MD_FL_original_glszm_HighGrayLevelZoneEmphasis'	0.0037
'MD_FL_original_glszm_SizeZoneNonUniformityNormalized'	0.0014
'fecvf_FL_original_firstorder_10Percentile'	0.0010
'fecvf_FL_original_firstorder_90Percentile'	5.0E-08
'fecvf_FL_original_glcmm_InverseVariance'	0.0121
'ficvf_FL_original_glszm_SizeZoneNonUniformityNormalized'	0.0225
'fiso_FL_original_firstorder_Entropy'	0.0138
'fiso_FL_original_glszm_LowGrayLevelZoneEmphasis'	1.4E-06
'odi_FL_original_firstorder_10Percentile'	3.6E-06
'odi_FL_original_firstorder_Kurtosis'	0.0404
'odi_FL_original_firstorder_Skewness',	0.0206
'odi_FL_original_ngtdm_Contrast'	0.0211

Table 5.7. Significant radiomic features discriminating human MES vs non-MES GBMs after data augmentation and VIF (3D-FLAIR mask). The first component of the feature indicates the diffusion map in which it was significant (FA, MD, fiso, fecv, ficv, odi), the second part which sequence mask was applied to the map (FL), the last part the actual radiomic feature. In bold are reported the features in common to xenografts (see Table 5.6). FA: fractional anisotropy (DTI), fecv: fraction of extraneurite volume (NODDI), ficv: fraction of intraneurite volume (NODDI), fiso: fraction of isotropic diffusion (NODDI), FL: FLAIR, MD: mean diffusivity (DTI), odi: orientation dispersion index (NODDI).

Radiomic feature	p-Value
'FA_T1_original_glcm_ClusterProminence'	5.6E-07
'FA_T1_original_glcm_Contrast'	0.0019
'FA_T1_original_glcm_JointEnergy'	0.0209
'FA_T1_original_glszm_SizeZoneNonUniformityNormalized'	0.0190
'MD_T1_original_firstorder_Skewness'	0.0025
'MD_T1_original_glcm_ClusterProminence'	5.4E-05
'fecvf_T1_original_firstorder_10Percentile'	0.0100
'fecvf_T1_original_firstorder_90Percentile'	5.3E-05
'fecvf_T1_original_glcm_ClusterProminence'	0.0236
'fecvf_T1_original_glcm_Contrast'	0.0461
'fecvf_T1_original_glszm_LowGrayLevelZoneEmphasis'	0.0466
'ficvf_T1_original_glszm_LowGrayLevelZoneEmphasis'	0,0024
'fiso_T1_original_glcm_DifferenceVariance'	8.3E-06
'fiso_T1_original_glcm_Imc1'	0.0159
'fiso_T1_original_glcm_InverseVariance'	6.5E-08
'fiso_T1_original_glszm_GrayLevelVariance'	1.0E-06
'fiso_T1_original_glszm_LargeAreaHighGrayLevelEmphasis'	0.0003
'fiso_T1_original_glszm_LowGrayLevelZoneEmphasis'	0.0207
'fiso_T1_original_glszm_SizeZoneNonUniformityNormalized'	0.0014
'odi_T1_original_firstorder_90Percentile'	0.0002
'odi_T1_original_glszm_LargeAreaLowGrayLevelEmphasis'	0.0117

Table 5.8. Significant radiomic features discriminating human MES vs non-MES GBMs after data augmentation and VIF (post-Gd T1 mask). The first component of the feature indicates the diffusion map in which it was significant (FA, MD, fiso, fecv, ficv, odi), the second part which sequence mask was applied to the map (T1), the last part the actual radiomic feature. FA: fractional anisotropy (DTI), fecv: fraction of extraneurite volume (NODDI), ficv: fraction of intraneurite volume (NODDI), fiso: fraction of isotropic diffusion (NODDI), MD: mean diffusivity (DTI), odi: orientation dispersion index (NODDI), T1: post-contrast T1.

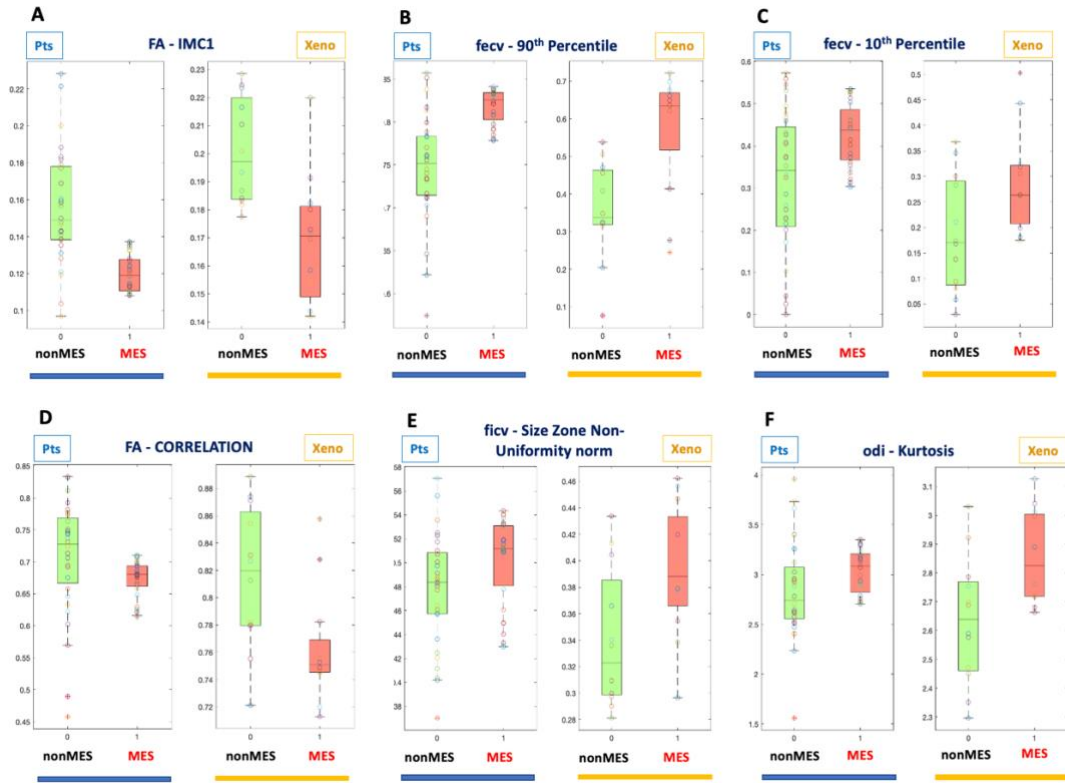


Figure 5.5. Significant radiomic features discriminating MES (red) from non-MES (green) tumors, both in patients (blue) and xenografts (orange) as isolated from augmented datasets. A, GLCM-IMC1 on FA DTI map. B, first-order 90th percentile on fecv NODDI map. C, first-order 10th percentile on fecv NODDI map. D, GLCM-Correlation on FA DTI map. E, GLSZM-Size Zone Non-Uniformity Normalized on ficv NODDI map. F, first-order Kurtosis on odi NODDI map. FA: fractional anisotropy, fecv: fraction of extraneurite volume, ficv: fraction of intraneurite volume, GLCM: grey levels co-occurrence matrix, GLSZM: grey level size zone matrix, MES: mesenchymal, non-MES: non mesenchymal, odi: orientation dispersion index, Pts: patients, Xeno: xenografts.

Lastly, we trained 4 different prediction models (Logistic Regression, Support Vector Machine, K-Nearest Neighbor, and Random Forest) with the selected six features and tuned parameters on xenograft dataset values. As a matter of fact, xenografts are more extremized than original tumors, and may provide more consistent thresholds for the model parameters. We then used the augmented patient dataset as a test and evaluated models' performances. Table 5.9 reports the confusion matrix for each model, while Table 5.10 compares the four performances. Figure 5.6 shows ROC curves of the prediction models. We obtained a great improvement in the results compared to previous prediction model, with the best performance being reached with Random Forest.

Model	Predicted MES		Predicted non-MES
	MES	non-MES	MES
LR	MES	24	8
	non-MES	4	28
SVM	MES	24	8
	non-MES	7	25
KNN	MES	20	12
	non-MES	0	32
RF	MES	31	1
	non-MES	0	32

Table 5.9. Confusion matrices for the 4 prediction models based on the 6 radiomic features identified. Models are tuned on augmented xenograft values (train dataset) and tested on augmented patients (test dataset). KNN: k-nearest neighbor, LR: linear regression, MES: mesenchymal, non-MES: non mesenchymal, RF: random forest, SVM: support vector machine.

Model	LR	SVM	KNN	RF
Sensitivity	0.750	0.750	0.625	0.969
Specificity	0.875	0.781	1.00	1.00
Accuracy	0.813	0.766	0.813	0.984
Precision	0.778	0.756	0.727	0.970

Table 5.10. Performances of the 4 prediction models based on the 6 radiomic features identified. Models are tuned on augmented xenograft values (train dataset) and tested on augmented patients (test dataset). KNN: k-nearest neighbor, LR: linear regression, RF: random forest, SVM: support vector machine.

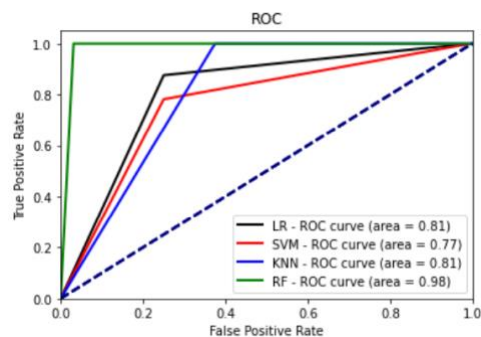


Figure 5.6. ROC curves for the 4 prediction models based on the 6 radiomic features identified on augmented datasets. KNN: k-nearest neighbor, LR: linear regression, RF: random forest, SVM: support vector machine.

5.2 Dynamics of transcriptional subgroup affiliation plasticity of GSCs and their derived xenografts.

5.2.1 Investigation of transcriptional subtype drift from the original tissue to the *in vitro* stem cell line by GSEA.

To identify whether *in vitro* culturing conditions could affect transcriptional subgroup affiliation, we subjected 9 pairs of original GBM tissues and their derivative GSCs to RNA Sequencing analysis and subsequent paired differential gene expression (DGE) analysis. GSEA analysis on the obtained ranked differential gene list, by using Verhaak's GBM subtype transcriptional signatures (Verhaak *et al*, 2010), showed a statistically significant enrichment of the MES signature in tissues as compared to their GSCs, as shown in Figure 5.7. On the contrary, none of the three signatures was significantly enriched in GSCs, suggesting a loss of the mesenchymal features under NSA culturing conditions.

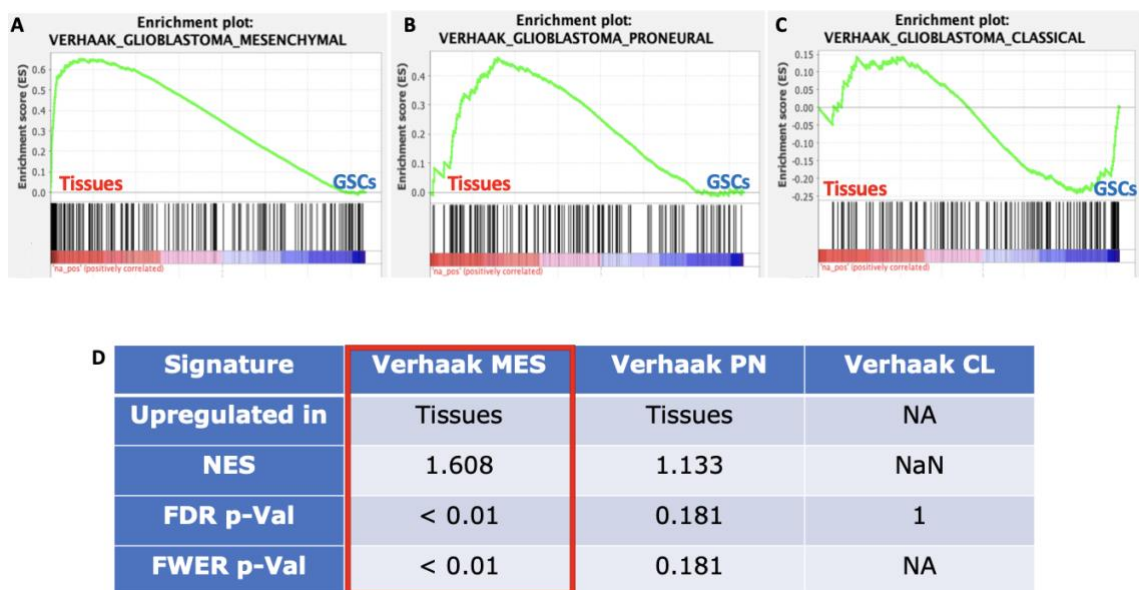


Figure 5.7. Gene Set Enrichment Analysis of Verhaak's transcriptional signatures in GBM tissues and their derivative GSCs. Panels A, B, C show enrichment plots of MES, PN and CL signatures, respectively. Panel D reports statistical parameters of each analysis. Only the MES signature is significantly enriched in Tissues. CL: Classical, FDR: false discovery rate, FWER: family-wise error rate, NA: not available, NaN: not a number, NES: Normalized enrichment score, MES: mesenchymal, PN: proneural.

5.2.2 GBM tissues and GSCs subgroup affiliation can be assessed by specific gene classifiers at the protein level.

As previously stated, RNA Sequencing is a time consuming and expensive technique to monitor subgroup affiliation of GBM tissues and their derived GSCs. Therefore, based on literature evidence and previous work from our lab (Verhaak *et al*, 2010; Park *et al*, 2017; Narayanan *et al*, 2019), we selected three genes identified as GBM subgroup classifiers, to assign sample affiliation based on their relative protein expression. EGFR is an indicator for CL subtype, Achaete-scute homolog 1 (**ASCL1**), a transcription factor related to neuronal differentiation and exit from cell cycle, for PN subclass, and N-Myc downstream-regulated gene 1 (**NDRG1**), an mTORC2-downstream member of the alpha/beta hydrolase superfamily negatively regulated by ASCL1, or its phosphorylated form (**pNDRG1**), for MES affiliation.

To test whether this minimal gene signature could be reasonably reliable for subgroup determination, we performed *in silico* analysis on R2 platform by assigning each of the TCGA GBM samples for which the actual subgroup affiliation was known to one of 3 clusters identified with a k-means algorithm, based on expression levels of the three aforementioned genes. Figure 5.8 shows the resulting sample clusters.

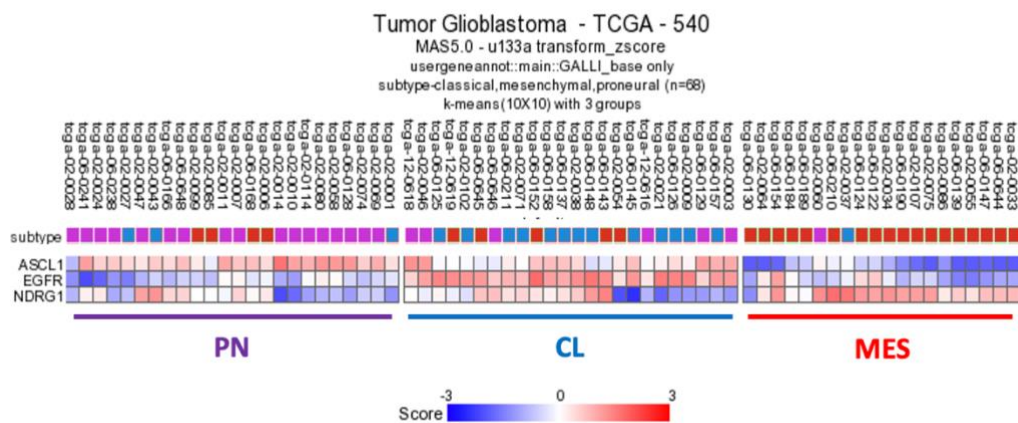


Figure 5.8. K-means clustering of TCGA GBM samples for which transcriptional subgroup was known, based on expression levels of minimal gene signature. Minimal gene signature consists of ASCL1 for PN (in purple), EGFR for CL (in blue) and NDRG1 for MES (in red). Minimal gene signature sensitivity and positive predictive value were respectively 70.8% and 70.8% for PN GBMs, 76.5% and 54.2% for CL GBMs, and 66.7% and 90% for MES GBMs. CL: classical, MES: mesenchymal, PN: proneural, TCGA: the cancer genome atlas.

The algorithm correctly reported 17/24 PN samples (sensitivity 70.8%, positive predictive value 70.8%), 13/17 CL samples (sensitivity 76.5%, positive predictive

value 54.2%), and 18/27 MES samples (sensitivity 66.7%, positive predictive value 90%), for a total sensitivity of 70.6% (Fisher's exact test p-value = 4.75e-9).

We then performed WB analysis on original tissues and their derivative GSCs at progressive culturing passages under NSA conditions and assigned subgroup affiliation based on the relative protein expression levels of EGFR, NDRG1 and ASCL1. Remarkably, we demonstrated a variable drift of our GSC lines as some of them upregulate the PN marker ASCL1 together with a downregulation of the other markers (see Figure 5.9) as in a sort of "proneuralization" of the lines, while others maintained or even upregulated the MES marker NDRG1, suggesting the occurrence of a mesenchymal shift (see Figure 5.10). Notably, line 160704 shows concomitant upregulation of ASCL1 and NDRG1, indicating a mixed PN/MES affiliation at later passages. As ASCL1 is known to repress NDRG1, this may indicate a mixture of two cell subpopulations, one expressing ASCL1 and the other NDRG1, conferring the mixed affiliation to the whole GSC line, but each maintaining its own identity.

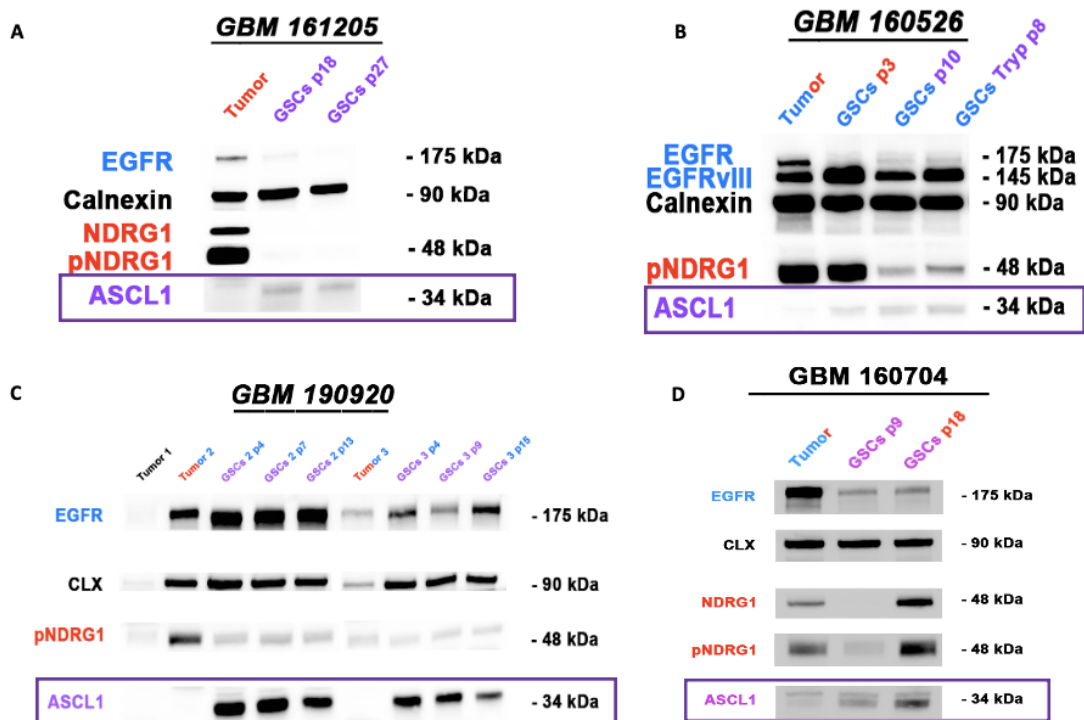


Figure 5.9. Minimal gene signature western blot analysis of GBM tissues and derived GSCs showing upregulation of the PN marker. Upregulation of the PN marker ASCL1 in addition to downregulation of both EGFR and NDRG1 (A), or only the MES marker NDRG1 (B and C). C shows generation of 2 lines from a multi-sampled GBM. Of note, D shows initial upregulation of ASCL1 with repression of both EGFR and NDRG1, while at later in vitro passages it shows also strong expression of NDRG1, indicating mesenchymalization. Subgroup affiliation is shown with the color code: CL blue, MES red, PN purple. CL: classical, MES: mesenchymal, PN: proneural.

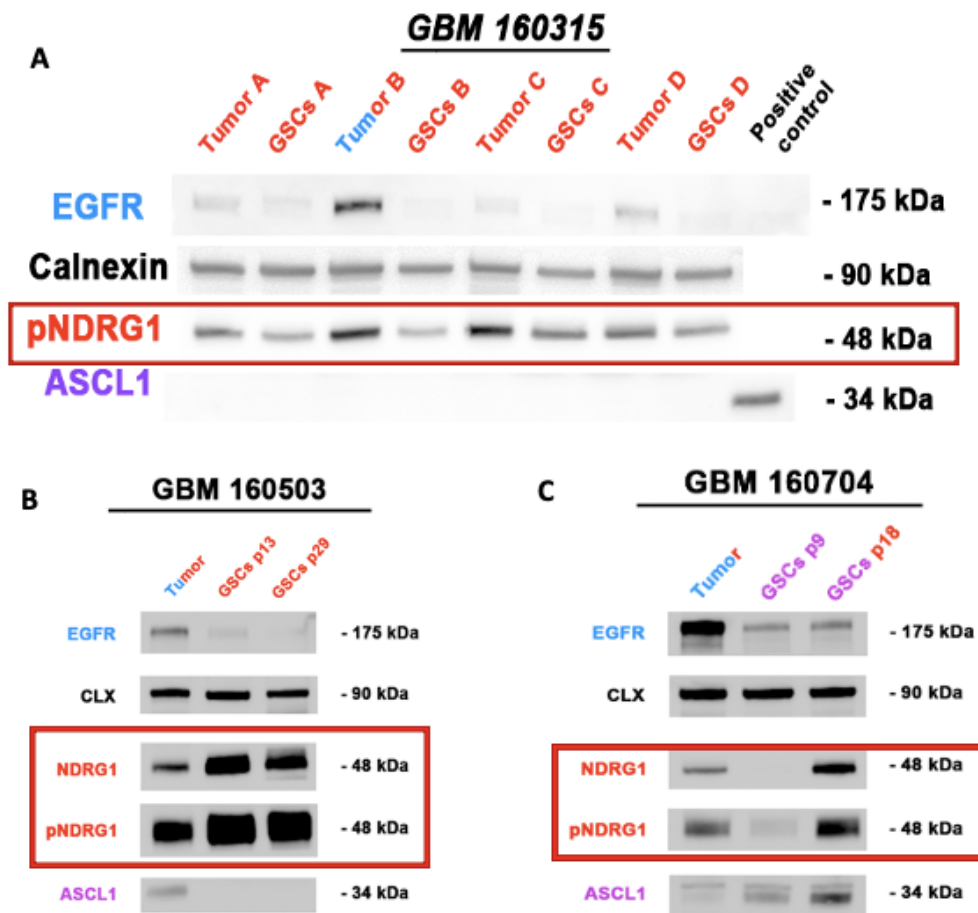


Figure 5.10. Minimal gene signature western blot analysis of GBM tissues and derived GSCs showing maintenance or upregulation of the MES marker. Panel A shows general maintenance of the MES marker NDRG1 in a GBM that underwent multiple sampling and generated 4 GSC lines with the same profile. Panel B shows progressive upregulation of NDRG1 with repression of EGFR and ASCL1 in a different GSC line. Notably, Panel C shows initial repression followed by strong upregulation of NDRG1, indicating a degree of mesenchymalization, even in the presence of increasing levels of ASCL1 in another GSC line. Subgroup affiliation is shown with the color code: CL blue, MES red, PN purple. CL: classical, MES: mesenchymal, PN: proneural.

Interestingly, in most cases (4 out of 6), the CL marker EGFR showed downregulation under NSA culturing conditions, likely due to the presence of high concentrations of EGF ligand in the culture medium (Figure 5.11). By means of WB, we were able to identify samples with the EGFRvIII mutation that induces the loss of exons 2-7 affecting the extracellular domain, rendering this variant constitutively active and independent of its ligand EGF. The presence of this variant was confirmed by immunohistochemistry in the original tumor samples. The presence of EGFRvIII led us to propend for a CL or mixed CL affiliation, weighing this component more than the other two, unless it showed a marked downregulation over the in vitro passages. Of note, in the majority of cases (3 out of 5), GSCs harboring EGFRvIII maintained

a very high level of the protein, at least in its mutated form. In the remaining cases, GSCs downregulated both the wild-type and mutated receptor, even though they did not completely turn off its expression (see Figure 5.12), therefore maintaining a CL component. Table 5.11 summarizes subgroup affiliations of the original tumors and their derived GSC lines based on WB analysis of minimal protein expression signature.

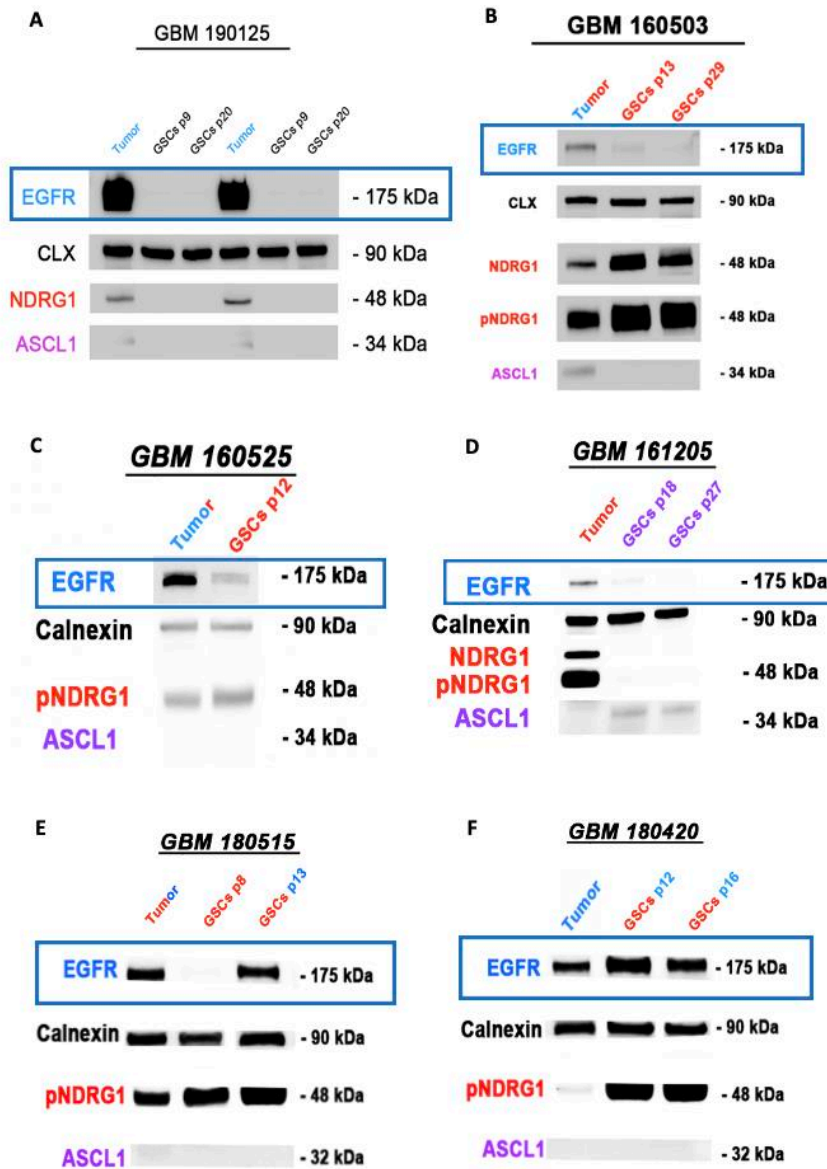


Figure 5.11. Minimal gene signature western blot analysis of GBM tissues and derived GSCs showing regulation of CL marker. Panel A, B, C and D show general downregulation of the CL marker EGFR in 3 different GSC lines. Notably, downregulation of all markers in the GSC line in panel A does not allow us to assign a subtype to the generated GSC line. On the contrary, Panel E and F show constant EGFR expression even if an earlier in vitro passage in E represses the receptor in other two GSC lines. Subgroup affiliation is shown with the color code: CL blue, MES red, PN purple. CL: classical, MES: mesenchymal, PN: proneural.

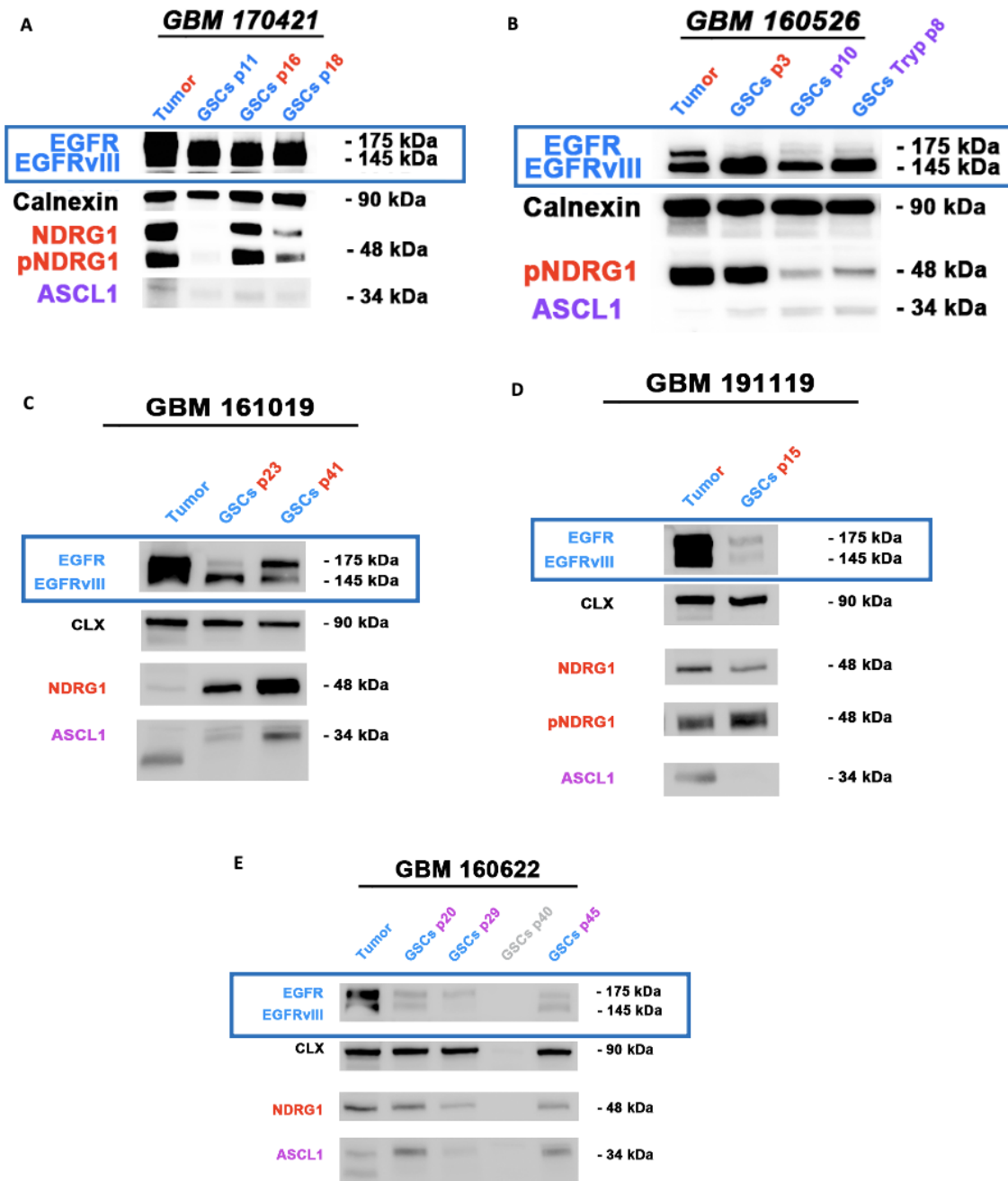


Figure 5.12. Minimal gene signature western blot analysis of GBM tissues and derived GSCs showing regulation of CL marker in EGFRvIII positive samples. Panel A, B, and C show general maintenance of the CL marker EGFR at least in its mutated, truncated form. On the contrary, D and E show downregulation of both wild-type and mutated EGFR. Subgroup affiliation is shown with the color code: CL blue, MES red, PN purple. CL: classical, MES: mesenchymal, PN: proneural.

Patient	Tumor WB	GSC WB early	GSC WB late
160315	MES	MES	MES
160526	CL	PN/CL*	PN/CL*
160704	CL	PN	PN/MES
160503	CL/MES	MES	MES
160525	CL	MES	MES
160622	CL	PN/CL	PN/CL
161019	CL	CL/MES*	CL/MES*
161205	MES	PN	PN
170421	CL/MES	CL*	CL/MES*
180420	CL	CL/MES	CL/MES
180515	CL/MES	MES	CL/MES
190125	CL	nd	nd
190920	CL/MES	PN/CL	PN/CL
191119	CL	CL/MES	.

Table 5.11. Subgroup affiliation of original GBM tissue and derived GSC lines at early and intermediate/late in vitro passages under NSA conditions, based on minimal gene signature classification. GSC lines that maintain EGFRvIII are indicated with an asterisk (*). CL: classical, MES: mesenchymal, nd: not defined, PN: proneural.

5.2.3 Comparison between IHC-based GBM subgroup definition and WB-based minimal signature affiliation.

As previously explained, for the subgroup affiliation of our patients' GBM specimens, we relied on immunohistochemical (IHC) definition of GBM subgroup according to the algorithm published by our collaborators in the Pathology Unit of University of Brescia based on the 8-gene signature (Orzan *et al*, 2020). Both IHC- and WB-based classifications rely on protein expression, as a proxy for transcriptional subgroup affiliation. In particular, the 8-gene IHC panel also includes the proteins that we previously identified as master switches for the PN (ASCL1) and MES (NDRG1) subgroup (Narayanan *et al*, 2019) and selected for our 3-gene minimal signature. Notwithstanding the good performance of our 3-gene signatures in classifying GBM samples by K-means (see Fig. 5.8), our minimal signature is smaller than the 8-gene IHC panel and generally more qualitative than quantitative; moreover, WB analysis doesn't let us distinguish between tumoral cells and

microenvironment contribution within the same GBM specimen, as would be easily carried out with a section staining. Additionally, WB analysis is performed on limited portions of the samples, which were not fully representative of the whole surgical samples. Conversely, IHC diagnosis was carried out on pathologist-reviewed representative sections of the gross tumor samples. Therefore, we used IHC subgroup diagnosis as a reference to which compare our WB panel. Table 5.12 summarizes subgroup affiliations of original tumors as defined by IHC and WB minimal signature.

Patient	Tumor IHC	Tumor WB
160315	MES	MES
160526	CL	CL
160704	CL	CL
160503	CL	CL/MES
160525	PN	CL
160622	CL	CL
161019	MIX	CL
161205	PN/CL	MES
170421	CL	CL/MES
180420	MES	CL
180515	PN/CL	CL/MES
190125	PN/CL	CL
190920	CL	CL/MES
191119	PN/CL	CL

Table 5.12. Subgroup affiliation of original GBM tissue based on IHC panel and WB minimal gene signature classification. CL: classical, MES: mesenchymal, MIX: balanced mix of the 3 subtypes, PN: proneural.

Remarkably, when comparing the two classifications, we had a perfect match in 4/14 cases (1 MES and 3 CL), while we added a MES component to 3 CL tumors, defining 3 mixed CL/MES GBMs. This may be explained by a contribution of infiltrated normal white matter in the tumor microenvironment that stains positive for NDRG1, which cannot be discerned by WB in the lysate of the whole sample. Accordingly, quantification of pNDRG1 by IHC on tumor slices of the same tumors was null in two cases and very weak in the remainder. Overweighing of the MES component is evident also in other two cases, in which a PN/CL GBM was classified as MES and a

PN/CL as CL/MES. In the former case, WB analysis showed high NDRG1 and pNDRG1, which was null in IHC, EGFR showed low expression in both techniques, while ASCL1 was consistently stronger in IHC and almost undetectable in WB. In the latter, IHC quantification of the three subgroups resulted in 54,8% PN, 42% CL and 3.2% MES, while WB revealed only high levels of EGFR and pNDRG1. Two PN/CL tumors were classified as CL by WB, as in case of GBM 190125 EGFR signal was extremely strong compared to NDRG1 and ASCL1, while IHC on the same patient revealed higher levels of ASCL1, in addition to positivity for Olig2 and PDGFRA. Tumor 191119 showed both extremely high levels of EGFR in addition to EGFRvIII mutation, therefore we did not consider the much lower signals of NDRG1 and ASCL1. On the other hand, in IHC ASCL1, Olig2 and PDGFRA all showed an intense signal, thus conferring a mixed PN/CL affiliation. This underestimation of the PN component by WB minimal signature may be due to a mixed effect of the known lower performance of ASCL1 antibody in WB, and of the smaller amount of tumoral material with respect to IHC. These effects would also explain the definition of 161019 as pure CL (high EGFR + EGFRvIII and low, yet detectable ASCL1) instead of a balanced mixture (32.4% PN, 20.8% MES, 46.8% CL according to IHC), and of 160525 as CL (high EGFR, very low pNDRG1 and undetectable ASCL1) instead of PN according to IHC.

5.3 Identification and testing of GSC-derived gene signatures for GBM transcriptional subgroup determination.

5.3.1 RNA Sequencing, unsupervised clustering of GSCs and differential gene expression.

In addition to paired comparison between tumor tissues and their derived GSCs (Fig. 5.7), we used RNA Sequencing data to perform unsupervised clustering of our cell lines and to compare the obtained clusters with the subgroup affiliation as determined by WB-based minimal signature. Figure 5.13A lists the GSC lines we analyzed with RNA Sequencing and their WB-based affiliation. We added a previously established GSC line (131210), deriving from a gliosarcoma, an extremely aggressive form of GBM, to implement our MES subset. Of note, GSC line 160704 was used at *in vitro* passage p9, when it expressed no MES marker and was therefore considered

as a pure PN line. Additionally, in case of mixed affiliation, prevalent subgroup was defined based on the relative levels of WB markers and their variation kinetics and tendency.

As evident in Figure 5.13B, unsupervised clustering based on transcriptome profiling returned two main clusters (A and B), the former consisting of 3 pure MES lines, while the latter comprising all the PN, CL and other mixed MES lines. Notably, among the MES lines in cluster A, there is one derived from a gliosarcoma and another (160315) isolated from a giant cell GBM, *i.e.*, two very aggressive tumors with marked mesenchymal features. However, some GSC lines that we classified as MES were classified as non-MES by RNAseq. Interestingly, Cluster A, is a subset of the WB-defined MES GSCs.

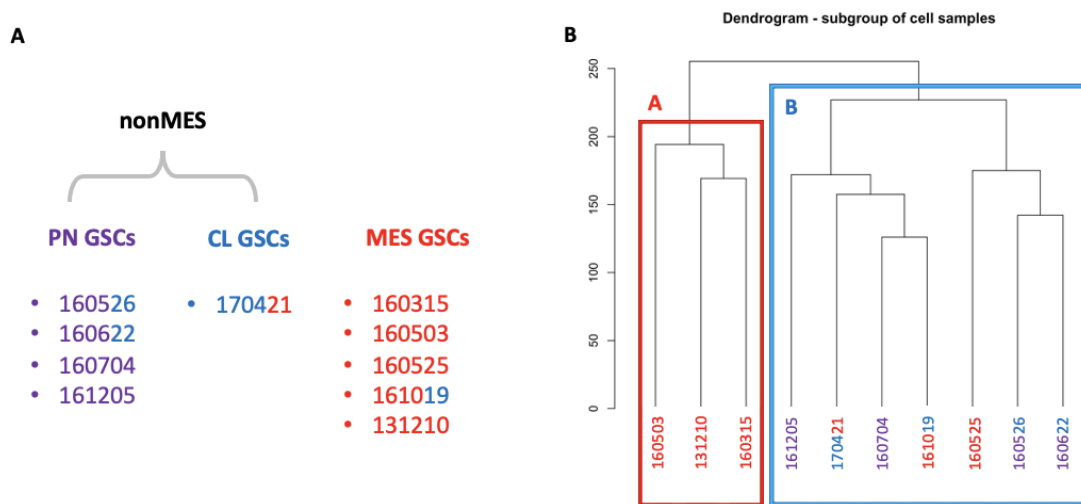


Figure 5.13. Scheme of GSC lines studied with RNA Sequencing. Panel A, List of lines and their subgroup affiliation according to WB-minimal signature. Line 160704 was used at *in vitro* passage p9, when it expressed no MES marker and was therefore considered as a pure PN line. B, Unsupervised clustering of GSCs based on transcriptome profiling. Subgroup affiliation is shown with the color code: CL blue, MES red, PN purple. CL: classical, MES: mesenchymal, PN: proneural, WB: western blot.

Notwithstanding the discrepancy with RNAseq based classification, to assess whether the 3-gene WB minimal signature was still consistent and predictive of transcriptional profiling of GSC lines, we performed differential gene expression on a supervised comparison, by opposing our WB-based MES lines to our WB-based PN, thus excluding the only CL line (170421). Next, we ran gene set enrichment with Verhaak's transcriptional signatures (Figure 5.14), which showed statistically significant enrichment of the MES signature in our WB-based MES lines and, conversely, of PN and CL signatures in our WB-based PN lines. Interestingly, ASCL1,

which is also part of Verhaak's PN signature, was the 11th differentially expressed proneural gene in our WB-defined PN lines compared to WB-defined MES GSCs.

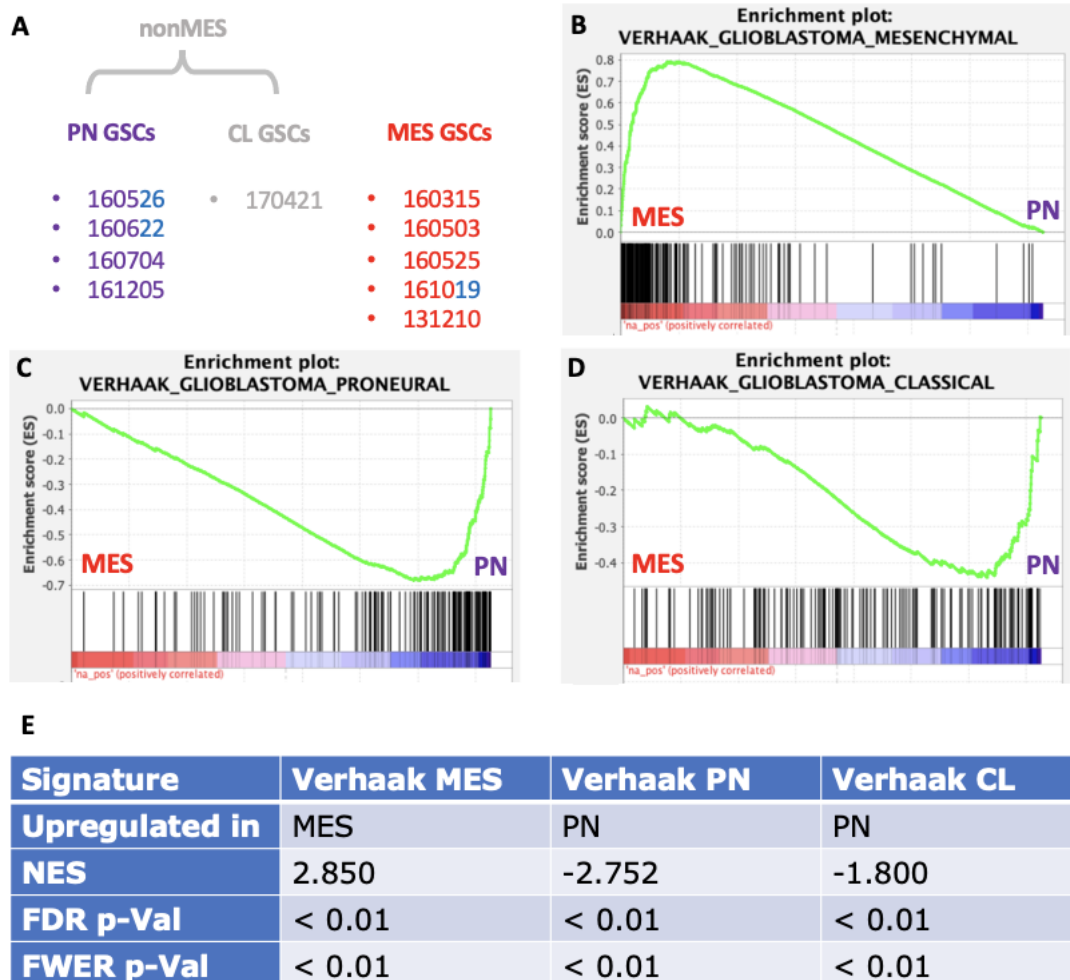


Figure 5.14. Gene set enrichment analysis on differential gene expression between WB-based MES and PN GSC lines. A, List of GSC lines in the analysis: the only classical one was left out for a cleaner comparison. Enrichment plots of Verhaak's signatures (MES in B, PN in C, CL in D). E, statistical parameters of each analysis. Subgroup affiliation is shown with the color code: CL blue, MES red, PN purple. CL: classical, FDR: false discovery rate, FWER: family-wise error rate, MES: mesenchymal, NES: normalized enrichment score, PN: proneural.

Parallely, we performed differential gene expression between the two unsupervised clusters resulting from RNA Sequencing and gene set enrichment with Verhaak's signatures (see Figure 5.15). As expected, MES signature was significantly enriched in cluster A and, conversely, PN and CL signatures in cluster B. Even more

strikingly, ASCL1 resulted the very first gene of the PN signature being differentially expressed between the two clusters. On the other hand, EGFR, which is also one of the genes in Verhaak's CL signature, was the 16th signature gene to be enriched in cluster B when compared to cluster A.

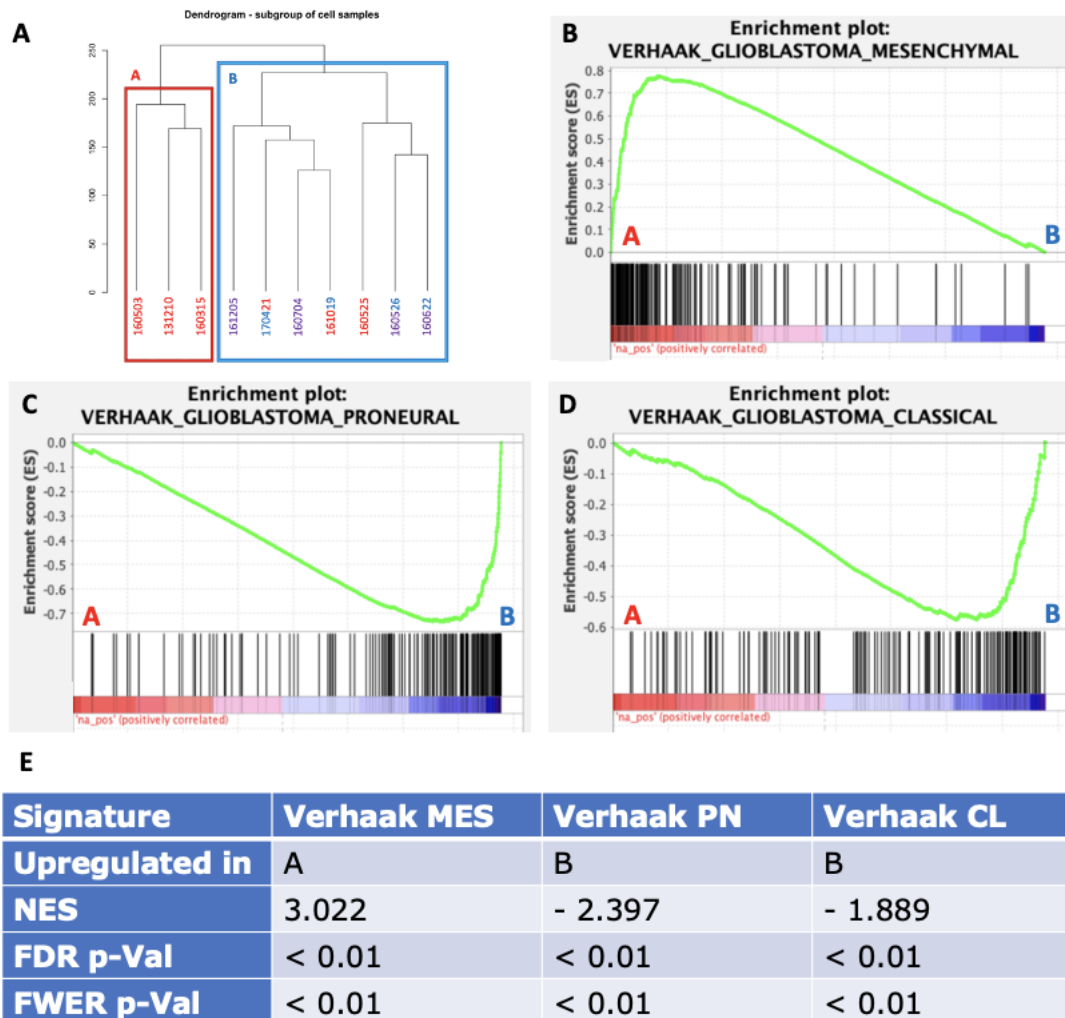


Figure 5.15. Gene set enrichment analysis on differential gene expression between RNA Sequencing unsupervised clusters A and B. A, Unsupervised clustering based on transcriptional profile; we performed differential gene expression analysis on A (red square) vs B (blue square). Enrichment plots of Verhaak's signatures (MES in B, PN in C, CL in D). E, statistical parameters of each analysis. Subgroup affiliation is shown with the color code: CL blue, MES red, PN purple. CL: classical, FDR: false discovery rate, FWER: family-wise error rate, MES: mesenchymal, NES: normalized enrichment score, PN: proneural.

5.3.2 Identification and validation of GSC-derived transcriptional signatures.

Since both our comparisons clearly reflected the difference between MES and PN or, at least, MES and non-MES transcriptional subgroups, we set out to generate GSC-derived transcriptional signatures that could be useful in classifying unrelated GBMs. As Verhaak's signatures consist of about 200 genes each, we wanted to propose possibly smaller but equally effective panels.

Therefore, when considering the DGE between our WB-defined MES and PN lines, we selected only significantly up- and downregulated genes according to SeqC filter in DESeq2 R package, set a multiple test adjusted p-value threshold of 0.05 and identified a log2FC threshold of 3.32 (indicating a ten-fold change between the groups). By this strategy, we generated an 81-gene WB-MES signature and a 43-gene WB-PN signature, as reported in Tables 5.13 and 5.14.

WB-MES Signature					
CDKN2B	MFAP2	IL6	LAMC2	PGF	PLXDC2
CDKN2A	MYO1B	GPR157	AMPD3	LINC00346	TMEFF2
PLIN2	AIM1	MICAL2	IGFBP5	COL7A1	ALDH1A3
COL6A3	CA9	PLEKHG4	PPL	AKR1B10	LIF
TGFB1	P4HA2	RCN3	FBLN2	LOXL4	PKP3
JAK3	DPT	KRT18	S100A4	ADAMTSL1	FBN2
MXRA5	SLC16A3	KRT8	KCNK3	GNA14	TUBA4A
RP11-152P17.2	COL1A1	STAT6	IRX3	SPINT1	ERMN
IGFBP3	ABI3BP	WSCD2	PTPRB	FAM46C	ITPR1
LOXL2	GJB2	NBL1	ELL2	ITGA11	C7orf57
MCTP2	MYLK	OLFML2A	PLAUR	KIAA1024	PCOLCE
IL4R	FER1L4	TMEM119	PCDHGA9	GNG11	
IER3	MT1L	G0S2	TMEM45A	SIRPB1	
TGM2	COL5A1	GFPT2	RP3-412A9.11	PCDHGA11	

Table 5.13. 81-gene MES signature derived from WB-based MES vs PN GSC differential gene expression. Parameters: multiple test adjusted p-value threshold ≤ 0.05 , log2FC threshold ≥ 3.32 .

WB-PN Signature				
ATP1A2	EPHB1	OMG	FRZB	COL9A1
ADAM23	KIF1A	AC018647.3	MX1	ID3
SLC4A10	EGFEM1P	ERBB4	NOL4	AC068535.2
MTRNR2L8	KCNJ6	ATP10B	CA2	Z83851.1
MPPED2	C1QTNF3	DLX6-AS1	PXDNL	FAM171B
CSMD1	SLITRK3	IFI27	SEPP1	CMPK2
CXCL10	POU3F4	SALL3	PTPRT	TNFSF13B
AQP4	RSAD2	MTRNR2L1	GRIA4	
GABRQ	TSPAN12	DUSP26	BRINP2	

Table 5.14. 43-gene PN signature derived from WB-based MES vs PN GSC differential gene expression. Parameters: multiple test adjusted p-value threshold ≤ 0.05 , \log_2FC threshold ≤ -3.32 .

In order to validate these signatures, we performed 10x10 rounds of k-means (Log2 Z-score transformation) on the TCGA GBM samples with known transcriptional affiliation that are available on the R2 platform, by dividing them into 3 clusters based on expression levels of the so defined 124 genes. We imposed 3 clusters based on the rationale that, by providing gene signatures for MES and PN, CL subgroup would be identified by exclusion.

As evident in Figure 5.16A, one cluster consisted of only MES tumors (10/27, sensitivity 37%, positive predictive value 100%), one aggregated almost all PN GBMs (21/24, sensitivity 87.5%, positive predictive value 63.6%) and the last one a mixture of the remaining 17 MES tumors together with few PN and CL. Also, CL GBMs were not precisely identified and resulted equally distributed across two clusters. The last "mixed MES" cluster may collect those tumors with a mixed MES component, while the first "pure MES" cluster may consist of the most severe and less heterogeneous MES lesions. Notwithstanding the imperfect identification of the real transcriptional affiliation, this panel shows a significant correlation with GBM subgroups (Fisher's exact test p-value = $9.9e-12$).

Since we generated this panel by comparing only MES and PN GSC lines, we decided to repeat the k-means clustering analysis but only on MES and PN tumors. We noticed an improvement in the accuracy and reliability of the test with still high

significance (Fisher's exact test p-value = 1.56e-11, see Figure 5.16B), as all PN GBMs were clustered together (24/24, sensitivity 100%, positive predictive value 88.9%), while 24/27 MES tumors were properly defined (sensitivity 88.9%, positive predictive value 100%).

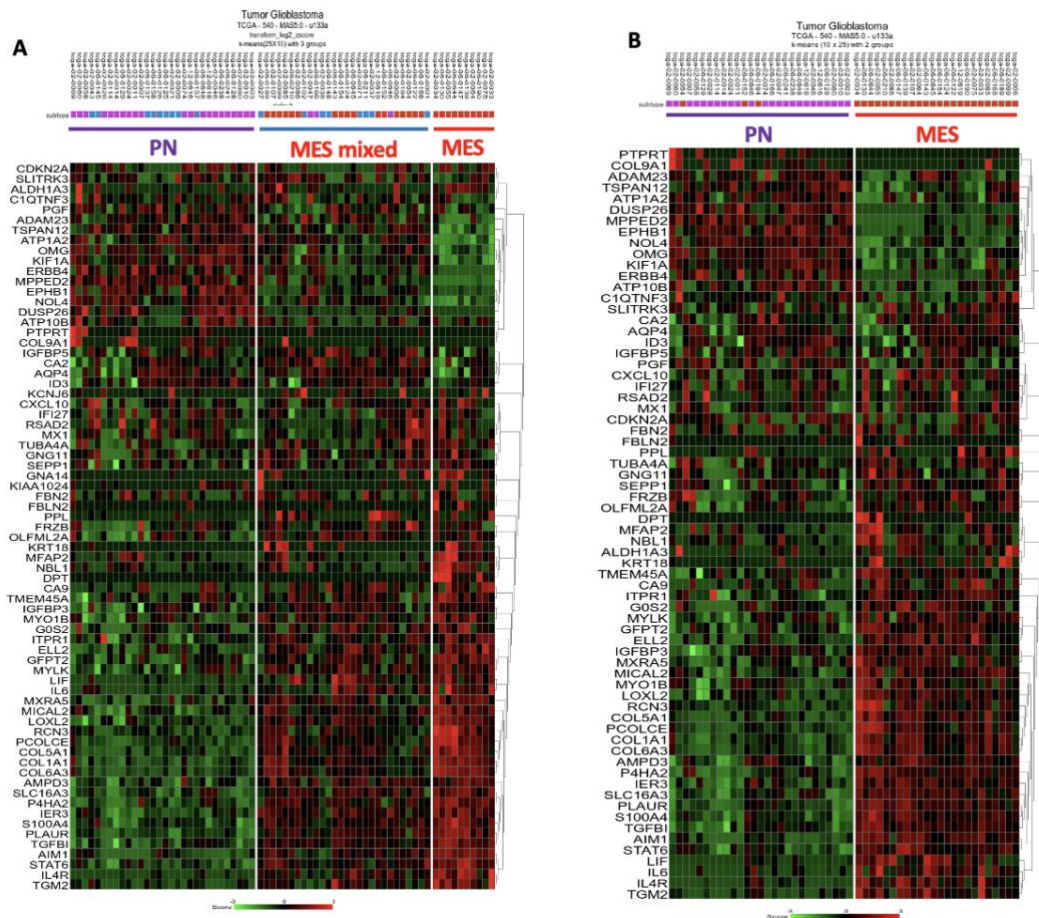


Figure 5.16. K-means clustering analysis of unrelated human GBMs, based on transcriptional signatures identified from the WB-MES vs WB-PN comparison. A, 3 clusters with MES, PN and CL tumors; cluster 1 identifies 21/24 PN GBMs, cluster 3 10/27 MES GBMs, while cluster 2 17/27 MES with additional tumors of the other subtypes; CL GBMs are equally distributed across cluster 1 and 2. B, 2 clusters with MES and PN tumors; all PN GBMs cluster together (plus 3 MES), while 24/27 MES segregate in the other cluster. Subgroup affiliation is shown with the color code: CL blue, MES red, PN purple. CL: classical, MES: mesenchymal, PN: proneural.

Parallely, we repeated the same procedure and considered the differential gene expression between unsupervised cluster A and B, selecting only significantly up- and downregulated genes according to SeqC filter. However, since in this comparison the differences between the two groups were much stronger, we could set a multiple test adjusted p-value threshold of 0.005 and a log2FC threshold of 5.64 (indicating a fifty-fold change between the groups). Thus, we generated a 49-gene Cluster A (MES) signature and a 98-gene Cluster B (non-MES) signature, as reported in Tables 5.15 and 5.16.

Cluster A (MES) Signature				
STAT6	RP11-58E21.3	CSF3	LCP1	NDNF
KCNK3	SLC16A6	IER3	AQP9	LPTM5
CA9	ATP8B4	SLC7A14	EDIL3	RP11-15012.1
IL6	GJB2	COL21A1	FCMR	PCSK1
FGF7	PTGS2	SPINT1	AQP1	MTUS2
CCL26	TGM2	RP11-91K9.1	DPT	CH25H
GDF5	FOXS1	AKR1B15	RP11-47I22.3	GPRC5C
MMP1	KRT8	MSC	GS1-600G8.5	SRGN
FGF5	RP11-213H15.3	IL1B	DMKN	ALDH1A3
PTGER4	POU2F2	ERMN	RP11-352D3.2	

Table 5.15. 49 gene - Cluster A (MES) signature derived from RNA Sequencing unsupervised cluster A vs B GSC differential gene expression. Parameters: multiple test adjusted p-value threshold ≤ 0.005 , log2FC threshold ≥ 5.64 .

Cluster B (non-MES) Signature					
NLGN3	C1orf61	FAM181B	DLL3	PLD5	DDX25
S100B	CCND2	HEPACAM	DCC	SLC6A1	ASIC4
IGSF11	NXPH1	ADAMTS3	SALL3	CACNG8	LINGO1
ATP1B2	BBOX1	LAMP5	ZDHHC22	ARHGAP25	COL20A1
KCNJ10	PMP2	KIF1A	TAGLN3	RPE65	RP11-449J21.5
TTYH1	SLITRK2	SNTG1	AGAP2	CD200	ADGRA1
NAT8L	FLRT3	FAT3	ATP13A4	RPH3A	SLITRK3
OLIG1	B3GAT1	CACNG4	FAM19A5	ATCAY	LRRTM3
OLIG2	SLIT1	SCG3	HAPLN1	DUSP26	CHST9
AIF1L	SEZ6L	GFAP	PALM	ADGRB1	ATP1A2
WSCD1	FGFBP3	KCNJ16	ASCL1	IGLON5	CXXC4
ZNF853	KIAA1456	GDPD2	LRRC4B	COL9A1	COL2A1
RP11-161M6.2	SOX8	ACTN2	ERBB3	CADM2	SCN3A
BCAN	C2orf80	DPP6	C2orf72	GSG1L	RLBP1
CACNG7	KCNJ8	FOXF2	HES5	KCNQ2	
BRSK2	MAL	OMG	CD74	NRSN1	
NCAN	ICA1	MYT1	CDH10		

Table 5.16. 98 gene Cluster B (non-MES) signature derived from RNA Sequencing unsupervised cluster A vs B GSC differential gene expression. Parameters: multiple test adjusted *p*-value threshold ≤ 0.005 , \log_2FC threshold ≤ -5.64 .

As done for signatures generated from the supervised differential gene expression, we performed 10x10 rounds of k-means (Z-score transformation) on the TCGA GBM samples with known transcriptional affiliation and available on the R2 platform, generating 3 clusters based on expression levels of these 147 genes. Again, we relied on the rationale that, by providing gene signatures for MES and PN, CL subgroup would be identified by exclusion. As a matter of fact, even though cluster B also includes one MES line and two CL/MES mixtures (in each of which either subtype prevailed), thus indicating more of a non-MES profile, GSEA analyses showed a clearer PN profile (NES -2.397 and denser signature hits on cluster B extreme) than

CL (NES -1.889 and more widespread signature hits). Therefore, we assumed that profile A and B might be able to also discern CL subgroup.

Figure 5.17 shows the clustering with correct identification of 25/27 MES tumors (sensitivity 92.6%, positive predictive value 92.6%), 20/24 PN GBMs (sensitivity 83.3%, positive predictive value 95.2%) and 15/17 CL lesions (sensitivity 88.2%, positive predictive value 75%), for a total panel sensitivity of 88.2%. Therefore, this panel is actually able to discern all three subgroups with a very high performance also on CL samples (Fisher's exact test p-value = $3.35e-21$) and shows an extreme improvement with respect to the signatures identified from the supervised DGE as well as to the 3-gene panel used in WB.

As done previously, if we repeat the k-means clustering (Log2 z-score transformation) only on MES and PN tumors, we obtain an 100% accuracy and reliability of the test (Fisher's exact test p-value = $4.36e-15$, see Figure 5.18A), as all PN and MES GBMs correctly clustered together. Notably, if we perform the k-means clustering (Log2 z-score transformation) only on CL and PN tumors and only using Cluster B signature, we correctly classify 21/24 PN (sensitivity 87.5%, positive predictive value 95.5%) and 16/17 CL GBMs (sensitivity 94.1%, positive predictive value 84.2%), with a very high significance (Fisher's exact test p-value = $4.36e-15$, see Figure 5.18B).

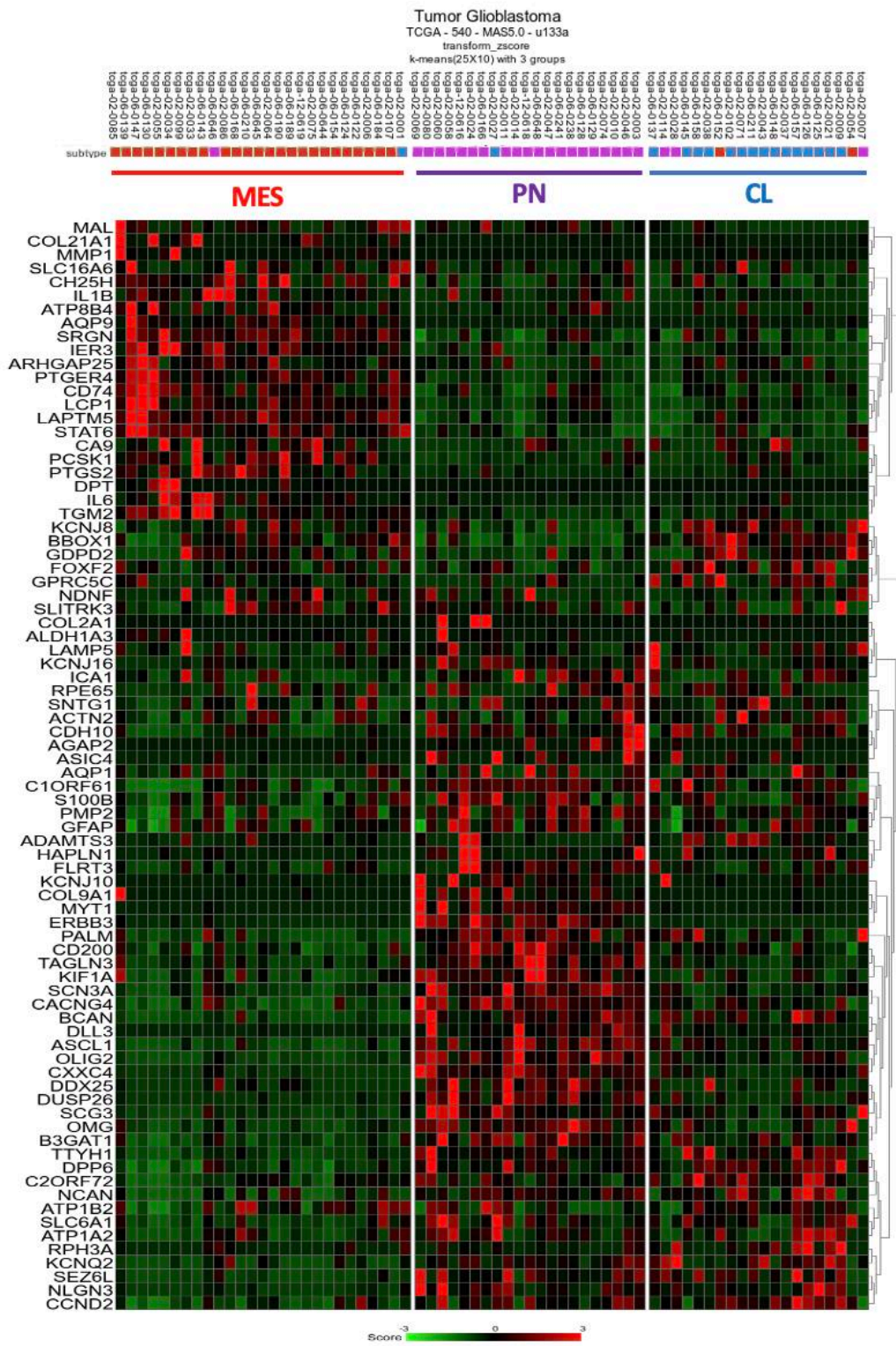


Figure 5.17. K-means clustering analysis of unrelated human GBMs, based on transcriptional signatures identified from the RNA Sequencing Unsupervised Clusters A and B. 3 clusters with MES, PN and CL tumors; cluster 1 identifies 25/27 MES GBMs, cluster 2 20/24 PN GBMs, cluster 3 15/17 CL GBMs. Subgroup affiliation is shown with the color code: CL blue, MES red, PN purple. CL: classical, MES: mesenchymal, PN: proneural.

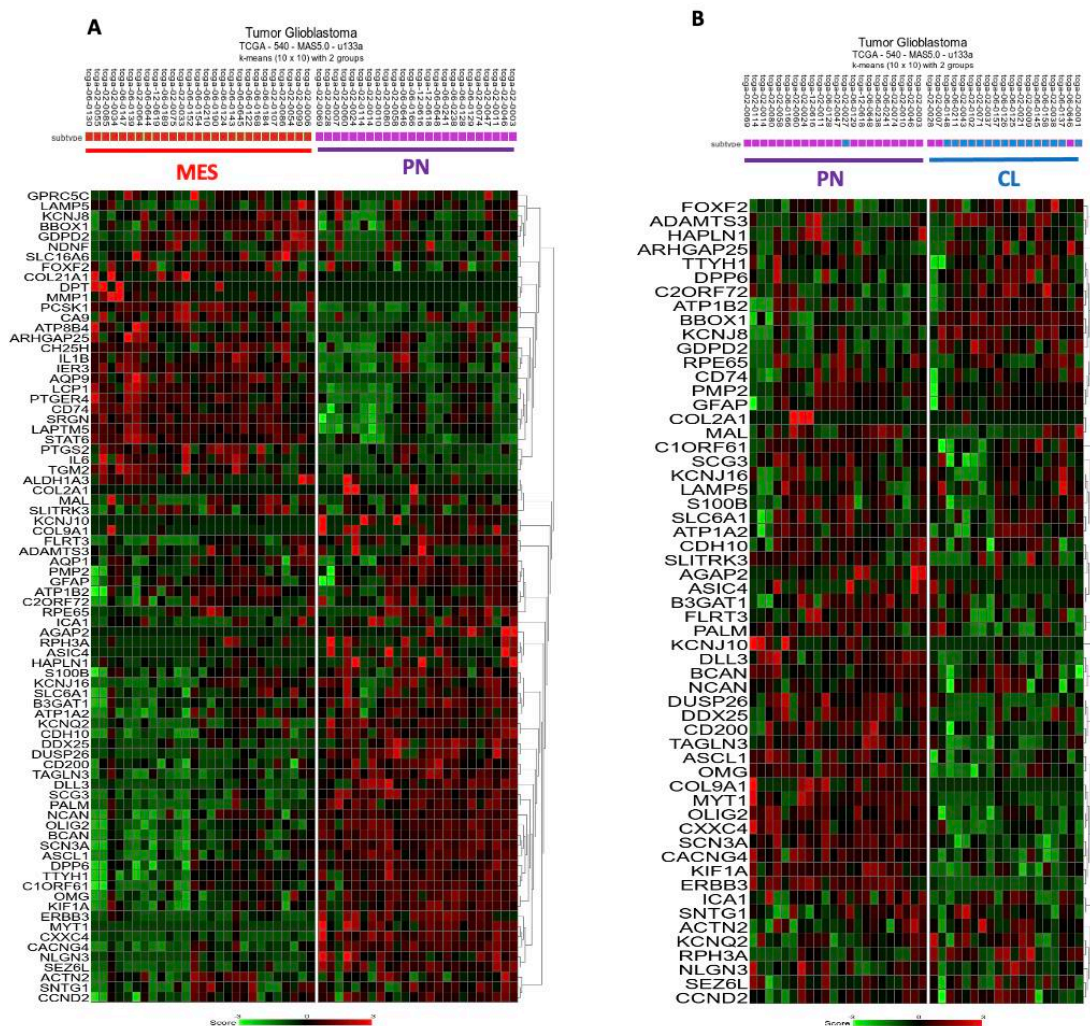


Figure 5.18. K-means clustering analysis of unrelated human GBMs, based on transcriptional signatures identified from the RNA Sequencing Unsupervised Clusters A and B. A, 2 clusters with MES and PN tumors; all PN and MES GBMs cluster together in the two perfectly segregated groups. B, 2 clusters with CL and PN tumors segregated solely on the base of Cluster B signature; 21/24 PN and 16/17 CL tumors are correctly grouped together. Subgroup affiliation is shown with the color code: CL blue, MES red, PN purple. CL: classical, MES: mesenchymal, PN: proneural.

We further validated our proposed signatures by gene set enrichment analyses (GSEA) performed on publicly available mRNA sequencing data of human samples. Briefly, we matched transcriptional affiliation of TCGA GBM samples from R2 platform with downloadable data from a cohort of the same patients, selecting only primary tumors, as described in the Methods section. We then performed pairwise differential gene expression analysis on the obtained 4 CL, 9 MES and 6 PN TCGA samples, and subsequent GSEA, interrogating the 3 contrasts with our signatures. Figures 5.19, 5.20 and 5.21 report the relative enrichment plots and statistics. Our signatures were

significantly able to discern MES from CL and from PN tumors. Interestingly, our MES signatures (Cluster A and WB-MES) were significantly enriched in PN tumors when compared to CL GBMs.

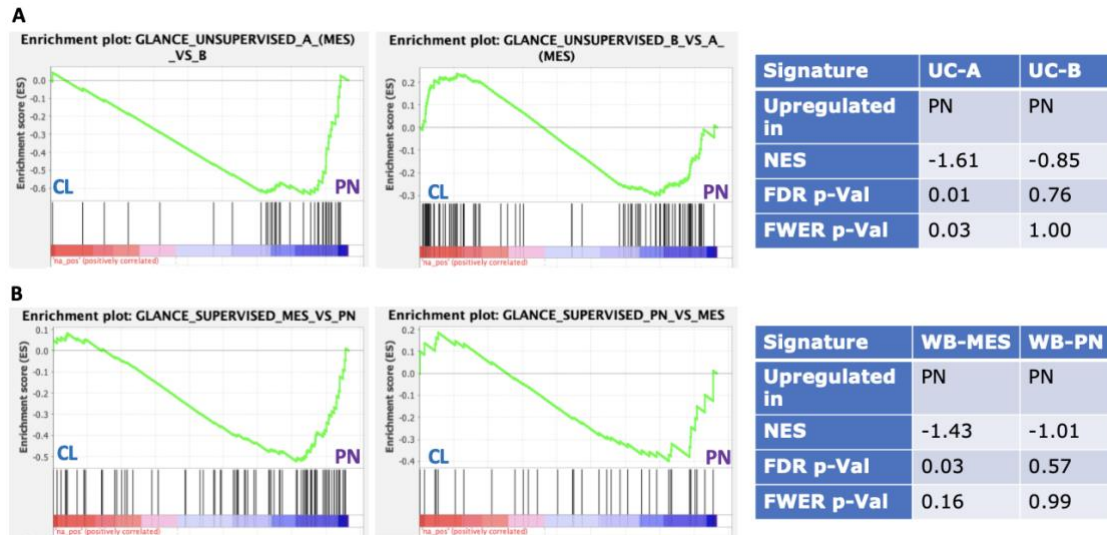


Figure 5.19. Enrichment plots and statistics of our transcriptional signatures in the TCGA CL vs PN GBMs. A, signatures derived from unsupervised clusters A and B. B, signatures derived from WB-panel defined PN and MES GSCs. Only MES signatures (UC-A and WB-MES) are enriched in PN vs CL tumors. CL: classical, FDR: false discovery rate, FWER: family-wise error rate MES: mesenchymal, NES: normalized enrichment score, PN: proneural, UC: unsupervised cluster, WB: western blot.

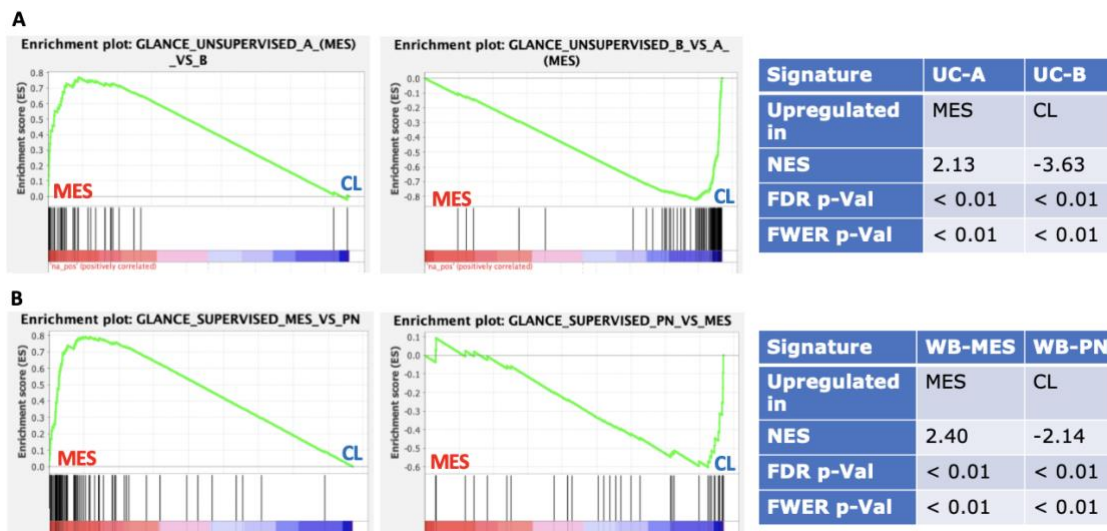


Figure 5.20. Enrichment plots and statistics of our transcriptional signatures in the TCGA MES vs CL GBMs. A, signatures derived from unsupervised clusters A and B. B, signatures derived from WB-panel defined PN and MES GSCs. MES signatures (UC-A and WB-MES) are enriched in MES vs CL tumors, while non-MES/PN signatures (UC-B and WB-PN) in CL. CL: classical, FDR: false discovery rate, FWER: family-wise error rate MES: mesenchymal, NES: normalized enrichment score, PN: proneural, UC: unsupervised cluster, WB: western blot.

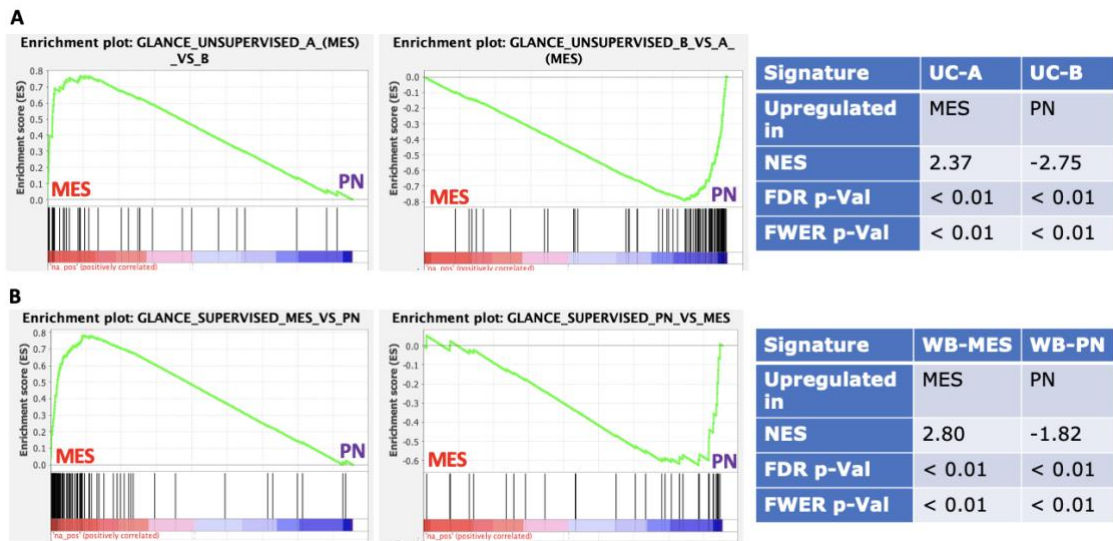


Figure 5.21. Enrichment plots and statistics of our transcriptional signatures in the TCGA MES vs PN GBMs. *A*, signatures derived from unsupervised clusters A and B. *B*, signatures derived from WB-panel defined PN and MES GSCs. MES signatures (UC-A and WB-MES) are enriched in MES vs PN tumors, while non-MES/PN signatures (UC-B and WB-PN) in PN. FDR: false discovery rate, FWER: family-wise error rate MES: mesenchymal, NES: normalized enrichment score, PN: proneural, UC: unsupervised cluster, WB: western blot.

To assess whether the signatures we identified could be a novel addition to what is already described in the literature, we evaluated the extent to which the 3 MES and the 3 PN signatures overlapped with each other. Considering MES signatures, only STAT6 was common to Verhaak’s MES, WB-MES, and Cluster A Signatures. Table 5.17 reports the genes common to the mesenchymal signatures. Jaccard Index was 1.5% for Verhaak’s MES and Cluster A, 3.8% for Verhaak’s MES and WB-MES, 10.2% and for Cluster A and WB-MES, indicating general independence of the 3 signatures, with a slightly higher similarity between the two MES signatures that we proposed.

Common to	#	Genes
Verhaak MES – Cluster A – WB MES	1	STAT6
Verhaak MES – Cluster A	4	LAPTM5, LCP1, PTGER4, STAT6
Verhaak MES – WB MES	11	IL4R, P4HA2, AIM1, S100A4, COL1A1, COL5A1, SLC16A3, TGFBI, PLAUR, AMPD3, STAT6
Cluster A – WB MES	12	DPT, ALDH1A3, IER3, KCNK3, ERMN, KRT8, TGM2, CA9, GJB2, SPINT1, IL6, STAT6

Table 5.17. Genes common to the three MES Signatures. MES: mesenchymal, WB: western blot, #: number of genes.

Table 5.18 reports the genes that are found in common among the 3 proneural signatures. Only DUSP26 was common to all of them. Jaccard Index was 5.8% for Verhaak’s PN and Cluster B, 1.9% for Verhaak’s PN and WB-PN, and 5.2% for Cluster B and WB-PN, denoting independence of the signatures. Figure 5.22 reports quantitative Venn diagrams of the two sets of signatures.

Common to	#	Genes
Verhaak PN – Cluster B – WB PN	1	DUSP26
Verhaak PN – Cluster B	15	SCG3, SCN3A, DLL3, TTYH1, SEZ6L, BCAN, CXXC4, C1orf61, ERBB3, ASCL1 , NLGN3, OLIG2, DPP6, MYT1, DUSP26
Verhaak PN – WB PN	4	MPPED2, NOL4, EPHB1, DUSP26
Cluster B – WB PN	7	KIF1A, COL9A1, ATP1A2, SLITRK3, SALL3, OMG, DUSP26

Table 5.18. Genes common to the three PN Signatures. Of note, ASCL1 is common to Verhaak’s PN and GSC Unsupervised Cluster B signatures. PN: proneural, WB: western blot, #: number of genes.

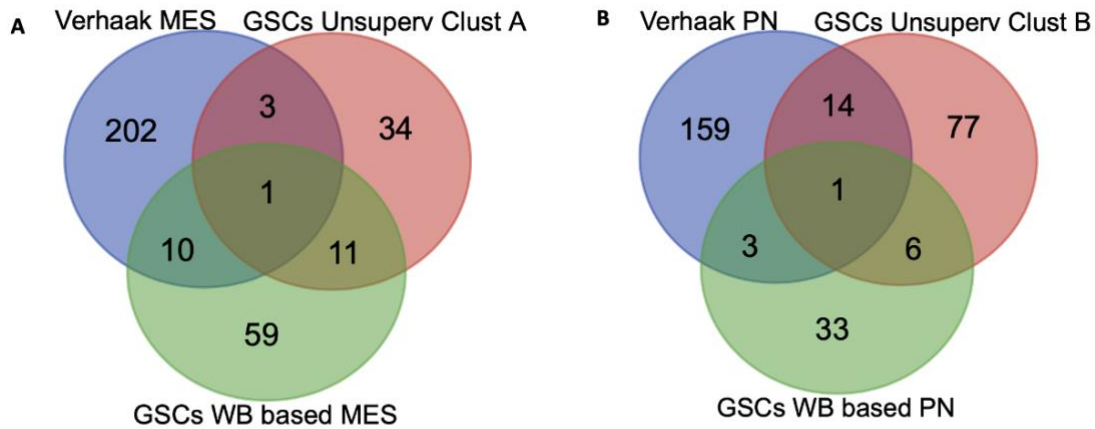


Figure 5.22. Venn diagrams showing the intersections between the MES (A) and PN (B) transcriptional signatures published by Verhaak and those identified from the RNA Sequencing of our GSCs. MES: mesenchymal, PN: proneural, WB: western blot.

We performed the same overlap analyses with the reduced subgroup-specific signatures proposed in the subsequent work by Verhaak's group, after removal of the NEU subtype (Wang *et al*, 2017). Considering MES signatures, only IER3 and IL6 were common to Wang's MES, WB-MES, and Cluster A Signatures. Table 5.19 reports the genes common to the mesenchymal signatures. Jaccard Index was 3.1% for Wang's MES and Cluster A and 5.6% for Wang's MES and WB-MES, confirming general independence of the 3 signatures. Table 5.20 reports the genes that are found in common among the 3 proneural signatures. Notably, no gene was common to all of them. Jaccard Index was 1.1% for Wang's PN and Cluster B and 1.7% for Wang's PN and WB-PN. Figure 5.23 reports quantitative Venn diagrams of the two sets of signatures.

Common to	#	Genes
Wang MES – Cluster A – WB MES	2	IER3, IL6
Wang MES – Cluster A	1	PTGS2
Wang MES – WB MES	5	COL5A1, P4HA2, TGFBI, PLAUR, COL1A1
Cluster A – WB MES	10	DPT, ALDH1A3, KCNK3, ERMN KRT8, TGM2, CA9, GJB2, SPINT1, STAT6

Table 5.19. Genes common to the three MES Signatures. MES: mesenchymal, WB: western blot, #: number of genes.

Common to	#	Genes
Wang PN – Cluster B – WB PN	0	
Wang PN – Cluster B	2	MYT1, ERBB3
Wang PN – WB PN	1	EPHB1
Cluster B – WB PN	7	KIF1A, COL9A1, ATP1A2, SLITRK3 SALL3, OMG, DUSP26

Table 5.20. Genes common to the three PN Signatures. PN: proneural, WB: western blot, #: number of genes.

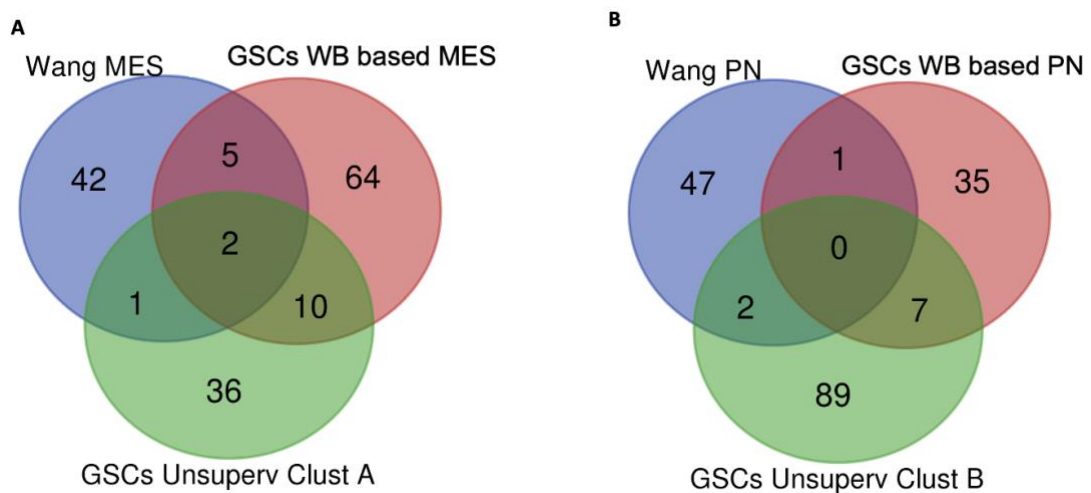


Figure 5.23. Venn diagrams showing the intersections between the MES (A) and PN (B) transcriptional signatures published by Wang and those identified from the RNA Sequencing of our GSCs. MES: mesenchymal, PN: proneural, WB: western blot.

5.4 Identification of subgroup-restricted altered pathways and phenotypes in MES and PN GSCs.

By exploiting all significant differentially expressed genes between our GSCs, without imposing any cut-off we took advantage of EnrichR webtools to identify transcriptional pathways, molecular pathways, ontologies and cell or tissue phenotypes, potentially playing a key role in our subgroup specific GSC lines. Tables 5.21, 5.22, 5.23 and 5.24 show the most relevant findings for both sides of each comparison between our GSCs.

Feature	ID	OR	Adj p-value
Transcriptomic	NFkB1 Activation	9.18	2.2e-31
Transcriptomic	HIF1A	7.30	2.8e-30
Transcriptomic	ASCL1 knock out	4.77	7.5e-18
Transcriptomic	YAP1	3.48	0.02232
Transcriptomic	cJUN	2.30	3.2e-13
Pathway	IL7R	3.95	0.00265
Pathway	PDGFRB	3.03	0.02403
Pathway	EGFR	2.17	0.00527
Pathway	Hypoxia	7.31	3.2e-18
Pathway	Epithelial-Mesenchymal Transition	6.84	8.1e-17
Pathway	TNFa signaling via NFkB	5.96	9.6e-14
Pathway	Glycolysis	4.72	1.8e-9
Pathway	mTORC1 signaling	3.40	2.2e-5
Pathway	Inflammatory response	3.05	2.0e-4
Pathway	Reactive O2 Species	4.52	0.00653
Pathway	IL signaling	4.42	0.00235
Pathway	Integrin signaling	3.35	0.00235
Pathway	ECM organization	4.78	2.0e-11
Pathway	Collagen formation	8.50	6.3e-9
Pathway	Beta1 integrin	8.79	1.6e-7
Pathway	Beta3 integrin	10.59	1.8e-6
Pathway	HIF1A network	6.07	1.8e-4
Cell types	U87MG CNS	5.27	1.4e-28

Cell types	U118MG CNS	3.53	3.0e-18
Cell types	Smooth muscle	9.53	2.9e-45
Cell types	Adipocyte	3.61	5.6e-5

Table 5.21. Characterizing features of GSC Unsupervised cluster A with respect to B. Odds ratios (OR) and relative multiple tests-adjusted p-values are reported. CNS: central nervous system, ECM: extracellular matrix, IL: interleukin.

Feature	ID	OR	Adj p-value
Transcriptomic	ASCL1	8.18	4.2e-36
Transcriptomic	OLIG2	16.53	6.7e-82
Transcriptomic	OLIG1	16.04	8.9e-80
Transcriptomic	NFkB1 DOWN	15.07	3.1e-75
Pathway	Neural crest differentiation	5.94	4.0e-6
Pathway	Oligodendrocyte specification	8.95	0.00102
Pathway	GBM signaling	4.31	0.00264
Pathway	Chemical synapses	3.65	4.7e-6
Pathway	Medulloblastoma	6.00	2.8e-7
Pathway	Ependymoma	8.10	0.00018
Pathway	Neuronal System	2.70	0.00083
Pathway	GBM PN	24.99	0.00072
Pathway	P53 pathway	4.97	0.00133
Pathway	Notch pathway	4.26	0.02845
Pathway	Hedgehog signaling	5.95	0.00473
Pathway	Rb DNA damage response	13.00	0.02430
Disease	Pilocytic astrocytoma	4.64	0.00022
Disease	Ependymoma	4.13	0.00117
Disease	Primary brain tumor	4.31	0.00661
Disease	Anaplastic astrocytoma	3.21	0.00662
Disease	Recurrent brain tumor	7.20	0.00662
Disease	Brainstem neoplasm	41.56	0.01864

Table 5.22. Characterizing features of GSC Unsupervised cluster B with respect to A. Odds ratios (OR) and relative multiple tests-adjusted p-values are reported. GBM: glioblastoma, PN: proneural.

Feature	ID	OR	Adj p-value
Transcriptomic	HIF1A	6.69	1.5e-6
Transcriptomic	cJUN	3.74	2.1e-9
Transcriptomic	SMAD2 & SMAD3	2.65	8.7e-5
Transcriptomic	YAP1	2.07	0.00396
Pathway	Epithelial-Mesenchymal Transition	21.61	2.9e-23
Pathway	Hypoxia	16.52	3.9e-17
Pathway	Glycolysis	7.99	1.9e-5
Pathway	TNF α signaling via NF κ B	7.24	0.00001
Pathway	Inflammatory response	4.40	0.00592
Pathway	Angiogenesis	6.97	0.08443
Pathway	ECM organization	9.39	1.7e-9
Pathway	ECM-receptor interaction	12.27	0.00008
Pathway	PI3K-Akt signaling	6.92	2.6e-7
Pathway	Focal adhesion	7.20	0.00008
Pathway	Fructose and mannose metabolism	11.93	0.02983
Disease	Li-Fraumeni syndrome MES stem cells	13.70	5.8e-26
Cell types	U87MG CNS	7.05	7.4e-12
Cell types	U118MG CNS	4.33	3.6e-7
Cell types	Smooth muscle	14.34	6.9e-23
Cell types	Adipocyte	8.05	8.0e-6

Table 5.23. Characterizing features of WB-defined MES GSCs with respect to WB-defined PN GSCs. Odds ratios (OR) and relative multiple tests-adjusted p-values are reported. CNS: central nervous system, ECM: extracellular matrix.

Feature	ID	OR	Adj p-value
Transcriptomic	ASCL1	15.11	5.7e-6
Transcriptomic	ADAMTS19	13.26	0.00002
Transcriptomic	OLIG1	8.21	0.00205
Pathway	IFNa response	21.58	0.00013
Pathway	IFNgamma response	10.13	0.00199
Pathway	Spinal cord injury	13.64	0.01712
Pathway	CMP phosphorylation	41.02	0.05174
Pathway	Superpathway of pyrimidine ribonucleotides de novo biosynthesis	28.39	0.05174
Ontology	Neuron projection	6.02	0.00954
Ontology	Dendrite	7.43	0.02108
Cell types	Cancer stem cells (skin)	25.59	0.00002
Cell types	Neuronal progenitors	25.59	0.00002
Cell types	Neural stem cells	20.46	0.00034
Cell types	Glial cells (brain)	17.11	0.00220
Cell types	Dorsal striatum	8.61	5.5e-12
Cell types	Prefrontal cortex	8.61	5.5e-12
Cell types	Cerebral cortex	7.44	1.4e-10
Cell types	Fetal brain	6.91	6.4e-10
Cell types	Oligodendrocyte	6.91	6.4e-10
Cell types	DAOY	3.90	0.00035
Cell types	ASCL1 UP	5.17	0.00009

Table 5.24. Characterizing features of WB-defined PN GSCs with respect to WB-defined MES GSCs. Odds ratios (OR) and relative multiple tests-adjusted p-values are reported.

Consistently, the two MES clusters (Cluster A and WB-MES GSCs), show similar phenotypes, as both groups result related to classical glioma cell lines (U87MG and U118MG) and to other normal mesenchyme-derived lineages (adipocytes and smooth muscle cells). Notably, Cluster A, which, as previously stated, is a subset of the WB-

MES GSCs, has much higher significance levels for these phenotypes with respect to WB-MES. On the other hand, epithelial to mesenchymal transition features are strikingly more significant in our WB-MES (OR 21.61 p-value 2.9×10^{-23}) than in Cluster A (OR 6.84 p-value 8.1×10^{-17}). This suggests that non-MES to MES transition is very similar to epithelial to mesenchymal transition and is a key process in GBM progression. Therefore, the classification of our MES lines in unsupervised Clusters A and B weakens the difference in these stigmata, while our WB-MES GSCs comprehend all cell lines with a significant degree of mesenchymalization.

On the contrary, Cluster A groups only MES lines that are most highly divergent from cluster B. In line with this observation, hypoxic (WB-MES OR 16.52 p-value 3.9×10^{-17} , Cluster A OR 7.31 p-value 3.2×10^{-18}), glycolytic (WB-MES OR 7.99 p-value 1.9×10^{-5} , Cluster A OR 7.72 p-value 1.8×10^{-9}) and inflammatory response pathways (WB-MES OR 4.40 p-value 5.9×10^{-3} , Cluster A OR 3.05 p-value 2.0×10^{-4}), as well as important master genes related to these pathways such as HIF1a and NF-kB, are only present or generally more enriched in Cluster A than in WB-MES.

Performing the same analysis on the two non-MES/PN clusters (Cluster B and WB-PN), both groups show significant similarity to normal brain components (e.g., dorsal striatum and prefrontal cortex) and to less malignant glial tumors or even benign tumors such as pilocytic astrocytoma and ependymoma. Notably, the actual pathway stigmata typical of PN GBMs are recognized in Cluster B, in addition to neuronal tumors (medulloblastoma) and their derived cell lines (DAOY) in WB-PN GSCs. Both groups show enrichment for pathways involved in CNS developmental stages (e.g., neural crest differentiation, fetal brain, neuronal progenitors, neural stem cells) as well as in normal glial counterparts (glial cells in brain, oligodendrocyte, oligodendrocyte specification) and in healthy neuronal components (dendrite, neuron projection, chemical synapses). Also, both groups, but especially Cluster B, show enrichment of known PN master genes such as ASCL1 (OR 8.18 p-value 4.2×10^{-36}), OLIG1 (OR 16.04 p-value 8.9×10^{-80}) and OLIG2 (OR 16.53 p-value 6.7×10^{-82}).

Therefore, MES GSCs are characterized by features of inflammatory response, epithelial to mesenchymal transition, hypoxia and glycolysis and show a more deranged and undifferentiated state, while non-MES/PN GSCs show features of differentiated glial or progenitor neural cells and of normal healthy components of neural cells.

5.5 Assessment of the *in vivo* abilities of MES and PN GSCs to model GBM subgroups.

As our GSC lines showed a marked subgroup drift under NSA culturing conditions, as previously reported (Bhat *et al*, 2013; Nakano, 2015), we further proceeded to assess their *in vivo* potential to replicate reliably subgroup-specific features in preclinical *in vivo* GBM models. As a matter of fact, MES GSCs as well as GCLs induce tumors that grow as lowly infiltrative, compact, expansive, and sometimes intraventricular masses, which grossly distort brain parenchyma and may cause hydrocephalus. On the contrary, PN GSCs usually grow more infiltratively, spreading through myelin fiber-bundles and resembling lower grade gliomas (see Figure 5.24).

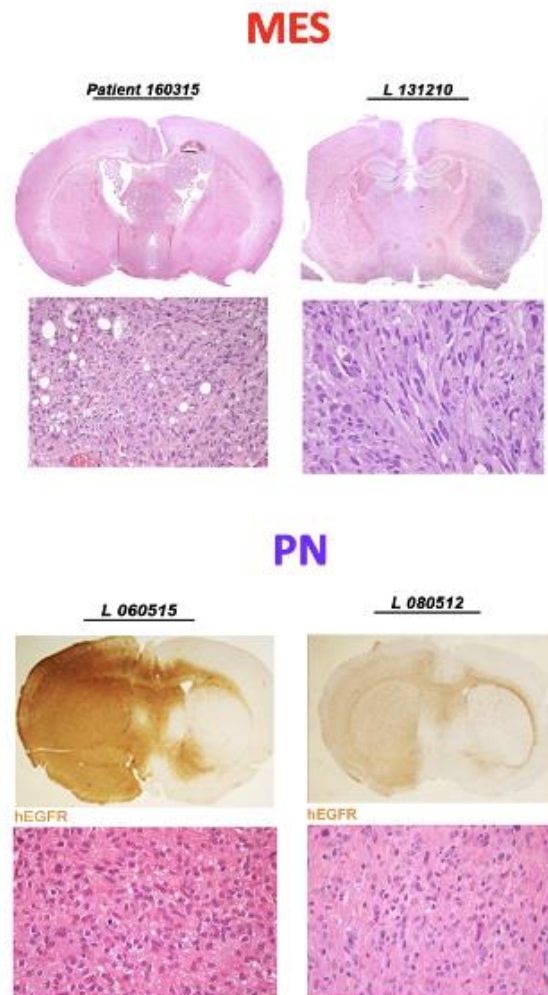


Figure 5.24. *In vivo* growth pattern of established MES and PN GSCs. MES lines grow as more compact masses, easily detectable by H&E staining and may cause hydrocephalus. PN lines variably infiltrate mouse brain parenchyma and may be difficult to detect by H&E staining; thus, IHC for human EGFR is necessary to detect infiltrating cells along white matter tracts. MES lines give rise to tumors that resemble actual human high-grade gliomas, while PN lines more often form tumors that mimic lower grade gliomas. H&E: hematoxylin and eosin, IHC: immunohistochemistry, MES: mesenchymal, PN: proneural.

Therefore, we transplanted all the 14 GSC lines that we isolated from our patients in the striatum of immunocompromised mice and followed the development of tumoral masses by serial MRI and evaluation of mice health status.

Nine out of the total 14 lines generated tumors that were detectable at MRI, while 5 failed to give rise to xenografts, as also confirmed by hematoxylin and eosin sections. GSC lines that gave rise to xenografts took from a minimum of 2 months to a maximum of 6 to generate full-blown lesions. Lines that failed to produce tumors were followed for up to 6 months before sacrificing the mice and confirming the negative result by pathological analysis. Out of the 5 GSC lines that did not generate full blown xenografts, lines 180515 and 160622 gave rise to grafts comprising a few tumor cells as detected by H&E staining; however, they were too scant for warranting subsequent investigations.

Based on MRI aspect of the tumors at T2-weighted, post-contrast T1 and diffusion/NODDI sequences and on the kinetics of tumor growth, we assigned a radiology-based subgroup pattern to the xenografts, as provided in Table 5.25. As a matter of fact, PN xenografts appear more infiltrating, do not take up contrast and grow more slowly, while MES xenografts are more compact, do not infiltrate, show contrast enhancement, and grow faster. CL tumors are variably infiltrating, with noticeable nodules and show variable enhancement. Figure 5.25 shows an example of radiologic PN and MES patterns.

Patient	GSC WB early	GSC WB late	Time to full blown xenograft	Xenograft MRI growth pattern	Xenograft radiologic subgroup
160315	MES	MES	6 months	Compact mass, contrast enhancement	MES
160526	PN/CL*	PN/CL*	4-5 months	Infiltrating mass, non-enhancing	PN
160704	PN	PN/MES	2 months	Compact mass, non-enhancing	CL
160503	MES	MES	No tumor	No tumor	No tumor
160525	MES	MES	No tumor	No tumor	No tumor
160622	CL/PN	CL/PN	No tumor	No tumor	No tumor
161019	CL/MES*	CL/MES*	2-3 months	Variably infiltrating, variably enhancing	PN/CL
161205	PN/CL	PN/CL	4 months	Variably infiltrating, non-enhancing	PN
170421	CL*	CL/MES*	2 months	Variably infiltrating, lowly enhancing	CL
180420	CL/MES	CL/MES	2-3 months	Compact mass, non-enhancing, high T2	MES
180515	MES	CL/MES	No tumor	No tumor	No tumor
190125	nd	nd	4-5 months	Compact mass, non-enhancing	CL
190920	PN/CL	PN/CL	2 months	Compact and infiltrating, lowly enhancing	CL
191119	CL/MES	.	No tumor	No tumor	No tumor

Table 5.25. Radiological growth pattern of GSC-derived xenografts compared to WB classification of the same lines. 9/14 lines generated MRI-detectable lesions. Radiological subgroup was defined based on MRI aspect of the tumors (infiltrating vs compact lesions, contrast uptake) and kinetics of tumor growth. GSC lines that maintain EGFRvIII are indicated with an asterisk (*). CL: classical, MES: mesenchymal, nd: not defined, PN: proneural.

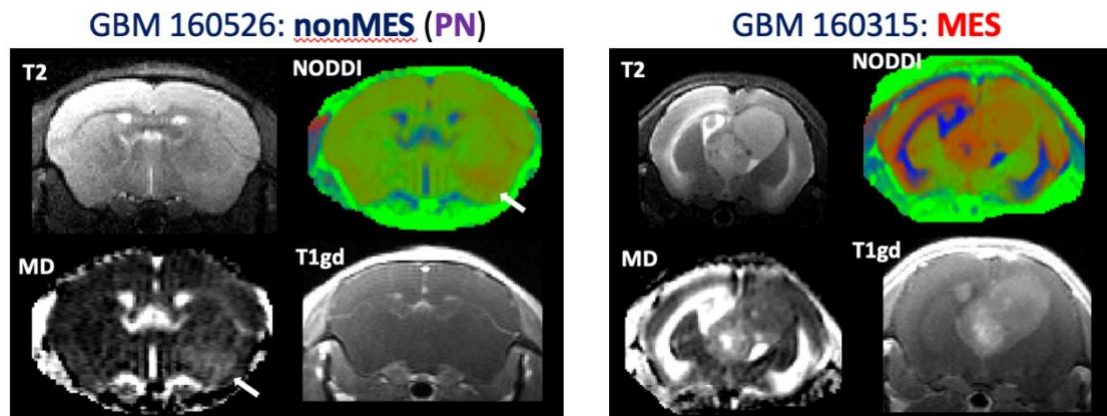


Figure 5.25. Radiological analysis of growth pattern of GSCs. GBM 160526 shows a very infiltrating lesion that is visible only in NODDI tricompartmental diffusion map and mean diffusivity map (red and white signals, respectively; white arrows), while T2 and post-contrast T1 sequences look unaffected. GBM 160315 generates a large compact mass that develops also intraventricularly causing hydrocephalus and showing gadolinium uptake. MD: mean diffusivity, MES: mesenchymal, PN: proneural, T1gd: T1+gadolinium.

We then performed the same IHC panel analysis to determine the transcriptional subgroup affiliation of the xenografts, as done with the parental human tumors. Figure 5.26 shows IHC analysis of patient samples 180420 and 161205 and their derived xenografts. Table 5.26 summarizes the transcriptional subgroups of the original GBMs and of their relative xenografts, as determined by IHC. As it is evident, most of the times the affiliation of the xenografts perfectly matches that of the original human tumor, and, when a variation is observed, is generally minor. Table 5.27 reports the percentage of each subgroup-specific cell component for the tumors that generated GSCs and for their derived xenografts.

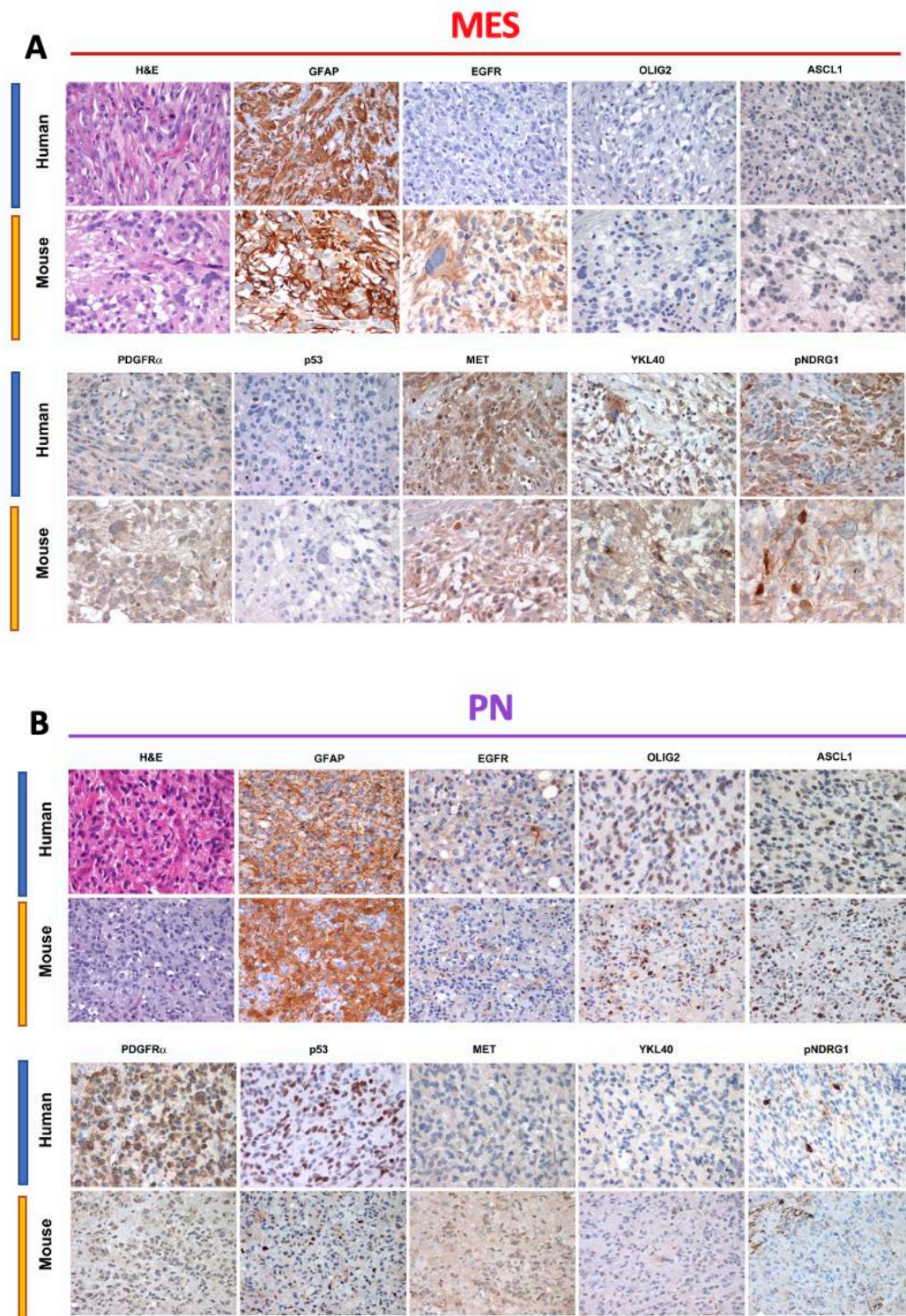


Figure 5.26. IHC panel comparing the expression of subgroup-specific classifiers in the original human GBM samples and their corresponding GSC-derived xenografts. A, GBM 180420 is a MES tumor that gave rise to a MES GSC line that generated a MES xenograft. B, GBM 161205 is a PN/CL tumor that gave rise to a PN GSC line that in turn formed a PN xenograft. Intensity of staining signal is generally very similar between the human and the murine counterparts with very limited divergences (see EGFR in A and PDGFR α in B). H&E: hematoxylin end eosin, MES: mesenchymal, PN: proneural.

Patient	Tumor IHC	Xenograft IHC
160315	MES	MES
160526	CL	CL
160704	CL	PN/CL
160503	CL	No tumor
160525	PN	No tumor
160622	CL	Only few cells
161019	MIX	CL
161205	PN/CL	PN
170421	CL	CL
180420	MES	MES
180515	PN/CL	Only few cells
190125	PN/CL	PN/CL + MIX
190920	CL	CL
191119	PN/CL	No tumor

Table 5.26. Transcriptional subgroup affiliation of parental human tumors and their GSC-derived xenografts as determined by IHC. 9/14 lines generated full-blown lesions, 2 showed only few cells that were too scant for panel analysis, 3 did not show tumoral cells at all. CL: classical, MES: mesenchymal, MIX: balanced mix of the 3 subtypes, PN: proneural.

Pts	%PN	%CL	%MES	Subtype	%PN	%CL	%MES	Subtype
	Human	Human	Human	Human	Xeno	Xeno	Xeno	Xeno
160315	9.2	1.4	89.4	MES	22	13	65	MES
160526	18.2	81.8	0	CL	17.5	82.5	0	CL
160704	21.8	77.8	0.4	CL	48	45	7	PN/CL
161019	32.4	46.8	20.8	MIX	16.5	82.5	1	CL
161205	53	44.6	2.4	PN/CL	68	30	2	PN
170421	9	91	0	CL	20	70	10	CL
180420	20.2	5.2	74.6	MES	16.5	8.5	75	MES
190125	56.8	41.6	1.6	PN/CL	51.6	46.2	2.2	PN/CL
					39.4	35	25.6	MIX
190920	33.6	65.4	1	CL	26.7	72.6	0.7	CL

Table 5.27. Percentage of transcriptional subgroup-specific components in parental human tumors and their GSC-derived xenografts as assessed by IHC (on tumors for which also xenografts were available). Of note, the line derived from patient 190125 generated two different types of xenografts, one PN/CL and a tumor that was a mixture of the 3 subgroups. CL: classical, MES: mesenchymal, MIX: balanced mix of the 3 subtypes, PN: proneural, Pts: patients, Xeno: xenografts.

5.6 Summary of subgroup evolution from original human GBM to *in vitro* GSC line, to *in vivo* xenograft.

Table 5.28 provides a comprehensive summary of the subgroup affiliation evolution from parental human GBM to *in vitro* GSCs, and finally to *in vivo* xenografts. By setting the IHC panel-based diagnosis as the standard reference for assessing the most reliable affiliation, we generated 1 stable MES GSC line from 1 MES tumor, while in the remaining 13 cases there was a variable shift from parental GBM to their derived cell lines. However, once transplanted *in vivo*, lines derived from 2 MES, 3 CL and 1 PN/CL human GBMs gave rise to xenografts identical to the parental patients' GBM specimen, while in 3 additional cases there was a slight change (from 1 CL GBM to PN/CL xenograft, from 1 PN/CL GBM to PN xenograft and from 1 mixed affiliation human tumor to CL xenograft). The remaining 5 cases (all CL and PN/CL original tumors) failed to generate xenografts, with only scant tumoral cells detectable in case of patients 160622 and 180515.

Taken together, this indicates a substantial reversion *in vivo* of the subgroup drift observed *in vitro*. As a matter of fact, all GSC lines showed a change in subgroup affiliation when transplanted into mice striatum, except for 1 MES, 1 PN and 1 CL (170421, with a very minor MES component) cell lines, which were derived respectively from a MES, a PN/CL and a CL tumor. Notably, the non-tumorigenic GSC lines were mostly MES, as out of five, 2 were purely MES, 2 balanced or MES-prevalent CL/MES mixes and only one was PN/CL.

Patient	Tumor IHC	Tumor WB	GSC WB early	GSC WB late	Xenograft radiologic subgroup	Xenograft IHC
160315	MES	MES	MES	MES	MES	MES
160526	CL	CL	PN/CL*	PN/CL*	PN	CL
160704	CL	CL	PN	PN/MES	CL	PN/CL
160503	CL	CL/MES	MES	MES	No tumor	No tumor
160525	PN	CL	MES	MES	No tumor	No tumor
160622	CL	CL	PN/CL	PN/CL	No tumor	Only few cells
161019	MIX	CL	CL/MES*	CL/MES*	PN/CL	CL
161205	PN/CL	MES	PN	PN	PN	PN
170421	CL	CL/MES	CL*	CL/MES*	CL	CL
180420	MES	CL	CL/MES	CL/MES	MES	MES
180515	PN/CL	CL/MES	MES	CL/MES	No tumor	Only few cells
190125	PN/CL	CL	nd	nd	CL	PN/CL + MIX
190920	CL	CL/MES	PN/CL	PN/CL	CL	CL
191119	PN/CL	CL	CL/MES	.	No tumor	No tumor

Table 5.28. Table summarizing the subgroup affiliation of the original GBM samples, their derived GSC lines at early and intermediate/late in vitro passages under NSA conditions, and of their corresponding xenografts. GSC lines that maintain EGFRvIII are indicated with an asterisk (*). CL: classical, MES: mesenchymal, nd: not defined, PN: proneural.

5.7 Analysis of metabolic pathways differentiating MES and PN GSCs.

We next took advantage of a collateral untargeted metabolomic analysis, performed on PN and MES xenografts generated from the transplantation of high-passages PN GSC lines (previously isolated in our lab), recently established MES lines (160315, 131210) and GCLs (U87). This experiment was performed for another lab project, primarily focusing on metabolic pathways specific to PN and MES tumors (Pieri *et al*, 2022).

Briefly, we integrated differentially expressed genes and differentially enriched metabolites to specifically interrogate the metabolic pathways potentially involved in subgroup affiliation. By means of bioinformatic analysis we identified 23 significantly upregulated metabolites in PN xenografts and 237 in MES xenografts. We subsequently integrated these findings with only the genes pertaining to the 4 signatures (respectively the 23 PN metabolites with Cluster B and WB-PN signatures, and the 237 MES metabolites with Cluster A and WB-MES signatures) by the MetaboAnalyst webtool. We identified Arginine and Proline metabolism (p-value 0.006), pentose and glucuronate interconversions (p-value 0.010), pantothenate and coenzyme A biosynthesis (p-value 0.011) and nitrogen metabolism (p-value 0.048) for PN GSCs. However, they did not remain significant at FDR test correction. On the contrary, we identified aminoacyl-tRNA biosynthesis (p-value $6.3e-6$), Valine, Leucine and Isoleucine biosynthesis (p-value $3.2e-5$), pyrimidine metabolism (p-value $5.4e-4$), Lysine degradation (p-value 0.007), purine metabolism (p-value 0.006), taurine and hypotaurine metabolism (p-value 0.017), Histidine metabolism (p-value 0.024), pantothenate and coenzyme A biosynthesis (p-value 0.030), Glycine, Serine and Threonine metabolism (p-value 0.031) and Phenylalanine metabolism (p-value 0.036) for MES GSCs. Out of these, only aminoacyl-tRNA biosynthesis (FDR p-value $5.2e-4$), Valine, Leucine and Isoleucine biosynthesis (FDR p-value 0.001), pyrimidine metabolism (FDR p-value 0.015) retained their significance in multiple test correction.

In order to be less stringent and possibly increase the number of significant metabolic pathways identified, we matched the metabolites with all the significant genes identified by RNA Sequencing analyses, without limiting them only to the top-ranking genes used to determine the signatures but including all differentially expressed genes. Therefore, we included 922 genes for Cluster B, 716 for Cluster A, 55 for WB-PN and 169 for WB-MES.

By matching significant PN metabolites and PN genes (all Cluster B and WB-PN), we identified Arginine and Proline metabolism (p-value 0.006), pentose and glucuronate interconversions (p-value 0.010), pantothenate and coenzyme A biosynthesis (p-value 0.011), glycerolipid metabolism (p-value 0.002), Alanine, Aspartate and Glutamate metabolism (p-value 0.005), Nitrogen metabolism (p-value 0.028), butanoate metabolism (p-value 0.042) and phosphatidylinositol signaling system (p-value 0.048), as being specific for PN GBMs. None of them, however, retained significance at multiple correction test. When we matched MES metabolites and MES genes (all Cluster A and WB-MES), we identified Valine, Leucine and Isoleucine metabolism (p-value $6.0e-5$), aminoacyl-tRNA biosynthesis (p-value $2.5e-$

5), pyrimidine metabolism (p-value 4.5e-4), Lysine degradation (p-value 0.008), thiamine metabolism (p-value 0.008), purine metabolism (p-value 0.009), glycolysis or gluconeogenesis (p-value 0.009), nicotinate and nicotinamide metabolism (p-value 0.011), taurine and hypotaurine metabolism (p-value 0.024), galactose metabolism (p-value 0.030), Histidine metabolism (p-value 0.037), pantothenate and coenzyme A biosynthesis (p-value 0.045), Phenylalanine metabolism (p-value 0.049), riboflavin metabolism (p-value 0.050), and amino sugar and nucleotide sugar metabolism (p-value 0.013). Only Valine, Leucine and Isoleucine metabolism (FDR p-value 0.003), aminoacyl-tRNA biosynthesis (p-value 0.002) and pyrimidine metabolism (p-value 0.013) maintained significance at multiple test adjustment.

5.8 Dynamics of transcriptional subgroup evolution of GBMs with respect to healthy tissue.

5.8.1 Inferring a possible evolutionary model of GBM transcriptional subgroup composition.

As we demonstrated a dynamic balance in the proportion and prevalence of the three subgroups promoted by various, reversible switches from the original tissue to *in vitro* culture to *in vivo* xenografts, we decided to investigate the relation between them and healthy tissue by exploiting transcriptional data of human tumors.

We matched transcriptional affiliation of TCGA GBM samples from R2 platform with downloadable mRNA sequencing data from a cohort of TCGA GBMs as described in the Methods section. Transcriptional data were available for primary tumors and their recurrences. Due to the possible change in affiliation from primary GBM to recurrence, we discarded data from relapsed tumors. Table 5.29 reports the so-obtained 4 CL, 9 MES, 6 PN and 5 Healthy TCGA samples. We then performed differential gene expression between each pathological state and the healthy counterparts and between each pathological state with the other two. Table 5.30 shows the absolute and percentual numbers of genes for each comparison and demonstrates an increasing number of significantly differentially expressed genes respectively for PN, CL, and MES GBMs compared to healthy controls. Comparing pathological states, instead, an increasing divergence is noted from CL vs PN, to MES vs CL, to MES vs PN samples.

CL	MES	PN	Healthy
TCGA.06.0125.01	TCGA.02.0055.01	TCGA.02.0047.01	TCGA.06.0675.11
TCGA.06.0125.02	TCGA.06.0130.01	TCGA.06.0129.01	TCGA.06.0678.11
TCGA.06.0157.01	TCGA.06.0152.02	TCGA.06.0238.01	TCGA.06.0680.11
TCGA.06.0158.01	TCGA.06.0168.01	TCGA.06.0646.01	TCGA.06.0681.11
TCGA.06.0211.01	TCGA.06.0184.01	TCGA.12.0616.01	TCGA.06.AABW.11
TCGA.06.0211.02	TCGA.06.0190.01	TCGA.12.0618.01	
	TCGA.06.0190.02		
	TCGA.06.0210.01		
	TCGA.06.0210.02		
	TCGA.06.0644.01		
	TCGA.06.0645.01		
	TCGA.12.0619.01		

Table 5.29. TCGA GBM samples for which both mRNA sequencing data and transcriptional affiliation were retrievable. The suffix indicates primary (.01) or recurrent (.02) tumor. Only primary tumors were retained for analyses. CL: classical, MES: mesenchymal, PN: proneural.

Comparison	# DEGs	# Significant DEGs	# Upregulated significant DEGs	# Downregulated significant DEGs
CL vs Healthy	18048	9642 (53.4%)	4555 (25.2%)	5087 (28.2%)
PN vs Healthy	18195	8440 (46.4%)	4308 (23.7%)	4132 (22.7%)
MES vs Healthy	18106	10514 (58.1%)	5195 (28.7%)	5319 (29.4%)
CL vs PN	17951	766 (4.3%)	285 (1.6%)	481 (2.7%)
MES vs CL	17857	2876 (16.1%)	1856 (10.4%)	1020 (5.7%)
MES vs PN	17981	3391 (18.9%)	1711 (9.5%)	1680 (9.4%)

Table 5.30. Number of total and significant differentially expressed genes for each comparison between pathologic subgroups and healthy state or within subgroups. Significance defined as FDR adjusted p-Value ≤ 0.05 . Comparing pathological and healthy states there is an increasing number of significant DEGs from PN vs healthy to CL vs Healthy, to MES vs Healthy. Comparing pathological states an increasing divergence is noted from CL vs PN to MES vs CL, to MES vs PN samples. CL: classical, DEGs: differentially expressed genes, FDR: false discovery rate, MES: mesenchymal, PN: proneural.

We then performed GSEA analysis interrogating each comparison with Verhaak's transcriptional signatures. In particular we were interested in testing the subtype signatures' behavior in contrast with healthy state, as the gene sets were determined only within the boundaries of the pathological states. (Verhaak *et al*, 2010) Figures 5.27, 5.28 and 5.29 show enrichment plots and statistics for the 3 subtypes as opposed to basal healthy state. Remarkably, neural signature resulted constantly enriched in healthy controls for all three pathological states, as previously reported (Wang *et al*, 2017). On the other hand, PN signature was consistently enriched in PN subtype as opposed to normal counterpart, while an opposite trend (i.e., enrichment in healthy tissues) was seen when interrogating the MES vs healthy and CL vs healthy DEGs for the same signature. Interestingly, both CL and MES signatures were invariably enriched in all pathologic states compared to controls, even though only MES reached statistical significance in all cases, while CL signature retained significance only in CL vs healthy and MES vs healthy controls.

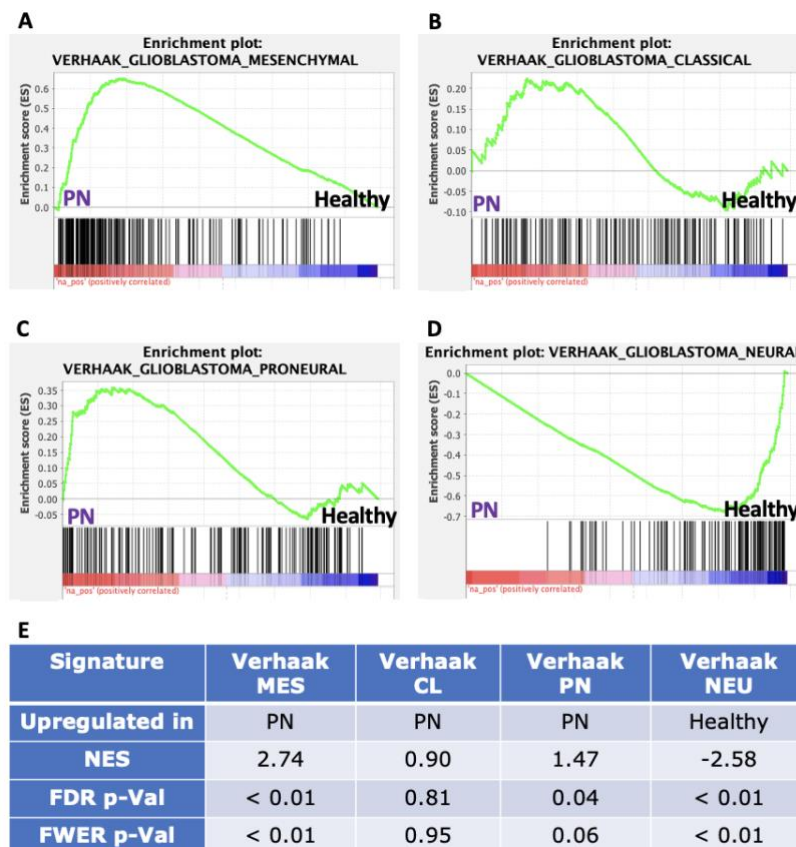


Figure 5.27. Enrichment plots and statistics of Verhaak's transcriptional signatures in the TCGA PN GBMs vs Healthy controls. A, MES, B, CL, C, PN, D, NEU signatures. E, table with enrichment statistics: CL signature is the only signature that does not reach significance. CL: classical, FDR: false discovery rate, FWER: family-wise error rate MES: mesenchymal, NES: normalized enrichment score, NEU: neural, PN: proneural.

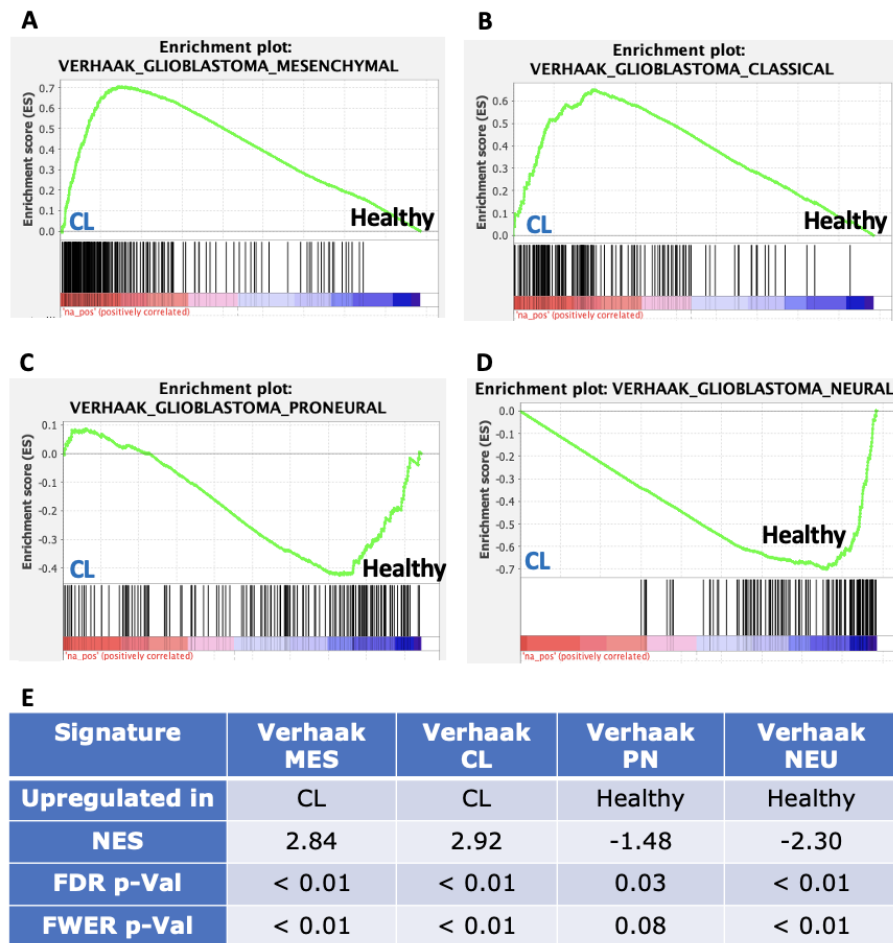


Figure 5.28. Enrichment plots and statistics of Verhaak's transcriptional signatures in the TCGA CL GBMs vs Healthy controls. A, MES, B, CL, C, PN, D, NEU signatures. E, table with enrichment statistics. CL: classical, FDR: false discovery rate, FWER: family-wise error rate MES: mesenchymal, NES: normalized enrichment score, NEU: neural, PN: proneural.

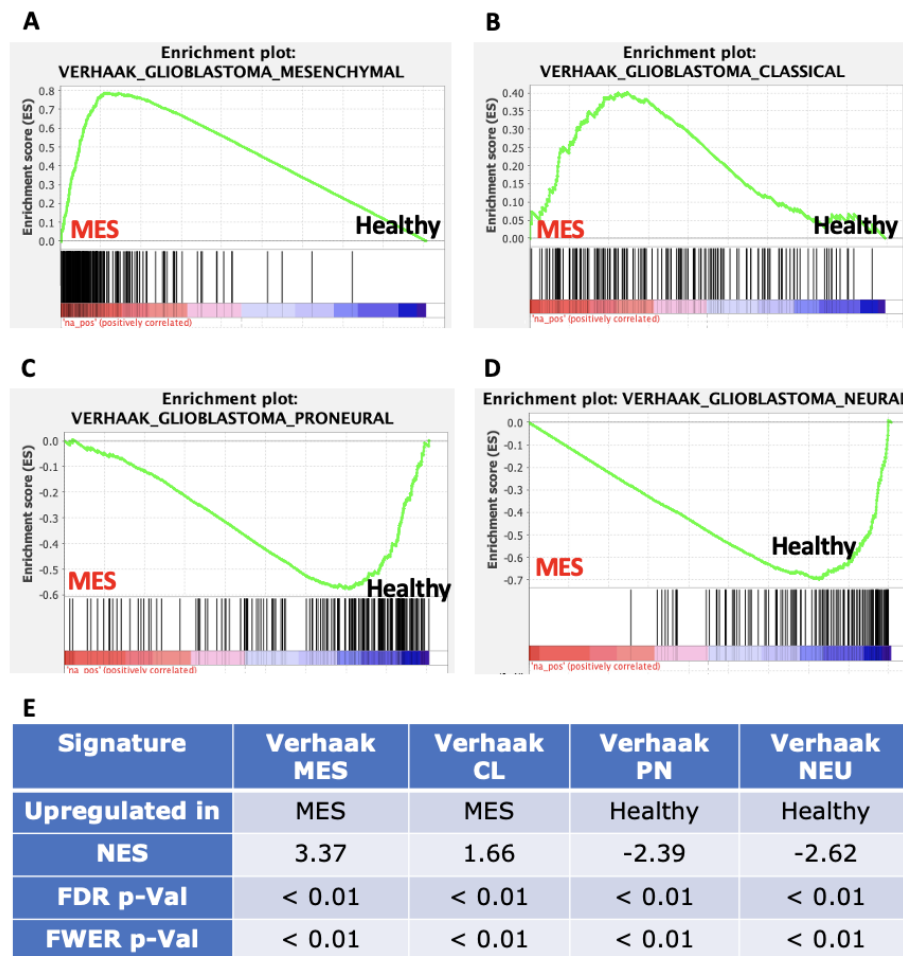


Figure 5.29. Enrichment plots and statistics of Verhaak's transcriptional signatures in the TCGA MES GBMs vs Healthy controls. A, MES, B, CL, C, PN, D, NEU signatures. E, table with enrichment statistics. CL: classical, FDR: false discovery rate, FWER: family-wise error rate MES: mesenchymal, NES: normalized enrichment score, NEU: neural, PN: proneural.

We also performed GSEA on the pairwise combinatorial contrasts between the three transcriptional subgroups of GBM: Figures 5.30, 5.31 and 5.32 show enrichment plots and statistics for these analyses. It is fundamental to notice that normalized enrichment score (NES) sign indicates only in which side of the contrast a particular gene set is enriched, while the actual entity of the enrichment is indicated by the absolute value of the NES itself. Notably, MES signature is enriched in MES vs PN and CL, and in CL vs PN, with increasing absolute value of NES from CL vs PN comparison to MES vs PN. Parallely, CL signature is enriched in CL vs PN and MES, and in MES vs PN, with the lowest absolute value of NES in MES vs PN and the highest in CL vs PN comparison. The PN signature is consistently enriched in PN vs CL and MES, and in CL vs MES, with increasing NES absolute values from PN vs CL to CL vs MES, to PN vs MES. Lastly, neural signature is significantly enriched in PN vs CL and

MES, while it did not reach statistical significance in the MES vs CL contrast. Table 5.31 provides a summary of the enrichment statistics in the pairwise contrasts.

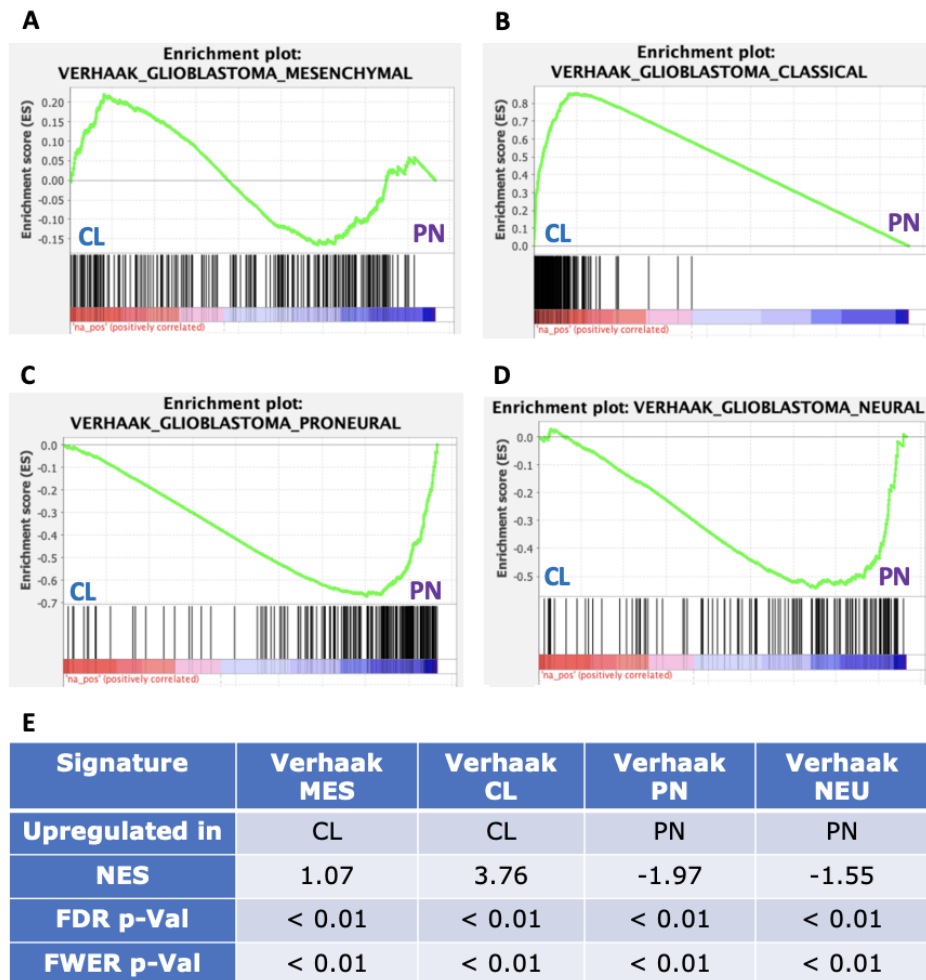


Figure 5.30. Enrichment plots and statistics of Verhaak's transcriptional signatures in the TCGA CL vs PN GBMs. A, MES, B, CL, C, PN, D, NEU signatures. E, table with enrichment statistics. CL: classical, FDR: false discovery rate, FWER: family-wise error rate MES: mesenchymal, NES: normalized enrichment score, NEU: neural, PN: proneural.

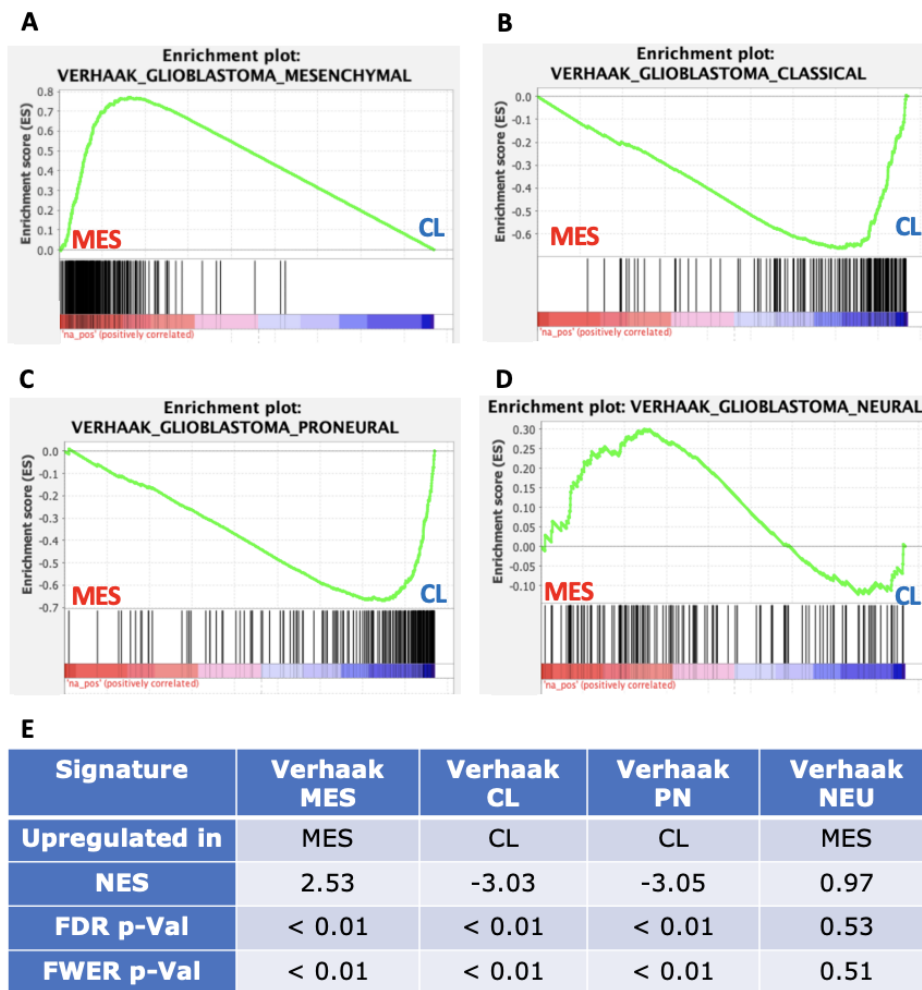


Figure 5.31. Enrichment plots and statistics of Verhaak's transcriptional signatures in the TCGA MES vs CL GBMs. A, MES, B, CL, C, PN, D, NEU signatures. E, table with enrichment statistics. NEU signature did not reach statistical significance. CL: classical, FDR: false discovery rate, FWER: family-wise error rate MES: mesenchymal, NES: normalized enrichment score, NEU: neural, PN: proneural.

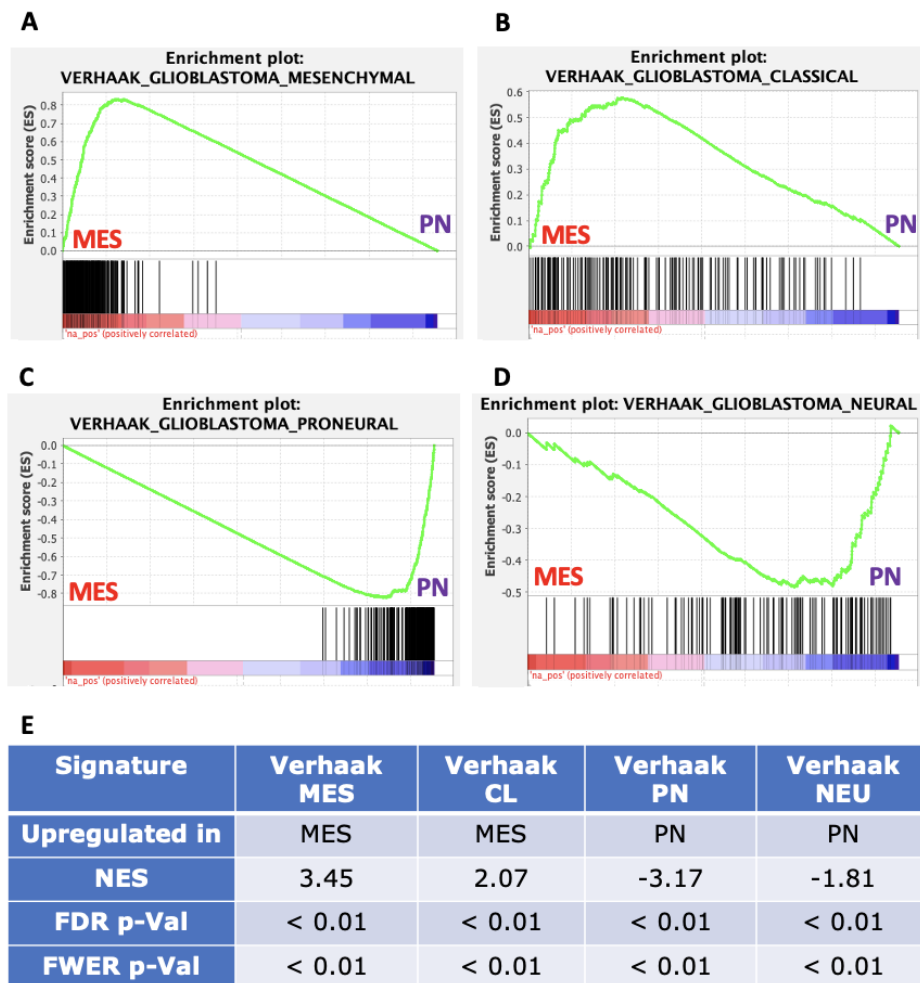


Figure 5.32. Enrichment plots and statistics of Verhaak's transcriptional signatures in the TCGA MES vs PN GBMs. A, MES, B, CL, C, PN, D, NEU signatures. E, table with enrichment statistics. CL: classical, FDR: false discovery rate, FWER: family-wise error rate MES: mesenchymal, NES: normalized enrichment score, NEU: neural, PN: proneural.

MES Signature	MES vs PN	MES vs CL	CL vs PN
Upregulated in	MES	MES	CL
Absolute value of NES	3.45	2.53	1.07
FDR p-Val	< 0.01	< 0.01	< 0.01
FWER p-Val	< 0.01	< 0.01	< 0.01
CL Signature	MES vs PN	MES vs CL	CL vs PN
Upregulated in	MES	CL	CL
Absolute value of NES	2.07	3.03	3.76
FDR p-Val	< 0.01	< 0.01	< 0.01
FWER p-Val	< 0.01	< 0.01	< 0.01
PN Signature	MES vs PN	MES vs CL	CL vs PN
Upregulated in	PN	CL	PN
Absolute value of NES	3.17	3.05	1.97
FDR p-Val	< 0.01	< 0.01	< 0.01
FWER p-Val	< 0.01	< 0.01	< 0.01
NEU Signature	MES vs PN	MES vs CL	CL vs PN
Upregulated in	PN	MES	PN
Absolute value of NES	1.81	0.97	1.55
FDR p-Val	< 0.01	0.53	< 0.01
FWER p-Val	< 0.01	0.51	< 0.01

Table 5.31. Enrichment statistics of Verhaak’s transcriptional signatures in pairwise contrasts between the three pathological subtypes of TCGA GBMs. CL: classical, FDR: false discovery rate, FWER: family wise error rate, MES: mesenchymal, NES: normalized enrichment score, NEU: neural, PN: proneural.

We also interrogated the differentially expressed genes between the pathological subtypes and the healthy controls with the signatures we derived from our GSCs, as shown in Table 5.32. Remarkably, the significance threshold was reached only for the WB-MES signature in all the comparisons, being enriched in the tumoral side of the contrast, in addition to Cluster A signature in MES tumors and Cluster B and WB-PN gene sets in healthy controls when opposed to MES GBMs. Figure 5.33 shows enrichment plots and statistics for the MES tumors vs healthy controls contrast. These results suggest that MES tumors are the most divergent from the basal normal state, as the pathological signatures can satisfactorily discern the disease from the normal brain in this comparison, but not when the pathological term of the comparison are

CL or PN GBMs. Additionally, the fact that WB-MES signature is enriched in all GBM subtypes vs healthy brain seems to suggest that a certain degree of mesenchymalization is implicated in GBM initiation. On the contrary, enrichment pattern of Verhaak's and our PN signatures may indicate the PN subtype as the ground state of GBM.

PN vs Healthy	Cluster A	Cluster B	WB MES	WB PN
Upregulated in	PN	PN	PN	PN
NES	1.01	0.85	1.53	1.01
FDR p-Val	0.75	0.81	0.03	0.61
FWER p-Val	0.84	0.98	0.03	0.84
CL vs Healthy	Cluster A	Cluster B	WB MES	WB PN
Upregulated in	Healthy	Healthy	CL	Healthy
NES	-1.22	-0.91	1.28	-0.82
FDR p-Val	0.22	0.83	0.08	0.82
FWER p-Val	0.64	1.00	0.08	1.00
MES vs Healthy	Cluster A	Cluster B	WB MES	WB PN
Upregulated in	MES	Healthy	MES	Healthy
NES	1.70	-2.36	2.44	-1.45
FDR p-Val	< 0.01	< 0.01	< 0.01	0.02
FWER p-Val	< 0.01	< 0.01	< 0.01	0.08

Table 5.32. Enrichment statistics of GSC-derived transcriptional signatures in contrasts between the three pathological subtypes of TCGA GBMs and healthy controls. Statistical significance was reached for all signatures only in the MES vs healthy contrast. CL: classical, FDR: false discovery rate, FWER: family wise error rate, MES: mesenchymal, NES: normalized enrichment score, PN: proneural.

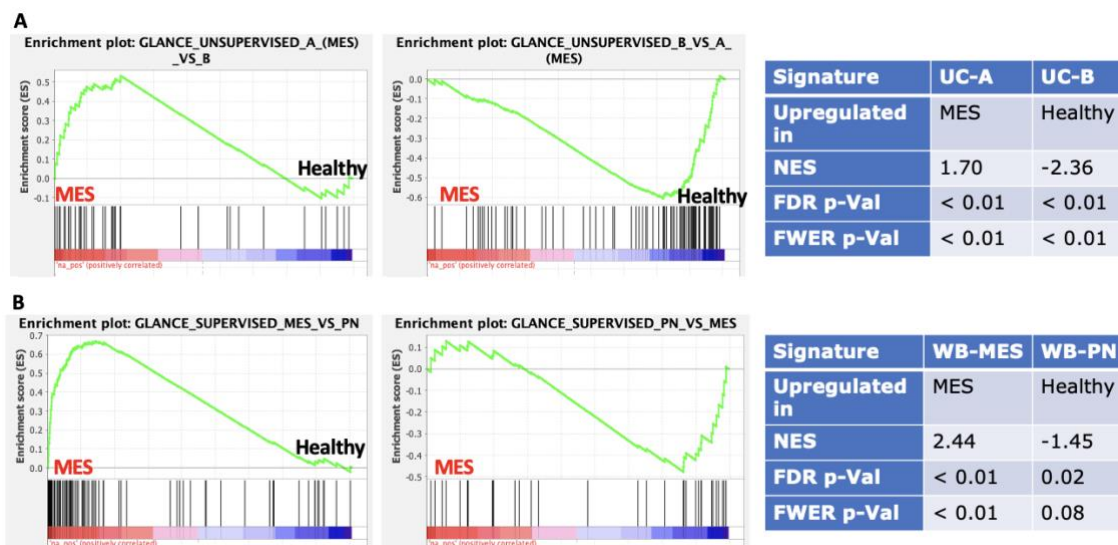


Figure 5.33. Enrichment plots and statistics of our transcriptional signatures in the TCGA MES GBMs vs Healthy controls. A, signatures derived from unsupervised clusters A and B. B, signatures derived from WB-panel defined PN and MES GSCs. MES signatures (UC-A and WB-MES) are significantly enriched in MES tumors while non-MES/PN signatures (UC-B and WB-PN) in healthy controls. CL: classical, FDR: false discovery rate, FWER: family-wise error rate MES: mesenchymal, NES: normalized enrichment score, PN: proneural, UC: unsupervised cluster, WB: western blot.

5.9 Identification of molecular pathways and phenotypes characterizing each GBM subgroup in contrast to healthy brain tissue.

We next exploited the differential gene expression between the pathological states of GBM and healthy brain to perform an extended pathway analysis to identify characterizing features of each subgroup. The rationale was to denote which features would be common to all subtypes (thus, possibly inherent to GBM initiation) and which would differentiate the comparisons, suggesting progressive divergence from basal healthy state. Tables 5.33, 5.34 and 5.35 show the EnrichR results of the most significant characterizing features of PN, CL and MES tumors with respect to normal brain.

Feature	ID	OR	Adj p-value
Transcriptomic	E2F4	3.97	1.4e-65
Transcriptomic	TP53	3.11	1.1e-9
Transcriptomic	MYCN	4.57	0.00003
Transcriptomic	NFkB1	2.17	5.4e-8
Transcriptomic	STAT3	2.54	9.3e-6
Pathway	Rb gene in cancer	8.19	1.7e-18
Pathway	DNA mismatch repair	10.36	6.3e-6
Pathway	E2F targets	7.60	6.1e-42
Pathway	IFNgamma response	5.04	3.2e-27
Pathway	Epithelial-Mesenchymal Transition	3.42	7.8e-16
Pathway	Angiogenesis	4.08	0.00012
Pathway	Inflammatory response	2.11	3.2e-6
Pathway	mTORC1 signaling	1.89	0.00010
Pathway	p53 pathway	1.85	0.00017
Pathway	Superpathway of pyrimidine synthesis	4.87	0.04448
Pathway	Integrin signaling	2.29	0.00009
Cell types	Microglia in cerebellum	3.46	6.5e-45
Cell types	Microglia in cerebrum	3.58	4.2e-37

Table 5.33. Characterizing features of TCGA PN GBMs with respect to normal brain tissue. Odds ratios (OR) and relative multiple tests-adjusted p-values are reported. PN: proneural.

Feature	ID	OR	Adj p-value
Transcriptomic	E2F4	3.51	2.8e-54
Transcriptomic	MYC	1.93	1.3e-11
Transcriptomic	SOX2	1.49	6.0e-6
Transcriptomic	KLF4	1.43	6.3e-6
Transcriptomic	NANOG	2.03	0.00075
Pathway	Rb gene in cancer	6.29	4.0e-14
Pathway	E2F targets	7.52	1.8e-41
Pathway	DNA mismatch repair	7.90	0.00009
Pathway	Epithelial-Mesenchymal Transition	3.96	5.0e-20
Pathway	IFNgamma response	4.21	1.5e-21
Pathway	Angiogenesis	4.32	0.00006
Pathway	Inflammatory response	1.91	0.00006
Pathway	mTORC1 signaling	1.59	0.00471
Pathway	p53 pathway	1.71	0.00105
Pathway	Superpathway pyrimidine synthesis	4.60	0.04036
Pathway	Integrin signaling	2.69	3.7e-7
Ontology	ECM organization	2.54	4.5e-11
Ontology	Focal adhesion	2.77	1.3e-18
Cell types	Microglia in cerebellum	3.46	6.5e-45
Cell types	Microglia in cerebrum	3.58	4.2e-37
Cell types	Vascular endothelial cells in cerebellum	2.21	4.5e-17
Cell types	Vascular endothelial cells in cerebrum	2.21	7.2e-16

Table 5.34. Characterizing features of TCGA CL GBMs with respect to normal brain tissue. Odds ratios (OR) and relative multiple tests-adjusted p-values are reported. CL: classical, ECM: extracellular matrix.

Feature	ID	OR	Adj p-value
Transcriptomic	E2F4	2.93	8.1e-41
Transcriptomic	MYC	2.22	3.8e-18
Transcriptomic	KLF4	1.53	7.0e-9
Transcriptomic	STAT3	1.55	2.1e-7
Transcriptomic	NFkB1	3.49	7.9e-24
Transcriptomic	HIF1A	3.40	1.3e-6
Pathway	TCR regulation of apoptosis	3.32	1.1e-42
Pathway	IFNgamma response	8.11	1.9e-43
Pathway	Epithelial-Mesenchymal Transition	6.05	1.3e-33
Pathway	E2F targets	5.52	7.2e-31
Pathway	IL6/JAK/STAT3 signaling	7.56	1.1e-18
Pathway	Inflammatory response	4.64	2.5e-25
Pathway	Hypoxia	3.07	3.3e-14
Pathway	Angiogenesis	5.72	8.8e-7
Pathway	Glycolysis	2.95	2.7e-13
Pathway	p53 pathway	2.45	2.1e-9
Pathway	mTORC1 signaling	2.17	3.2e-7
Pathway	Immune system	1.91	2.3e-27
Pathway	Integrin signaling	3.03	1.3e-9
Pathway	Chemokine/cytokine signal inflammation	2.32	1.1e-6
Pathway	IL signaling	2.27	0.00258
Pathway	SLC2A4	2.32	3.1e-21
Pathway	IL7R	3.11	1.1e-8
Cell types	Microglia in cerebellum	4.98	7.7e-79
Cell types	Microglia in cerebrum	4.42	1.2e-52
Cell types	U118MG CNS	2.04	3.2e-19

Table 5.35. Characterizing features of TCGA MES GBMs with respect to normal brain tissue. Odds ratios (OR) and relative multiple tests-adjusted p-values are reported. CNS: central nervous system, IL: interleukin, MES: mesenchymal, TCR: T-cell receptor.

As expected, many features were conserved in all the comparisons, as each contrast was carried out comparing a transcriptional subgroup of the same disease (i.e., GBM) with healthy tissue. Notably, however, significance values changed drastically from one comparison to the other. As a matter of fact, the enrichment in the transcription factor NF- κ B1 had a p-value ranging from $5.4e-8$ in PN tumors to $7.9e-24$ in MES GBMs. Additionally, enrichment for fundamental tumoral or tumor-exploited dynamics such as epithelial to mesenchymal transition and angiogenesis consistently characterized all subtypes, with progressive increase in significance from PN to CL, to MES (p-values respectively $7.8e-16$ to $5.0e-20$, to $1.3e-33$ for the former and $1.2e-4$ to $6.0e-5$, to $8.8e-7$ for the latter). Conversely, E2F4 decreased its significance from $1.4e-65$ in PN to $2.8e-54$ in CL, to $8.1e-41$ in MES, or E2F targets from $6.1e-42$ in PN to $1.8e-41$ in CL, to $7.2e-31$ in MES tumors.

Not surprisingly, enrichment for cell type profile of vascular endothelial cells in cerebellum (p-value $4.5e-17$) and in cerebrum (p-value $7.2e-16$) was peculiar of CL GBMs. As a matter of fact, even though angiogenesis was slightly more significant in MES than CL tumors, CL GBMs are known to be highly vascularized as part of the pleiotropic effect of EGFR, probably the most important master gene for the CL subgroup. Quite peculiar, instead, was the behavior of the profile of microglia in cerebrum and cerebellum, which reaches its nadir in CL tumors (p-value respectively $1.3e-12$ and $5.0e-17$), an intermediate significance in PN (p-value $4.2e-37$ and $6.5e-45$) and its zenith in MES GBMs (p-value $1.2e-52$ and $7.7e-79$). However, this very last observation was predictable and in line with other characterizing features of MES tumors, which present a high contribution of immune and inflammatory response (Hara *et al*, 2021; Gangoso *et al*, 2021). As a matter of fact, MES GBMs show a much higher activation of inflammatory and immune pathways, as shown by the general inflammatory response pathway enrichment, which is equally significant in PN and CL subgroups (p-value $3.2e-6$ and $6.0e-5$) but to a strikingly lower extent than in MES subtype (p-value $2.5e-25$), which is extremely enriched also in interleukins and chemokines pathways.

Additionally, we performed pathway enrichment analysis on the contrast between MES and PN GBMs, as reported in Tables 5.36 and 5.37. According to what found in the same studies conducted on PN and MES GSCs, human PN tumors show activation of ASCL1 and downregulation of NF- κ B1, as well as activation of nervous system developmental processes and affinity to neuron-derived tumors. MES tumors, instead, show activation of genes related to metastasization, inflammation and immune response, and even higher significance in the enrichment for pathways

involved in epithelial to mesenchymal transition (p-value 7.3e-55) and inflammatory response (p-value 5.9e-31). Enrichment for hypoxic features maintain the same very high significance when compared to PN tumors or to healthy tissue (respectively p-value 3.4e-13 and 3.3e-14). Interestingly, when compared to PN tumors, MES GBMs, besides showing high affinity to mesenchymal tissues such as muscle, fat, or blood, show significant activation of mannose metabolism and proteoglycans in cancer.

Feature	ID	OR	Adj p-value
Transcriptomic	NFkB1 downregulation	4.15	1.2e-18
Transcriptomic	ASCL1 knockout downregulation	4.02	1.4e-10
Transcriptomic	DACH2	4.74	3.2e-24
Transcriptomic	ASCL1	4.36	4.2e-21
Pathway	Neural crest differentiation	4.18	0.00002
Pathway	Axon guidance	2.47	0.00001
Pathway	Developmental biology	2.19	0.00002
Pathway	Neuronal system	2.26	0.00047
Pathway	Hedgehog signaling	3.65	0.03815
Pathway	Ionotropic Glutamate receptors	3.99	0.02187
Ontology	Nervous system development	2.37	5.1e-7
Ontology	Synapse organization	3.61	0.00003
Ontology	Axon genesis regulation	5.50	0.00028
Ontology	Neuron generation	2.56	0.00073
Disease	Neuroblastoma	11.89	1.5e-27
Disease	DNET	3.05	0.03870
Cell types	Fetal brain	6.52	8.4e-43
Cell types	Oligodendrocyte precursor	12.33	0.00050

Table 5.36. Characterizing pathway enrichment features of TCGA PN vs MES GBMs. Odds ratios (OR) and relative multiple tests-adjusted p-values are reported. DNET: dysembryoplastic neuroepithelial tumor, MES: mesenchymal, PN: proneural.

Feature	ID	OR	Adj p-value
Transcriptomic	NFkB1	3.83	1.5e-16
Transcriptomic	STAT3	5.07	5.0e-13
Transcriptomic	TWIST2	6.04	0.00184
Pathway	Epithelial-Mesenchymal Transition	12.01	7.3e-55
Pathway	Inflammatory response	7.11	5.9e-31
Pathway	TNFalpha signaling through NFkB1	6.23	3.5e-26
Pathway	IL2 signaling	4.98	6.1e-19
Pathway	IL6 signaling	6.01	1.1e-11
Pathway	IFNgamma response	4.05	9.7e-14
Pathway	Complement	4.05	9.7e-14
Pathway	Hypoxia	3.95	3.4e-13
Pathway	Angiogenesis	6.08	0.00001
Pathway	Mannose metabolism	21.43	0.00944
Pathway	CK-CK receptor interaction	3.77	1.0e-15
Pathway	Proteoglycans in cancer	3.24	9.0e-9
Pathway	IL7R	3.06	0.00012
Pathway	IL signaling pathway	4.18	4.3e-6
Ontology	Immune system	3.97	7.8e-57
Disease	Metastasis	2.24	8.8e-42
Cell types	Smooth muscle	9.19	4.1e-73
Cell types	Adipocyte	4.92	7.1e-17
Cell types	CD14+ Monocytes	5.04	3.6e-35
Cell types	CD133+ Myeloid	3.26	2.0e-27
Cell types	U118MG CNS	4.05	7.3e-48
Cell types	U87MG CNS	3.63	2.0e-29

Table 5.37. Characterizing pathway enrichment features of TCGA MES vs PN GBMs. Odds ratios (OR) and relative multiple tests-adjusted p-values are reported. IL: interleukins, MES: mesenchymal, PN: proneural.

5.10 Exploiting transcriptional analyses for pinpointing subgroup-specific drug vulnerability.

As a final step in the comparison between GBM subgroups and normal healthy brain tissue, we performed a preliminary study aimed at identifying potential drugs to be exploited for repurposing. Basically, putative effective molecules might be identified through an integrated transcriptomic analysis to narrow down the set of compounds to subsequently screen in “wet” experimental settings. The rationale is that, just as a pathological process induces the activation of a specific transcriptional profile, also exposure to selected compounds induces alterations in different cell populations’ transcriptome. By selecting those compounds that would induce a transcriptional modification opposite to the pathological process, we may identify putative useful drugs to revert the pathology.

A publicly screenable database of compounds tested on different cell lineages is available at the Connectivity Map (cMap) consortium. As described in the Methods section, we selected the top 150 significant genes up- and downregulated in each GBM subgroup in relation to healthy tissue, and that were also included in the L1000 gene panel used to infer transcriptional signatures of each compound. Next, we performed the query on cMap webtool and obtained a drug-sensitivity fingerprint for each subtype, as reported in Table 5.38. The more negative the enrichment score, the more the compound is capable to revert the pathological transcriptional signature, thus suggesting a potentially high efficacy of the drug. Interestingly, the three subgroups show variable sensitivity spectra that only partially overlap. As a matter of fact, PN GBMs show possible sensitivity to drug classes that are not apparently effective for CL and MES tumors; conversely CL tumors have peculiar weaknesses that are not shared with the other subtypes. On the other hand, MES GBMs, which are known to be more resistant to pharmacological treatment, appear sensitive to only 3 drug classes, all in common with PN and CL subgroups, and all with less negative enrichment scores, suggesting less efficacy. This trend is even more accentuated if we lower the opposing enrichment score threshold from <-90 to <-85 ; in this case we identify 97 possibly effective drugs for PN tumors, 85 for CL, and only 63 for MES. Of note, this prediction is carried out based on the mean transcriptional response to the screened compounds of 9 cell lines (i.e., skin cancer, 2 lung cancers, liver cancer, colon cancer, breast cancer, 2 prostate cancers, and

normal kidney), none of which unfortunately derives from GBM cell lines that were not available in the Connectivity Map (cMap) database.

Drug class	PN		CL		MES	
	#	ES	#	ES	#	ES
IGF-1 inhibitor	3	-93.50				
Leucine-rich repeat kinase inhibitor	3	-94.72	3	-95.72	3	-91.88
DNA dependent protein kinase inhibitor	4	-96.05				
HDAC inhibitor	20	-99.12	20	-99.39	20	-96.55
Bromodomain inhibitor	6	-99.55	6	-99.41	6	-99.00
Proteasome inhibitor			3	-90.02		
HIF activator			4	-90.65		
Total	36		36		29	

Table 5.38. Summary of drugs putatively effective against each GBM subgroup based on cMap analyses. Averaged compound-induced transcriptional alterations analyzed on 9 cell lines (skin cancer, 2 lung cancers, liver cancer, colon cancer, breast cancer, 2 prostate cancers, and normal kidney) are tested against transcriptional modifications induced by the tumoral state. Opposing enrichment scores identify drugs that may revert tumor-induced alterations. CL: classical, ES: enrichment score, HDAC: histone deacetylase, MES: mesenchymal, PN: proneural #: number.

5.11 Identification of GBM-related mutations in GSCs by whole exome sequencing (WES) analysis.

To further study our GBM models, we submitted our GSC lines to whole exome sequencing as reported in the Methods section. The main aim of this analysis was to pinpoint GBM-associated mutations in our GSC lines. We performed our analyses on a subgroup of 8 GSC lines isolated from our patients, in addition to MES line 131210 and 3 additional GSC lines (2 PN 0104 and 0605 and 1 CL 0627) previously established in our lab and cultured for years, obtaining a stable phenotype both *in vitro* and *in vivo*. As presented in Figure 5.34A, unsupervised clustering of GSC lines subjected to WES generated two main clusters, which consistently separated WB-defined PN/CL and MES lines. Notably, unsupervised clustering of RNA Sequencing (Fig. 5.34B) does not perfectly match that of WES. However, the two analyses were not conducted on the same collection of GSCs, as lines 0605, 0104 and 0627 were not present in RNAseq and line 170421 was not sent for WES. Line 160704 was submitted to both WES and RNAseq at lower *in vitro* passages, thus with a WB-based PN affiliation, while at increasing passages it demonstrated a PN/MES affiliation as previously reported. Interestingly, in RNAseq analysis, this line clustered with mostly non-MES lines, even though was more similar to mixed MES/CL lines 161019 and 170421, both showing a MES component, while in WES unsupervised clustering it was located in the MES subset. This might suggest that line 160704 might present from its establishment some stable exomic alterations (possibly still to be identified) that would induce a MES shift which, at early *in vitro* passages, are not yet reflected at a transcriptional (RNAseq) and protein (WB) level, rendering the MES shift evident only at later passages.

However, all GSC lines pertaining to RNAseq unsupervised cluster A, i.e., those with a particularly striking MES profile, locate in the MES cluster of WES.

Both main WES clusters are further subdivided in 5 total subsets of cell lines identified by 52 high-confidence somatic variants. Mutational analysis performed on the 5 subsets identified the presence of already known GBM-related mutations in all clusters, as reported in Table 5.39.

Remarkably, the last subset, indicated with the letter E in Figure 5.34A, consists of the two MES GSC lines that grow *in vitro* as adherent monolayer even under NSA conditions, and fail to generate xenografts in immunocompromised mice. Mutational analysis identified the presence of already known GBM-related mutations also in

these two cell lines (Table 5.39), thus discarding the hypothesis that lines 160503 and 160525 could be contaminating healthy mesenchymal cells isolated by chance from the relative patients and confirming that they consist of proper tumor cells. The same observation was made for line 160622, a PN/CL line that properly forms neurospheres *in vitro* but fails to generate tumors *in vivo*.

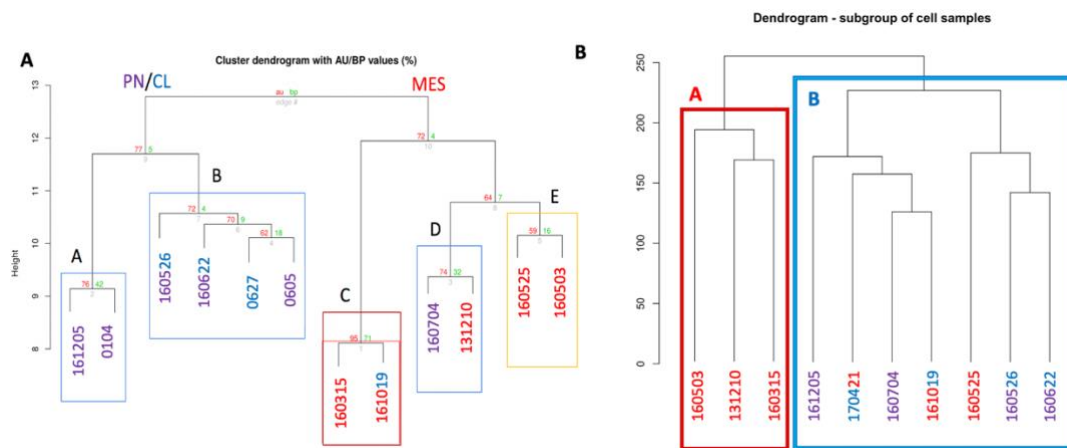


Figure 5.34. Comparison between unsupervised clustering of GSC lines by means of whole exome sequencing (WES, A) and RNA sequencing analysis (RNAseq, B). Color code (CL blue, PN purple, MES red) indicates the WB-panel subgroup affiliation. The two analyses were conducted on partially overlapping collections of samples, as lines 0605, 0104 and 0627 were present in WES but not in RNAseq and line 170421 was sent for RNAseq but not for WES. Line 160704 was submitted for both analyses at lower *in vitro* passages, thus with a WB-based PN affiliation, while at increasing passages it demonstrated a PN/MES affiliation. Interestingly, in RNAseq, this line clustered with non-MES lines, while in WES it was in the MES subset. WES cluster E consists of lines with identical *in vitro* growth pattern and unable to generate xenografts. CL: classical, MES: mesenchymal, PN: proneural Color code (CL blue, PN purple, MES red) indicates the WB-panel subgroup affiliation.

A	B	C	D	E
MLH3	ATRX	NF1	FUBP1	FGFR3
PIK3CA	EGFR	PIK3CA	LZTR1	NF1
RB1	MSH2	PTEN	PDGFRA	PTEN
	PIK3R1	RB1	PIK3R1	TP53
	PTEN	TP53	RB1	
	TP53		TP53	
	BRAF			
	LZTR1			
	NF1			
	POLD1			
	RB1			

Table 5.39. Known GBM-related mutations for each WES subcluster of GSC lines. Name of the clusters refers to those in Figure 5.34A. Notably, lines 160503 and 160525 (cluster E) are characterized by mutations known to be pathogenic for GBM, thus confirming them as being true cancer cells even if they grow in vitro as adherent monolayers under NSA conditions and fail to generate xenografts.

5.12 Identification and characterization of IL7R as a putative marker for GBMs.

Reactome enriched pathway analysis conducted on exome investigation of non-tumorigenic MES GSC lines pointed out the IL7 pathway as the first deranged signaling system. Notably, EnrichR queries carried on RNAseq cluster A (MES) vs cluster B (non-MES) indicated IL7R as significantly enriched hub protein in the MES subgroup (OR 3.95, adjusted p-value 2.7e-3), as well as in TCGA MES GBMs vs healthy brain (OR 3.11, adjusted p-value 1.1e-8) and vs PN GBMs (OR 3.06, adjusted p-value 1.2e-4). Parallely, previous microarray data from our lab, performed by confronting MES GCLs with PN GSCs, identified IL7R as the 31st significantly upregulated gene in GCLs with respect to PN GSCs (log2FC 2.99, adjusted p-value 0.002), as well as its ligand, IL7, ranking 101st (log2FC 2.08, adjusted p-value 9.5e-3). These observations led us to investigate IL7R as a putative novel player in GBM pathogenesis and subgroup specification.

5.12.1 In silico preliminary analysis of IL7R correlation with GBMs and transcriptional subgroups.

In silico analysis on TCGA GBM samples publicly available on the R2 platform showed a weak but significant positive correlation between levels of IL7R expression and those of our MES marker NDRG1 (Pearson's R coefficient 0.26, p-value 1.4e-9) and a moderate negative correlation with our PN marker ASCL1 (Pearson's R coefficient -0.40, p-value 1.0e-21). Overall survival was significantly lower in patients bearing tumors with higher levels of IL7R (Bonferroni correction of p-value 0.025). Progress-free survival was shorter in GBMs with higher expression of IL7R (raw p-value 4.0e-3); however, the test was not significant after Bonferroni correction. Overall survival analysis based on IL7R expression in TCGA GBMs (carried out on the Human Protein Atlas) quantified the 3-year OS in the subset with high IL7R to be 0%, as opposed to 16% in the subset with low IL7R (p-value 0.039). Considering expression levels across subgroups, MES GBMs showed significantly higher levels of IL7R (ANOVA p-value 4.2e-8 for gross comparison). Welch p-values for pairwise comparisons were 4.1e-7 for MES vs PN, 1.5e-3 for MES vs CL and non-significant for CL vs PN tumors. Figure 5.35 summarize in silico evidence for IL7R in TCGA GBMs.

Further investigation on IL7R cancer and cancer stem cells specificity was conducted on TCGA collection using the Human Protein Atlas webtool. Figure 5.36A reports IL7R expression levels across tumors of different origins present in the dataset. Besides showing no particular cancer specificity, IL7R scores generally low in the glioma subset. Figure 5.36B instead depicts gene transcript quantification across cancer cell lines deriving from a variety of tumors. Consistently, there is generally low expression of the gene except for cancerous lymphoid cell lines. Notably, high levels of the transcript are detected in mesenchymal cancer cell lines of non-lymphoid origin, and low, but constantly detectable levels in lines derived from gliomas.

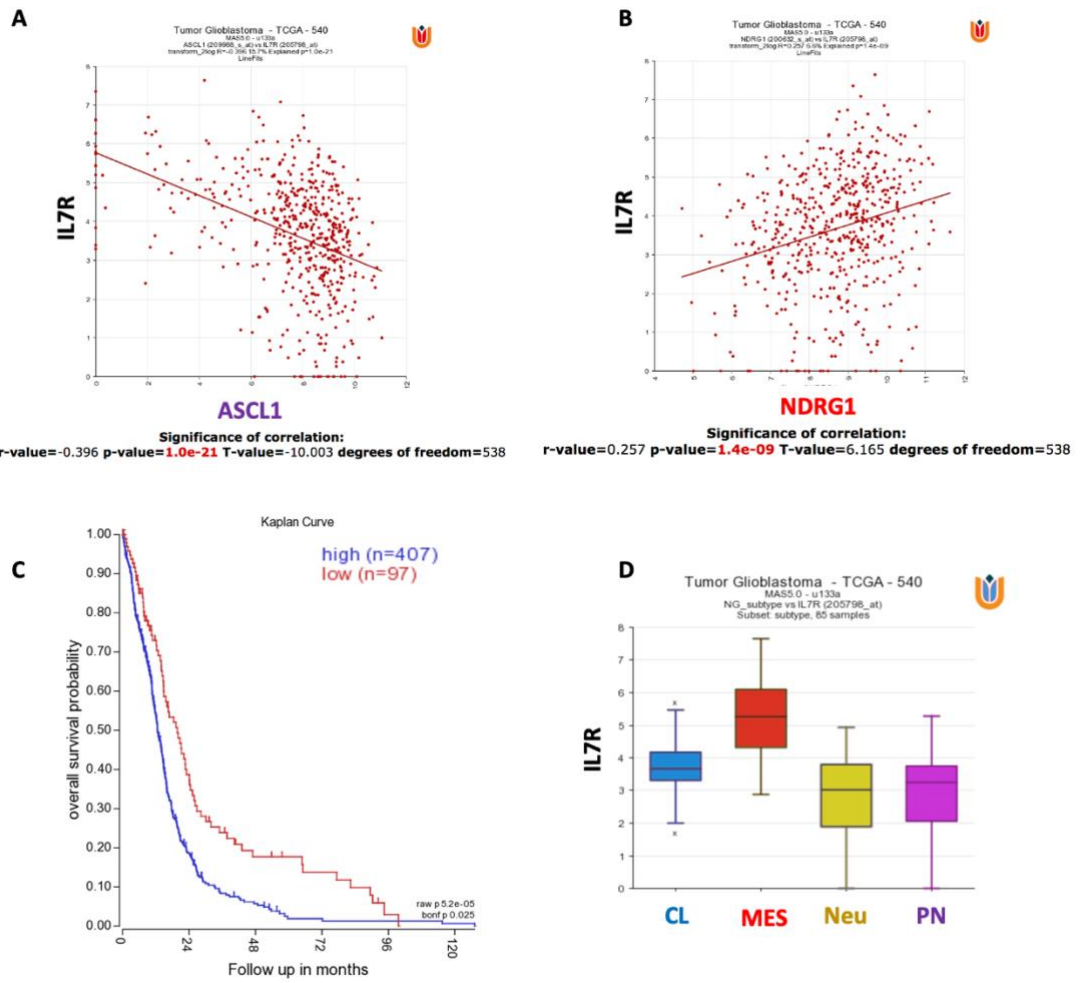


Figure 5.35. In silico analysis of IL7R expression in TCGA GBMs. A, significant moderate negative correlation between PN marker ASCL1 and IL7R in human GBMs (Pearson's coefficient -0.40). B, significant weak positive correlation between MES marker NDRG1 and IL7R in human GBMs (Pearson's coefficient 0.26). C, Kaplan Meier curve for OS in IL7R-high (blue) and low (red) tumors, showing significant lower OS in high expression patients. D, IL7R expression in GBM transcriptional subgroups, depicting significantly higher levels in MES. CL: classical, MES: mesenchymal, Neu: neural, OS: overall survival, PN: proneural.

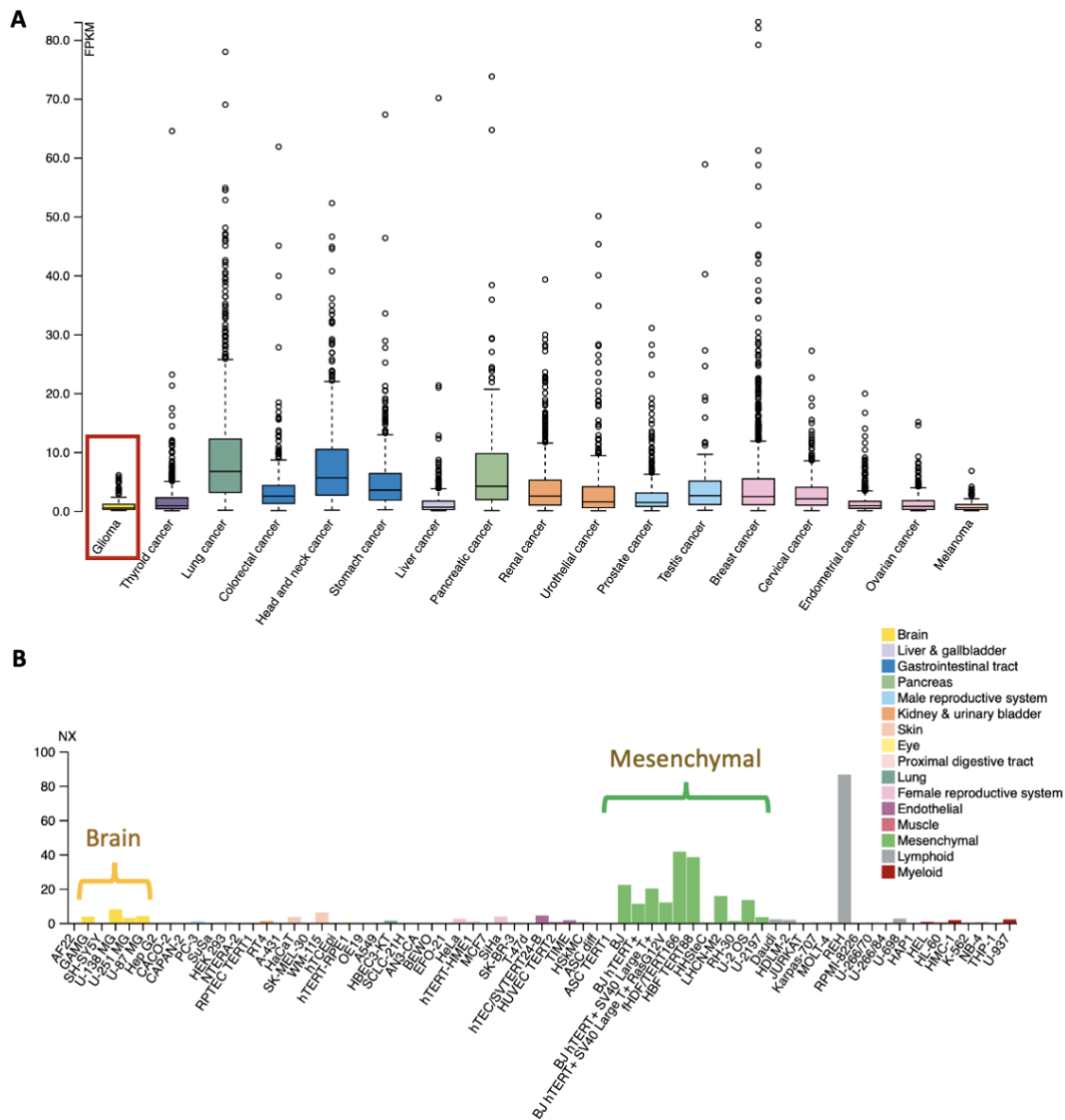


Figure 5.36. In silico analysis of IL7R expression in TCGA cancers and in cancer cell lines. A, low cancer specificity of IL7R transcript quantification; low levels are detected also in gliomas. B, IL7R transcript levels in cancer cell lines; highest levels detected in lymphoid cells (grey) as expected due to the normal activity of IL7R in the lymphoid compartment. Notably, very high levels of the transcript are detected also in mesenchymal cancer cell lines of non-lymphoid origin, or in transformed/immortalized mesenchyme-derived cell lineages. The third group of cell lines with consistently detectable transcript levels is that of glioma-derived cancer cells.

5.12.2 Transcriptional analysis of interleukin-receptors and IL7 pathway.

Given the evidence collected from *in silico* query on IL7R, we decided to interrogate the differentially expressed genes obtained by comparing TCGA GBM subgroups against healthy controls with module signatures for multiple IL-receptors and, more specifically, for IL7 pathway.

The comprehensive interleukin receptors module was significantly enriched in all comparisons with the healthy tissue (in PN with NES 2.03, FDR and FWER adjusted p-value < 0.01, in CL with NES 2.12, FDR and FWER adjusted p-value < 0.01, in MES with NES 2.37, FDR and FWER adjusted p-value < 0.01) and in MES against PN GBMs (NES 2.30, FDR and FWER adjusted p-value < 0.01). Interestingly, as reported in Table 5.40, IL7R was differentially expressed in all the aforementioned contrasts and was in the core enriched gene subsets with increasing ranking for PN vs normal brain (11th gene of the gene set, 2405th in the DEG list), MES vs normal brain (5th gene of the gene set, 697th in the DEG list) and MES vs PN tumors (3rd gene of the gene set, 414th in the DEG list). On the contrary, it was not in the core enrichment genes for the CL vs healthy brain comparison (19th gene of the gene set, 5114th in the DEG list). Figure 5.37 shows enrichment plot for this gene set in the comparisons on the TCGA.

Comparison	Rank in gene set	Rank in DEG list	Core enrichment
PN vs Healthy	11	2405	Yes
CL vs Healthy	19	5114	No
MES vs Healthy	5	697	Yes
MES vs PN	3	414	Yes

Table 5.40. IL7R ranking and core enrichment status for gene set enrichment analysis on differentially expressed genes of TCGA GBMs and healthy controls with comprehensive interleukin receptor gene set. IL7R is always differentially expressed and in the significant core enrichment except for the comparison of CL tumors against healthy brain tissue. CL: classical, DEG: differentially expressed genes, MES: mesenchymal, PN: proneural.

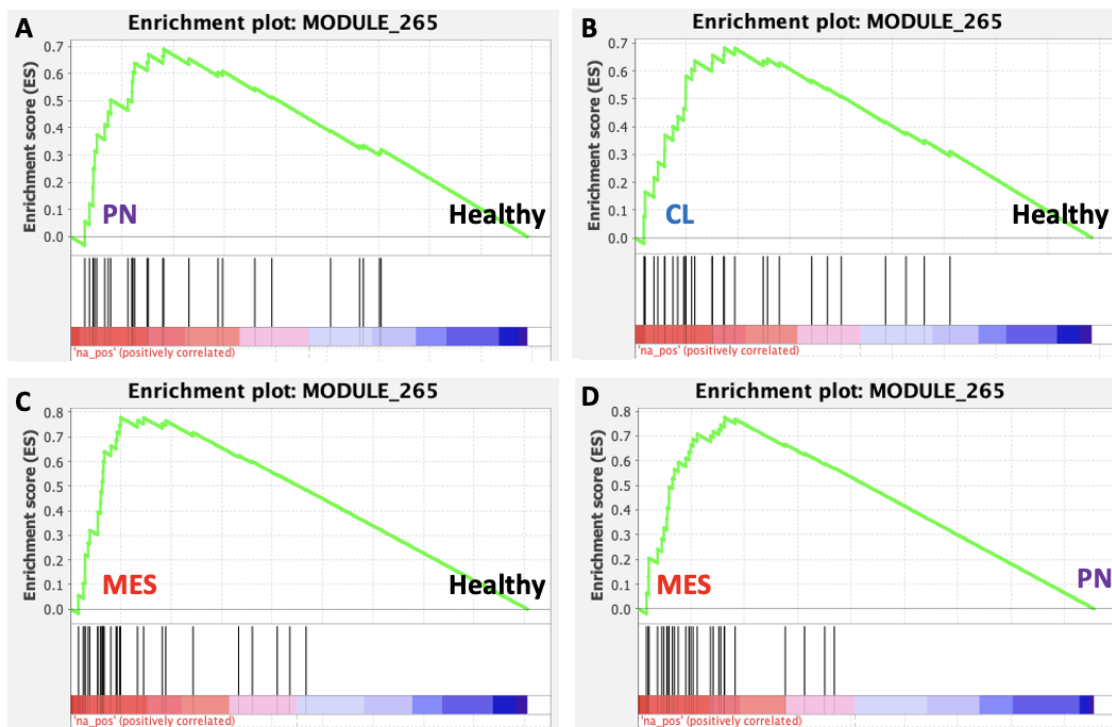


Figure 5.37. Comprehensive Interleukin receptors gene set enrichment analysis on TCGA GBM samples and healthy tissues. A, PN vs healthy brain tissue. B, CL vs healthy brain tissue. C, MES vs healthy brain tissue. D, MES vs PN GBMs. The IL receptors module is significantly enriched in all comparisons against healthy tissue and in MES tumors compared to PN GBMs. CL: classical, MES: mesenchymal, PN: proneural.

Next, we performed the same analysis with the signature genes of IL7 pathway. Parallely to what observed for the previous analysis, IL7 pathway was enriched in all pathological subtypes compared to healthy controls and in MES GBMs vs PN tumors. However, statistical significance was reached only for the MES vs normal brain comparison (NES 1.44, FDR-adjusted p-value 0.07, FWER-adjusted p-value 0.03) and only approached significance in the CL vs normal brain (NES 1.19, FWER-adjusted p-value 0.06), MES vs PN tumors (NES 1.24, FWER-adjusted p-value 0.09), and PN vs normal brain DGEs (NES 1.19, FWER-adjusted p-value 0.12). Nevertheless, interesting qualitative indications could be inferred. As a matter of fact, IL7R behaved in the same manner as in the above analysis, being in the core enrichment of PN vs healthy brain, MES vs healthy brain and MES vs PN tumors, with increasing ranking in the gene set (respectively 5th, 2nd, and 1st), while it was not present in the core enrichment of the CL vs normal brain contrast (7th gene in the gene set). Notably, IL7 was differentially expressed in all comparisons and was always part of the core enrichment genes, even in CL tumors. Table 5.41 shows the rankings of IL7 in the analyses. Remarkably, IL7 was more predominant than IL7R

in PN and CL tumors, while it was the opposite in MES GBMs. Figure 5.38 shows enrichment plot for IL7 pathway gene set in the comparisons on the TCGA.

Comparison	Rank in gene set	Rank in DEG list	Core enrichment	IL7R rank in gene set
PN vs Healthy	1	1131	Yes	5
CL vs Healthy	1	688	Yes	7
MES vs Healthy	3	841	Yes	2
MES vs PN	4	2023	Yes	1

Table 5.41. IL7 ranking and core enrichment status for gene set enrichment analysis on differentially expressed genes of TCGA GBMs and healthy controls with IL7 pathway gene set. IL7 is always differentially expressed and in the significant core enrichment genes. Notably, it is predominant onto IL7R in CL and PN tumors, while it comes after its receptor in MES GBMs. CL: classical, DEG: differentially expressed genes, MES: mesenchymal, PN: proneural.

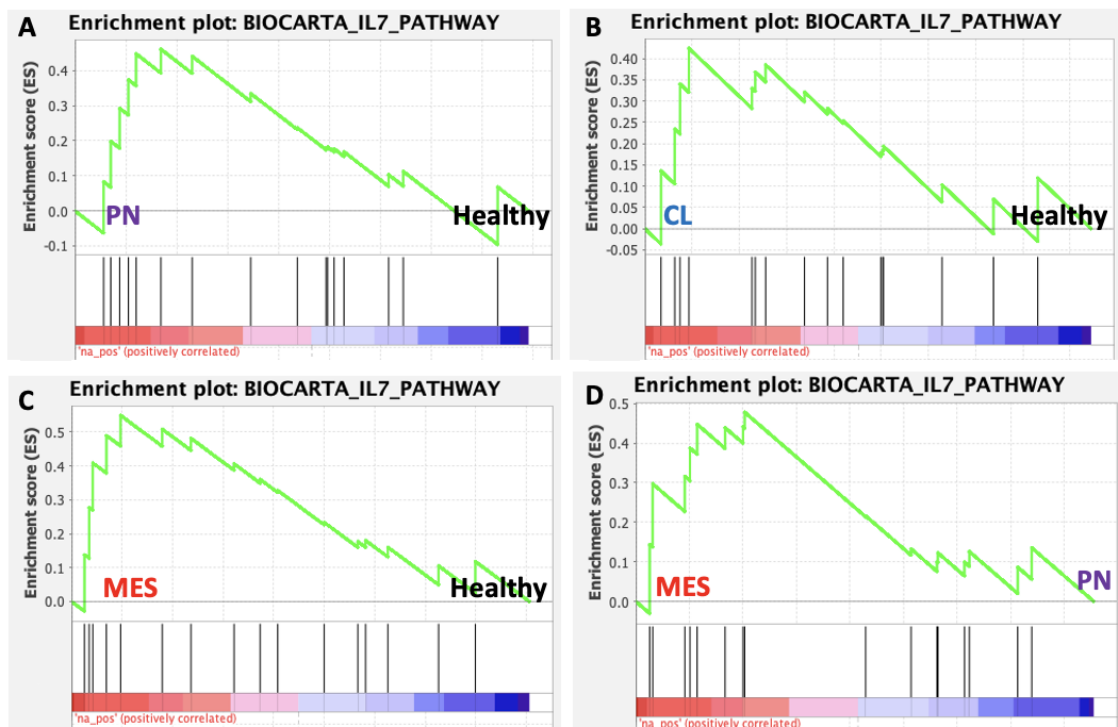
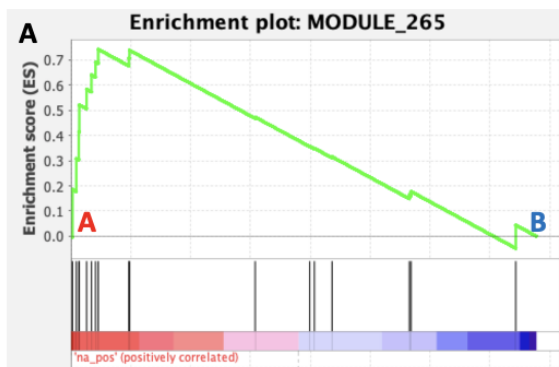


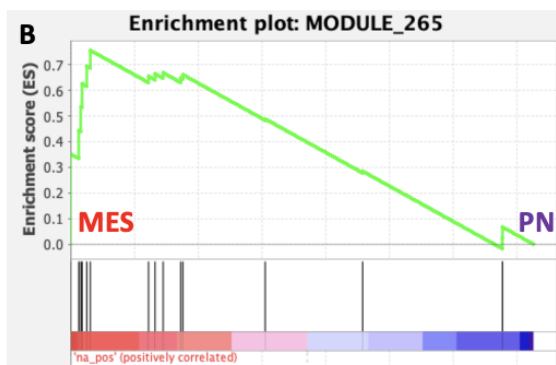
Figure 5.38. IL7 pathway gene set enrichment analysis on TCGA GBM samples and healthy tissues. A, PN vs healthy brain tissue. B, CL vs healthy brain tissue. C, MES vs healthy brain tissue. D, MES vs PN GBMs. The IL7 pathway module is significantly enriched only in MES GBMs vs healthy brain comparison, while only approaching significance in the other contrasts. CL: classical, MES: mesenchymal, PN: proneural.

As a following step, we carried out both analyses also on the differentially expressed genes between our MES and PN GSCs. Interestingly, the comprehensive interleukin receptors module was significantly enriched by GSEA in our MES GSCs (either identified by RNA Sequencing unsupervised clustering, i.e., Cluster A, or based on our WB panel) when compared to non-MES/PN cell lines (Cluster B and WB-PN), as reported in Figure 5.39. However, IL7R was not differentially expressed between these GSC subsets. We then performed differential gene expression analysis between the two non-tumorigenic MES GSC lines (160503 and 160525) and the other tumorigenic MES GSCs (160315, 161019 and 131210) and run another gene set enrichment analysis for interleukin receptors. Quite remarkably, this module resulted enriched in the tumorigenic MES lines, even though it missed statistical significance. However, even more strikingly, IL7R was now differentially expressed between the two subsets and was the 1st gene of the gene set to be enriched in the tumorigenic cells, and the 14815th out of 15013 total genes in the DEG list, thus being the 198th gene from the tumorigenic MES side of the comparison.

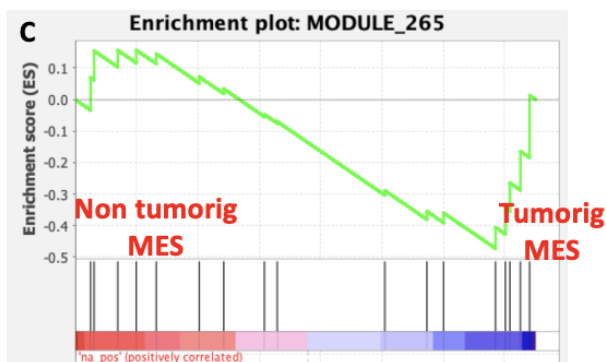
The same analysis conducted with the IL7 pathway gene set returned consistent results, in which IL7 was differentially expressed in Cluster A vs Cluster B and WB-MES vs WB-PN GSC lines but not in the core enrichment. This analysis, however, failed to reach statistical significance. When the DEGs between non-tumorigenic and tumorigenic MES GSCs was interrogated with IL7 pathway gene set, IL7R was the 1st gene of the gene set to be enriched in the tumorigenic cells, and IL7 the 4th in the same side of the comparison, even though it did not reach statistical significance.



Upregulated in	A
NES	1.78
FDR p-Val	< 0.01
FWER p-Val	< 0.01



Upregulated in	MES
NES	1.78
FDR p-Val	< 0.01
FWER p-Val	< 0.01



Upregulated in	Tumorigenic MES
NES	-1.12
FDR p-Val	0.32
FWER p-Val	0.18

Figure 5.39. Comprehensive Interleukin receptors gene set enrichment analysis on GSCs. A, RNA Sequencing unsupervised Cluster A (MES) vs Cluster B (non-MES). B, WB-panel defined MES vs PN. C, WB-panel defined non tumorigenic MES (160503 and 160525) vs WB-panel defined tumorigenic MES (160315, 161019 and 131210). Significant enrichment only in the first two comparisons. MES: mesenchymal, PN: proneural, Tumorig: tumorigenic.

We also measured IL7R transcripts in our GSC lines by means of RT-qPCR. As reported in Figure 5.40, the transcript resulted much higher in tumorigenic MES GSCs than in non-tumorigenic MES or non-MES GSCs, even though not statistically significant, probably due to the limited number of samples per group. Quite remarkably, when we performed RT-qPCR on parental tumor lysates, highest levels of IL7R transcripts were detected in tumors generating non tumorigenic MES GSC lines.

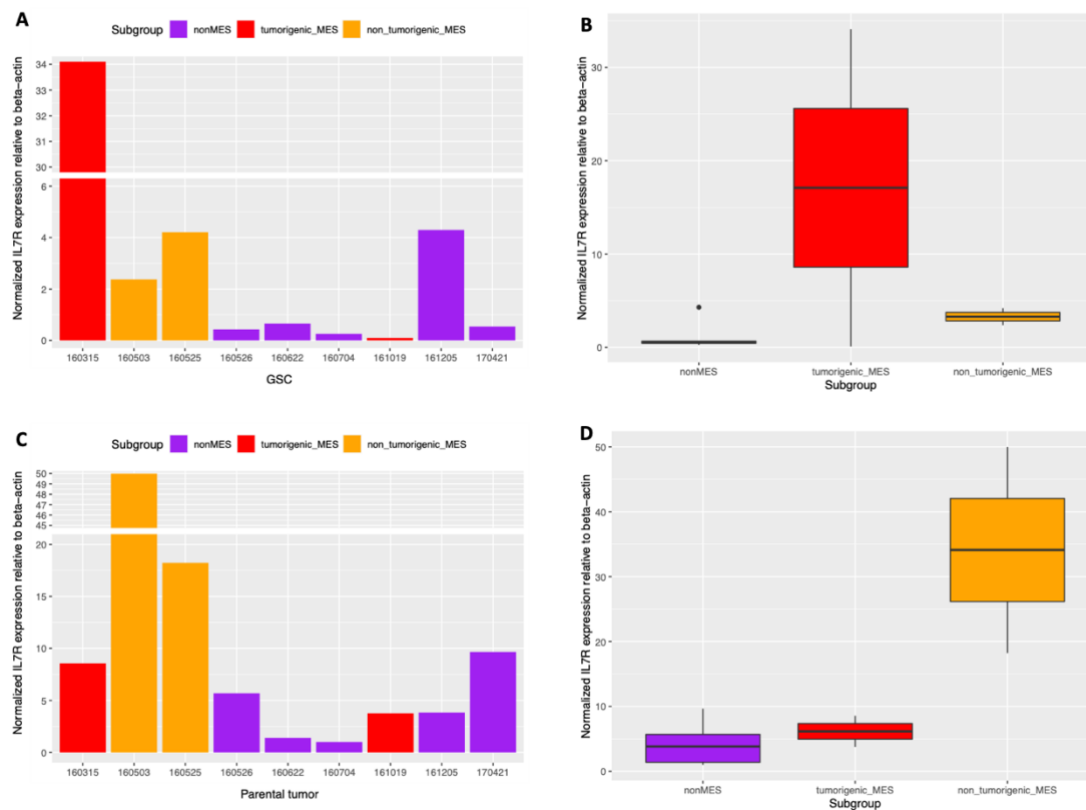


Figure 5.40. RT-qPCR analysis of IL7R in GSCs (A and B) and their parental human tumors (C and D). GSC lines showed higher levels of IL7R mRNA in the tumorigenic MES (red) with respect to non-tumorigenic MES (orange) and non-MES (purple) lines. Parental human GBMs showed instead an opposite trend, with higher levels in those generating non-tumorigenic MES lines.

5.12.3 Protein expression analysis of IL7R in vitro.

As a further step, we investigated the accumulation of IL7R protein in our GSC lines. We first performed WB analysis and relative quantification on 2 of our MES GSCs and GCLs vs 3 previously established PN lines (Figure 5.41A and B). Significantly higher levels of IL7R were detected in MES lines as opposed to PN (p-value 0.021). We then repeated the experiment and quantifications on our MES and on more recently established non-MES GSC lines. As delineated in Figure 5.41C and D, the difference was not significant anymore, even though a trend for IL7R being higher in MES relative to non-MES GSC lines was still present. Notably, we included GSC lines 161019 and 180420 to the MES group of GSCs, as both showed much higher NDRG1 than EGFR, thus being CL/MES mixtures with a higher MES component.

On the contrary, line 170421, still being a CL/MES mix, had higher EGFR and EGFRvIII levels, being more of a CL subtype, and thus being included in the non-MES GSC subset.

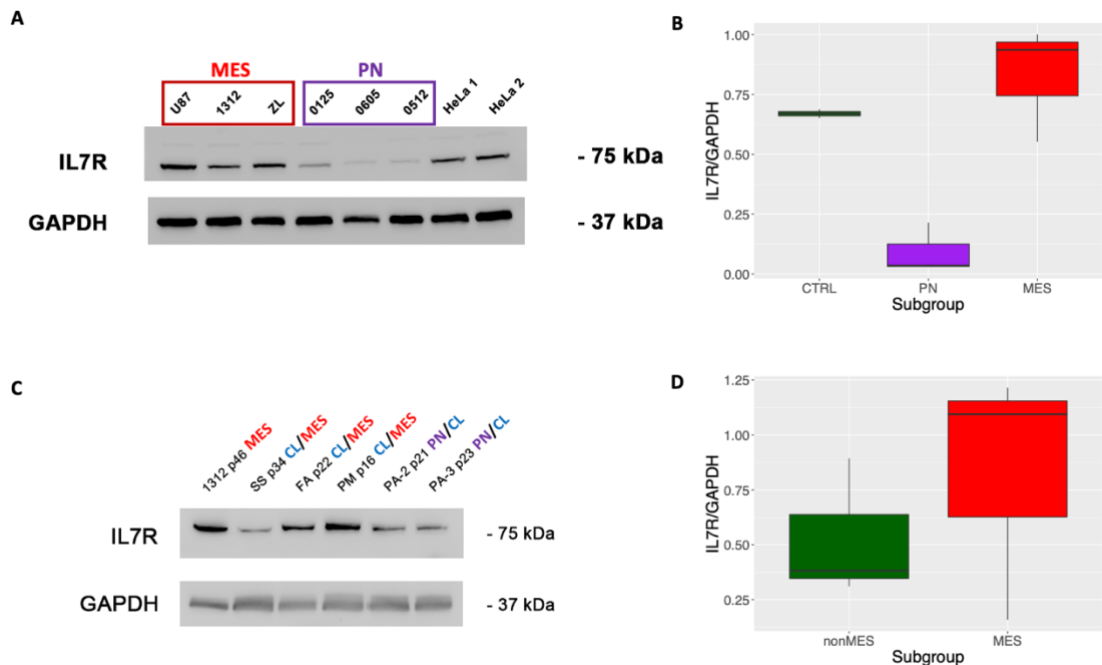


Figure 5.41. Western blot quantification of IL7R in MES and PN/non-MES cell lines. A, WB panel comparing IL7R protein expression in MES (160315 = ZL) and long-term cultured PN GSCs. HeLa cells were used as a IL7R positive control as this was the first assay for IL7R detection. B, quantification of IL7R relative to the housekeeping protein GAPDH for A. There is a significantly higher IL7R expression in MES cells (in red) than in PN cells (purple). C, WB panel comparing IL7R protein expression in MES and more recently established non-MES GSCs. Among the MES lines we included 161019 (SS) and 180420 (PM), as both showed much higher expression of the mesenchymal marker NDRG1 than EGFR, thus being considered more MES. On the contrary, line 170421 (FA), still a CL/MES mix, had higher EGFR and EGFRvIII levels, pertaining to CL subtype, and thus included in the non-MES GSC subset. D, quantification of IL7R relative to the housekeeping protein GAPDH for C. Despite showing a trend for higher IL7R expression in MES cells (in red) than in non-MES cells (green), statistical significance is not reached anymore. CL: classical, MES: mesenchymal, non-MES: non mesenchymal, PN: proneural.

These findings might be explained by regulation of IL7R expression during the initial stages of GSC line establishment, as observed for subgroup markers EGFR, NDRG1 and ASCL1. As a matter of fact, Figure 5.42 shows variable levels of the protein at varying culturing passages of the same cell line. Therefore, differences among transcriptional subgroups may be initially inconstant and become increasingly divergent as the cell line is progressively cultured *in vitro*. Additionally, as demonstrated earlier, GSC lines tend to have mixed affiliations at least up until intermediate culturing passages.

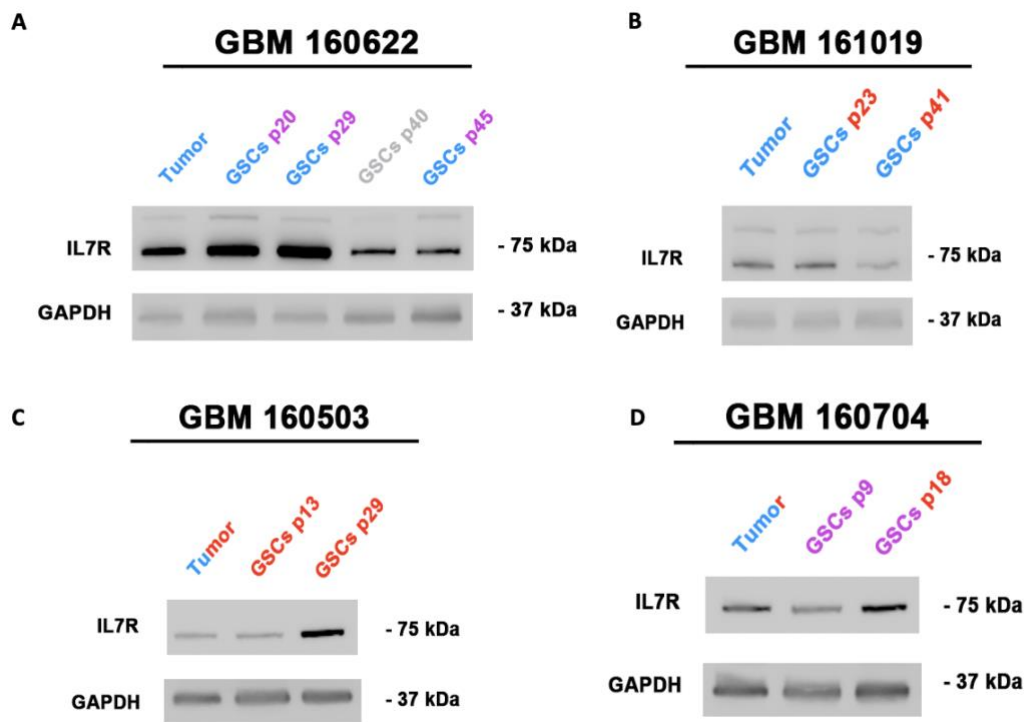


Figure 5.42. IL7R protein expression modulation at progressive culturing passages in each GSC line. Transcriptional subgroup affiliation is indicated by color code: CL blue, PN purple, MES red. Passage 40 of 160622 showed scant detection of the 3 subgroup markers and was then set to gray to distinguish it. CL: classical, MES: mesenchymal, PN: proneural.

To further assess the entity of IL7R expression in our GSCs, we performed flow cytometry analysis on our MES lines, particularly considering tumorigenic vs non-tumorigenic cell lines. According to what observed in western blot assays, line 131210 showed high IL7R signal, found in circa 80% of the cells. On the contrary, line 160503, which was investigated at earlier passages, was almost negative, while line 160525 demonstrated intermediate signal; both findings are in concordance with WB analysis on the same lines. Quite peculiarly, the tumorigenic MES line 160315 showed very low signal, as opposed to what detected by western blot.

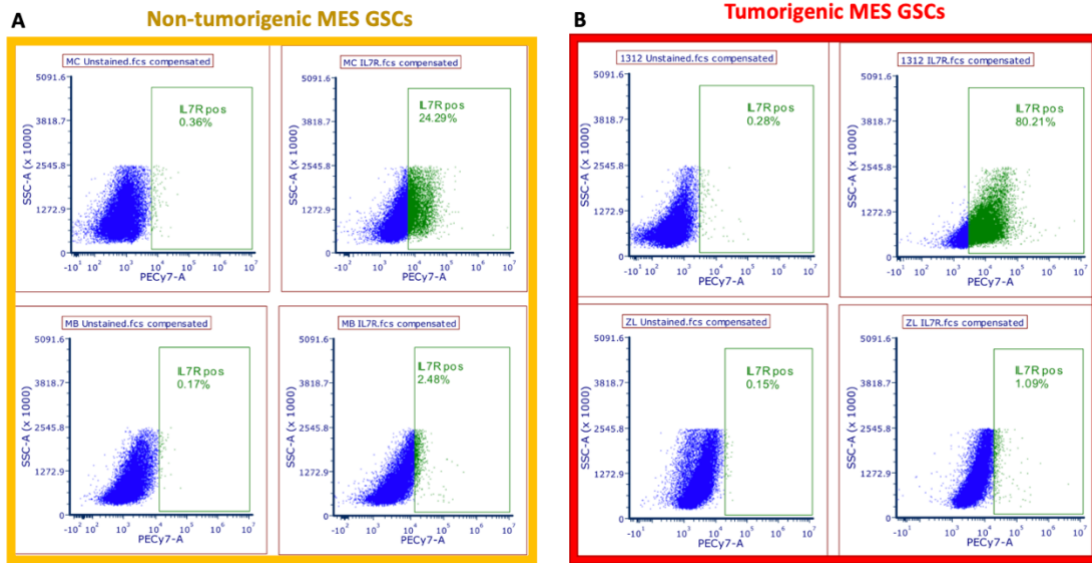


Figure 5.43. IL7R protein expression in tumorigenic and non-tumorigenic MES GSC lines as measured by flow cytometry. A, non-tumorigenic MES GSCs (160525 top, 160503 bottom), left unstained, right with IL7R staining. B, tumorigenic MES GSCs (131210 top, 160315 bottom), left unstained, right with IL7R staining. MES: mesenchymal.

5.12.4 Protein expression analysis of IL7R *in vivo*.

As a final step in our validation of IL7R as a putative new player in MES GBM pathology, we studied its expression *in vivo*. We performed western blot analysis on lysates of xenografts derived from transplantation of GSCs and GCL in the striatum of immunocompromised mice. Notably, as reported in Figure 5.44, IL7R was detected in all cases, even in a PN GSC xenograft. As a positive control we used cultured cells of the MES line 131210, whose xenograft showed slightly lower signal, suggesting remodulation of the expression *in vivo*. Next, we performed immunohistochemical assessment on the very same xenografts, by optimizing staining protocol on human tonsils as a positive control (see Figure 5.44). Quite remarkably, the xenografts resulted negative for the staining, even though they resulted positive at western blot.

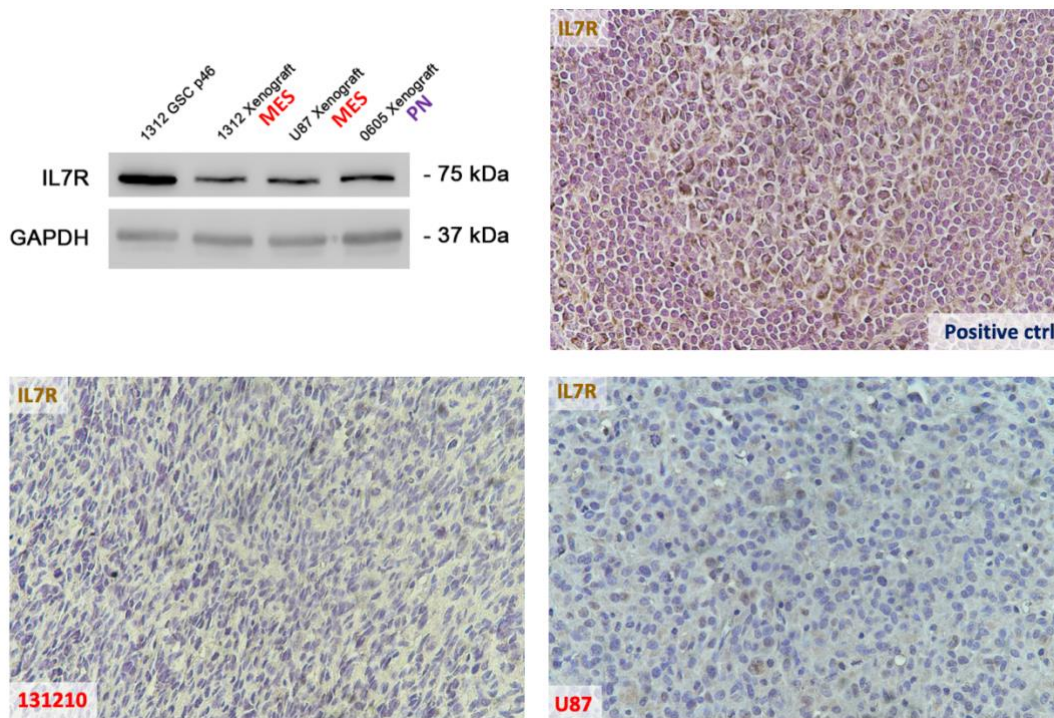


Figure 5.44. IL7R protein expression in GSC and GCL-derived orthotopic xenografts in immunocompromised mice. Western blot detects IL7R in MES and PN xenografts, even if with a lower signal than what seen for GSCs. IHC staining was optimized on human tonsils as positive control; however, MES GSC (131210) and GCL (U87) xenografts stained negative. ctrl: control, MES: mesenchymal, PN: proneural.

We thus performed IL7R staining on primary human GBM specimen, as shown in Figure 5.45. Notably, tumors of each subgroup presented variable expression of IL7R, with a generally stronger signal in MES tumors and variably low to negative signal in PN and CL GBMs. Even more remarkably, IL7R was able to discern a MES component (identifiable also by fusiform cell morphology) in the context of a prevalent PN tumor (Figure 5.46).

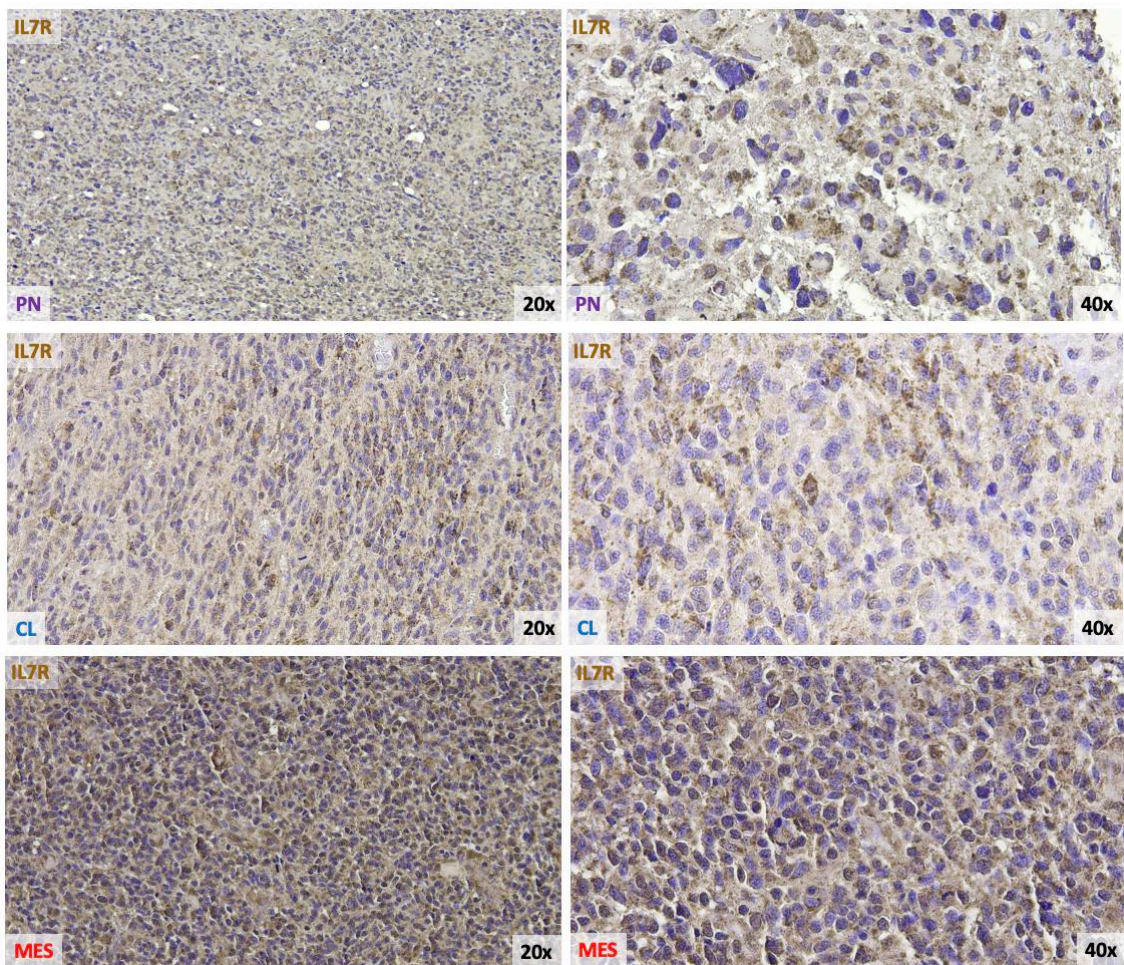


Figure 5.45. IL7R IHC on human GBMs of different transcriptional subgroup: PN (top), CL (middle), and MES (bottom). On the left lower magnification (20x); on the right, higher magnification (40x) of the same sample. CL: classical, MES: mesenchymal, PN: proneural.

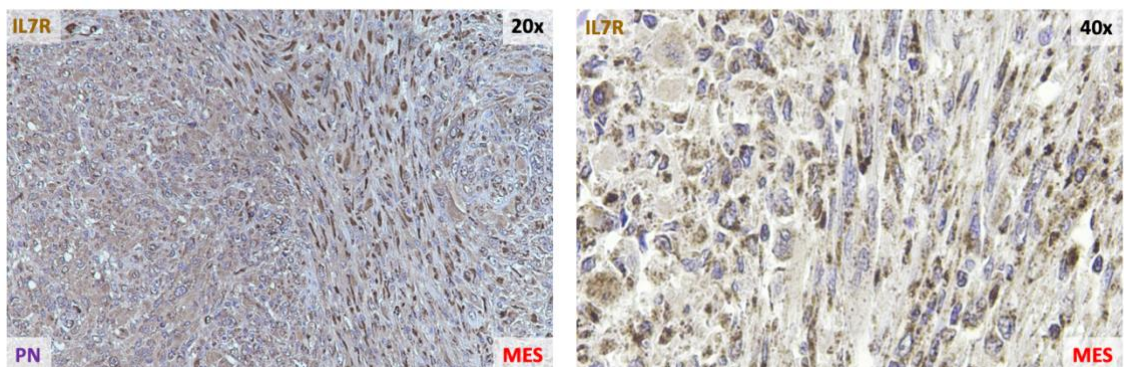


Figure 5.46. IL7R IHC on a human PN GBM with a MES component. The MES component is morphologically recognizable on the 20x image on the left as the strain of fusiform cells. On the right, higher magnification (40x) on the MES component, with strong granular signal. MES: mesenchymal, PN: proneural.

Given the evidence in human tumors and in *in vitro* cultured cells, the negativity of GSC-derived xenografts was striking. Since IL7R is a protein implicated in adaptive immunity, we wondered whether the absence of adaptive immune response in NSG mice could impair IL7R expression in our xenografts. To investigate this hypothesis, we set out to transplant a murine glioma cell line (GL261) into immunocompetent mice. Western blot analysis showed comparable IL7R levels in this cell line and our MES GSCs (180420-PM and 131210), as reported in Figure 5.47A. We then orthotopically transplanted GL261 cells both into immunodeficient NSG and syngeneic immunocompetent C57bl/6 mice and monitored them with MRI to detect tumor formation. Upon sacrifice of the mice, we performed IL7R IHC staining. As shown in Figure 5.47B, tumors that originated in immunocompromised mice were IL7R-negative, as opposed to tumors that engrafted in immunocompetent mice. Therefore, expression of IL7R on tumor cells may require a functioning immune system and may be mediated by the tumor-microenvironment crosstalk, which will be the subject of further investigation.

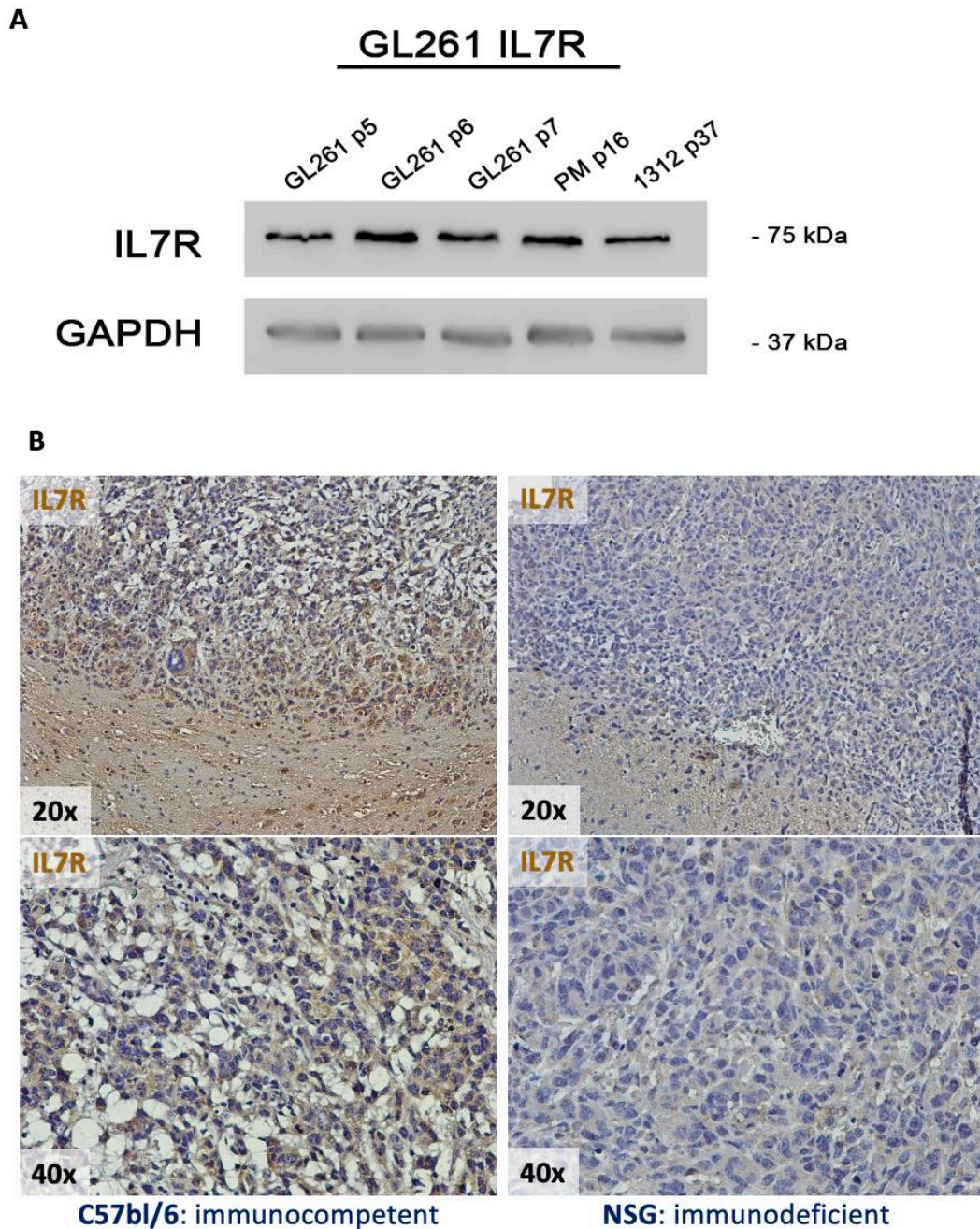


Figure 5.47. IL7R expression in the GL261 murine glioma cell line and in tumors derived from its implantation. A, GL261 cells express IL7R levels comparable to our human MES GSCs (GSC line 180420-PM, and GSC line 131210) by western blot assay. B, Upon transplantation of GL261 in the striatum of either immunocompromised (NSG) or immunocompetent (C57bl/6) mice, tumors are formed. However, IL7R is expressed in tumor developing in immunocompetent (on the left) but staining is negative in those forming in immunodeficient mice (on the right), indicating the requirement of normal immune system and tumor-microenvironment interaction for proper IL7R expression in tumor cells.

6 DISCUSSION

As explained in the aim section, this is a modular work where, starting from a principal research flow (i.e., radiogenomic study of GBM transcriptional subtypes), we took advantage of our GBM modeling tools (i.e., GSCs) to better characterize their dynamic profile and perform additional studies. Therefore, the whole work can be subdivided into 4 concatenated research flows stemming from each other by exploitation of data or analyses shared between them.

6.1 Radiogenomic analysis of GBM transcriptional subtypes.

Current progress in imaging and nuclear medicine techniques made available a huge number of quantitative parameters that cannot be captured visually but can still provide useful information for diagnosis and treatment management. Radiomics and radiogenomics currently represent an extremely hot topic in neurooncology. In particular, radiogenomics correlates radiomic features with genetic anomalies, configuring a potential virtual biopsy. Therefore, the contribution deriving from radiomics and artificial intelligence will likely increase in neurooncology practice in the near future (Galldiks *et al*, 2020). Basic radiological and advanced radiomic features can be extracted from a variety of MRI protocols.

The primary project that we focused on is the investigation of radiomic features correlating to transcriptional subtypes of GBMs. Therefore, we prospectively enrolled 56 CNS WHO grade 4 gliomas to have a consistent representation of PN, CL and MES lesions, based on the inferred prevalence of the transcriptional subtypes (Huse *et al*, 2011; Lin *et al*, 2014). However, due to a recent revision of CNS tumor taxonomy, GBMs have been restricted to only IDHwt tumors. Thus, we discarded 6 IDH-mutant lesions, now defined as grade 4 astrocytomas, maintaining only 50 IDHwt GBMs. Of these, advanced MRI studies were available for 36 cases, on which we performed the radiogenomic analyses. Transcriptional affiliation was defined by an IHC panel of gene

classifiers as published by our collaborators (Orzan *et al*, 2020). Notably, contrary to what reported in the literature, our cohort contained only 4 MES lesions (5 if considering also tumors for which advanced diffusion MRI were not available). Such a discordance may be partially explained by using IHC as a surrogate for transcriptomic classification. However, as reported by Orzan *et al.*, concordance between IHC and transcriptional classifications reached its highest for MES tumors (Orzan *et al*, 2020). In concordance with the literature (Bhat *et al*, 2013; Kim *et al*, 2021; Wang *et al*, 2017), survival analysis demonstrated MES GBMs as having the worst prognosis among all subtypes, even though it did not reach significance, probably due to the scarcity of MES samples as opposed to others (Figure 5.2). Since the study population consisted in 4 MES, 13 CL, 11 PN, 6 PN/CL and 2 balanced mixes of the three subtypes, we decided to dichotomize the patients in MES vs non-MES tumors, as the numbers were too small to infer reliable data on all the affiliations. Thus, we focused on the MES subtype, as it is generally associated with shortest survival and increased treatment resistance. Tumor ROIs were delineated on FLAIR and post-contrast T1 sequences and coregistered to diffusion (DTI) and advanced diffusion (NODDI) studies to extract radiomic features. Significant features distinguishing the MES from the non-MES cohort were identified with a univariate filter method (Tables 5.3 and 5.4).

To further select radiomic significant discriminating features and build a predictive model, we generated xenografts from patient derived GSCs of different affiliations and confirmed their molecular subgroup by means of the same IHC panel used in patients. As a matter of fact, xenografts tend to show more extremized features than human tumors. Moreover, xenograft cohorts were more balanced, with 9 MES and 12 non-MES tumors. Therefore, we hypothesized that features identified as discriminating also the murine models would strongly associate with molecular affiliation. Notably, contrast uptake was not detectable in all xenografts, especially in non-MES tumors, and a few MES lesions only showed a modest uptake. Consequently, xenografts' ROIs were constantly delineable only on B0 sequences (Table 5.5), making them comparable to the FLAIR ROIs in humans.

By comparing the significant discriminating features between humans and mice, we selected 3 common features, two of which (IMC1 and IMC2) derived from DTI and one (90th percentile) from NODDI. IMC1 and IMC2 provide an indication on texture complexity, with values approaching to 0 indicating more heterogeneous textures and increasingly positive or negative values signifying more homogeneous qualities. Both IMC1 and IMC2 are significantly closer to 0 in MES than non-MES tumors, suggesting MES as more heterogeneous than CL or PN tumors. Parallely, higher 90th

percentile values in the NODDI fecv map indicates that MES lesions portend higher local tumoral infiltration, since, under pathological conditions, the anisotropic gaussian diffusion expressed by the fecv map provides an approximation of infiltrative edema.

Subsequent generation of a simplified prediction model with the identified three features, with hyperparameters tuned on the xenografts' values, allowed a quite accurate classification of the human samples.

A major limitation of this model is the extreme imbalance between the MES and non-MES cohorts. Additionally, both IMC1 and IMC2 provide essentially equivalent information, possibly rendering the prediction computationally expensive with no substantial contribute. To circumvent this, we reapplied the same pipeline after data augmentation and correction for codependency of the radiomic features. Basically, by applying SMOTE algorithm, we created fictitious human and murine MRI studies to balance the two cohorts (64 human, 32 MES and 32 non-MES) and 24 xenografts (12 MES and 12 non-MES) with the same value distributions of the original populations. Correction for codependence was then applied after feature extraction by discarding elements with a correlation coefficient higher than 0.7 with other retained features (Tables 5.6, 5.7 and 5.8). When compared to the initial analysis, IMC2 was correctly ignored, as fundamentally overlapping to IMC1. On the contrary, the 2 previously found features (IMC1 on the FA map and the 90th percentile on the NODDI fecv map) even showed increased significance. Notably, we identified 4 more features, 3 coming from NODDI (fecvf_firstorder_10Percentile, odi_firstorder_Kurtosis and ficvf_glszm_SizeZoneNonUniformityNormalized) and 1 from DTI (GLCM_Correlation).

Remarkably, these additional features corroborate our previous inference, as higher values in the 10th percentile of extraneurite diffusion in MES tumors seem to suggest that they are more locally infiltrative than non-MES lesions. Of note, at least for xenografts, which are more extreme than their human counterparts, MES lesions appear more circumscribed while CL and especially PN lesions tend to be more widespread along white matter tracts. This might appear in contrast with our observation. However, higher values in the 10th and 90th percentile of extraneurite diffusion compartment only describe the immediate vicinity to the core nodule, where important FLAIR/T2 anomalies might be due to either cell infiltration or vasogenic edema, with no indications on what happens farther away. Therefore, MES lesions might be locally more invasive due to higher cell proliferation in the proximity of the necrotic nodule, or due to stagnation of malignant cells in the tumor surroundings, with a progressively smaller difference in this trend the more distant from the core.

This occurrence would also be in line with what is reported by some authors, who demonstrated that distinct GBM cell subpopulations have distinct proliferative or infiltrative profiles, participating in a cooperative model of tumor progression that, once more, synergistically relies on tumor heterogeneity (Tari *et al*, 2022; Vinci *et al*, 2018). Therefore, MES GBMs may have an overrepresentation of fast proliferating, slow migrating, and less invasive cells. Notably, we extracted radiomic features from the entire FLAIR altered signal region. However, some authors suggest that features extracted from the very first centimeter around the contrast-enhancing nodule correlate better with cancer invasion than features extracted from the whole edematous region (Liu *et al*, 2022). Another possible explanation is that MES GBMs overproduce components of extracellular matrix, which accumulate and orient themselves, thereby reducing isotropy. Similar findings were reported in differentiating GBMs from brain metastases (Mao *et al*, 2020). Additionally, reports showed that both the isotropic (fiso) and the hindered (fecv) water diffusion are significantly higher in the tumoral regions respect to normal brain, due to vasogenic and infiltrative edema. Parallely, the restricted (ficv) component is decreased due to edema and neuronal loss. As the NODDI deconvolutes the voxel unitary diffusion into the three compartments, if one of them prevails, the influence of the remainder is overpowered (Kadota *et al*, 2020). Therefore, in the perinodular area in the context of FLAIR alterations, the influence of the hindered fraction is likely stronger than that of isotropic or restricted components.

Parallely, like IMC1, correlation ranges from 0 (uncorrelated) to 1 (perfectly correlated). The Size Zone Non-Uniformity Normalized feature measures the variability of zonal volumes throughout the image, with lower values indicating homogeneity. Lastly, higher kurtosis indicates a flatter distribution of the grey values with more consistent representation of the tails. Profile of these latter features in MES tumors indicate them as having a more heterogeneous texture than non-MES lesions. While at first sight these findings seem to corroborate our previous inference, we must note that they are obtained from a modification of the original datasets that does not introduce any bias. However, the correspondence between the initial and final value distributions might have affected the scope of characteristics identified as significant at both analyses.

As a final step, we trained a predictive model with the six features identified, tuning the hyperparameters on the xenografts' augmented dataset and testing it on the augmented patients' dataset. Accuracy was globally increased, reaching the highest performance with the random forest model.

Notably, proper validation of both models (obtained from analyses on original and augmented datasets) should be carried out on human datasets not implicated in generation of the models themselves. This could be achieved either by acquiring more diffusion studies on new patients to following classify them as MES or non-MES, or by testing the model on public datasets used for other studies. Unfortunately, however, NODDI sequences are not yet routinely used in clinics, and radiomic studies on NODDI are still extremely few, with no published reports correlating both DTI and NODDI to GBM transcriptional subgroups to our knowledge.

As a matter of fact, the few studies in the literature to adopt both diffusion techniques (DTI and NODDI) to characterize brain tumors did not apply radiomic pipelines and focused on different tasks. Kadota et al. utilized diffusion metrics to differentiate GBMs from metastases. They reported a significantly higher signal in the NODDI extraneurite component of diffusion in the perinodular area of GBMs, while the isotropic component was increased in metastases, though not statistically significant. Notably, ADC and FA may aid in the diagnosis, but the fecv fraction proved to be stronger, making NODDI more sensitive (Kadota *et al*, 2020). In a similar study, Mao et al. described parallel findings, though the isotropic component of the diffusion showed the greatest discriminative power in the peritumoral surroundings (Mao *et al*, 2020). Figini et al. compared DTI and NODDI performances in characterizing grade 2, 3 and 4 gliomas. They found that IDHwt grade 2 and 3 gliomas had significantly lower minimum MD values, and maximal FA and NODDI restricted diffusion component compared to IDH-mutant counterparts. They demonstrated a significant correlation between DTI parameters and IDH mutations in grade 2 and 3 gliomas. However, NODDI did not significantly improved diagnostic accuracy, thus not justifying the additional computational costs (Figini *et al*, 2018). Notably, higher NODDI odi values indicate more dispersed fibers in IDHwt gliomas, whereas increased extracellular water in IDH-mutant lesions may reduce their FA signal (Figini *et al*, 2018). Parallely, Zhao et al. found no differences in NODDI metrics pertaining IDH status. However, they reported that gliomas with high restricted component (ficv) in the tumor parenchyma and low restricted diffusion in the surrounding periphery were more likely to be high-grade, as opposed to lower grade gliomas which showed a significant opposite trend (Zhao *et al*, 2018). As a matter of fact, fiber disruption and tumor isotropic tissue lead to a general reduction of the intracellular diffusivity and increased odi signal. Higher-grade gliomas show increased cellularity and nuclear pleomorphism, endothelial and vascular proliferation, all concurring in destroying the fiber tracts architecture to a greater extent than lower-grade gliomas (Zhao *et al*,

2018). Further studies investigated NODDI and DTI ability to distinguish normal tissue, edematous white matter and tumor nodule (Masjoodi *et al*, 2018).

The likely rationale explaining the outperformance of NODDI with respect to other diffusion techniques is that brain and tumor microarchitecture affects water molecule diffusion into a non-Gaussian probability distribution. Therefore, non-Gaussian diffusion models may better approximate this complex tissue microenvironment than Gaussian diffusion models (Mao *et al*, 2020).

On the contrary, many studies applied the radiomic pipeline to conventional, routine MRI acquisitions to study brain tumors. By exploiting a radiomic-based model, Cluceru *et al*. provided an accurate and simultaneous prediction of IDH and 1p/19q codeletion status in 147 patients (Cluceru *et al*, 2021). Liu *et al*. correlated the immune cell infiltrate profile of GBMs to radiomic features. In particular, they elaborated an efficacious 11-feature model able to predict high immune infiltrate scores, which correlate with worse prognosis and higher expression of immunosuppressive molecules, limiting therapies acting against immune checkpoint inhibitors (Liu *et al*, 2022). Le *et al*. applied radiomic analyses to conventional post-contrast T1 and T2/FLAIR sequences, extracting features from enhancing nodule, non-enhancing tumor core, and peripheral edema. They identified 13 features with which they built a model predicting transcriptional affiliation with reported accuracy of 70.9%, 73.3%, and 88.4% for CL, MES, and PN tumors respectively (Le *et al*, 2021). Parallel findings following similar analyses on conventional MRI have been reported by others (Rathore *et al*, 2019). However, both groups did not include diffusion-weighted sequences, neither DTI, nor the more advanced NODDI. Additional works demonstrated how prediction models based on radiomic features derived from MRI or combined PET/MRI acquisitions could determine proliferation index, tumor grades, and molecular markers (e.g., IDH, MGMT, ATRX, 1p/19q codeletion, EGFR, PTEN, RB1, TP53, TERT) and correlate with prognosis in gliomas, besides differentiating tumor recurrences from radionecrosis and predicting response to immunotherapy (Lohmann *et al*, 2021; Liu *et al*, 2021; Moassefi *et al*, 2022; Beig *et al*, 2020).

Despite encouraging initial applications of radiomics and deep learning in brain tumor patients, most prediction models have not yet become integral to clinical practice. One of the main motivations is the general lack of method standardization, which hampers replicability of the findings and prevents the widespread diffusion of radiomic models to different institutions. (Galldiks *et al*, 2020) Radiomics features are affected by many factors, from the quality of acquisitions to differences in contrast, voxel resolution and slice thickness, to magnetic field strengths, up to image

pre- and postprocessing. First-order features, which correspond to image histograms, are more reproducible than higher-order shape and textural features (Lohmann *et al*, 2021). Also, tumor segmentation is a major concern. On one hand, manual techniques are labor intensive and suffer from inter-observer variability. On the other hand, automatic or semiautomatic segmentation may be imperfect and over- or underestimate tumor burden, thereby impacting on the subsequent analyses (Beig *et al*, 2020). Their interpretability is an additional issue that may prevent radiomics from becoming routinely adopted in clinical practice, as features and their derived models are still perceived as too abstract.

6.1.1 Limitations and conclusions

As outlined before, the main limitation of our study is the small number of patients investigated, which is further impaired by the scarcity of MES samples. Many radiomic and radiogenomic studies base their findings on much greater numbers, as allowed also by increasing potential of computational and deep learning approaches. However, such studies generally rely on image acquisitions systematically integrated in clinical routine and, sometimes, on merging datasets from different institutions. To our knowledge, we were the first to apply the radiogenomic pipeline to an advanced diffusion MRI technique (NODDI) and correlate resulting features to GBM transcriptional subgroups. Therefore, our numbers, though limited, provide interesting preliminary results, suitable for a cutting-edge single-center study. Still, such results need to be confirmed by more advanced analyses relying on conspicuous numbers, permitting to properly delineate *ab initio* a training, validation, and test populations. Further limitations of this study are the use only of diffusion-weighted images. Perfusion MRI might catch other crucial differences among the three subgroups, but its inclusion in the pipeline was prevented by too variable contrast uptake by xenografts and inconstant acquisitions on patients. The scant enhancement of xenografts also made us discard any analyses on the simple T1 contrast-enhancing lesions, limiting the study to FLAIR anomalies, which suit better the perinodular area than the proper nodule. Lastly, the definition of the transcriptional subgroup affiliation was determined by an IHC surrogate, which, even if demonstrated to be very accurate, was far from reaching perfect statistics.

In the end, by applying the radiomic pipeline to advanced diffusion MRI techniques, we identified novel features that proved to correlate and possibly predict mesenchymal affiliation of GBMs and provide insight into pathologic characteristics of transcriptional subtypes. Still, radiomics should not be used as a stand-alone

technique but should implement other clinical and pathological information allowing additional low cost, automated data evaluation (Lohmann *et al*, 2021). Improvement of these critical aspects will render radiomics and radiogenomics more acceptable, aiding precision medicine and patient-tailored treatment strategies (Liu *et al*, 2021).

6.2 GSC evolution and model reliability.

As previously reported (Bhat *et al*, 2013; Stringer *et al*, 2019), culturing conditions may affect transcriptional affiliation of GSC lines propagated over a high number of passages. Therefore, primary cultures may better reflect the parental tumors at early or intermediate *in vitro* passages, although these cultures may be composed not only by stem cells but also by short-term proliferating progenitors that are still responsive to mitogen stimuli. On the contrary, when performed for many subculturing passages, the NSA allows for bona fide isolation of cells endowed with stem properties, with the more committed progenitors exhausting their replication and being selected away. As a matter of fact, NSA is a very efficient functional assay for establishing GSC lines. Additionally, stabilized GSC lines are essential for reproducibility and subsequent replication of experimental results and thus cannot be used only at early passages.

Given this, a systematic investigation of GSC lines' dynamic transcriptional drift or stability is of particular interest. To date, very few proper studies have addressed this question, while most of others concentrated on the sole comparison between tumor tissues and derived stable lines or on the spatial and temporal heterogeneity of the tumor (Nakano, 2015; Cusulin *et al*, 2015; Stringer *et al*, 2019; Neftel *et al*, 2019).

Since for our radiogenomic investigation we established new primary cultures from enrolled patients, we decided to parallelly screen them for their molecular profile under progressive *in vitro* passages. We were able to isolate 14 new GSC lines from a total of 48 GBM samples, all of which were cultured up to at least intermediate (around p20) or intermediate-late (around p40) passages. In fact, at these passages we considered our lines to be fully established, though other authors consider them already stabilized at passages between 8 and 15 (Stringer *et al*, 2019). Notably, we subjected 9 pairs of parental tissues and derived GSC line to RNAseq. GSEA analysis with Verhaak's transcriptional signatures confirmed that parental tissues were

significantly more mesenchymal than their matched derived lines, suggesting an actual drift in culture toward non-MES phenotypes.

Transcriptional classification relies on expensive methods that would be unrealistic to perform multiple times for longitudinal screening on several samples; in addition, transcriptional differences are not always maintained at the protein level. Therefore, we proposed a cost-effective, protein-based surrogate for our classification. We selected 3 genes that were demonstrated by our and other groups to regulate transcriptional affiliation (EGFR for CL, ASCL1 for PN and NDRG1 for MES) (Verhaak *et al*, 2010; Park *et al*, 2017; Narayanan *et al*, 2019), whose expression levels can easily be detected by simple western blot assay.

We first evaluated the reliability of this minimal signature by clustering TCGA GBMs based on transcript levels of these genes (Figure 5.8), showing an acceptable accuracy (70.6%) and reaching up to 90% of positive predictive value for MES samples. Parental tumor and GSCs at progressive passages were subjected to this WB panel and affiliation was determined based on relative expression of the 3 proteins and their modulation in culture (Figures 5.9-5.12). Remarkably, we also compared our WB-based to IHC-based tumor classification, both of which exploit EGFR, ASCL1 and NDRG1 as classifiers (Table 5.12). Interestingly, the classifications perfectly matched in 4/14 cases (1 MES and 3 CL), while 3 CL tumors showed a MES component at WB, defining 3 mixed CL/MES GBMs. This may be due to contaminating normal white matter which is NDRG1-positive, and is not discernible by WB, as opposed to IHC. This overweighing of the MES component affected also two PN/CL tumors that resulted as MES and CL/MES at WB. Additionally, Two IHC-based PN/CL tumors were classified as CL by WB. This underestimation of the PN component by WB may be due to a mixed effect of the lower performance of ASCL1 antibody in WB, and of the smaller amount of tumoral material with respect to IHC, that may be insufficient to catch the total spatial tumor heterogeneity. Given these differences and the accuracy of IHC (Orzan *et al*, 2020) and WB classifications (both protein surrogates of transcriptional classification), we selected IHC as our reference, but also maintained WB affiliation for consistent comparison in the context of the analysis upon progressive *in vitro* subculturing of GSCs.

Intriguingly, minimal signature analysis at increasing culturing passages showed different scenarios. On one hand, some lines demonstrated increasing signal of the PN marker, generally coupled with decreased intensity of CL and MES markers, consistent with GSEA findings that tumor tissues are more mesenchymal than their matched lines. On the contrary, in a few lines, the MES marker remained expressed at high levels or even progressively increased its expression, indicating stabilization

of a MES profile. As for the CL marker, most lines showed extreme downregulation or even abolition of its signal, inducing an extreme underrepresentation of the CL component, which is generally maintained only in the presence of EGFRvIII. This might be due to presence of EGF in the medium, which modulates its receptor through negative feedback (Mazzoleni *et al*, 2010; Stringer *et al*, 2019). This is concordant with the disappearance of wild-type EGFR and the retention of EGFRvIII, which is ligand-independent and therefore not modulable. However, loss of the truncated form of EGFR in at least two cases (GBM 160622 and GBM 191119) seems to indicate that other mechanisms may take place in reducing the classical component, which is not simply explained by a biased inference, where the CL counterpart is defined solely by a master regulator (although being the strongest for this affiliation) that is necessarily downregulated by high concentrations of its ligand. Such evidence have also been reported in previous works showing *in vitro* downregulation of EGFRvIII (De Bacco *et al*, 2016).

Notably, we also performed multiple sampling from the same tumor and managed to establish cell lines from multiple areas in 2 cases. Both times, lines generated from the same tumor showed same affiliation, which was identical to the original tumor in 1 case (a MES line from a MES tumor) and showed a “proneuralization” in the latter (a PN/CL line from a CL tumor). This is similar to what demonstrated by Laks *et al.*, who generated lines from multiply sampled tumors in 4 cases. In 3 of them affiliation was the same for each line established and equivalent to the parental tumor (MES), while in the remainder they had a drift from CL to MES profile (Laks *et al*, 2016).

Remarkably, one GSC line (GBM 190125) was not assigned to any affiliation, as WB panel showed a general downregulation of all 3 protein markers, hence representing a flaw of this method. One of the most notable evolutions, instead, was that of GBM 160704. Though starting from a CL tumor, as assessed both by WB and IHC, early passages of the established line showed a proneuralization with increased ASCL1 and decreased EGFR. At intermediate passages, the same GSC line showed even stronger ASCL1 signal, but surprisingly also a very strong NDRG1 signal, thus delineating a mixed PN/MES line. This is very intriguing as at p18 contamination by normal white matter can be ruled out and our group demonstrated that NDRG1 is transcriptionally repressed by ASCL1 (Narayanan *et al*, 2019), thus suggesting the development of two subclones of this line, one more PN and one more MES.

This complex picture seems to be in line with many reports in the literature. Laks *et al.* reported a stable, though weak correlation between GSCs and their original tumors in 60% of their lines, especially for TCGA classification. Samples that changed affiliation did not show a peculiar trend in switching subgroup (Laks *et al*, 2016). On

one hand, some other authors suggest a general “proneuralization” of the cancer cells in culture, as the NSA was designed to isolate normal neural stem cells, whose profile is more similar to PN profile (Verhaak *et al*, 2010; Bhat *et al*, 2013; Xie *et al*, 2015). On the other hand, other authors report establishment of a consistent number of MES lines in addition to PN (Nakano, 2015; Stringer *et al*, 2019; Mack *et al*, 2019). Lastly, some authors describe virtual absence of CL cell lines (Nakano, 2015; Stringer *et al*, 2019). Whether such affiliation drift is the result of thriving pre-existing clones or due to inherent single cell plasticity is still to be fully understood. While human GBMs are proven to harbor cells of differing subtypes that can be variably selected in culture, *in vivo* evidence suggests that some cells may shape their phenotype according to signals from the microenvironment (Stringer *et al*, 2019; Li *et al*, 2022).

As explained previously, we performed RNAseq on 9 GSC lines (4 MES, 1 CL and 4 PN) whose affiliation was determined by WB minimal signature, to which we added a previously established line known to bear an extreme mesenchymal profile (deriving from a gliosarcoma, as described also by other authors (Laks *et al*, 2016)), to perfectly balance MES and non-MES cohorts. Unsupervised clustering discriminated two main groups, one consisting of 3 WB-defined MES lines and the other containing all the PN, CL and the remaining 2 MES lines. Consistently, GSEA showed significant enrichment of MES profile in the first cluster and of CL and PN profile in the second one (Figure 5.15). Of note, we investigated our cell lines also by means of whole exome sequencing, which returned two unsupervised clusters which perfectly distinguished MES lines from PN and CL ones (Figure 5.34). These findings are further corroborated by other authors, who showed that transcriptome-based clustering divides GSCs into a MES and a non-MES group, the latter comprising PN and CL lines (Cusulin *et al*, 2015; Laks *et al*, 2016).

We also performed a supervised transcriptional analysis by comparing our WB-defined PN and MES GSC lines (Figure 5.14). Remarkably, MES profile was very significantly enriched in our MES lines just as the PN profile in our PN GSCs. Also, CL profile was significantly enriched in our WB-defined PN GSCs, though consistently to a lesser extent than PN signature. This is an extremely interesting finding as it confirms our minimal protein signature in a proper transcriptional setting, which, together with what reported for IHC by other groups (Conroy *et al*, 2014) and our collaborators (Orzan *et al*, 2020), suggests that a classification based on protein expression might have a comparable impact to that based on transcript levels. This is reasonable since the proper molecular effectors are usually the products of transcript translation, therefore actual proteins, even though molecular functions (beside translation regulation) have been recently identified for RNA fragments.

Evidence of affiliation drift from parental tumor to GSC lines since the very earliest culturing passages raises the question whether these divergences might render these cells an unreliable tool for modeling GBM transcriptional heterogeneity. To address this issue, we exploited differential gene expression between our WB-defined PN and MES lines to obtain defining gene signatures for each affiliation. By selecting stringent log₂FC and multiple test adjusted p-value thresholds, we identified an 81-gene signature for MES subgroup and a 43-gene signature for PN (Tables 5.13 and 5.14). We then performed k-means clustering of TCGA human GBMs based on the expression of these signatures and obtained an almost perfect accuracy in assigning each sample to the correct affiliation (Figure 5.16). Notably, this diagnostic power further improved when we derived signatures from the comparison of our unsupervised clusters, where Cluster A retained probably the most divergent MES lines, while Cluster B grouped all the non-MES and, likely, two less extremized MES lines. As a matter of fact, with this comparison, we were able to render log₂FC and p-value thresholds even more stringent, identifying 49 characterizing genes for Cluster A/MES and 98 for Cluster B/non-MES (Tables 5.15 and 5.16). By performing k-means on human unrelated TCGA GBMs, these signatures separated MES from PN tumors with a 100% accuracy (Figure 5.18). Most notably, if we also included CL samples and imposed to cluster tumors in 3 groups these signatures managed to correctly identify the transcriptional affiliation with an accuracy of 88.2% (Figure 5.17), and the sole Cluster B signature was able to accurately discern PN from CL tumors (Figure 5.18). These findings were also confirmed by performing differential gene expression on available raw transcriptomic data of TCGA human GBMs. Basically, we isolated RNAseq data of known-affiliation tumors from a publicly available repository and performed pairwise comparison between MES and PN, MES and CL, and CL and PN tumors. We then interrogated the 3 resulting gene lists with our GSC-derived signatures by means of GSEA (Figures 5.19-5.21) and demonstrated correct enrichment of our genesets. This suggests that, even though GSC lines start drifting away from the molecular profile of their parental tumor from the very first culturing passages, their transcriptional affiliations still approximate very well the corresponding affiliations of unrelated human full-blown GBMs and may reliably be used to model GBM heterogeneity.

Notably, our signatures do not significantly overlap with each other, nor with the more conspicuous signatures deposited by Verhaak et al. (Verhaak *et al*, 2010) (Tables 5.17 and 5.18) or subsequently by Wang et al. (Wang *et al*, 2017) (Tables 5.19 and 5.20). Thus, the genes included in our signatures may provide interesting insights in identifying novel signaling and metabolic pathways that concur in driving

different subgroups' behavior and that may be underrepresented when considering GBM tumor samples as a whole. Remarkably, other authors exploited machine learning strategies to propose an even more condensed transcriptional signature, consisting of only 5 genes, able to predict the 3 GBM subtypes with a reported accuracy of 80.12%, therefore amenable to be used also by qRT-PCR (Tang *et al*, 2021). However, caution should be paid with classifications relying on too small transcriptional genesets. As a matter of fact, Stringer *et al*. reported that using the reduced subgroup-specific signatures published by Wang *et al*. (Wang *et al*, 2017) to subtype IDHwt GBMs, tumors often obtained non-significant enrichment scores. Consequently, it still may be better to rely on the original, larger Verhaak's signatures (Stringer *et al*, 2019). Accordingly, we chose Verhaak's genesets as our standard reference. Moreover, condensed transcriptional signatures have been proposed by other groups and correlated to GBM prognosis (Johnson *et al*, 2020).

To better delineate our MES and PN GSC profiles, we utilized the significantly upregulated genes in each group to identify molecular pathways, cell phenotypes and gene ontologies potentially playing subgroup-specific key roles. In particular, gene ontologies are a basilar component of computational biology, based on genes functionally interconnected in defined biological and developmental processes (Beig *et al*, 2020). The two MES clusters (Cluster A and WB-MES GSCs) showed similar phenotypes, resembling other mesenchymal lineages. Known mesenchymal players (Lv *et al*, 2022; Behnan *et al*, 2019), as well as stigmata of hypoxia, ECM deposition and inflammation were detected, with generally higher significance scores in Cluster A. As a matter of fact, Unsupervised Cluster A is a subset of the WB-MES GSCs and likely collects the most extreme mesenchymal lines. Notably, however, epithelial to mesenchymal transition (EMT) was strikingly more significant in WB-defined MES GSCs. This suggests that a mesenchymal transition very similar to EMT might be a common, basal process in GBM progression. Therefore, comparing only the 3 most MES lines against all others diminishes such difference, which is better detected by collecting altogether the lines with a significant (though maybe not prominent) MES component. Remarkably, our MES lines had significant enrichment of fructose and mannose metabolism, a metabolic pathway that our group very recently identified as crucial for MES GBMs (Pieri *et al*, 2022). Parallely, the two non-MES/PN clusters (Cluster B and WB-PN GSCs) shared profiles reminding of differentiated glial or progenitor neural cells, normal brain compartments and less malignant or benign glial tumors. Notably, recent evidence showed how gliomas hijack neuronal signaling by forming synapses with healthy neurons to drive and promote invasion (Venkataramani *et al*, 2022, 2019). Enrichment in ontologies linked to

synaptogenesis and synaptic transmission in PN GBMs and GSCs is therefore consistent with their more invasive nature than MES tumors and cell lines. Also, our non-MES/PN clusters shared pathways typical of neuronal tumors like medulloblastoma.

As previously described, we orthotopically transplanted our GSCs in murine recipients to generate xenografts for our radiogenomic pipeline. We took advantage of these transplants to assess the tumorigenic potentials of the lines established, which is an integral part of the functional definition of cancer stem cells. Notably, high passages GSC lines of different affiliations tend to stably extremize different aspects of human GBMs, i.e., invasion for the PN and nodular mass and angiogenesis for the MES. Remarkably, very few CL lines have been stably established over the years, as described above.

Given the mixed and dynamic transcriptional affiliation of our freshly established lines, we postulated that their derived xenografts would reproduce more closely the heterogeneous phenotype of human lesions both at MRI and at histopathology. Nine out of the total 14 lines generated tumors that were detectable at MRI, while 5 apparently failed to give rise to xenografts. Of the 5 GSC lines that did not generate full blown xenografts, two demonstrated a few tumor cells at H&E staining, without forming full-blown tumors. As hypothesized, qualitative MRI evaluation of T2, post-contrast T1 and diffusion/NODDI sequences allowed us to assign a radiological-based subgroup pattern to the xenografts (Table 5.25) that comprised all the subgroups. Nonetheless, radiological phenotypes were already assuming the extreme features of long-established lines, although showing more heterogeneity.

Upon IHC analysis of GSC-derived xenografts, by adopting the same panel used for human tumors, we surprisingly noticed that most of the times xenografts perfectly matched the affiliation of the original GBM from which they were derived, independent of the drift demonstrated *in vitro* by the corresponding GSC line (Figure 5.26 and Table 5.26). When a variation was observed, this was generally minor. Surely, a possible confounding factor is the determination of transcriptional affiliation by WB minimal signature for GSCs and by IHC for human and murine tumors. Nevertheless, EGFR, ASCL1 and NDRG1 are present in both classifications, making them somehow comparable and not completely disjointed. Interestingly, other authors recently reported a similar behavior of their GSC lines, which generated xenografts of the same transcriptional affiliation as the parental human tumor, while showing a different subgroup *in vitro* (Stringer *et al*, 2019). In particular, Stringer *et al*. reported only two cases for which transcriptional subgroup was the same in the human tumor, its derived GSC line and subsequent xenograft. The two cases, in fact,

were respectively MES and CL (Stringer *et al*, 2019). In line with this, we found perfect correspondence between original GBM, GSC line and xenografts in 1 MES case (GBM 160315) out of the 14 total new lines, plus the previously isolated line 131210, derived from a gliosarcoma. Both these GSC lines belong to the abovementioned Cluster A. This indicates that extreme MES profile is the most stable and resistant to influence of the microenvironment and of the culture conditions.

Lastly, the absence of tumor formation should incline towards a “non-cancer stem” profile of the failing cell lines. Particularly, the 2 MES lines 160503 and 160525 displayed a peculiar growth pattern *in vitro*. In fact, they replicated fast but grew as single fusiform cells in adherent monolayers, even under NSA conditions. However, also other authors reported generation of GSC lines that displayed peculiar cell morphologies and inability to form neurospheres (Stringer *et al*, 2019). Additionally, whole exome sequencing analyses of 3 of our 5 non-tumorigenic cell lines (the latter 2 were not established yet) demonstrated that they all bore known GBM-related mutations. Moreover, unsupervised clustering grouped these cell lines together with other tumorigenic GSC lines both in exomic and transcriptomic analyses, suggesting they are actual cancer stem cells (Figure 5.34 and Table 5.39). As a matter of fact, if such lines constituted healthy stem cells (which indefinitely propagate in culture, but do not generate tumors upon transplantation) we would expect them to cluster away from tumor cells in a very early tree-branch. Most notably, other authors reported that not all GSC lines are tumorigenic *in vivo*, with a higher incidence of failure in MES lines (Xie *et al*, 2015). In line with this observation, 2 of our non-tumorigenic lines were purely MES and other 2 had a very well represented MES component admixed to CL.

6.2.1 Limitations and conclusions

As for the limitation of this study, the small number of GSC lines established is a concern. Our findings should be replicated on a more conspicuous group of possibly balanced PN, CL and MES lines. Also, transcriptional profiling should be carried out on all samples and at various stages (parental human tumors, early, intermediate, and late *in vitro* passages, and xenografts) to better compare the protein-based classifications. Parallely, for proper insight to be inferred from transcriptomic and exomic studies, both should be carried on the same samples. Our RNAseq and WES analyses only partially overlapped, as we included in WES 3 historical GSC lines (2 PN and 1 CL) and did not include line 170421 (which had not been isolated yet). Besides, protein surrogates adopted for transcriptional affiliation were not univocally

tested by the same methodologies, as GSCs were analyzed only by WB and xenografts only by IHC, with the two methods not perfectly superimposable. Furthermore, accuracy of WB minimal signature is quite high, but still far from perfection.

In conclusion, GSC lines are a powerful tool that can be exploited to model GBM heterogeneity notwithstanding an almost unavoidable *in vitro* drift in transcriptional affiliation from the parental tumor. As a whole, however, PN lines efficiently model biological aspects of PN GBMs, as do MES GSCs for MES GBMs. Additionally, lines at early and intermediate passages are efficacious in modeling human GBM heterogeneity with a dynamic balance of each subgroup inside the same line. Moreover, the molecular affiliation drift observed *in vitro* is completely or partially reversed *in vivo*, proving to be a reliable model for *in vivo* testing. A final interesting finding is the evidence that much less expensive and labor-intensive protein-based categorizations effectively approximate transcriptional classifications that need more dispendious techniques. This also suggests that protein profile may be more important than transcriptional counterpart. Therefore, if consistent resources must be allocated, they should probably be more indicated for single cell proteomic than single cell transcriptomic. Given all this evidence, GSC use as a modeling tool is extensively supported, though many questions still need to be answered.

6.3 GBM transcriptional subtype evolution model.

GBM evolution and plasticity has been the subject of several studies on multiple levels, from genomic to transcriptomic. Körber et al. reconstructed the evolutionary processes of matched pairs of primary and recurrent IDHwt GBMs based on deep whole-genome-sequencing data. Tumor initiation was demonstrated to likely happen up to 7 years before diagnosis, following an early common tumorigenic path, involving chromosome 7 gain, 9p loss, or 10 loss. At this stage, the neoplasm likely remains clinically silent and MRI undetectable. Subsequent stabilizing mutations, such as those in TERT promoter, occur later and are necessary for accelerated expansive growth. (Körber *et al*, 2019)

Additionally, while some studies reported only few mutations common to primary and recurrent GBMs, suggesting selection of clones that branched off early during tumorigenesis (Kim *et al*, 2015), other studies demonstrated a much higher genetic

concordance between primary and relapsing lesions (Körber *et al*, 2019). However, interestingly, both observations might hold true, the former characterizing more frequently distant tumor relapses, while the latter being more typical of local recurrences (Körber *et al*, 2019).

Also for GBMs, an integral role in tumor plasticity is played by epithelial-mesenchymal transition (EMT), which is a well-recognized phenotypic progression of solid tumors and predicts dismal prognosis and treatment resistance (Lv *et al*, 2022). EMT allows a differentiated epithelial cell to assume a mesenchymal state by reducing cellular adhesion and increasing ECM proteases production and cytoskeletal activity. Such modifications induce migratory and invasive capabilities. Also, mesenchymal features have been associated with the gain of a stem cell program in gliomas (Ortensi *et al*, 2013), suggesting an invasive tumor front of migrating GSCs which is further corroborated by reports in the literature (Li *et al*, 2022). Remarkably, the mechanisms of EMT are driven by transcription factors such as SNAI, TWIST, and ZEB, which are altered in GBMs and are differently involved in the regulation of their proliferation, invasion and migration (Ortensi *et al*, 2013).

As EMT is a multiple-level process, reminiscent features influencing the phenotype transition can be detected at genomic, as well as transcriptomic, proteomic and metabolomic level. On a transcriptional level, PN GBMs have been demonstrated to become treatment-resistant and more aggressive, and to attain angiogenic and hypoxic potential by transitioning to MES affiliation. Therefore, from a transcriptional point of view, EMT may be analogously declined as "PN to MES transition", especially at recurrence after chemoradiation (Kim *et al*, 2021). Notably, although this phenomenon was initially described as frequent (Wood *et al*, 2016), other authors recently reported that 55% of IDHwt GBM retained their original transcriptional subtype at recurrence and transition towards a MES profile was not significantly higher than that towards PN or CL subtypes (Wang *et al*, 2017). This may be explained by the fact that transcriptional profile is naturally variable to finely tune the final phenotype and, thus, far from being crystallized in a single direction. Therefore, as far from being an invariantly proven dynamics, transcriptional evolution of GBMs may still offer interesting insights to be eviscerated.

As described above, we utilized transcriptional data of human TCGA GBMs from a publicly available repository to validate our GSC-derived gene signatures. Notably, such repository also contained data from healthy brains, giving us the chance to derive transcriptional profiles peculiar for each subgroup with respect to normal tissue. This is very interesting in light of the fact that the reference molecular profiles described by Verhaak and Wang were derived from transcriptional comparison

between each subgroup and all the others, thus considering only pathological states (Verhaak *et al*, 2010; Wang *et al*, 2017). By comparing only samples of a specific affiliation with healthy tissue and singularly with the other subtypes, we hypothesized to derive insights into a likely transcriptional evolution of the bulk pathological states, where increasing divergence from healthy tissue could act as a metrics to build a potential evolutionary tree.

We subjected RNAseq data from 4 CL, 9 MES, 6 PN and 5 Healthy TCGA samples to paired differential gene expression (Table 5.30). Though the absolute number of genes differentially expressed (DEGs) was almost equivalent for each combinatorial comparison, we found an increasing number of significant DEGs respectively for PN, CL, and MES GBMs compared to healthy controls. Considering paired comparisons between the pathological states, the significant DEGs were consistently lower than comparisons with healthy tissue, and increasing divergence was detected for CL vs PN, to MES vs CL, to finally MES vs PN GBMs.

Upon GSEA analysis with Verhaak's transcriptional signatures (Figures 5.27, 5.28 and 5.29) NEU profile was constantly enriched in healthy controls, consistent with the notion that NEU samples were predominantly healthy tissue infiltrated by tumor leading edge (Wang *et al*, 2017). Interestingly, as expected, PN profile was enriched in PN subtype opposed to normal tissue. However, in the comparison of MES and CL tumors with normal tissue, it was enriched on the healthy side. On the contrary, both CL and MES signatures were invariably enriched in all pathologic states compared to controls. Together with these findings, enrichment profile of Verhaak's signatures in pairwise combinatorial contrasts between the three transcriptional subgroups (Figures 5.30, 5.31 and 5.32, and Table 5.31) seems to suggest that MES tumors are the most divergent from the basal normal state, while the PN subtype is the ground state of GBM. Consequently, our results seem to propend towards a more pronounced PN to MES transition, as it looks quite unlikely that progression of a deranged condition like a malignant tumor may become increasingly more similar to normal healthy tissue. The CL phenotype seems to locate instead in between the two extremes, much closer either to PN profile as indicated by overall number of DEGs (Table 5.30), or to MES as suggested by GSEA enrichment of CL signature.

Notably, these profiles correspond to bulk transcriptomic signal, with recent works unveiling correlations between cellular states and bulk RNAseq profiles. As a matter of fact, Neftel *et al*. reported different frequencies of cellular states combinations, with some more frequent patterns and others virtually absent (Neftel *et al*, 2019). AC-like state (determining a CL affiliation) is recurrently detected together with MES1/2-like states, likely inducing a mixed CL/MES affiliation that can lean towards

extremized CL or MES profiles in a dynamic change of the two states percentual representations. However, other authors have reported how the MES profile is generally stably maintained once developed and selected for (Bhat *et al*, 2013; Gangoso *et al*, 2021). This propends more for an additional drift as a further evolution step in transcriptional profile, rather than a bidirectional balance between CL and MES features. Overall, the extreme complexity of transcriptional characterization of such a heterogeneous pathology may locate the truth somewhere in between these two distinct scenarios.

We next interrogated the DEGs between each pathological states and healthy brain to identify characterizing ontologies of each subgroup. The rationale was to denote which features would be common to all subtypes (thus, possibly inherent to GBM initiation) and which peculiar to single subtypes, suggesting progressive divergence from basal healthy state (Tables 5.33, 5.34 and 5.35). As expected, several features were detected in all the comparisons, though with dramatic changes in significance values. For instance, processes necessary for tumor progression such as EMT and angiogenesis showed progressively increasing significance from PN to CL, to MES tumors. Conversely, the enrichment for E2F4 (a transcription factor expressed in quiescent cells to promote G0-phase) decreased its significance from PN to CL, to MES, suggesting a possible effect on rising proliferation from PN to MES tumors.

Vascular endothelium profile was peculiar of CL GBMs. This is consistent with the fact that, even though angiogenesis was slightly more significant in MES tumors, CL GBMs are highly vascularized due to the pleiotropic effect of EGFR. Microglia profile scored the highest in MES GBMs, together with inflammatory and immune pathways and interleukins and chemokines pathways, in line with a high contribution to this subtype of immune and inflammatory response (Hara *et al*, 2021; Gangoso *et al*, 2021; Kim *et al*, 2021). Notably, features of hypoxia and glycolytic pathway were significantly expressed in MES lesions, indicating a more aggressive course of the disease and in line with other authors reporting a stronger correlation between MES profile and glycolytic metabolism (Neftel *et al*, 2019; Garofano *et al*, 2021). On their turn, hypoxic cells activate pro-angiogenic factors and induce the recruitment of inflammatory cells in addition to prompt PN to MES transition. Additionally, upon oxygen restriction, healthy astrocytes release cytokines that upregulate HIF1 α , helping GBM cells to adapt to hypoxia (Kim *et al*, 2021).

Ontology enrichment analysis on the contrast between MES and PN GBMs (Tables 5.36 and 5.37) showed that ASCL1 pathway was significantly enriched in PN tumors, further corroborating our previously described minimal signature, together with activation of nervous system developmental processes and affinity to neuron-derived

tumors. MES tumors, instead, showed once more stigmata of metastasization, inflammatory and immune response, and EMT. Remarkably, as previously reported, gliomas have been demonstrated to establish functional synapses with healthy neurons and to hijack neurotransmitter signals to drive invasion (Venkataramani *et al*, 2019; Venkatesh *et al*, 2019; Venkataramani *et al*, 2022). Considering this, actual GBM infiltration may be characterized by two very different processes. A first mechanism would be exploited by the PN compartment, infiltrating along the paths indicated by neurochemical signals (and very evident in our murine PN xenografts). A second mechanism would be adopted by the MES compartment, more linked to canonical types of invasion, marked by EMT and metastasization ontologies. Additionally, MES subtype showed also features of extracellular matrix deposition, reorganization, and interaction with it (Tables 5.21 and 5.23). This suggests an invasive profile more similar to metastases, which would also explain findings of our radiogenomic study, where increased hindered diffusion (fecv) signal of MES lesions could be related to more local, less widespread cellularity, as well as to deposition of oriented matrix fibers. Interestingly, when compared to PN tumors, MES GBMs, show significant activation of mannose metabolism and proteoglycans in cancer, in accordance with what recently demonstrated by our group (Pieri *et al*, 2022).

Transcriptional affiliation gained much attention due to the observation of MES GBMs as more aggressive than other subtypes, though significant survival difference was observed only in case of low transcriptional heterogeneity. Additionally, intratumoral heterogeneity raises a critical treatment concern. In fact, different subtypes might have different sensitivities and eradication of cells with a certain affiliation may favor the progression of other phenotypes (Kim *et al*, 2021). Therefore, identification of drugs with possible differential efficacy according to transcriptional subtype is paramount.

Drug repositioning allows to expand therapeutic indications of approved drugs by identifying novel usages for other diseases. Such process is faster and more advantageous than novel drug discovery since both the pharmacokinetics and safety have already been profiled. The rationale relies on the ability of small molecules to target distinct proteins by exploiting off-target effects, which is basically opposed to the traditional goal of one drug-one target. Various computational techniques allow identification of candidate drugs. In particular, systemic perturbation signatures can be used to repurpose molecules, relying on -omics profiles. (Tan *et al*, 2018).

Therefore, as a last step in our analysis, we took advantage of the DEG analyses on GBM subgroups against healthy brain tissue to computationally identify potential drug sensitivities. One of the most comprehensive datasets is the LINCS Connectivity

Map (cMap) (Subramanian *et al*, 2017), which collects data in a highly standardized fashion. Basically, different cell lines are exposed to different molecules, which perturb their transcriptional landscape and allow to infer an average transcriptional response to that specific perturbation. Based on this, drugs inducing a transcriptional alteration specular to that detected by comparing a pathological state to its normal counterpart might be efficacious against the disease itself. We screened our subgroup specific DEGs vs healthy tissue against a publicly available cMap database of compounds (Table 5.38). The more negative the enrichment score, the more the compound is specular to the pathological profile, hence suggesting a potential efficacy of the drug. Notably, GBM subgroups show only partially overlapping sensitivity spectra, with some putative drug responses restricted to either PN or CL subtype. As extensively reported in the literature, the MES subtype has been linked to treatment resistance and favored by selective pressure of therapeutic regimens. For instance, anti-angiogenic agents cause excessive pruning of new vessels, exacerbating hypoxia and subsequent inflammation. TMZ-induced hypermutator phenotype is associated with increased CD8 T-cells, which is typical of MES GBMs. Lastly, radiotherapy leads to activation of NF- κ B pathway and increased immune infiltrate due to augmented vascular permeability. All these effects together strongly favor the development of a MES profile (Kim *et al*, 2021). In line with this, also in our computational analysis, MES subgroup proved to be the most resistant to treatment with a limited sensitivity spectrum that consists only of drugs shared with both the other subtypes. Additionally, the putatively effective drugs show less negative enrichment scores in MES subtype, advocating less efficacy against it.

A recent study using solely transcriptional GBM and perturbation profiles identified 14 drug candidates that were preliminarily tested *in vitro* for efficacy. Eight of them demonstrated significant anti-proliferative activity and growth inhibition at useful concentrations. However, the authors did not differentiate GBM transcriptional subtypes. Notably, none of the perturbation profiles were generated using glioma cell lines, thus confirming the *in silico* prediction as valid across different cancer types, probably as a pondered result of averaged consensus signatures (Lee *et al*, 2016). Even more sophisticatedly, Stathias *et al*. developed a novel platform to pinpoint synergistic drug combinations against GBMs. Preliminary wet experiments demonstrated actual efficacy of the predicted drug combinations in reverting the disease phenotype, suggesting the integration of LINCS and TCGA transcriptional data as very promising (Stathias *et al*, 2018). Once again, however, they did not differentiate GBMs according to transcriptional affiliation. Notably, Stathias *et al*. reported that several molecule classes clustered independently as effective drugs

(Stathias *et al*, 2018). Among them, histone deacetylase (HDAC) inhibitors were also detected by our subtype-stratified analysis as the only drug class (in addition to bromodomain inhibitors) to be efficacious against all 3 subtypes.

This combined evidence underlines the utility of data-driven transcriptional analyses to find potentially useful drugs against GBMs. Although single repurposed compounds may play important roles, much more promising strategies likely rely on combinations of drugs, possibly selecting them based on non-overlapping mechanisms of action (Tan *et al*, 2018).

6.3.1 Limitations and conclusions

Several limitations affect our analysis. First, the small number of TCGA samples used, due to a scarce overlap between samples with downloadable transcriptomic raw data and samples with known affiliation. TCGA data have been used for several years and by many research groups. Therefore, data and metadata can be scattered and partially retrievable from different repositories, depending on the specific aim of each study that re-elaborated TCGA data. Integration of different repositories and several studies, matched with deeper bioinformatics scoping techniques, might implement more acceptable study cohorts. Additionally, healthy controls used for our analyses, despite being downloaded from TCGA, were unmatched to pathological specimens, thus introducing potential biases. Still, averaged transcriptional alterations can significantly be deduced. Also, due to the infiltrative nature of gliomas, it is very difficult to certainly exclude a biasing cancerous infiltration of apparently normal white matter and brain specimens from actual healthy people are obviously very difficult to obtain.

Secondly, putative drug sensitivity prediction was carried out by means of a webtool that is currently based on the mean perturbation-induced transcriptional response of 9 cell lines, none of which derives from GBMs. Notably, the cMap consortium has recently updated its library with several additional cell lines, a number of which are actually derived from gliomas. Unfortunately, the webtool has not yet been fully updated and deeper bioinformatics skills are required to access the library, scavenge it, select, and download the correct data and then locally perform an equivalent, lineage-restricted analysis.

In conclusion, GBM transcriptional profile is extremely variable with possible, apparently contrasting findings and opposing dynamics. Still, various cell-intrinsic, microenvironmental and treatment cues seem to induce a PN to MES transition, according to the contingent selective pressure that dynamically reshapes the

neoplasm in a fitter version. The CL phenotype positions itself in between the two PN and MES extremes. MES GBMs are characterized by a more hypoxic and angiogenic milieu and show overrepresentation of ECM metabolism and interaction with it. On the contrary, PN tumors exploit neuronal ontologies, possibly to establish functional synapses with healthy neurons. This may distinguish two different invasive processes: one typically proneural, following neurochemical signals along white matter tracts, and the other typically mesenchymal, linked to more canonical processes involving ECM. Remarkably, computational analyses confirmed divergent potential GBM drug sensitivity according to transcriptional affiliation, with MES subgroup consistently more resistant than the others.

6.4 Identification of novel putative players in determining GBM transcriptional subtypes.

As a very final step in our analysis, we applied a multistep approach to identify novel putative, functionally relevant targets for future research and possible therapy. As GSC cultures consist of pure tumor cells with no contaminant stroma, we hypothesized that new candidate players inferred from recurring evidence derived from such cultures, once confirmed also in human samples, might have a high degree of functional significance. We also assumed that alterations detected on an exomic level would be more stable and less subject to modulation, thus representing a reliable starting point.

Considering WES findings on our GSC lines, clustering of lines 160503 and 160525 was quite interesting. As a matter of fact, as explained above, these two lines grew as adherent, fusiform, single-cell monolayers even in non-adhesive NSA conditions and failed to generate xenografts in immunocompromised recipients. However, they demonstrated to harbor proper GBM-related mutations and clustered together with other MES GSC lines (Figure 5.34 and Table 5.39) indicating them as true GSCs. Additionally, line 160503 grouped in unsupervised Cluster A at RNAseq, i.e., that containing cells with a marked MES profile. Consequently, we addressed these two GSC lines as “non-tumorigenic MES” lines. WES-based reactome enrichment pathway analysis carried out on them identified the IL7 pathway as the first deranged signaling system. Additionally, transcriptional queries carried on RNAseq-derived GSC cluster A (MES) vs cluster B (non-MES) indicated IL7R as a significant hub protein in the MES

subgroup. These data were also confirmed by pivotal transcriptional analyses on proper human tumors, where IL7R proved to be significantly more expressed in TCGA MES GBMs than in PN tumors and healthy brain. Lastly, previous microarray data from our lab, performed by confronting MES GCLs with PN GSCs, identified both IL7R and IL7 as high rankings in the MES side of the DEG list, suggesting a possible autocrine/paracrine effect.

Most remarkably, as a surface protein, IL7R might represent an ideal GBM biomarker. Hence, we decided to investigate IL7R as a putative novel player in GBM pathogenesis and subgroup specification.

IL7R is an important molecule of the innate and adaptive immune responses, which is critical in a variety of T cells processes (e.g., cell development, survival, homeostasis, establishment of memory cells, and differentiation). It promotes B cell development through JAK/STAT and PI3K/AKT downstream signaling and may induce thymic progenitor cells to differentiate into dendritic cells or macrophages (Kim *et al*, 2020; Barata *et al*, 2019). Splice variants of IL7R were documented in developing human brains and in normal human neuronal progenitor cells. Such isoforms were shown to influence neural progenitor differentiation by inducing differential gene expression and altering neurodevelopmental pathways and glia formation (Moors *et al*, 2010). Peculiarly, IL7R seems not to be circumscribed to neural progenitor cells, as it was detected also in mature neurons and astrocytes, with downstream activation of JAK/STAT driving to astrogliogenesis and playing a role in reactive gliosis (Moors *et al*, 2010).

Sound evidence of IL7R tumor initiating role in hematologic malignancies has been reported (Almeida *et al*, 2021). Increased IL7R expression was linked with CNS involvement in pediatric B cell acute lymphoblastic leukemia (Mohme *et al*, 2020). Additionally, IL7 proved to sustain growth of breast cancer derived cells in culture (Vitiello *et al*, 2018). Multiple evidence shows how epithelial cancers aberrantly express IL7R and IL7 in a likely relevant activation loop. Neoplasms may in fact derive benefit from their induced chronic activation of immune and inflammatory response, inducing pro-tumoral effects. High expression of IL7 and IL7R has been detected in lung, pancreatic, and esophageal cancers. Increased IL7/IL7R levels have been correlated to metastases in colon, breast, and lung tumors, and to poor prognosis in breast, lung, and prostate carcinomas (Barata *et al*, 2019). However, expression patterns and roles in solid tumors remain to be fully elucidated, even though multiple evidence is being collected for its active role in tumor cell proliferation, migration, epithelial to mesenchymal transition, resistance to chemotherapy, lymphangiogenesis, maintenance of stem properties, and osteoclastogenesis and

bone metastasization (Vitiello *et al*, 2018; Jian *et al*, 2019; Kim *et al*, 2018; Barata *et al*, 2019).

Our work revealed that patients harboring IL7R-high GBMs have a significantly worse prognosis than those with low IL7R expression and that there is a significant association between the MES subgroup and IL7R levels in TCGA patients (Figure 5.35). Notably, transcriptional analysis carried on multiple tumor and cell lines showed no specificity for solid cancers, with generally low levels in gliomas (Figure 5.36). GSEA analyses for comprehensive interleukin receptors and IL7 pathway conducted on DGE data obtained from human TCGA samples showed significant expression of IL7R with increasing ranking from PN vs Healthy contrast to MES vs Healthy, and MES vs PN GBMs (Table 5.40). In addition, also IL7 was significantly identified in all contrasts (Table 5.41), but interestingly to a lesser extent than its receptor in the MES subtype. However, these findings might be contaminated by the expression in the tumor immune microenvironment.

To better discern the situation, we investigated IL7R expression in our GSCs. Interestingly, upon GSEA analysis, the comprehensive interleukin receptors module was significantly enriched in MES GSCs with respect to non-MES/PN lines; however, IL7R was not detected as differentially expressed (Figure 5.39). Most remarkably, when comparing the two non-tumorigenic MES GSC lines to the other tumorigenic MES GSCs, IL7R was detected as differentially expressed, ranking very high on the tumorigenic MES side of the comparison. WB quantifications of IL7R demonstrated a significant difference between MES GSCs and GCLs vs PN lines (Figure 5.41). Intriguingly, such difference became a nonsignificant trend when the compared lines were at intermediate passages since their stabilization, when they still showed variably heterogeneous transcriptional affiliation and modulation of defining markers. Also, dynamic IL7R regulation during the early and intermediate stages of GSC line establishment cannot be excluded (Figure 5.42). Therefore, differences among transcriptional subgroups may be initially inconstant and become increasingly divergent as the cell line is progressively cultured.

Parallely, we performed immunostaining on human specimens of each GBM subgroup, with the evidence of stronger signal in MES samples (Figure 5.45). Also, the protein was clearly expressed by tumoral cells, discarding the hypothesis that such difference was mostly due to immune infiltrate. Peculiarly, IL7R staining starkly demarcated the MES component in a mixed PN-MES tumor (Figure 5.46), suggesting a possible role as a diagnostic marker for identifying the most aggressive component in this pathology. Most remarkably, when we analyzed rodent xenografts generated by implanting IL7R-positive GSCs in immunocompromised mice, we could not detect

any signal (Figure 5.44). However, since IL7R is a protein implicated in adaptive immunity, we wondered whether the absence of adaptive immune response (and more specifically of T-cells) in NSG mice could impair IL7R expression in our orthotopic xenografts. To investigate this hypothesis, we selected the murine glioma cell line GL261, which proved to harbor IL7R at WB assay, and transplanted it in immunocompetent and immunocompromised mice (Figure 5.47). Xenografts generated in the former resulted IL7R-positive, while those in the latter were IL7R-negative. Therefore, expression of IL7R on tumoral cells may require a functioning immune system and may represent a tumor-microenvironment crosstalk which is worth further investigation.

Further insight into IL7R mechanism of action provides the rationale by which it may favor a MES affiliation. As a matter of fact, one of its downstream effectors is STAT3, which has been demonstrated to reprogram neural stem cells along a mesenchymal lineage (Ortensi *et al*, 2013). Most notably, there is evidence in the literature of a hybrid interleukin-7/hepatocyte growth factor (IL7/HGF) which can signal through both IL7R and c-Met. In fact, IL7R has been demonstrated to heterodimerize with c-Met in B-cell precursors (Lai *et al*, 2006). This is quite interesting, as c-Met is the product of the gene MET, which plays a fundamental role in MES GBMs (De Bacco *et al*, 2012; Boccaccio & Comoglio, 2013). Therefore, a similar interaction with IL7R might play a part also in gliomas.

Remarkably, there are no systematic previous reports of IL7R in GBMs. Up to now, very few studies only utilized IL7 in an antitumoral setting, either to improve efficacy and safety of CAR T-cell treatments against GBMs (Shum *et al*, 2017; Huang *et al*, 2021), or to modulate the immunosuppressive microenvironment through locally delivered mesenchymal stem cells expressing IL7 (Mohme *et al*, 2020). To our knowledge, the only published pivotal work to relate GBMs and IL7R is an *in vitro* study performed on canonical GCLs that demonstrated how addition of IL7 to culturing medium increases cell resistance to the chemotherapeutic agent cisplatin. (Cui *et al*, 2012) However, the work by Cui *et al*. is flawed on multiple levels. First, it is performed on canonical cell lines that were used in the past as GBM models, but are no more considered reliable, as they more liken sarcomatous metastases than actual GBMs. Therefore, nowadays patient-derived GSCs are considered the proper modeling tool for the parental pathology. Second, the assays were performed only *in vitro*, with no actual *in vivo* validation, and by testing a chemotherapeutic agent that is not commonly used for GBMs in clinical routine.

6.4.1 Limitations and conclusions

Some limitations affect our study. First, most of the evidence is derived by *in silico* analyses and the *in vitro* and *in vivo* findings, though quite indicative, are drawn from very small sample subsets, thus conferring only a preliminary profile to our study. *In vitro* and especially *in vivo* assays should be replicated on larger cohorts. Besides, additional functional experiments should be performed, for instance trying supplementing GSC culture media with human recombinant IL7 and see if an actual mesenchymalization is observed. Overexpression or RNA-interference induced silencing of IL7R should be carried out, assessing proliferation, tumorigenic potential, or resistance to temozolomide as read outs. Interestingly, IL7 pathway rescue in non-tumorigenic MES lines would prove of particular value if it restored their ability to engraft in rodent brains. Remarkably, in this setting, simple overexpression of IL7R may not suffice, as no mutations in either IL7 or IL7R were detected by WES and the pathway may be affected by modifications in downstream effectors. Lastly, the interplay between host immune system and IL7R-expressing GSCs should be better evaluated before inferring dependence of IL7R tumoral expression on immune system. In this sense, though complex, xenotransplantation of GSCs into immunocompromised and humanized mice (i.e., replaced for a functioning human immune system) may be clarifying.

In conclusion, to our knowledge, we are the first to describe the role of IL7R in GBMs, though only preliminarily. In our investigation of transcriptional subgroups, we identified IL7R as an interesting candidate for characterizing and addressing the most malignant MES affiliation. We also provided evidence for this protein to be a link in the crosstalk between the immune microenvironment and tumoral cells. If confirmed by further studies, IL7R may be used as a marker to distinguish the aggressive, MES component in this protean pathology. Additionally, due to its nature of surface signaling protein, it may also represent an exploitable therapeutic target for GBM treatment.

7 MATERIALS AND METHODS

7.1 Patient enrollment

Over the 5-year period from March 1st 2016 to February 28th 2020, before the COVID 19 outbreak, all the patients accessing the Neurosurgical Department of San Raffaele Hospital for suspected high grade intra-axial primary lesions, either on elective or emergency basis, were prospectively screened.

Patients were considered eligible for protocol enrollment when they had no previous oncological history.

Informed consent to participate was regularly collected from all patients, according to the protocol 01CSC07, which was revised and approved by IRCCS San Raffaele Hospital Ethical Committee. A copy of the consent, describing the rationale of the study, was given to each patient. The signed copies collected from all patients were stored and secured in a safe place in our laboratory, inaccessible to third parties.

Conventional preoperative MRI studies were conducted on all patients, while advanced imaging for research intent was organized whenever possible, without postponing the surgery date, for the patients' sake.

Basal, preoperative advanced MR imaging consisting of both diffusion and perfusion-weighted acquisitions or, at least, diffusion studies were deemed necessary inclusion criteria. Therefore, patients for whom such MRI were not available were excluded from the study, unless they presented clinical features suggesting particularly aggressive disease behavior or evolution from lower-grade gliomas. In such cases surgical samples were collected anyway to isolate GSCs, even if the parental tumor would have not been included in the radiogenomic evaluation.

Patients whose lesion resulted to be a metastasis or a glial tumor other than Glioblastoma or CNS WHO grade IV Astrocytoma were excluded from the study.

Data regarding clinical presentation, neurological evaluation on admission and surgical procedure were collected for each patient enrolled.

7.2 Patients' MRI acquisition

Datasets were acquired on a 3T Ingenia CX scanner (Philips Healthcare, Best, The Netherlands), using a 32-channel head coil.

MRI acquisitions were carried out as described by Pieri et al. and here following reported: conventional MRI protocol included an axial 3D fluid attenuated inversion recovery (3D-FLAIR) (TR/TE/TI 9,000/290/2,500 ms; flip angle, 40; 204 slices; thickness, 0.7/−0.5 mm gap; matrix, 204 × 197; SENSE reduction factor R = 2; acquisition time 7 min 30 s) that was used to manually segment the tumor considering hyperintense regions.

NODDI protocol consisted in a two-shell acquisition based on axial single-shot spin-echo echo planar imaging with an anterior–posterior phase-encoding direction that included:

- HARDI acquisition: 60 diffusion-weighted volumes (diffusion gradients were applied along 60 noncollinear directions; b-value, 3,000 s/mm²).
- DTI acquisition: 35 diffusion-weighted volumes (diffusion gradients were applied along 35 noncollinear directions; b-value, 711 s/mm²).
- 11 “B0” volumes without diffusion-weighting (b-value, 0 s/mm²), whose acquisition was placed in between the diffusion-weighted volumes.

Finally, a “reverse B0” volume without diffusion-weighting was acquired (b-value, 0 s/mm²), which shared with the NODDI sequence all the geometrical features but the phase-encoding direction, that was posterior–anterior, to allow for the subsequent correction of susceptibility artifacts (Pieri *et al*, 2021).

Manual segmentation of tumoral lesions was performed with the software ITK-SNAP (v.3.8.0) on the FLAIR and post-contrast T1 images.

7.3 Surgical sample collection

Surgical sample was collected either directly in the operating theater at the time of surgical excision of tumor mass, or in the Anatomopathological lab where it was sent in sterile saline solution. Fragment allocation was decided by an expert pathologist so to represent the whole heterogeneity of the sample, including more peripheral as well as more central tissue. The extent of the sample allocated for research purposes was always evaluated in order not to subtract necessary material for proper histological and molecular diagnosis for clinical routine and subsequent patient management.

In case the surgical sample was not considered enough for research allocation, the whole specimen was paraffin-embedded for clinical routine. White sections were then requested from the Pathology department for immunohistochemical molecular subgroup classification.

7.4 GBM specimen processing

Patients' GBM samples were minced with a scalpel and tissue fragments were stored as following: 1) partly in 4% PFA overnight at 4°C; 2) partly as "dry" pellet at -80°C for Western Blot; 3) partly in RNA-Later stabilizing buffer for 24 hours at 4°C and then shifted to -80°C; 4) partly in 90% FBS/10% DMSO and transferred at -80°C for long term storage.

A fifth part of the fragments was digested at 37°C for 60 minutes in EBSS containing 16 U/ml of papain (Worthington, Lakewood, NJ), 0.2 mg/ml of cysteine and EDTA and 1% DNase. After incubation, tumor tissues were mechanically dissociated and purified from debris by means of differential centrifugation, as described by Galli et al. (Galli *et al*, 2004)

The so-obtained single cells suspension was plated in complete stem cell medium (DMEM-F12 supplemented with hormone mix, EGF and FGF2).

7.5 GSCs culture propagation

GSC lines were cultured using DMEM-F12 (Invitrogen) supplemented with 2% Glutamax (Invitrogen), penicillin/streptomycin (100 units/ml penicillin, 100 µg/ml streptomycin - Invitrogen), glucose 0.6% (Sigma-Aldrich), HEPES 5 mM (Invitrogen), NaHCO₃ 0.11% (Invitrogen), Bovine Serum Albumin 2 mg/ml (BSA - Sigma-Aldrich) Hormone Mix (B27 analogue, homemade), Heparin 0.2% (Sigma-Aldrich), EGF 20 ng/ml (Peprotech), FGF2 10 ng/ml (Peprotech).

Hormone mix is constituted by DMEM-F12, glucose 0.6% (Sigma-Aldrich), HEPES 5 mM (Invitrogen), NaHCO₃ 0.11% (Invitrogen), 77.2 mg/ml putrescine (Sigma-Aldrich), 0.5 mg/ml insulin (Sigma-Aldrich), 100 mg/ml apo-transferrin (Sigma-Aldrich), sodium selenite 3 mM (Sigma-Aldrich), progesterone 2 mM (Sigma-Aldrich).

Upon GSC neurosphere formation, they were harvested and centrifuged, and the pellet mechanically dissociated to obtain single cell suspension that was again plated in complete culture medium to allow for cell lines propagation.

7.6 Orthotopic implantation of GSCs

GSC tumorigenicity was assessed by orthotopically injecting the distinct cell preparations in either nu/nu, NSG or C57bl/6 mice striatum. Either 300.000 or 400.000 cells, as appropriate for in vitro growth kinetics, resuspended respectively in 3 or 4 μL of DMEM supplemented with 1:500 DNase (Sigma-Aldrich) was loaded into a HPLC micro-syringe (Hamilton) and delivered into the right striatum (0.5 $\mu\text{L}/\text{min}$) by stereotactic injection. Mice were anesthetized by means of Avertin® and fixed on the stereotaxis instrument (Kopf) using two ear bars. After accurate disinfection, a cut was done on the animal head by means of sterile scalpel to make the skull visible, and the coordinates were fixed from bregma (antero-posterior: 0; medio-lateral: 2.5 mm; dorso-ventral: 3 mm). A burr hole was performed with an automatic drill, the syringe was gradually inserted and let sit for a minute to allow for parenchyma adjustment, then starting with cell injection (0.5 $\mu\text{L}/\text{min}$). After surgical procedure the skin was again disinfected and stitched up with 5-0 non resorbable nylon suture. Mice were then kept on warm pad and wrapped in clean paper foil to prevent hypothermia.

7.7 Xenograft MRI acquisition

Xenograft MRI acquisitions were carried out, as thoroughly reported by Esposito *et al.*, on a 7 Tesla preclinical magnetic resonance scanner (Bruker, BioSpec 30/70 USR, Paravision 5.1, Germany) at OSR Experimental Imaging Center (EIC), equipped with 450/675 mT/m gradients (slew-rate: 3400-4500T/m/s; rise-time: 140ms). A phased-array mouse-head coil with four internal preamplifiers was used as receiver, coupled with a 72 mm linear-volume coil as transmitter (Esposito *et al*, 2013).

Mice were anaesthetized as described by Visigalli *et al.* with Sevoflurane® (5% for induction and 2% for maintenance), in a 95-98% O₂ mixture. During acquisition, mice were positioned prone on a dedicated temperature control apparatus to prevent hypothermia (Visigalli *et al*, 2009).

The same imaging protocol used for patients, adapted for small animals, was applied. Manual segmentation of tumoral lesions was performed with the software ITK-SNAP (v.3.8.0) on the DWI B₀ image to avoid coregistration errors.

7.8 MRI image preprocessing

This step was conducted in collaboration with Dr. Antonella Castellano and Ing. Nicolo Pecco from the Department of Neuroradiology, San Raffaele Scientific Institute, Milan, Italy and as thoroughly described by Pieri et al. and here following reported. All NODDI volumes were corrected for movement and eddy-current distortions, using the “eddy” tool of FMRIB Software Library (FSL, University of Oxford, <https://fsl.fmrib.ox.ac.uk/fsl/>). Once preprocessing was completed, the Watson-NODDI model was fitted to the two-shell dMRI datasets (NODDI acquisition: 60 directions at b-value 3,000 s/mm², 35 directions at b-value 711 s/mm², 11 B0 volumes) using the MATLAB NODDI toolbox (<http://mig.cs.ucl.ac.uk/Tutorial.NODDI Matlab>) to extract the following NODDI maps: voxel fraction of Gaussian anisotropic diffusion (extracellular volume fraction [fecv]), voxel fraction of non-Gaussian anisotropic diffusion (intraneurite volume fraction [ficv]), voxel fraction of isotropic Gaussian diffusion (fiso), and orientation dispersion index (odi) maps (Pieri *et al*, 2021). More in detail, the NODDI toolbox outputs the isotropic and intraneurite compartments of each voxel, as well as the odi map, which quantifies angular variation of neurite orientation: the more coherently oriented are the fibers, the lower is the odi value. Then, the output compartments were reparameterized to derive the extraneurite compartment, so that the sum of ficv, fecv, and fiso equaled 1 in each voxel (Caverzasi *et al*, 2016; Pieri *et al*, 2021).

NODDI compartment maps were also combined into a single 4D RGB image (red for fecv, green for ficv, and blue for fiso) for visualization and quality-check purposes.

FSL built-in “dtifit” tool was separately applied to DTI shell (DTI: 35 directions at b-value 711 s/mm²) to estimate the diffusion tensor, and to generate FA and MD tensorial maps (Pieri *et al*, 2021).

FSL Brain Extraction Tool (bet) was used to skull-strip the 3D-Flair image as well as the DWI b=0 image. Skull-stripped 3D-Flair and DWI b=0 images were coregistered through an affine transformation. Output transformation matrix was then applied to the 3D-Flair segmentation-mask to achieve the best overlap between the DWI tumor lesion and the 3D-Flair segmentation-mask (Pieri *et al*, 2021).

7.9 Radiomic features extraction

This step was conducted in collaboration with Dr. Antonella Castellano and Ing. Nicolo Pecco from the Department of Neuroradiology, San Raffaele Scientific Institute, Milan, Italy.

Image pre-processing steps except segmentation are not a requirement for 3-dimensional texture analysis but help enhancing texture features and maintain uniformity and standardization. Image spatial resolution is one of the most influential parameters in texture analysis and images interpolation can be a way to obtain higher resolution. It is well known that second order features could be affected by image characteristics such as image contrast or brightness (Schad, 2004). Hence, a grey-level normalization procedure, $\pm 3SD$ normalization (Collewet *et al*, 2004), was implemented to increase robustness and reproducibility of the texture features so that gray-levels located outside the range $[\mu-3SD, \mu+3SD]$ were excluded from the analysis. Finally, quantization of gray levels was performed by decreasing the number of gray-levels to 6 bits/pixel meaning 64 levels of gray for the calculation of GLCM and GLRLM features (Larroza *et al*, 2016).

Patients were analyzed and interpolation was applied to same-type images to obtain the closest isotropic voxel. DWI scans average in-plane voxel resolution was $1.95 \times 1.95 \text{ mm}^2$ while average slice thickness was 2.16 mm. The smallest voxel spacing over same-type images was used to obtain isotropic voxels (Depeursinge *et al*, 2014), therefore all NODDI and DTI-derived maps were resampled to $2 \times 2 \times 2 \text{ mm}^3$ to enhance reproducibility and standardize the texture analysis in all 3 directions. This way, texture features will not be affected by uneven voxel size. Differently, mice image acquisition DWI scans in-plane voxel resolution was $0.11 \times 0.11 \text{ mm}^2$ while slice thickness was 0.75 mm. We assume that oversampling of axial plane resolution, which contain relevant information for molecular class, could lead to loss of this information. Therefore, mice maps were resampled to isotropic voxels of dimension 0.11 mm^3 .

Texture features were extracted from each diffusion map by using the 3D-FLAIR and post-Gd T1 masks in patients and the B0 image in mice. However, we concentrated on human 3D-FLAIR masks (comparable to mice B0) as post-contrast T1 was not available in most mice, due to lack in contrast uptake by some xenografts. Radiomic features were extracted using Pyradiomics software (v2.2.0, <http://www.radiomics.io/pyradiomics.htm>) from the original image and customizing the extraction (Van Griethuysen *et al*, 2017).

A total of 91 features were extracted for each map, leading to 546 features for each mask: 2 masks for patients (3D-FLAIR and post-Gd T1) and 1 for mice (B0). In particular, (I) 18 histogram features, based on the count of pixels in the ROI that possess a given grey-level value, (II) 22 grey-level-co-occurrence matrix (GLCM), contain probabilities of co-occurrence of pixel pairs with given grey levels, (III) 16 grey-level-run-length matrix (GLRLM), represent runs of pixels having the same grey-level value, (IV) 16 size zone matrix, characterize pixels zones in the images, (V) 14 grey-level-dependence matrix (GLDM), quantify the level of dependency in an image and finally (VI) 5 neighboring-grey-tone difference matrix (NGTDM), explaining the difference between a grey value and its neighbors. Shape-based features were not calculated due to different dimension of ROIs and manual lesions segmentation.

7.10 Radiomic analysis and feature selection

This step was conducted in collaboration with Dr. Antonella Castellano and Ing. Nicolo Pecco from the Department of Neuroradiology, San Raffaele Scientific Institute, Milan, Italy.

Consecutive feature selection and dimension reduction was performed on both the mice and patient datasets throughout a Univariate feature selection approach. Statistical analyses were conducted with MATLAB2021. Testing for group differences was performed using Wilcoxon sum-rank test or Unpaired T-test after testing for normality of data distribution with the Shapiro-Wilks test. A two-tailed p-Value < 0.05 was considered statistically significant.

Significative radiomic features common to patients and mice and with the highest discrimination ability among the MES and non-MES sets were used to train a K-Nearest Neighbor (KNN) algorithm. Mice dataset was selected as the train set due to the balanced numbers of the MES and non-MES samples, and their extremized tumor characteristics. This dataset was equally divided into train and validation subsets for hyperparameters tuning. KNN algorithm parameters were set to default except the number of neighbors (K: from 1 to 8, with resulting best value as 3) and the metric used to compute the distance between samples (euclidean, manhattan, chebyshev and minkowski, with chebyshev resulting best metric).

All features were then normalized through standard score normalization in Python. Patient dataset was used as the test set. The algorithm performance was described by means of accuracy, precision, recall, F1-score as well as sensitivity and specificity.

7.11 Radiomic analysis and feature selection with data augmentation

This step was conducted in collaboration with Dr. Antonella Castellano and Ing. Nicolo Pecco from the Department of Neuroradiology, San Raffaele Scientific Institute, Milan, Italy.

Patient and mice data augmentation for class balances was performed with the Synthetic Minority Oversampling Technique (SMOTE) tool. As thoroughly described on <https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification/?cv=1>, SMOTE selects examples that are close in the feature space, drawing a line between the examples in the feature space and drawing a new sample along that line. Specifically, a random example from the minority class is first chosen. Then k of the nearest neighbors for that example are found ($k=4$). A randomly selected neighbor is chosen, and a synthetic example is created at a randomly selected point between the two examples in feature space.

Consecutive feature selection and dimension reduction was performed on both mice and patient datasets to select common features with the most discriminative power between the two classes. Dimensionality reduction was achieved by using VIF (Variance Inflation Factor) algorithm to drop features explaining the same amount of information with a correlation coefficient higher than 0.7.

Univariate feature selection was achieved using statistical algorithm (ANOVA-ScikitLearn Fclassif) providing most important features in discriminating the two classes and common to patients and mice. Significant features found by the statistical algorithm were then normalized (Z-score normalization) and used as an input for model evaluations.

Different machine learning approaches such as logistic regression, K-nearest-neighbor, support vector machines and random forest classifiers were tested. We trained each classifier in Leave-One-Out cross validation (LOO) to achieve balanced trade-off between performance and robustness. A further nested 4-fold cross-validation was performed during hyperparameter tuning. We used mice dataset as training and patient dataset as test.

Accuracy, precision, recall, F1-score, Area under the curve (AUC) as well as sensitivity and specificity were used as metrics to describe the algorithm performances. Statistical analyses were carried out with ScikitLearn Fclassif algorithm that performs a test for group differences. A two-tailed p -value of < 0.05 was regarded as statistically significant.

7.12 RNA Sequencing (RNAseq) analysis of GBM tissues and GSCs

Total RNA from GBM tumor specimens and from GSC lines was extracted using the RNeasy Mini kit (Qiagen, Chatsworth, CA, USA) according to the manufacturer's protocol and sent to San Raffaele Scientific Institute -Omics facility for subsequent processing.

As we proceeded in a previous work, the cDNA was synthesized starting from total RNA by QuantSeq 3' mRNA-Seq Library Prep Kits (Lexogen). After barcoding, the RNA libraries were pooled, denatured, and diluted to 2.4 pM final concentration. RNAseq was performed using NextSeq 550 (Illumina) with SMART-Seq protocol, yielding an average of 15×10^6 clusters for each sample (Conti *et al*, 2021).

Sequences were trimmed with trimmomatic tool and subsequently aligned using STAR (version 2.5.3a) on the reference genome hg19; association between reads and genes was performed by featureCounts, using GENCODE (version M13) basic annotation as reference. Normalization and analysis of count data was performed using the R package DESeq2 (version 1.0.19) (differential gene expression analysis based on the negative binomial distribution of counts data). The independent filtering of genes with low counts was set to a mean of 9 raw counts between all samples. The cutoff imposed for differential gene expression was the one suggested by the Sequencing Quality Control Consortium, which defines a gene as differentially expressed when it has an associated FDR value lower than 0.1 (adjusted p-value < 0.1, Benjamini and Hochberg correction) and, at the same time, the absolute value of its log₂ fold change is greater than 1 (log₂FC >1 or log₂FC <-1) (Conti *et al*, 2021).

7.13 Differentially Expressed Genes (DEG) in TCGA dataset

TCGA GBM mRNA normalized (Level 3) raw data publicly available at <http://firebrowse.org/?cohort=GBM> were downloaded and prepped for further analyses with the R script 1_Raw_counts_prep.R (see Appendix). Transcriptional subgroup affiliation of the available samples was obtained by cross-matching sample IDs on R2 platform (<https://hgserver1.amc.nl/cgi-bin/r2/main.cgi>).

Samples data for subsequent steps were cut down to subsets with R script 2a_Subsetting_raw_counts.R (see Appendix). Recurrent samples were removed from the analysis. Raw counts were prefiltered for genes with extremely low counts (< total samples) and rounded to the nearest integer. Differentially expressed genes

analysis was conducted by means of DESeq command in the Bioconductor - DESeq2 R package (see Appendix – R Scripts: 3a_DGE.R). The automatic independent filtering was based on mean normalized counts for each gene. FDR cutoff alpha was set to 0.05 and an adjusted p-value < 0.05 was considered for significance of DEGs.

Analyses were executed with R (v4.1.2), RStudio (Ghost Orchid release v2021.09.01) and appropriate R packages.

7.14 Gene Set Enrichment Analysis (GSEA)

Differentially expressed genes (DEGs) either provided by the San Raffaele Scientific Institute -Omics Facility or obtained from in silico analyses of TCGA GBM samples raw data, were filtered to eliminate transcripts with adjusted p-value NA. Gene transcripts were reordered in a ranked list for descending Log2FC and saved as .rnk file (0_Getting_preranked_no_NA.R, 4_xlsx_to_rnk_no_NA.R).

Gene signatures were either downloaded from publicly available repositories (<https://www.gsea-msigdb.org/gsea/index.jsp>), as done for Verhaak's GBM subtype signatures and predefined molecular pathways, or transformed to .gmx files with ad hoc R script (0_Signatures_to_gmx.R).

R conversions were executed with R (v4.1.2), RStudio (Ghost Orchid release v2021.09.01) and appropriate R packages (see Appendix – R Scripts).

GSEA analysis was performed with the Broad institute application GSEA (v4.1.0), downloadable at <https://www.gsea-msigdb.org/gsea/index.jsp>, by importing preranked gene lists and gene signatures. Analyses parameters were as follows: 1000 permutations, No_collapse (use dataset as it is), exclude signature genesets larger than 500 and smaller than 5. The remaining parameters were set as default for the function Run_GSEA_preranked.

Enrichment plot, Normalized Enrichment Score (NES), False discovery rate (FDR)-adjusted p-value and Family-wise error rate (FWER)-adjusted p-value were considered for each comparison.

7.15 Western blot analysis (WB)

Lysates from GBM tissues, GSCs and GCLs were prepared using RIPA buffer supplemented with protease and phosphatase inhibitors (Roche). Loading dye 4X containing β -mercaptoethanol were then added to the volume of lysates

corresponding to the quantity of protein desired and all were denatured for 5 minutes at 95°C. Lysates were then loaded on a 10% Stain-Free FastCast Acrylamide gel cast (BioRAD) and then transferred on nitrocellulose by means of electroblotting (7 minutes at 2.5A, 25V using BioRAD TransBlot Turbo transfer system v1.02). After protein transfer, the membrane was blocked with TBS (140 mM sodium chloride, 20 mM Tris-HCl pH 7.6) supplemented with 5% of milk for 1 hour. The primary antibody was then added, diluted at the appropriate concentration in 0.1% of TBS Tween-20 (TBS-T) supplemented with either 5% of milk or 3% BSA and incubated overnight under gentle shaking.

Primary antibodies used were: 1:1000 rabbit anti-EGFR (Cell Signaling, Beverly, MA, USA), 1:1000 rabbit anti-NDRG1 (Cell Signaling, Beverly, MA, USA), 1:1000 rabbit anti-phosphoNDRG1 (Cell Signaling, Beverly, MA, USA), 1:250 mouse anti-ASCL1 (anti-MASH1, BD Pharmingen, San Diego, CA, USA), 1:1000 rabbit anti-MET (Cell Signaling, Beverly, MA, USA), and 1:1000 rabbit anti-IL7R (Abcam, Cambridge, UK). As loading control, a 1:3000 rabbit anti-Calnexin (Genetex, Irvine, CA, USA) or a 1:5000 mouse anti-GADPH antibody was used (Sigma-Aldrich, St Louis, MO, USA).

Membranes were then washed three times with TBS-T (15 minutes each wash) and incubated with secondary antibodies (1:3000; BioRAD) for 1 hour at room temperature. After incubation, membrane was washed for two times with TBS-T and one time with TBS (10 minutes each wash). Peroxidase activity was detected using Clarity Western ECL assay (BioRAD) and Alliance Mini HD9 development (Uvitec, Cambridge, UK).

Quantifications of bands relative to reporter protein was performed with ImageJ v1.8.0_172 and R script QuantoPlot.R (see Appendix – R Scripts).

7.16 k-Means Clustering analysis based on transcriptional profile

Human GBM samples clustering analysis was performed in silico on the publicly available R2 platform (<https://hgserver1.amc.nl/cgi-bin/r2/main.cgi>) by selecting samples for which the transcriptional affiliation was known in the TCGA-540-MAS5.0-u133a dataset. The software was requested to identify either 2 or 3 sample clusters as appropriate, based on the transcript levels of user-defined gene lists. The total rounds for which the computation was repeated, and number of passes were set to 10, the transformation applied was set to Z-score or Log2(Z-score) as appropriate for consistent, reproducible, best-fitting results.

7.17 Immunohistochemical (IHC) analysis and subgroup affiliation

Bright field immunohistochemical analyses were performed by Prof. Pietro Luigi Poliani and Dr. Manuela Cominelli from the Pathology Unit, Department of Molecular and Translational Medicine, University of Brescia, Italy, and by Ilaria Pagano in our laboratory.

For immunohistochemical staining we proceeded as described in (Cominelli *et al*, 2015): 2 µm sections were cut from paraffin blocks, deparaffinized and rehydrated. The endogenous peroxidase activity was blocked with 0.3% H₂O₂ in methanol for 20 minutes. Antigen retrieval was performed using a 0,05% protease type XIV (Sigma-Aldrich) solution at 37°C for 20 minutes or a microwave-oven or a thermostatic bath in either 1.0 mM EDTA buffer (pH 8.0) or 1mM Citrate buffer (pH 6.0). Then slides were washed in TBS (pH 7.4) and incubated for an hour or overnight in the specific primary antibody diluted in TBS 1% Bovine Serum Albumin (BSA) or in Antibody Diluent with Background Reducing Components (DAKO). Then sections were washed in TBS and signal revealed using Envision+System-HRP Labelled Polymer Anti-mouse or Anti-Rabbit (DAKO) or Novolink™ Polymer Detection System (Novocastra™) or Rabbit specific HRP/DAB Detection IHC Detection Kit-Micro-polymer (abcam), followed by Diaminobenzidine (DAB) as chromogen and Hematoxylin as counterstain.

Images were acquired with a Nikon DS-Ri2 camera (4908x3264 full-pixel) mounted on a Nikon Eclipse 50i microscope equipped with Nikon Plan lenses (x10/0.25; x20/0.40; x40/0.65; x100/1.25) using NIS-Elements 4.3 imaging software (Nikon Corporation).

Primary antibodies used for the staining were: mouse anti-EGFR (clone E30) #M7239 (Dako, Glostrup, Denmark, USA 1:20, 1 hour), 0,05% protease type XIV (Sigma-Aldrich) solution at 37°C for 20 minutes, Envision+System-HRP Labelled Polymer Anti-Mouse (DAKO); mouse anti-p53 Ab-5 (clone DO-7) #MS-186-P (Thermo Scientific, Waltham, MA, USA 1:2, 1 hour), thermostatic bath in EDTA, Envision+System-HRP Labelled Polymer Anti-Mouse (DAKO); rabbit anti-PDGFRα #PA5-16742 (Thermo Scientific, Waltham, MA, USA 1:50, 1 hour), thermostatic bath in EDTA, Envision+System-HRP Labelled Polymer Anti-Rabbit (DAKO); rabbit anti-Olig2 #AB9610 (Chemicon, Darmstadt, Germany 1:600), thermostatic bath in EDTA, Envision+System-HRP Labelled Polymer Anti-Rabbit (DAKO); mouse anti-ASCL1 (clone 24B72D11.1) #556604 (BD Biosciences, Franklin Lakes, NJ, USA 1:50 overnight), thermostatic bath in EDTA, Novolink™ Polymer Detection System (Novocastra™); rabbit anti-pNDRG1 (Thr346) (clone D98G11) #5482 (Cell Signaling,

Beverly, MA, USA 1:100 overnight), thermostatic bath in EDTA, Envision+System-HRP Labelled Polymer Anti-Rabbit (DAKO); rabbit anti-YKL40 #4815 (Quidel, San Diego, CA, USA 1:100 overnight), microwave-oven in EDTA, Envision+System-HRP Labelled Polymer Anti-Rabbit (DAKO); rabbit anti-MET (C-12) #sc-10 (Santa Cruz Biotechnology, Dallas, Texan, USA 1:50 overnight), thermostatic bath in Citrate, Novolink™ Polymer Detection System (Novocastra™); rabbit anti-CD127 (IL7R, EPR2955(2)) (Abcam, Cambridge, UK 1:100 overnight), thermostatic bath in EDTA, Envision+System-HRP Labelled Polymer Anti-Rabbit (DAKO).

Transcriptional subgroup affiliation of human GBM samples and GSC-derived xenografts was determined according to the protocol published by our collaborators in Brescia (Orzan *et al*, 2020). Briefly, expression levels of EGFR, PDGFRA, OLIG2, ASCL1, pNDRG1, YKL40, MET, and TP53 were assigned a score to calculate respectively the CL, PN, and MES percentual component of the sample. A clear subgroup affiliation was defined if the prevalent component would exceed the second-prevalent by > 20%, otherwise it would be defined as a mixture of the two highest components over the third, or even mixed if all the three subgroups were equally represented as ~33-33-33%.

7.18 Determination of GSC-derived subtype signatures

To identify putative GSC-derived specific subtype signatures, DEG lists obtained from different supervised and unsupervised clusters comparisons performed by the San Raffaele Scientific Institute -Omics Facility were individually screened. Only significantly up- and downregulated genes according to the SeqC filter were retained for each comparison. Appropriate significance (adjusted p-value) and log2FC thresholds were selected for each DEG list so to provide a final geneset of at least 40 either up- or down-regulated genes that were considered characterizing for the GSCs pertaining to each side of the comparisons.

Analyses were executed with R (v4.1.2), RStudio (Ghost Orchid release v2021.09.01) and appropriate R packages (see Appendix – R Scripts: Getting_signature.R).

Similarity of the various MES and PN signatures was investigated by visualizing Venn diagrams for intersections between them (<http://bioinformatics.psb.ugent.be/webtools/Venn/>).

7.19 Analysis of enriched pathways in GSCs clusters

DEGs provided by San Raffaele Scientific Institute -Omics Facility and those obtained from TCGA GBM subtypes and Healthy controls analyses were subdivided into significantly upregulated genes for each side of the comparison by means of R (v4.1.2), RStudio (Ghost Orchid release v2021.09.01) and appropriate R packages (see Appendix – R Scripts: Upregulated_genes.R).

The obtained gene lists were singularly uploaded into EnrichR platform (<https://maayanlab.cloud/Enrichr/>) to identify enriched transcriptional and molecular pathways, and ontology similarity in each GSC subset. Odds ratios (OR) and adjusted p-values for multiple tests were reported for the considered pathways and ontologies.

7.20 Metabolomic-transcriptomic integrated analyses

Determination of differentially enriched metabolites between MES and PN xenografts was performed with R (v4.1.2), RStudio (Ghost Orchid release v2021.09.01) and appropriate R packages.

PN and MES GSC-derived xenografts (n=4 GCLs/GSCs, n=3-4 biological replicates for each condition) were subjected to untargeted metabolomics (Metabolon, Morrisville, NC, USA). Samples were prepared using the automated MicroLab STAR® system from Hamilton Company. The resulting extract was analyzed by reverse phase (RP)/UPLC-MS/MS methods with either positive or negative ion mode electrospray ionization (ESI). Raw data were extracted, peak-identified and quality-control processed using Metabolon's hardware and software (Pieri *et al*, 2022). After mass-normalization and missing values imputation, raw metabolite counts were rescaled by multiplying by 1000 and rounded to nearest integer. Differential metabolite enrichment was carried out with `omu::omu_summary` function and non-parametric Mann Whitney U test (see Appendix – R Scripts: 3_Differential_metabolites_MannWhitney.R).

Significant metabolites were selected below a nominal p-value threshold of 0.05 and were divided in subsets enriched in PN or in MES xenografts.

Lists of significantly upregulated transcripts in each term of GSCs RNAseq comparisons or GSCs derived signatures (see Appendix – R Scripts: Upregulated_genes.R and Getting_signature.R) were then uploaded with consistent differentially enriched metabolites in xenografts on the MetaboAnalyst platform

(<https://www.metaboanalyst.ca/MetaboAnalyst/ModuleView.xhtml>) to perform joint metabolic pathway analysis.

Metabolic pathway nominal p-value, FDR-adjusted p-value and impact coefficient were reported for each query.

7.21 Transcriptional identification of drug sensitivity

Once obtained DEGs between pathological GBM subtypes and healthy controls from TCGA data, only significantly up- or downregulated genes (adjusted p-value < 0.05) were retained for each comparison (see Appendix – R Scripts: 5_Get_significant_sig_for_DrugRepo.R).

Gene lists were filtered so to keep only those present in the transcriptomic L1000 platform (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL20573>) (Subramanian, 2017), which was used to implement perturbagen screening in the Connectivity Map project (<https://clue.io>) (Lamb *et al*, 2006). After rearranging entries for decreasing significance, the top 150 up- and downregulated L1000 genes for each comparison were retained (see Appendix – R Scripts: 6b_Get_Genes_for_CMAP.R).

The Connectivity Map webtool (<https://clue.io>) was interrogated with the so-obtained lists to generate a set of compounds that would induce an opposite L1000 transcriptional signature. Enrichment Scores either < -90 or < -85 were considered significantly opposing to the pathological state.

7.22 Whole Exome Sequencing (WES) analysis of GSCs

Total DNA from GSCs was extracted using the DNeasy Blood & Tissue kit (Qiagen, Chatsworth, CA, USA) and sent to University of Padova for subsequent processing in collaboration with Dr. Alessandra Gasperini and Dr. Stefano Indraccolo, Basic and Translational Oncology Unit, Veneto Institute of Oncology IOV, Padova, Italy, in the context of GBM taskforce – Alliance Against Cancer.

WES analysis was performed with NextSeq. Sequences reads (FASTQ) were aligned to the reference human genome (hg19). Aligned BAM files were preprocessed with MarkDuplicated, IndelRealigner, BaseQualityScoreRecalibration to limit alignment errors. Following, variant calling was carried out with Mutect2 (GATK4.6), by comparing tumoral samples and a panel of normal individuals provided by another

institution and a germline variant set (provided by GATK website) to exclude healthy variants.

Variants were then filtered with FilterMutectCalls and annotated with ANNOVAR. Considered annotations were refGene (FASTA sequences for all annotated transcripts in RefSeq Gene), avsnp150 (dbSNP150 with allelic splitting and left-normalization), gnomad_genome (gnomAD allelic frequencies), genomicSuperDups (duplications of > 1000 Bases of Non-RepeatMasked Sequence, so >90 percent similar), clinvar_20170130 (CLINVAR database with Variant Clinical Significance such as unknown, untested, non-pathogenic, probable-non-pathogenic, probable-pathogenic, pathogenic, drug-response, histocompatibility, other, and Variant disease name), cosmic70 (COSMIC known variants), dbnsfp33a (pathogenic variant prediction), ICGC (International Cancer Genome Consortium annotation).

Subsequent analyses considered only exonic functional or splicing variants with VAF (variant allele frequency) > 5% and alternative allele depth \geq 5, therefore excluding frequent SNPs (found in dbSNP and gnomAD) and synonymous variants.

Variants were further filtered according to the following GBM-relevant gene lists: driver genes (Frattini *et al*, 2013), MMR genes (Wang *et al*, 2016), IRCCS genes (genes selected for Alliance Against Cancer – GBM gene panel design) and IRCCS SNPs (included in the abovementioned panel) (Dubbink *et al*, 2016).

7.23 In silico IL7R analysis

In silico scavenging analysis of IL7R was performed with two-gene correlation, view gene across samples and survival study tools on R2 platform (<https://hgserver1.amc.nl/cgi-bin/r2/main.cgi>) and survival study and transcript quantification tools in human tumors and cancer stem cells on the Human Protein Atlas (<https://www.proteinatlas.org>).

7.24 Quantitative real time PCR (qRT-PCR)

Total RNA was isolated from GSCs by means of the RNeasy Mini kit (Qiagen, Chatsworth, CA, USA); 1 μ g of total RNA was reverse-transcribed to cDNA with first strand synthesis kit Superscript III RNaseH-Reverse Transcriptase (Invitrogen, Carlsbad, CA) and OligodT primers. Each cDNA was diluted 1:3 and 1 μ l was used for each real-time PCR.

Quantitative RT-PCR was carried out with IQ SybrGreen (Biorad, Hercules, CA, USA) according to manufacturer's instructions. All cDNAs were normalized to β -actin levels. Human-specific primers for IL7R were obtained from Sigma (KiCqStart™ Primers). Real-time PCR runs were performed on the Stratagene MX3000P thermal cycler.

Δ Ct of the gene was calculated matched to its β -actin. Data were analyzed by the $\Delta\Delta$ Ct method.

7.25 Flow cytometry

Neurospheres were harvested and mechanically dissociated to obtain single cells, which were re-suspended in blocking solution (PBS supplemented with BSA 5 mg/ml and EDTA pH 8 2 mM) and put on ice for 20 minutes. The blocking solution was removed, and the PeCy7-anti-CD127 (IL7R) antibody (Invitrogen) was added after 1:50 dilution in the blocking solution. The antibody was incubated for 10 minutes at room temperature as the manufacturer's instructions suggested. The cells were washed with blocking solution and the acquisition was performed on BD FACS CANTO™ II instrument. Acquired data were analyzed with FCS Express 6.0 software.

7.26 Statistical analyses

Statistical analyses were conducted with R (v4.1.2), RStudio (Ghost Orchid release v2021.09.01) and appropriate R packages.

8 REFERENCES

- Aftab K, Aamir FB, Mallick S, Mubarak F, Pope WB, Mikkelsen T, Rock JP & Enam SA (2022) Radiomics for precision medicine in glioblastoma. *J Neurooncol* 156: 217–231
- Albert NL, Weller M, Suchorska B, Galldiks N, Soffietti R, Kim MM, La Fougère C, Pope W, Law I, Arbizu J, *et al* (2016) Response Assessment in Neuro-Oncology working group and European Association for Neuro-Oncology recommendations for the clinical use of PET imaging in gliomas. *Neuro Oncol* 18: 1199–1208
- Allen M, Bjerke M, Edlund H, Nelander S & Westermarck B (2016) Origin of the U87MG glioma cell line: Good news and bad news. *Sci Transl Med* 8
- Almeida ARM, Neto JL, Cachucho A, Euzébio M, Meng X, Kim R, Fernandes MB, Raposo B, Oliveira ML, Ribeiro D, *et al* (2021) Interleukin-7 receptor α mutational activation can initiate precursor B-cell acute lymphoblastic leukemia. *Nat Commun* 12: 1–16
- Anido J, Sáez-Borderías A, González-Juncà A, Rodón L, Folch G, Carmona MA, Prieto-Sánchez RM, Barba I, Martínez-Sáez E, Prudkin L, *et al* (2010) TGF- β Receptor Inhibitors Target the CD44^{high}/Id1^{high} Glioma-Initiating Cell Population in Human Glioblastoma. *Cancer Cell* 18: 655–668
- Armstrong TS & Gilbert MR (2012) Practical strategies for management of fatigue and Sleep Disorders in People With Brain. 65–72
- Armstrong TS, Shade MY, Breton G, Gilbert MR, Mahajan A, Scheurer ME, Vera E & Berger AM (2017) Sleep-wake disturbance in patients with brain tumors. *Neuro Oncol* 19: 323–335
- Arvanitis CD, Ferraro GB & Jain RK (2020) The blood–brain barrier and blood–tumour barrier in brain tumours and metastases. *Nat Rev Cancer* 20: 26–41
- De Bacco F, Casanova E, Medico E, Pellegatta S, Orzan F, Albano R, Luraghi P, Reato G, D’Ambrosio A, Porrati P, *et al* (2012) The MET oncogene is a functional marker of a glioblastoma stem cell subtype. *Cancer Res* 72: 4537–4550
- De Bacco F, D’Ambrosio A, Casanova E, Orzan F, Neggia R, Albano R, Verginelli F,

- Cominelli M, Poliani PL, Luraghi P, *et al* (2016) MET inhibition overcomes radiation resistance of glioblastoma stem-like cells. *EMBO Mol Med* 8: 550–568
- Bae S, Choi YS, Ahn SS, Chang JH, Kang SG, Kim EH, Kim SH & Lee SK (2018) Radiomic MRI phenotyping of glioblastoma: Improving survival prediction. *Radiology* 289: 797–806
- Bagley SJ, Desai AS, Linette GP, June CH & O'Rourke DM (2018) CAR T-cell therapy for glioblastoma: Recent clinical advances and future challenges. *Neuro Oncol* 20: 1429–1438
- Bao S, Wu Q, McLendon RE, Hao Y, Shi Q, Hjelmeland AB, Dewhirst MW, Bigner DD & Rich JN (2006a) Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature* 444: 756–760
- Bao S, Wu Q, Sathornsumetee S, Hao Y, Li Z, Hjelmeland AB, Shi Q, McLendon RE, Bigner DD & Rich JN (2006b) Stem cell-like glioma cells promote tumor angiogenesis through vascular endothelial growth factor. *Cancer Res* 66: 7843–7848
- Barata JT, Durum SK & Seddon B (2019) Flip the coin: IL-7 and IL-7R in health and disease. *Nat Immunol* 20: 1584–1593
- Bardella C, Al-Dalahmah O, Krell D, Brazauskas P, Al-Qahtani K, Tomkova M, Adam J, Serres S, Lockstone H, Freeman-Mills L, *et al* (2016) Expression of Idh1R132H in the Murine Subventricular Zone Stem Cell Niche Recapitulates Features of Early Gliomagenesis. *Cancer Cell* 30: 578–594
- Barthel FP, Wesseling P & Verhaak RGW (2018) Reconstructing the molecular life history of gliomas. *Acta Neuropathol* 135: 649–670
- Baskaran S, Mayrhofer M, Kultima HG, Bergström T, Elfineh L, Cavelier L, Isaksson A & Nelander S (2018) Primary glioblastoma cells for precision medicine: A quantitative portrait of genomic (in)stability during the first 30 passages. *Neuro Oncol* 20: 1080–1091
- Basser PJ, Mattiello J & LeBihan D (1994) MR diffusion tensor spectroscopy and imaging. *Biophys J* 66: 259–267
- Behnan J, Finocchiaro G & Hanna G (2019) The landscape of the mesenchymal signature in brain tumours. *Brain* 142: 847–866
- Beig N, Bera K & Tiwari P (2020) Introduction to radiomics and radiogenomics in neuro-oncology: Implications and challenges. *Neuro-Oncology Adv* 2: IV3–IV14
- Beig N, Patel J, Prasanna P, Hill V, Gupta A, Correa R, Bera K, Singh S, Partovi S, Varadan V, *et al* (2018) Radiogenomic analysis of hypoxia pathway is predictive of overall survival in Glioblastoma. *Sci Rep* 8: 1–11

- Ben-David U, Ha G, Tseng YY, Greenwald NF, Oh C, Shih J, McFarland JM, Wong B, Boehm JS, Beroukheim R, *et al* (2017) Patient-derived xenografts undergo mouse-specific tumor evolution. *Nat Genet* 49: 1567–1575
- Van Den Bent MJ, Gao Y, Kerkhof M, Kros JM, Gorlia T, Van Zwieten K, Prince J, Van Duinen S, Sillevs Smitt PA, Taphoorn M, *et al* (2015) Changes in the EGFR amplification and EGFRvIII expression between paired primary and recurrent glioblastomas. *Neuro Oncol* 17: 935–941
- Bhat KPL, Balasubramaniyan V, Vaillant B, Ezhilarasan R, Hummelink K, Hollingsworth F, Wani K, Heathcock L, James JD, Goodman LD, *et al* (2013) Mesenchymal Differentiation Mediated by NF- κ B Promotes Radiation Resistance in Glioblastoma. *Cancer Cell* 24: 331–346
- Bi J, Chowdhry S, Wu S, Zhang W, Masui K & Mischel PS (2020) Altered cellular metabolism in gliomas — an emerging landscape of actionable co-dependency targets. *Nat Rev Cancer* 20: 57–70
- Boccaccio C & Comoglio PM (2013) The MET oncogene in glioblastoma stem cells: Implications as a diagnostic marker and a therapeutic target. *Cancer Res* 73: 3193–3199
- Brandes AA, Bartolotti M, Tosoni A & Franceschi E (2016a) Nitrosoureas in the Management of Malignant Gliomas. *Curr Neurol Neurosci Rep* 16: 1–7
- Brandes AA, Finocchiaro G, Zagonel V, Reni M, Caserta C, Fabi A, Clavarezza M, Maiello E, Eoli M, Lombardi G, *et al* (2016b) AVAREG: A phase II, randomized, noncomparative study of fotemustine or bevacizumab for patients with recurrent glioblastoma. *Neuro Oncol* 18: 1304–1312
- Brat DJ, Aldape K, Colman H, Holland EC, Louis DN, Jenkins RB, Kleinschmidt-DeMasters BK, Perry A, Reifenberger G, Stupp R, *et al* (2018) cIMPACT-NOW update 3: recommended diagnostic criteria for “Diffuse astrocytic glioma, IDH-wildtype, with molecular features of glioblastoma, WHO grade IV”. *Acta Neuropathol* 136: 805–810
- Brennan C, Momota H, Hambarzumyan D, Ozawa T, Tandon A, Pedraza A & Holland E (2009) Glioblastoma subclasses can be defined by activity among signal transduction pathways and associated genomic alterations. *PLoS One* 4
- Brennan CW, Verhaak RGW, McKenna A, Campos B, Nounshmehr H, Salama SR, Zheng S, Chakravarty D, Sanborn JZ, Berman SH, *et al* (2013) The somatic genomic landscape of glioblastoma. *Cell* 155: 462
- Brescia P, Ortensi B, Fornasari L, Levi D, Broggi G & Pelicci G (2013) CD133 is essential for glioblastoma stem cell maintenance. *Stem Cells* 31: 857–869
- Brescia P, Richichi C & Pelicci G (2012) Current strategies for identification of

- glioma stem cells: Adequate or unsatisfactory? *J Oncol* 2012
- Brigliadori G, Foca F, Dall'Agata M, Rengucci C, Melegari E, Cerasoli S, Amadori D, Calistri D & Faedi M (2016) Defining the cutoff value of MGMT gene promoter methylation and its predictive capacity in glioblastoma. *J Neurooncol* 128: 333–339
- Brodbelt A, Greenberg D, Winters T, Williams M, Vernon S & Collins VP (2015) Glioblastoma in England: 2007-2011. *Eur J Cancer* 51: 533–542
- Brown TJ, Brennan MC, Li M, Church EW, Brandmeir NJ, Rakszawski KL, Patel AS, Rizk EB, Suki D, Sawaya R, *et al* (2016) Association of the extent of resection with survival in glioblastoma a systematic review and meta-Analysis. *JAMA Oncol* 2: 1460–1469
- Butler M, Pongor L, Su YT, Xi L, Raffeld M, Quezado M, Trepel J, Aldape K, Pommier Y & Wu J (2020) MGMT Status as a Clinical Biomarker in Glioblastoma. *Trends in Cancer* 6: 380–391
- Caverzasi E, Papinutto N, Castellano A, Zhu AH, Scifo P, Riva M, Bello L, Falini A, Bharatha A & Henry RG (2016) Neurite Orientation Dispersion and Density Imaging Color Maps to Characterize Brain Diffusion in Neurologic Disorders. *J Neuroimaging* 26: 494–498
- Ceccarelli M, Barthel FP, Malta TM, Sabedot TS, Salama SR, Murray BA, Morozova O, Newton Y, Radenbaugh A, Pagnotta SM, *et al* (2016) Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell* 164: 550–563
- Chaddad A, Desrosiers C, Hassan L & Tanougast C (2016) A quantitative study of shape descriptors from glioblastoma multiforme phenotypes for predicting survival outcome. *Br J Radiol* 89
- Chaddad A, Kucharczyk MJ, Daniel P, Sabri S, Jean-Claude BJ, Niazi T & Abdulkarim B (2019) Radiomics in glioblastoma: Current status and challenges facing clinical implementation. *Front Oncol* 9: 1–9
- Chakravarty D, Pedraza AM, Cotari J, Liu AH, Punko D, Kokroo A, Huse JT, Altan-Bonnet G & Brennan CW (2017) EGFR and PDGFRA co-expression and heterodimerization in glioblastoma tumor sphere lines. *Sci Rep* 7: 1–10
- Chang SM, Parney IF, Huang W, Anderson FA, Asher AL, Bernstein M, Lillehei KO, Brem H, Berger MS & Laws ER (2005) Patterns of care for adults with newly diagnosed malignant glioma. *J Am Med Assoc* 293: 557–564
- Chen J, Li Y, Yu TS, McKay RM, Burns DK, Kernie SG & Parada LF (2012) A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature* 488: 522–526

- Cheng L, Huang Z, Zhou W, Wu Q, Donnola S, Liu JK, Fang X, Sloan AE, Mao Y, Lathia JD, *et al* (2013) Glioblastoma stem cells generate vascular pericytes to support vessel function and tumor growth. *Cell* 153: 139–152
- Chinot OL, Wick W, Mason W, Henriksson R, Saran F, Nishikawa R, Carpentier AF, Hoang-Xuan K, Kavan P, Cernea D, *et al* (2014) Bevacizumab plus Radiotherapy–Temozolomide for Newly Diagnosed Glioblastoma. *N Engl J Med* 370: 709–722
- Cho J, Pastorino S, Zeng Q, Xu X, Johnson W, Vandenberg S, Verhaak R, Cherniack AD, Watanabe H, Dutt A, *et al* (2011) Glioblastoma-derived epidermal growth factor receptor carboxyl-terminal deletion mutants are transforming and are sensitive to EGFR-directed therapies. *Cancer Res* 71: 7587–7596
- Clarke JL, Ennis MM, Yung WKA, Chang SM, Wen Y, Cloughesy TF, Deangelis LM, Robins HI, Lieberman FS, Fine H a, *et al* (2011) Is surgery at progression a prognostic marker for improved 6-month progression-free survival or overall survival for patients with recurrent glioblastoma? *J Neurooncol* 13: 1118–1124
- Cloughesy TF, Yoshimoto K, Nghiemphu P, Brown K, Dang J, Zhu S, Hsueh T, Chen Y, Wang W, Youngkin D, *et al* (2008) Antitumor activity of rapamycin in a phase I trial for patients with recurrent PTEN-deficient glioblastoma. *PLoS Med* 5: 0139–0151
- Cluceru J, Interian Y, Phillips JJ, Molinaro AM, Luks TL, Alcaide-leon P, Marram P, Nair D, Lafontaine M, Shai A, *et al* (2021) Improving The Noninvasive Classification of Glioma Genetic Subtype With Deep Learning and Diffusion-Weighted Imaging. *Front Oncol* 24: 1–14
- Colen R, Foster I, Gatenby R, Giger ME, Gillies R, Gutman D, Heller M, Jain R, Madabhushi A, Madhavan S, *et al* (2014) NCI workshop report: Clinical and computational requirements for correlating imaging phenotypes with genomics signatures. *Transl Oncol* 7: 556–569
- Collewet G, Strzelecki M & Mariette F (2004) Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magn Reson Imaging* 22: 81–91
- Cominelli M, Grisanti S, Mazzoleni S, Branca C, Buttolo L, Furlan D, Liserre B, Bonetti MF, Medicina D, Pellegrini V, *et al* (2015) EGFR amplified and overexpressing glioblastomas and association with better response to adjuvant metronomic temozolomide. *J Natl Cancer Inst* 107: 1–13
- Conroy S, Kruyt FAE, Joseph J V., Balasubramaniyan V, Bhat KP, Wagemakers M, Enting RH, Walenkamp AME & Den Dunnen WFA (2014) Subclassification of newly diagnosed glioblastomas through an immunohistochemical approach.

PLoS One 9: 1–21

- Conti Nibali M, Gay LG, Sciortino T, Rossi M, Caroli M, Bello L & Riva M (2021) Surgery for Glioblastoma in Elderly Patients. *Neurosurg Clin N Am* 32: 137–148
- Conti V, Cominelli M, Pieri V, Gallotti AL, Pagano I, Zanella M, Mazzoleni S, Pivetta F, Patanè M, Scotti GM, *et al* (2021) mTORC1 promotes malignant large cell/anaplastic histology and is a targetable vulnerability in SHH-TP53 mutant medulloblastoma. *JCI Insight* 6
- Cui L, Fu J, Chung-Sean Pang J, Qiu Z-K, Liu X-M, Chen F-R, Shi H-L, Ng H-K & Chen Z-P (2012) Overexpression of IL-7 enhances cisplatin resistance in glioma. *Cancer Biol Ther* 13(7): 496–503
- Cusulin C, Chesnelong C, Bose P, Bilenky M, Kopciuk K, Chan JA, Cairncross JG, Jones SJ, Marra MA, Luchman HA, *et al* (2015) Precursor States of Brain Tumor Initiating Cell Lines Are Predictive of Survival in Xenografts and Associated with Glioblastoma Subtypes. *Stem Cell Reports* 5: 1–9
- Czarnek N, Clark K, Peters KB & Mazurowski MA (2017) Algorithmic three-dimensional analysis of tumor shape in MRI improves prognosis of survival in glioblastoma: a multi-institutional study. *J Neurooncol* 132: 55–62
- Dardis C, Donner D, Sanai N, Xiu J, Mittal S, Michelhaugh SK, Pandey M, Kesari S, Heimberger AB, Gatalica Z, *et al* (2021) Correction to: Gliosarcoma vs. glioblastoma: a retrospective case series using molecular profiling (BMC Neurology, (2021), 21, 1, (231), 10.1186/s12883-021-02233-5). *BMC Neurol* 21: 1–13
- Darmanis S, Sloan SA, Croote D, Mignardi M, Chernikova S, Samghababi P, Zhang Y, Neff N, Kowarsky M, Caneda C, *et al* (2017) Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma. *Cell Rep* 21: 1399–1410
- Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, Gephart MGH, Barres BA & Quake SR (2015) A survey of human brain transcriptome diversity at the single cell level. *Proc Natl Acad Sci U S A* 112: 7285–7290
- Davis FG, Smith TR, Gittleman HR, Ostrom QT, Kruchko C & Barnholtz-Sloan JS (2020) Glioblastoma incidence rate trends in Canada and the United States compared with England, 1995-2015. *Neuro Oncol* 22: 301–302
- Decarvalho AC, Kim H, Poisson LM, Winn ME, Mueller C, Cherba D, Koeman J, Seth S, Protopopov A, Felicella M, *et al* (2018) Discordant inheritance of chromosomal and extrachromosomal DNA elements contributes to dynamic disease evolution in glioblastoma. *Nat Genet* 50: 708–717

- Depeursinge A, Foncubierta-Rodriguez A, Van De Ville D & Müller H (2014) Three-dimensional solid texture analysis in biomedical imaging: Review and opportunities. *Med Image Anal* 18: 176–196
- Diehn M, Nardini C, Wang DS, McGovern S, Jayaraman M, Liang Y, Aldape K, Cha S & Kuo MD (2008) Identification of noninvasive imaging surrogates for brain tumor gene-expression modules. *Proc Natl Acad Sci U S A* 105: 5213–5218
- Draaisma K, Chatzipli A, Taphoorn M, Kerkhof M, Weyerbrock A, Sanson M, Hoeben A, Lukacova S, Lombardi G, Leenstra S, *et al* (2020) Molecular evolution of IDH wild-type glioblastomas treated with standard of care affects survival and design of precision medicine trials: A report from the EORTC 1542 study. *J Clin Oncol* 38: 81–99
- Driggers PH, Ennist DL, Gleason SL, Mak WH, Marks MS, Levi BZ, Flanagan JR, Appella E & Ozato K (1990) An interferon γ -regulated protein that binds the interferon-inducible enhancer element of major histocompatibility complex class I genes. *Proc Natl Acad Sci U S A* 87: 3743–3747
- Duan S, Yuan G, Liu X, Ren R, Li J, Zhang W, Wu J, Xu X, Fu L, Li Y, *et al* (2015) PTEN deficiency reprogrammes human neural stem cells towards a glioblastoma stem cell-like phenotype. *Nat Commun* 6
- Dubbink HJ, Atmodimedjo PN, Kros JM, French PJ, Sanson M, Idbaih A, Wesseling P, Enting R, Spliet W, Tijssen C, *et al* (2016) Molecular classification of anaplastic oligodendroglioma using next-generation sequencing: A report of the prospective randomized EORTC Brain Tumor Group 26951 phase III trial. *Neuro Oncol* 18: 388–400
- Dunn J, Baborie A, Alam F, Joyce K, Moxham M, Sibson R, Crooks D, Husband D, Shenoy A, Brodbelt A, *et al* (2009) Extent of MGMT promoter methylation correlates with outcome in glioblastomas given temozolomide and radiotherapy. *Br J Cancer* 101: 124–131
- Ellingson BM, Chung C, Pope WB, Boxerman JL & Kaufmann TJ (2017) Pseudoprogression, radionecrosis, inflammation or true tumor progression? challenges associated with glioblastoma response assessment in an evolving therapeutic landscape. *J Neurooncol* 134: 495–504
- Engebraaten O, Hjortland GO, Hirschberg H & Fodstad Ø (1999) Growth of precultured human glioma specimens in nude rat brain. *J Neurosurg* 90: 125–132
- Esposito A, Campana L, Palmisano A, De Cobelli F, Canu T, Santarella F, Colantoni C, Monno A, Vezzoli M, Pezzetti G, *et al* (2013) Magnetic Resonance Imaging at 7T Reveals Common Events in Age-Related Sarcopenia and in the Homeostatic

- Response to Muscle Sterile Injury. *PLoS One* 8: 1–9
- Faletti S, Osti D, Ceccacci E, Richichi C, Costanza B, Nicosia L, Noberini R, Marotta G, Furia L, Faretta MR, *et al* (2021) LSD1-directed therapy affects glioblastoma tumorigenicity by deregulating the protective ATF4-dependent integrated stress response. *Sci Transl Med* 13: 1–17
- Fan QW, Cheng CK, Gustafson WC, Charron E, Zipper P, Wong RA, Chen J, Lau J, Knobbe-Thomsen C, Weller M, *et al* (2013) EGFR Phosphorylates Tumor-Derived EGFRvIII Driving STAT3/5 and Progression in Glioblastoma. *Cancer Cell* 24: 438–449
- Fathi Kazerooni A, Bakas S, Saligheh Rad H & Davatzikos C (2020) Imaging signatures of glioblastoma molecular characteristics: A radiogenomics review. *J Magn Reson Imaging* 52: 54–69
- Felsberg J, Thon N, Eigenbrod S, Hentschel B, Sabel MC, Westphal M, Schackert G, Kreth FW, Pietsch T, Löffler M, *et al* (2011) Promoter methylation and expression of MGMT and the DNA mismatch repair genes MLH1, MSH2, MSH6 and PMS2 in paired primary and recurrent glioblastomas. *Int J Cancer* 129: 659–670
- Figini M, Riva M, Graham M, Castelli GM, Fernandes B, Grimaldi M, Baselli G, Pessina F, Bello L, Zhang H, *et al* (2018) Prediction of isocitrate dehydrogenase genotype in brain gliomas with MRI: Single-shell versus multishell diffusion models. *Radiology* 289: 788–796
- Finocchiaro G & Pellegatta S (2014) Perspectives for immunotherapy in glioblastoma treatment. *Curr Opin Oncol* 26: 608–614
- Francis JM, Zhang CZ, Maire CL, Jung J, Manzo VE, Adalsteinsson VA, Homer H, Haidar S, Blumenstiel B, Pedomallu CS, *et al* (2014) EGFR variant heterogeneity in glioblastoma resolved through single-nucleus sequencing. *Cancer Discov* 4: 956–971
- Frattini V, Pagnotta SM, Tala, Fan JJ, Russo M V., Lee SB, Garofano L, Zhang J, Shi P, Lewis G, *et al* (2018) A metabolic function of FGFR3-TACC3 gene fusions in cancer. *Nature* 553: 222–227
- Frattini V, Trifonov V, Chan JM, Castano A, Lia M, Abate F, Keir ST, Ji AX, Zoppoli P, Niola F, *et al* (2013) The integrated landscape of driver genomic alterations in glioblastoma. *Nat Genet* 45: 1141–1149
- Friedman HS, Prados MD, Wen PY, Mikkelsen T, Schiff D, Abrey LE, Yung WKA, Paleologos N, Nicholas MK, Jensen R, *et al* (2009) Bevacizumab alone and in combination with irinotecan in recurrent glioblastoma. *J Clin Oncol* 27: 4733–4740

- Gabrusiewicz K, Rodriguez B, Wei J, Hashimoto Y, Healy LM, Maiti SN, Thomas G, Zhou S, Wang Q, Elakkad A, *et al* (2016) Glioblastoma-infiltrated innate immune cells resemble M0 macrophage phenotype. *JCI Insight* 1: 0–19
- Gagliardi F, Narayanan A, Gallotti AL, Pieri V, Mazzoleni S, Cominelli M, Rezzola S, Corsini M, Brugnara G, Altabella L, *et al* (2020) Enhanced SPARCL1 expression in cancer stem cells improves preclinical modeling of glioblastoma by promoting both tumor infiltration and angiogenesis. *Neurobiol Dis* 134
- Galldiks N, Zadeh G & Lohmann P (2020) Artificial Intelligence, Radiomics, and Deep Learning in Neuro-Oncology. *Neuro-Oncology Adv* 2: IV1–IV2
- Galli R, Binda E, Orfanelli U, Cipelletti B, Gritti A, De Vitis S, Fiocco R, Foroni C, Dimeco F & Vescovi A (2004) Erratum: Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma (Cancer Research (October 2004) 64 (7011–7021)). *Cancer Res* 64: 8130
- Gangoso E, Southgate B, Bradley L, Rus S, Galvez-Cancino F, McGivern N, Güç E, Kapourani CA, Byron A, Ferguson KM, *et al* (2021) Glioblastomas acquire myeloid-affiliated transcriptional programs via epigenetic immunoediting to elicit immune evasion. *Cell* 184: 2454–2470.e26
- Garnett J, Chumbalkar V, Vaillant B, Gururaj AE, Hill KS, Latha K, Yao J, Priebe W, Colman H, Elferink LA, *et al* (2013) Regulation of HGF expression by δ EGFR-mediated c-Met activation in glioblastoma cells. *Neoplasia (United States)* 15: 73–84
- Garofano L, Migliozzi S, Oh YT, D'Angelo F, Najac RD, Ko A, Frangaj B, Caruso FP, Yu K, Yuan J, *et al* (2021) Pathway-based classification of glioblastoma uncovers a mitochondrial subtype with therapeutic vulnerabilities. *Nat Cancer* 2: 141–156
- Gehring K, Taphoorn MJB, Sitskoorn MM & Aaronson NK (2015) Predictors of subjective versus objective cognitive functioning in patients with stable grades II and III glioma. *Neuro-Oncology Pract* 2: 20–31
- Gerritsen JKW & Vincent AJPE (2020) Response to letter to the editor: "Impact of intraoperative stimulation mapping on high-grade glioma surgery outcome: a meta-analysis". *Acta Neurochir (Wien)* 162: 429–431
- Gevaert O, Mitchell LA, Achrol AS, Xu J, Echegaray S, Steinberg GK, Cheshier SH, Napel S, Zaharchuk G & Plevritis SK (2014) Glioblastoma multiforme: Exploratory radiogenomic analysis by using quantitative image features. *Radiology* 273: 168–174
- Gharzeddine K, Hatzoglou V, Holodny AI & Young RJ (2019) MR Perfusion and MR Spectroscopy of Brain Neoplasms. *Radiol Clin North Am* 57: 1177–1188

- Gilbert MR, Dignam JJ, Armstrong TS, Wefel JS, Blumenthal DT, Vogelbaum MA, Colman H, Chakravarti A, Pugh S, Won M, *et al* (2014) A Randomized Trial of Bevacizumab for Newly Diagnosed Glioblastoma. *N Engl J Med* 370: 699–708
- Gillies RJ, Kinahan PE & Hricak H (2016) Radiomics: Images are more than pictures, they are data. *Radiology* 278: 563–577
- Gimple RC, Bhargava S, Dixit D & Rich JN (2019) Glioblastoma stem cells: Lessons from the tumor hierarchy in a lethal cancer. *Genes Dev* 33: 591–609
- Gómez-Oliva R, Domínguez-García S, Carrascal L, Abalos-Martínez J, Pardillo-Díaz R, Verástegui C, Castro C, Nunez-Abades P & Geribaldi-Doldán N (2021) Evolution of Experimental Models in the Study of Glioblastoma: Toward Finding Efficient Treatments. *Front Oncol* 10: 1–16
- Grabowski MM, Recinos PF, Nowacki AS, Schroeder JL, Angelov L, Barnett GH & Vogelbaum MA (2014) Residual tumor volume versus extent of resection: Predictors of survival after surgery for glioblastoma. *J Neurosurg* 121: 1115–1123
- Van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin JC, Pieper S & Aerts HJWL (2017) Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 77: e104–e107
- Gritti A, Parati E, Cova L, Frolichsthal P, Galli R, Wanke E, Faravelli L, Morassutti D, Roisen F, Nickel D, *et al* (1996) Multipotential stem cells from the adult mouse brain proliferate and self-renew in response to basic fibroblast growth factor. *J Neurosci* 16
- Gutman DA, Cooper LAD, Hwang SN, Holder CA, Gao JJ, Aurora TD, Dunn WD, Scarpace L, Mikkelsen T, Jain R, *et al* (2013) MR imaging predictors of molecular profile and survival: Multi-institutional study of the TCGA glioblastoma data set. *Radiology* 267: 560–569
- Hambardzumyan D, Gutmann DH & Kettenmann H (2015) The role of microglia and macrophages in glioma maintenance and progression. *Nat Neurosci* 19: 20–27
- Hara T, Chanoch-Myers R, Mathewson ND, Myskiw C, Atta L, Bussema L, Eichhorn SW, Greenwald AC, Kinker GS, Rodman C, *et al* (2021) Interactions between cancer cells and immune cells drive transitions to mesenchymal-like states in glioblastoma. *Cancer Cell* 39: 779–792.e11
- Hayashida Y, Hirai T, Morishita S, Kitajima M, Murakami R, Korogi Y, Makino K, Nakamura H, Ikushima I, Yamura M, *et al* (2006) Diffusion-weighted imaging of metastatic brain tumors: Comparison with histologic type and tumor cellularity. *Am J Neuroradiol* 27: 1419–1425

- Heffron TP (2018) Challenges of developing small-molecule kinase inhibitors for brain tumors and the need for emphasis on free drug levels. *Neuro Oncol* 20: 307–312
- Hegi ME, Diserens A-C, Gorlia T, Hamou M-F, de Tribolet N, Weller M, Kros JM, Hainfellner JA, Mason W, Mariani L, *et al* (2005) MGMT Gene Silencing and Benefit from Temozolomide in Glioblastoma. *N Engl J Med* 352: 997–1003
- Hegi ME, Genbrugge E, Gorlia T, Stupp R, Gilbert MR, Chinot OL, Burt Nabors L, Jones G, Van Criekinge W, Straub J, *et al* (2019) MGMT promoter methylation cutoff with safety margin for selecting glioblastoma patients into trials omitting temozolomide: A pooled analysis of four clinical trials. *Clin Cancer Res* 25: 1809–1816
- Hemmati HD, Nakano I, Lazareff JA, Masterman-Smith M, Geschwind DH, Bronner-Fraser M & Kornblum HI (2003) Cancerous stem cells can arise from pediatric brain tumors. *Proc Natl Acad Sci U S A* 100: 15178–15183
- Henker C, Kriesen T, Glass Ä, Schneider B & Piek J (2017) Volumetric quantification of glioblastoma: experiences with different measurement techniques and impact on survival. *J Neurooncol* 135: 391–402
- Henriksen OM, del Mar Álvarez-Torres M, Figueiredo P, Hangel G, Keil VC, Nechifor RE, Riemer F, Schmainda KM, Warnert EAH, Wiegers EC, *et al* (2022) High-Grade Glioma Treatment Response Monitoring Biomarkers: A Position Statement on the Evidence Supporting the Use of Advanced MRI Techniques in the Clinic, and the Latest Bench-to-Bedside Developments. Part 1: Perfusion and Diffusion Techniques. *Front Oncol* 12: 1–27
- Hidalgo M, Amant F, Biankin A V., Budinská E, Byrne AT, Caldas C, Clarke RB, de Jong S, Jonkers J, Mælandsmo GM, *et al* (2014) Patient-derived Xenograft models: An emerging platform for translational cancer research. *Cancer Discov* 4: 998–1013 doi:10.1158/2159-8290.CD-14-0001 [PREPRINT]
- Higano S, Yun X, Kumabe T, Watanabe M, Mugikura S, Umetsu A, Sato A, Yamada T & Takahashi S (2006) Malignant astrocytic tumors: clinical importance of apparent diffusion coefficient in prediction of grade and prognosis. *Radiology*: 839–46
- Huang J, Zheng M, Zhang Z, Tang X, Chen Y, Peng A, Peng X, Tong A & Zhou L (2021) Interleukin-7-loaded oncolytic adenovirus improves CAR-T cell therapy for glioblastoma. *Cancer Immunol Immunother* 70: 2453–2465
- Hubert CG, Rivera M, Spangler LC, Wu Q, Mack SC, Prager BC, Couce M, McLendon RE, Sloan AE & Rich JN (2016) A three-dimensional organoid culture system derived from human glioblastomas recapitulates the hypoxic gradients and

- cancer stem cell heterogeneity of tumors found in vivo. *Cancer Res* 76: 2465–2477
- Hunter C, Smith R, Cahill D, Stephens P, Stevens C & et al (2006) A Hypermutation Phenotype and Somatic MSH6 Mutations in Recurrent Human Malignant Gliomas after Alkylator Chemotherapy. *Cancer Res* 66: 3987–3991
- Huse JT, Phillips HS & Brennan CW (2011) Molecular subclassification of diffuse gliomas: Seeing order in the chaos. *Glia* 59: 1190–1199
- Incekara F, Smits M, van der Voort SR, Dubbink HJ, Atmodimedjo PN, Kros JM, Vincent AJPE & van den Bent M (2020) The Association Between the Extent of Glioblastoma Resection and Survival in Light of MGMT Promoter Methylation in 326 Patients With Newly Diagnosed IDH-Wildtype Glioblastoma. *Front Oncol* 10: 1–8
- Indraccolo S, Lombardi G, Fassan M, Pasqualini L, Giunco S, Marcato R, Gasparini A, Candiotto C, Nalio S, Fiduccia P, et al (2019) Genetic, epigenetic, and immunologic profiling of MMR-deficient relapsed glioblastoma. *Clin Cancer Res* 25: 1828–1837
- Ismail M, Hill V, Statsevych V, Huang R, Prasanna P, Correa R, Singh G, Bera K, Beig N, Thawani R, et al (2018) Shape features of the lesion habitat to differentiate brain tumor progression from pseudoprogression on routine multiparametric MRI: A multisite study. *Am J Neuroradiol* 39: 2187–2193
- Jackson CM, Choi J & Lim M (2019) Mechanisms of immunotherapy resistance: lessons from glioblastoma. *Nat Immunol* 20: 1100–1109
- Jamshidi N, Diehn M, Bredel M & Kuo MD (2014) Illuminating radiogenomic characteristics of glioblastoma multiforme through integration of MR imaging, messenger RNA expression, and DNA copy number variation. *Radiology* 270: 1–2
- Jian M, Yunjia Z, Zhiying D, Yanduo J & Guocheng J (2019) Interleukin 7 receptor activates PI3K/Akt/mTOR signaling pathway via downregulation of Beclin-1 in lung cancer. *Mol Carcinog* 58: 358–365
- Johnson B, Mazor T, Hong C, Barnes M, Aihara K & et al (2014) Mutational Analysis Reveals the Origin and Therapy-Driven Evolution of Recurrent Glioma. *Science* (80-) 343: 189–194
- Johnson RM, Phillips HS, Bais C, Brennan CW, Cloughesy TF, Daemen A, Herrlinger U, Jenkins RB, Lai A, Mancao C, et al (2020) Development of a gene expression-based prognostic signature for IDH wild-type glioblastoma. *Neuro Oncol* 22: 1742–1756
- Kadota Y, Hirai T, Azuma M, Hattori Y, Khant ZA, Hori M, Saito K, Yokogami K &

- Takeshima H (2020) Differentiation between glioblastoma and solitary brain metastasis using neurite orientation dispersion and density imaging. *J Neuroradiol* 47: 197–202
- Karschnia P, Vogelbaum MA, van den Bent M, Cahill DP, Bello L, Narita Y, Berger MS, Weller M & Tonn JC (2021) Evidence-based recommendations on categories for extent of resection in diffuse glioma. *Eur J Cancer* 149: 23–33
- Keenan TE, Burke KP & Van Allen EM (2019) Genomic correlates of response to immune checkpoint blockade. *Nat Med* 25: 389–402
- Kickingereder P, Burth S, Wick A, Götz M, Eidel O, Schlemmer HP, Maier-Hein KH, Wick W, Bendszus M, Radbruch A, *et al* (2016) Radiomic profiling of glioblastoma: Identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models. *Radiology* 280: 880–889
- Kickingereder P, Isensee F, Tursunova I, Petersen J, Neuberger U, Bonekamp D, Brugnara G, Schell M, Kessler T, Foltyn M, *et al* (2019) Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study. *Lancet Oncol* 20: 728–740
- Kickingereder P, Wiestler B, Sahm F, Heiland S, Roethke M, Schlemmer H, Wick W, Bendszus M & Radbruch A (2014) Primary central nervous system lymphoma and atypical glioblastoma: multiparametric differentiation by using diffusion-, perfusion-, and susceptibility-weighted MR imaging. *Radiology*: 843–50
- Kim H & Verhaak RGW (2015) Transcriptional mimicry by tumor-associated stroma. *Nat Genet* 47: 307–309
- Kim H, Zheng S, Amini SS, Virk SM, Mikkelsen T, Brat DJ, Grimsby J, Sougnez C, Muller F, Hu J, *et al* (2015) Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. *Genome Res* 25: 316–327
- Kim MJ, Choi SK, Hong SH, Eun JW, Nam SW, Han JW & You JS (2018) Oncogenic IL7R is downregulated by histone deacetylase inhibitor in esophageal squamous cell carcinoma via modulation of acetylated FOXO1. *Int J Oncol* 53: 395–403
- Kim S jae, Chang HJ, Volin M V., Umar S, Van Raemdonck K, Chevalier A, Palasiewicz K, Christman JW, Volkov S, Arami S, *et al* (2020) Macrophages are the primary effector cells in IL-7-induced arthritis. *Cell Mol Immunol* 17: 728–740
- Kim Y, Varn FS, Park SH, Yoon BW, Park HR, Lee C, Verhaak RGW & Paek SH (2021) Perspective of mesenchymal transformation in glioblastoma. *Acta*

Neuropathol Commun 9: 1–20

- Kirby S & Purdy RA (2014) Headaches and brain tumors. *Neurol Clin* 32: 423–432
- Körber V, Yang J, Barah P, Wu Y, Stichel D, Gu Z, Fletcher MNC, Jones D, Hentschel B, Lamszus K, *et al* (2019) Evolutionary Trajectories of IDHWT Glioblastomas Reveal a Common Path of Early Tumorigenesis Instigated Years ahead of Initial Diagnosis. *Cancer Cell* 35: 692–704.e12
- Kreso A & Dick JE (2014) Evolution of the cancer stem cell model. *Cell Stem Cell* 14: 275–291
- Kreth FW, Thon N, Simon M, Westphal M, Schackert G, Nikkhah G, Hentschel B, Reifenberger G, Pietsch T, Weller M, *et al* (2013) Gross total but not incomplete resection of glioblastoma prolongs survival in the era of radiochemotherapy. *Ann Oncol* 24: 3117–3123
- Labreche K, Kinnersley B, Berzero G, Di Stefano AL, Rahimian A, Detrait I, Marie Y, Grenier-Boley B, Hoang-Xuan K, Delattre JY, *et al* (2018) Diffuse gliomas classified by 1p/19q co-deletion, TERT promoter and IDH mutation status are associated with specific genetic risk loci. *Acta Neuropathol* 135: 743–755
- Lai L, Zeff RA & Goldschneider I (2006) A recombinant single-chain IL-7/HGFβ hybrid cytokine induces juxtacrine interactions of the IL-7 and HGF (c-Met) receptors and stimulates the proliferation of CFU-S12, CLPs, and pre-pro-B cells. *Blood* 107: 1776–1784
- Laks DR, Crisman TJ, Shih MYS, Mottahedeh J, Gao F, Sperry J, Garrett MC, Yong WH, Cloughesy TF, Liao LM, *et al* (2016) Large-scale assessment of the gliomasphere model system. *Neuro Oncol* 18: 1367–1378
- Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet J, Subramanian A, Ross KN, *et al* (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. 313: 1929–1935
- Lambin P, Leijenaar RTH, Deist TM, Peerlings J, De Jong EEC, Van Timmeren J, Sanduleanu S, Larue RTHM, Even AJG, Jochems A, *et al* (2017) Radiomics: The bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 14: 749–762
- Lamborn KR, Yung WKA, Chang SM, Wen PY, Cloughesy TF, DeAngelis LM, Robins HI, Lieberman FS, Fine HA, Fink KL, *et al* (2008) Progression-free survival: An important end point in evaluating therapy for recurrent high-grade gliomas. *Neuro Oncol* 10: 162–170
- Lan X, Jörg DJ, Cavalli FMG, Richards LM, Nguyen L V., Vanner RJ, Guilhamon P, Lee L, Kushida MM, Pellacani D, *et al* (2017) Fate mapping of human

glioblastoma reveals an invariant stem cell hierarchy

- Larroza A, Bodì V & Moratal D (2016) Texture Analysis in Magnetic Resonance Imaging: Review and Considerations for Future Applications. In *Assessment of Cellular and Organ Function and Dysfunction using Direct and Derived MRI Methodologies* IntechOpen
- Lathia J, Mack S, Mulkearns-Hubert E, Valentim C & Rich J (2015) Cancer stem cells in glioblastoma. *Genes Dev* 3: 1203–1217
- Lathia JD, Gallagher J, Heddleston JM, Wang J, Eyler CE, MacSwords J, Wu Q, Vasanji A, McLendon RE, Hjelmeland AB, *et al* (2010) Integrin Alpha 6 regulates glioblastoma stem cells. *Cell Stem Cell* 6: 421–432
- Law M, Yang S, Babb JS, Knopp EA, Golfinos JG, Zagzag D & Johnson G (2004) Comparison of cerebral blood volume and vascular permeability from dynamic susceptibility contrast-enhanced perfusion MR imaging with glioma grade. *Am J Neuroradiol* 25: 746–755
- Lawler SE, Speranza MC, Cho CF & Chiocca EA (2017) Oncolytic viruses in cancer treatment a review. *JAMA Oncol* 3: 841–849
- Le NQK, Hung TNK, Do DT, Lam LHT, Dang LH & Huynh TT (2021) Radiomics-based machine learning model for efficiently classifying transcriptome subtypes in glioblastoma patients from MRI. *Comput Biol Med* 132: 104320
- Ledur PF, Onzi GR, Zong H & Lenz G (2017) Culture conditions defining glioblastoma cells behavior: what is the impact for novel discoveries? *Oncotarget* 8: 69185–69197
- Lee B, Park JE, Bjørnerud A, Kim JH, Lee JY & Kim HS (2018a) Clinical Value of Vascular Permeability Estimates Using Dynamic Susceptibility Contrast MRI: Improved Diagnostic Performance in Distinguishing Hypervascular Primary CNS Lymphoma from Glioblastoma. *Am J Neuroradiol*
- Lee H, Kang S & Kim W (2016) Drug repositioning for cancer therapy based on large-scale drug-induced transcriptional signatures. *PLoS One* 11: 1–17
- Lee J, Kotliarova S, Kotliarov Y, Li A, Su Q, Donin NM, Pastorino S, Purow BW, Christopher N, Zhang W, *et al* (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
- Lee JH, Lee JE, Kahng JY, Kim SH, Park JS, Yoon SJ, Um JY, Kim WK, Lee JK, Park J, *et al* (2018b) Human glioblastoma arises from subventricular zone cells with low-level driver mutations. *Nature* 560: 243–247
- Leece R, Xu J, Ostrom QT, Chen Y, Kruchko C & Barnholtz-Sloan JS (2017) Global

- incidence of malignant brain and other central nervous system tumors by histology, 2003-2007. *Neuro Oncol* 19: 1553-1564
- Li J, Ek F, Olsson R, Belting M & Bengzon J (2022) Glioblastoma CD105+ cells define a SOX2- cancer stem cell-like subpopulation in the pre-invasive niche. *Acta Neuropathol Commun* 10: 1-17
- Li YM, Suki D, Hess K & Sawaya R (2016) The influence of maximum safe resection of glioblastoma on survival in 1229 patients: Can we do better than gross-total resection? *J Neurosurg* 124: 977-988
- Ligon KL, Huillard E, Mehta S, Kesari S, Liu H, Alberta JA, Bachoo RM, Kane M, Louis DN, DePinho RA, *et al* (2007) Olig2-Regulated Lineage-Restricted Pathway Controls Replication Competence in Neural Stem Cells and Malignant Glioma. *Neuron* 53: 503-517
- Lim M, Xia Y, Bettgowda C & Weller M (2018) Current state of immunotherapy for glioblastoma. *Nat Rev Clin Oncol* 15: 422-442
- Lin N, Yan W, Gao K, Wang Y, Zhang J & You Y (2014) Prevalence and clinicopathologic characteristics of the molecular subtypes in malignant glioma: A multi-institutional analysis of 941 cases. *PLoS One* 9: 1-5
- Liu D, Chen J, Ge H, Yan Z, Luo B, Hu X, Yang K, Liu Y, Liu H & Zhang W (2022) Radiogenomics to characterize the immune-related prognostic signature associated with biological functions in glioblastoma. *Eur Radiol*
- Liu D, Chen J, Hu X, Yang K, Liu Y, Hu G, Ge H, Zhang W & Liu H (2021) Imaging-Genomics in Glioblastoma: Combining Molecular and Imaging Signatures. *Front Oncol* 11: 1-13
- Liu G, Yuan X, Zeng Z, Tunici P, Ng H, Abdulkadir IR, Lu L, Irvin D, Black KL & Yu JS (2006) Analysis of gene expression and chemoresistance of CD133+ cancer stem cells in glioblastoma. *Mol Cancer* 5: 1-12
- Lohmann P, Galldiks N, Kocher M, Heinzl A, Filss CP, Stegmayr C, Mottaghy FM, Fink GR, Jon Shah N & Langen KJ (2021) Radiomics in neuro-oncology: Basics, workflow, and applications. *Methods* 188: 112-121
- Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P & Ellison DW (2016) The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol* 131: 803-820
- Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, Hawkins C, Ng HK, Pfister SM, Reifenberger G, *et al* (2021) The 2021 WHO classification of tumors of the central nervous system: A summary. *Neuro Oncol* 23: 1231-1251

- Lu-Emerson C, Snuderl M, Kirkpatrick N, Goveia J, Davidson C, Huang Y, Riedemann L, Taylor J, Ivy P, Duda D, *et al* (2013) Increase in tumor-associated macrophages after antiangiogenic therapy is associated with poor survival among patients with recurrent glioblastoma. *Neuro Oncol* 15: 1079–1087
- Lu X, Xu W, Wei Y, Li T, Gao L, Fu X, Yao Y & Wang L (2019) Diagnostic performance of DWI for differentiating primary central nervous system lymphoma from glioblastoma: a systematic review and meta-analysis. *Neurol Sci* 40: 947–956
- Lv X, Li Q, Liu H, Gong M, Zhao Y, Hu J, Wu F & Wu X (2022) JUN activation modulates chromatin accessibility to drive TNF α -induced mesenchymal transition in glioblastoma. *J Cell Mol Med* 26: 4602–4612
- Mack SC, Singh I, Wang X, Hirsch R, Wu Q, Villagomez R, Bernatchez JA, Zhu Z, Gimple RC, Kim LJ, *et al* (2019) Chromatin landscapes reveal developmentally encoded transcriptional states that define human glioblastoma. *J Exp Med* 216: 1071–1090
- Mansouri A, Hachem LD, Mansouri S, Nassiri F, Laperriere NJ, Xia D, Lindeman NI, Wen PY, Chakravarti A, Mehta MP, *et al* (2019) MGMT promoter methylation status testing to guide therapy for glioblastoma: Refining the approach based on emerging evidence and current challenges. *Neuro Oncol* 21: 167–178
- Mao J, Zeng W, Zhang Q, Yang Z, Yan X, Zhang H, Wang M, Yang G, Zhou M & Shen J (2020) Differentiation between high-grade gliomas and solitary brain metastases: a comparison of five diffusion-weighted MRI models. *BMC Med Imaging* 20: 1–11
- Marko NF, Weil RJ, Schroeder JL, Lang FF, Suki D & Sawaya RE (2014) Extent of resection of glioblastoma revisited: Personalized survival modeling facilitates more accurate survival prediction and supports a maximum-safe-resection approach to surgery. *J Clin Oncol* 32: 774–782
- Di Mascolo D, Palange AL, Primavera R, Macchi F, Catelani T, Piccardi F, Spanò R, Ferreira M, Marotta R, Armirotti A, *et al* (2021) Conformable hierarchically engineered polymeric micromeshes enabling combinatorial therapies in brain tumours. *Nat Nanotechnol* 16: 820–829
- Masjoodi S, Hashemi H, Oghabian M & Sharifi G (2018) Differentiation of Edematous, Tumoral and Normal Areas of Brain Using Diffusion Tensor and Neurite Orientation Dispersion and Density Imaging. *J Biomed Phys Eng* 8
- Mazzoleni S, Politi LS, Pala M, Cominelli M, Franzin A, Sergi LS, Falini A, De Palma M, Bulfone A, Poliani PL, *et al* (2010) Epidermal growth factor receptor

- expression identifies functionally and molecularly distinct tumor-initiating cells in human glioblastoma multiforme and is required for gliomagenesis. *Cancer Res* 70: 7500–7513
- McFaline-Figueroa JL, Braun CJ, Stanciu M, Nagel ZD, Mazzucato P, Sangaraju D, Cerniauskas E, Barford K, Vargas A, Chen Y, *et al* (2015) Minor changes in expression of the mismatch repair protein MSH2 exert a major impact on glioblastoma response to temozolomide. *Cancer Res* 75: 3127–3138 doi:10.1158/0008-5472.CAN-14-3616 [PREPRINT]
- McKinnon C, Nandhabalan M, Murray SA & Plaha P (2021) Glioblastoma: Clinical presentation, diagnosis, and management. *BMJ* 374
- McLendon R, Friedman A, Bigner D, Van Meir EG, Brat DJ, Mastrogiannakis GM, Olson JJ, Mikkelsen T, Lehman N, Aldape K, *et al* (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455: 1061–1068
- Melin BS, Barnholtz-Sloan JS, Wrensch MR, Johansen C, Il'yasova D, Kinnersley B, Ostrom QT, Labreche K, Chen Y, Armstrong G, *et al* (2017) Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors. *Nat Genet* 49: 789–794
- Moassefi M, Faghani S, Conte GM, Kowalchuk RO, Vahdati S, Crompton DJ, Perez-Vega C, Cabreja RAD, Vora SA, Quiñones-Hinojosa A, *et al* (2022) A deep learning model for discriminating true progression from pseudoprogression in glioblastoma patients. *J Neurooncol*
- Mohme M, Maire CL, Geumann U, Schliffke S, Duhrsen L, Fita K, Akyuz N, Binder M, Westphal M, Guenther C, *et al* (2020) Local Intracerebral Immunomodulation Using Interleukin-Expressing Mesenchymal Stem Cells in Glioblastoma. *Clin Cancer Res* 26: 2626–2639
- Molinaro AM, Hervey-Jumper S, Morshed RA, Young J, Han SJ, Chunduru P, Zhang Y, Phillips JJ, Shai A, Lafontaine M, *et al* (2020) Association of Maximal Extent of Resection of Contrast-Enhanced and Non-Contrast-Enhanced Tumor with Survival Within Molecular Subgroups of Patients with Newly Diagnosed Glioblastoma. *JAMA Oncol* 6: 495–503
- Moon WJ, Choi JW, Roh HG, Lim SD & Koh YC (2012) Imaging parameters of high grade gliomas in relation to the MGMT promoter methylation status: The CT, diffusion tensor imaging, and perfusion MR imaging. *Neuroradiology* 54: 555–563
- Moors M, Vudattu NK, Abel J, Krämer U, Rane L, Ulfing N, Ceccatelli S, Seyfert-

- Margolies V, Fritsche E & Maeurer MJ (2010) Interleukin-7 (IL-7) and IL-7 splice variants affect differentiation of human neural progenitor cells. *Genes Immun* 11: 11–20
- Murat A, Migliavacca E, Gorlia T, Lambiv WL, Shay T, Hamou MF, De Tribolet N, Regli L, Wick W, Kouwenhoven MCM, *et al* (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
- Murphy ES, Xie H, Merchant TE, Yu JS, Chao ST & Suh JH (2015) Review of cranial radiotherapy-induced vasculopathy. *J Neurooncol* 122: 421–429
- Nakano I (2015) Stem cell signature in glioblastoma: Therapeutic development for a moving target. *J Neurosurg* 122: 324–330
- Narayanan A, Gagliardi F, Gallotti ALAL, Mazzoleni S, Cominelli M, Fagnocchi L, Pala M, Piras ISIS, Zordan P, Moretta N, *et al* (2019) The proneural gene ASCL1 governs the transcriptional subgroup affiliation in glioblastoma stem cells by directly repressing the mesenchymal gene NDRG1. *Cell Death Differ* 26: 1813–1831
- Neftel C, Laffy J, Filbin MG, Hara T, Shore ME, Rahme GJ, Richman AR, Silverbush D, Shaw ML, Hebert CM, *et al* (2019) An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* 178: 835–849.e21
- Niendorf T, Dijkhuizen RM, Norris DG, Van Lookeren Campagne M & Nicolay K (1996) Biexponential diffusion attenuation in various states of brain tissue: Implications for diffusion-weighted imaging. *Magn Reson Med* 36: 847–857
- Nishikawa R, Ji XD, Harmon RC, Lazar CS, Gill GN, Cavenee WK & Huang HJS (1994) A mutant epidermal growth factor receptor common in human glioma confers enhanced tumorigenicity. *Proc Natl Acad Sci U S A* 91: 7727–7731
- Noble M, Pröschel C & Mayer-Pröschel M (2004) Getting a GR(i)P on oligodendrocyte development. *Dev Biol* 265: 33–52
- Norden AD, Bartolomeo J, Tanaka S, Drappatz J, Ciampa AS, Doherty LM, Lafrankie DC, Ruland S, Quant EC, Beroukhir R, *et al* (2012) Safety of concurrent bevacizumab therapy and anticoagulation in glioma patients. *J Neurooncol* 106: 121–125
- Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, Pan F, Pelloski CE, Sulman EP, Bhat KP, *et al* (2010) Identification of a CpG Island Methylator Phenotype that Defines a Distinct Subgroup of Glioma. *Cancer Cell* 17: 510–522
- Nowakowski TJ, Bhaduri A, Pollen AA, Alvarado B, Mostajo-Radji MA, Di Lullo E,

- Haeussler M, Sandoval-Espinosa C, Liu SJ, Velmeshev D, *et al* (2017) Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* (80-) 358: 1318–1323
- Ogawa J, Pao GM, Shokhirev MN & Verma IM (2018) Glioblastoma Model Using Human Cerebral Organoids. *Cell Rep* 23: 1220–1229
- Olar A & Aldape KD (2014) Using the molecular classification of glioblastoma to inform personalized treatment. *J Pathol* 232: 165–177
- Ortensi B, Setti M, Osti D & Pelicci G (2013) Cancer stem cell contribution to glioblastoma invasiveness. *Stem Cell Res Ther* 4: 1–11
- Orzan F, Pagani F, Cominelli M, Triggiani L, Calza S, De Bacco F, Medicina D, Balzarini P, Panciani PP, Liserre R, *et al* (2020) A simplified integrated molecular and immunohistochemistry-based algorithm allows high accuracy prediction of glioblastoma transcriptional subtypes. *Lab Investig* 100: 1330–1344
- Osborn AG, Hedlund GL & Salzman KL (2017) Astrocytomas. In *Osborn's Brain - Imaging, pathology, and anatomy* pp 509–552. Elsevier
- Ostrom QT, Cioffi G, Gittleman H, Patil N, Waite K, Kruchko C & Barnholtz-Sloan JS (2019a) CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2012–2016. *Neuro Oncol* 21: V1–V100
- Ostrom QT, Fahmideh MA, Cote DJ, Muskens IS, Schraw JM, Scheurer ME & Bondy ML (2019b) Risk factors for childhood and adult primary brain tumors. *Neuro Oncol* 21: 1357–1375
- Ozawa M, Brennan PM, Zienius K, Kurian KM, Hollingworth W, Weller D, Grant R, Hamilton W & Ben-Shlomo Y (2019) The usefulness of symptoms alone or combined for general practitioners in considering the diagnosis of a brain tumour: A case-control study using the clinical practice research database (CPRD) (2000–2014). *BMJ Open* 9: 1–8
- Ozawa T, Riester M, Cheng YK, Huse JT, Squatrito M, Helmy K, Charles N, Michor F & Holland EC (2014) Most human non-GCIMP glioblastoma subtypes evolve from a common proneural-like precursor glioma. *Cancer Cell* 26: 288–300
- Pace A, Dirven L, Koekkoek JAF, Golla H, Fleming J, Rudà R, Marosi C, Rhun E Le, Grant R, Oliver K, *et al* (2017) European Association for Neuro-Oncology (EANO) guidelines for palliative care in adults with glioma. *Lancet Oncol* 18: e330–e340
- Palucka K & Banchereau J (2012) Cancer immunotherapy via dendritic cells. *Nat Rev Cancer* 12: 265–277

- Parada LF, Dirks PB & Wechsler-Reya RJ (2017) Brain tumor stem cells remain in play. *J Clin Oncol* 35: 2428–2431
- Park NI, Guilhamon P, Desai K, McAdam RF, Langille E, O'Connor M, Lan X, Whetstone H, Coutinho FJ, Vanner RJ, *et al* (2017) ASCL1 Reorganizes Chromatin to Direct Neuronal Fate and Suppress Tumorigenicity of Glioblastoma Stem Cells. *Cell Stem Cell* 21: 209-224.e7
- Park YW, Han K, Ahn SS, Bae S, Choi YS, Chang JH, Kim SH, Kang SG & Lee SK (2018) Prediction of IDH1-mutation and 1p/19q-codeletion status using preoperative MR imaging phenotypes in lower grade gliomas. *Am J Neuroradiol* 39: 37–42
- Parsons DW, Jones S, Zhang X, Lin JCH, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL, *et al* (2008) An integrated genomic analysis of human glioblastoma multiforme. *Science* (80-) 321: 1807–1812
- Pastrana E, Silva-Vargas V & Doetsch F (2011) Eyes wide open: A critical review of sphere-formation as an assay for stem cells. *Cell Stem Cell* 8: 486–498
- Patel A, Tirosh I, Trombetta J, Shalek A, Gillespie S, Wakimoto H, Cahill D, Nahed B, Curry W, Martuza R, *et al* (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* (80-) 344: 1396–1402
- Pece S, Disalvatore D, Tosoni D, Vecchi M, Confalonieri S, Bertalot G, Viale G, Colleoni M, Veronesi P, Galimberti V, *et al* (2019) Identification and clinical validation of a multigene assay that interrogates the biology of cancer stem cells and predicts metastasis in breast cancer: A retrospective consecutive study. *EBioMedicine* 42: 352–362
- Pellegatta S, Eoli M, Frigerio S, Antozzi C, Bruzzone MG, Cantini G, Nava S, Anghileri E, Cuppini L, Cuccarini V, *et al* (2013) The natural killer cell response and tumor debulking are associated with prolonged survival in recurrent glioblastoma patients receiving dendritic cells loaded with autologous tumor lysates. *Oncoimmunology* 2
- Perry JR, Bélanger K, Mason WP, Fulton D, Kavan P, Easaw J, Shields C, Kirby S, Macdonald DR, Eisenstat DD, *et al* (2010) Phase II trial of continuous dose-intense temozolomide in recurrent malignant glioma: RESCUE study. *J Clin Oncol* 28: 2051–2057
- Pessina F, Navarria P, Cozzi L, Ascolese AM, Simonelli M, Santoro A, Clerici E, Rossi M, Scorsetti M & Bello L (2017) Maximize surgical resection beyond contrast-enhancing boundaries in newly diagnosed glioblastoma multiforme: is it useful and safe? A single institution retrospective experience. *J Neurooncol* 135: 129–139

- Philips A, Henshaw DL, Lamburn G & O'Carroll MJ (2018) Brain tumours: Rise in glioblastoma multiforme incidence in England 1995-2015 Suggests an Adverse Environmental or Lifestyle Factor. *J Environ Public Health* 2018
- Phillips HS, Kharbanda S, Chen R, Forrest WF, Soriano RH, Wu TD, Misra A, Nigro JM, Colman H, Soroceanu L, *et al* (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
- Piao Y, Liang J, Holmes L, Henry V, Sulman E & De Groot JF (2013) Acquired resistance to anti-VEGF therapy in glioblastoma is associated with a mesenchymal transition. *Clin Cancer Res* 19: 4392–4403 doi:10.1158/1078-0432.CCR-12-1557 [PREPRINT]
- Piccirillo SGM, Reynolds BA, Zanetti N, Lamorte G, Binda E, Broggi G, Brem H, Olivi A, Dimeco F & Vescovi AL (2006) Bone morphogenetic proteins inhibit the tumorigenic potential of human brain tumour-initiating cells. *Nature* 444: 761–765
- Piccirillo SGM, Spiteri I, Sottoriva A, Touloumis A, Ber S, Price SJ, Heywood R, Francis NJ, Howarth KD, Collins VP, *et al* (2015) Contributions to drug resistance in glioblastoma derived from malignant cells in the sub-ependymal zone. *Cancer Res* 75: 194–202
- Pieri V, Gallotti A, Drago D, Cominelli C, Pagano I, Conti V, Valtorta S, Coliva A, Lago S, Michelatti D, *et al* (2022) Aberrant L-Fucose accumulation and increased core fucosylation are specific metabolic liabilities in mesenchymal glioblastoma. *Cancer Res - Under Rev*
- Pieri V, Sanvito F, Riva M, Petrini A, Rancoita PMV, Cirillo S, Iadanza A, Bello L, Castellano A & Falini A (2021) Along-tract statistics of neurite orientation dispersion and density imaging diffusion metrics to enhance MR tractography quantitative analysis in healthy controls and in patients with brain tumors. *Hum Brain Mapp* 42: 1268–1286
- Pitter KL, Tamagno I, Alikhanyan K, Hosni-Ahmed A, Pattwell SS, Donnola S, Dai C, Ozawa T, Chang M, Chan TA, *et al* (2016) Corticosteroids compromise survival in glioblastoma. *Brain* 139: 1458–1471
- Platten M, Nollen EAA, Röhrig UF, Fallarino F & Opitz CA (2019) Tryptophan metabolism as a common therapeutic target in cancer, neurodegeneration and beyond. *Nat Rev Drug Discov* 18: 379–401
- Pollard SM, Yoshikawa K, Clarke ID, Danovi D, Stricker S, Russell R, Bayani J, Head R, Lee M, Bernstein M, *et al* (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
doi:10.1016/j.stem.2009.03.014 [PREPRINT]
- Pollen AA, Nowakowski TJ, Chen J, Retallack H, Sandoval-Espinosa C, Nicholas CR, Shuga J, Liu SJ, Oldham MC, Diaz A, *et al* (2015) Molecular Identity of Human Outer Radial Glia during Cortical Development. *Cell* 163: 55–67
- Posti JP, Bori M, Kauko T, Sankinen M, Nordberg J, Rahi M, Frantzén J, Vuorinen V & Sipilä JOT (2015) Presenting symptoms of glioma in adults. *Acta Neurol Scand* 131: 88–93
- Prasanna P, Patel J, Partovi S, Madabhushi A & Tiwari P (2017) Radiomic features from the peritumoral brain parenchyma on treatment-naïve multi-parametric MR imaging predict long versus short-term survival in glioblastoma multiforme: Preliminary findings. *Eur Radiol* 27: 4188–4197
- Price SJ, Jena R, Burnet NG, Carpenter TA, Pickard JD & Gillard JH (2007) Predicting patterns of glioma recurrence using diffusion tensor imaging. *Eur Radiol* 17: 1675–1684
- Price SJ, Jena R, Burnet NG, Hutchinson PJ, Dean AF, Peña A, Pickard JD, Carpenter TA & Gillard JH (2006) Improved delineation of glioma margins and regions of infiltration with the use of diffusion tensor imaging: An image-guided biopsy study. *Am J Neuroradiol* 27: 1969–1974
- Rahman M, Reyner K, Deleyrolle L, Millette S, Azari H, Day BW, Stringer BW, Boyd AW, Johns TG, Blot V, *et al* (2015) Neurosphere and adherent culture conditions are equivalent for malignant glioma stem cell lines. *Anat Cell Biol*: 25–35
- Ranger AM, Patel YK, Chaudhary N & Anantha R V. (2014) Familial syndromes associated with intracranial tumours: A review. *Child's Nerv Syst* 30: 47–64
- Rathore S, Akbari H, Bakas S, Pisapia JM, Shukla G, Rudie JD, Da X, Davuluri R V., Dahmane N, O'Rourke DM, *et al* (2019) Multivariate Analysis of Preoperative Magnetic Resonance Imaging Reveals Transcriptomic Classification of de novo Glioblastoma Patients. *Front Comput Neurosci* 13: 1–9
- Rathore S, Akbari H & Doshi J (2018) Radiomic signature of infiltration in peritumoral edema predicts subsequent recurrence in glioblastoma: implications for personalized radiotherapy planning. *J Med Imaging* 5: 1
- Reynolds BA, Tetzlaff W & Weiss S (1992) A multipotent EGF-responsive striatal embryonic progenitor cell produces neurons and astrocytes. *J Neurosci* 12: 4565–4574
- Le Rhun E, Preusser M, Roth P, Reardon DA, van den Bent M, Wen P, Reifenberger G & Weller M (2019) Molecular targeted therapy of glioblastoma. *Cancer Treat*

Rev 80: 101896

- Richards LM, Whitley OKN, MacLeod G, Cavalli FMG, Coutinho FJ, Jaramillo JE, Svergun N, Riverin M, Croucher DC, Kushida M, *et al* (2021) Gradient of Developmental and Injury Response transcriptional states defines functional vulnerabilities underpinning glioblastoma heterogeneity. *Nat Cancer* 2: 157–173
- Richichi C, Osti D, Del Bene M, Fornasari L, Patanè M, Pollo B, DiMeco F & Pelicci G (2016) Tumor-initiating cell frequency is relevant for glioblastoma aggressiveness. *Oncotarget* 7: 71491–71503
- Ringel F, Pape H, Sabel M, Krex D, Bock HC, Misch M, Weyerbrock A, Westermaier T, Senft C, Schucht P, *et al* (2016) Clinical benefit from resection of recurrent glioblastomas: Results of a multicenter study including 503 patients with recurrent glioblastomas undergoing surgical resection. *Neuro Oncol* 18: 96–104
- Riva M, Lopci E, Gay LG, Nibaldi MC, Rossi M, Sciortino T, Castellano A & Bello L (2021) Advancing Imaging to Enhance Surgery: From Image to Information Guidance. *Neurosurg Clin N Am* 32: 31–46
- Romano A, Calabria LF, Tavanti F, Minniti G, Rossi-Espagnet MC, Coppola V, Pugliese S, Guida D, Francione G, Colonnese C, *et al* (2013) Apparent diffusion coefficient obtained by magnetic resonance imaging as a prognostic marker in glioblastomas: Correlation with MGMT promoter methylation status. *Eur Radiol* 23: 513–520
- Rossi M, Gay L, Ambrogi F, Conti Nibaldi M, Sciortino T, Puglisi G, Leonetti A, Mocellini C, Caroli M, Cordera S, *et al* (2021) Association of supratotal resection with progression-free survival, malignant transformation, and overall survival in lower-grade gliomas. *Neuro Oncol* 23: 812–826
- Rutkowska A, Stoczyńska-Fidelus E, Janik K, Włodarczyk A & Rieske P (2019) EGFRvIII: An Oncogene with Ambiguous Role. *J Oncol* 2019
- Sahm F, Capper D, Jeibmann A, Habel A, Paulus W, Troost D & Von Deimling A (2012) Addressing diffuse glioma as a systemic brain disease with single-cell analysis. *Arch Neurol* 69: 523–526
- Sampson JH, Gunn MD, Fecci PE & Ashley DM (2020) Brain immunology and immunotherapy in brain tumours. *Nat Rev Cancer* 20: 12–25
- Sanai N, Polley MY, McDermott MW, Parsa AT & Berger MS (2011) An extent of resection threshold for newly diagnosed glioblastomas: Clinical article. *J Neurosurg* 115: 3–8
- Sandmann T, Bourgon R, Garcia J, Li C, Cloughesy T, Chinot OL, Wick W, Nishikawa

- R, Mason W, Henriksson R, *et al* (2015) Patients with proneural glioblastoma may derive overall survival benefit from the addition of bevacizumab to first-line radiotherapy and temozolomide: Retrospective analysis of the AVAglio trial. *J Clin Oncol* 33: 2735–2744
- Sanghani P, Ti AB, Kam King NK & Ren H (2019) Evaluation of tumor shape features for overall survival prognosis in glioblastoma multiforme patients. *Surg Oncol* 29: 178–183
- Sarkaria JN, Hu LS, Parney IF, Pafundi DH, Brinkmann DH, Laack NN, Giannini C, Burns TC, Kizilbash SH, Laramy JK, *et al* (2018) Is the blood-brain barrier really disrupted in all glioblastomas? A critical assessment of existing clinical data. *Neuro Oncol* 20: 184–191
- Schad LR (2004) Problems in texture analysis with magnetic resonance imaging. *Dialogues Clin Neurosci* 6: 235–242
- Schiff D, Lee EQ, Nayak L, Norden AD, Reardon DA & Wen PY (2015) Medical management of brain tumors and the sequelae of treatment. *Neuro Oncol* 17: 488–504
- Schumacher TN & Schreiber RD (2015) Neoantigens in cancer immunotherapy. *Science* (80-) 348: 69–74
- Scoccianti S, Francolini G, Carta GA, Greto D, Detti B, Simontacchi G, Visani L, Baki M, Poggese L, Bonomo P, *et al* (2018) Re-irradiation as salvage treatment in recurrent glioblastoma: A comprehensive literature review to provide practical answers to frequently asked questions. *Crit Rev Oncol Hematol* 126: 80–91
- Shibue T & Weinberg RA (2017) EMT, CSCs, and drug resistance: The mechanistic link and clinical implications. *Nat Rev Clin Oncol* 14: 611–629
- Shu Q, Wong KK, Su JM, Adesina AM, Yu LT, Tsang YTM, Antalffy BC, Baxter P, Perlaky L, Yang J, *et al* (2008) Direct Orthotopic Transplantation of Fresh Surgical Specimen Preserves CD133+ Tumor Cells in Clinically Relevant Mouse Models of Medulloblastoma and Glioma. *Stem Cells* 26: 1414–1424
- Shum T, Omer B, Tashiro H, Kruse RL, Wagner DL, Parikh K, Yi Z, Sauer T, Liu D, Parihar R, *et al* (2017) Constitutive signaling from an engineered IL7 receptor promotes durable tumor elimination by tumor-redirectioned T cells. *Cancer Discov* 7: 1238–1247
- Singh S, Clarke I, Terasaki M, Bonn V, Hawkins C, Squire J & Dirks P (2003) Identification of a cancer stem cell in human brain tumors. *Cancer Res* Sep 15: 5821–8
- Singh SK, Hawkins C, Clarke ID, Squire JA, Bayani J, Hide T, Henkelman RM, Cusimano MD & Dirks PB (2004) Identification of human brain tumour initiating

- cells. *Nature* 432: 396–401
- Son MJ, Woolard K, Nam DH, Lee J & Fine HA (2009) SSEA-1 Is an Enrichment Marker for Tumor-Initiating Cells in Human Glioblastoma. *Cell Stem Cell* 4: 440–452
- Sottoriva A, Spiteri I, Piccirillo SGM, Touloumis A, Collins VP, Marioni JC, Curtis C, Watts C & Tavaré S (2013) Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc Natl Acad Sci U S A* 110: 4009–4014
- Stathias V, Jermakowicz AM, Maloof ME, Forlin M, Walters W, Suter RK, Durante MA, Williams SL, Harbour JW, Volmar CH, *et al* (2018) Drug and disease signature integration identifies synergistic combinations in glioblastoma. *Nat Commun* 9
- Straube C, Kessel KA, Zimmer C, Schmidt-Graf F, Schlegel J, Gempt J, Meyer B & Combs SE (2019) A Second Course of Radiotherapy in Patients with Recurrent Malignant Gliomas: Clinical Data on Re-irradiation, Prognostic Factors, and Usefulness of Digital Biomarkers. *Curr Treat Options Oncol* 20
- Strauss SB, Meng A, Ebani EJ & Chiang GC (2019) Imaging Glioblastoma Posttreatment: Progression, Pseudoprogression, Pseudoresponse, Radiation Necrosis. *Radiol Clin North Am* 57: 1199–1216
- Stringer BW, Day BW, D'Souza RCJ, Jamieson PR, Ensbey KS, Bruce ZC, Lim YC, Goasdoué K, Offenhäuser C, Akgül S, *et al* (2019) A reference collection of patient-derived cell line and xenograft models of proneural, classical and mesenchymal glioblastoma. *Sci Rep* 9: 1–14
- Stummer W, Pichlmeier U, Meinel T, Wiestler OD, Zanella F & Reulen HJ (2006) Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase III trial. *Lancet Oncol* 7: 392–401
- Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJ, Janzer RC, Ludwin SK, Allgeier A, Fisher B, Belanger K, *et al* (2009) Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol* 10: 459–466
- Stupp R, Mason W, van den Bent M & *et al* (2005) Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. *N Engl J Med*: 987–996
- Su Huang HJ, Nagane M, Klingbeil CK, Lin H, Nishikawa R, Ji XD, Huang CM, Gill GN, Wiley HS & Cavenee WK (1997) The enhanced tumorigenic activity of a mutant epidermal growth factor receptor common in human cancers is

- mediated by threshold levels of constitutive tyrosine phosphorylation and unattenuated signaling. *J Biol Chem* 272: 2927–2935
- Subramanian A (2017) A next generation Connectivity Map: L1000 Platform. *Cell* 171: 1437–1452 [PREPRINT]
- Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, Gould J, Davis JF, Tubelli AA, Asiedu JK, *et al* (2017) A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* 171: 1437-1452.e17
- Suchorska B, Weller M, Tabatabai G, Senft C, Hau P, Sabel MC, Herrlinger U, Ketter R, Schlegel U, Marosi C, *et al* (2016) Complete resection of contrast-enhancing tumor volume is associated with improved survival in recurrent glioblastoma - Results from the DIRECTOR trial. *Neuro Oncol* 18: 549–556
- Suh CH, Kim HS, Jung SC, Park JE, Choi CG & Kim SJ (2019) MRI as a diagnostic biomarker for differentiating primary central nervous system lymphoma from glioblastoma: A systematic review and meta-analysis. *J Magn Reson Imaging* 50: 560–572
- Sunwoo L, Choi SH, Park CK, Kim JW, Yi KS, Lee WJ, Yoon TJ, Song SW, Kim JE, Kim JY, *et al* (2013) Correlation of apparent diffusion coefficient values measured by diffusion MRI and MGMT promoter methylation semiquantitatively analyzed with MS-MLPA in patients with glioblastoma multiforme. *J Magn Reson Imaging* 37: 351–358
- Suvà ML, Rheinbay E, Gillespie SM, Patel AP, Wakimoto H, Rabkin SD, Riggi N, Chi AS, Cahill DP, Nahed B V., *et al* (2014) Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell* 157: 580–594
- Suvà ML & Tirosh I (2020) The Glioma Stem Cell Model in the Era of Single-Cell Genomics. *Cancer Cell* 37: 630–636
- Svendsen CN, Ter Borg MG, Armstrong RJE, Rosser AE, Chandran S, Ostenfeld T & Caldwell MA (1998) A new method for the rapid and long term growth of human neural precursor cells. *J Neurosci Methods* 85: 141–152
- Tan SK, Jermakowicz A, Mookhtiar AK, Nemeroff CB, Schürer SC & Ayad NG (2018) Drug repositioning in glioblastoma: A pathway perspective. *Front Pharmacol* 9
- Tang Y, Qazi M, Brown K, Mikolajewicz N, Moffat J, Singh S & McNicholas P (2021) Identification of five important genes to predict glioblastoma subtypes. *Neuro-Oncology Adv* 3: 1–11
- Tari H, Kessler K, Trahearn N, Werner B, Vinci M, Jones C & Sottoriva A (2022) Quantification of spatial subclonal interactions enhancing the invasive phenotype of pediatric glioma. *Cell Rep* 40: 111283
- Taylor JW & Schiff D (2015) Treatment Considerations for MGMT-Unmethylated

- Glioblastoma. *Curr Neurol Neurosci Rep* 15: 1–6
- Tesileanu CMS, Dirven L, Wijnenga MMJ, Koekkoek JAF, Vincent AJPE, Dubbink HJ, Atmodimedjo PN, Kros JM, Van Duinen SG, Smits M, *et al* (2020) Survival of diffuse astrocytic glioma, IDH1/2 wildtype, with molecular features of glioblastoma, WHO grade IV: A confirmation of the cIMPACT-NOW criteria. *Neuro Oncol* 22: 515–523
- Thiery JP (2002) Epithelial–mesenchymal transitions in tumour progression. *Nat Rev Cancer* 2: 442–454 (<https://www.nature.com/articles/nrc822>) [PREPRINT]
- Thust SC, van den Bent MJ & Smits M (2018) Pseudoprogression of brain tumors. *J Magn Reson Imaging* 48: 571–589
- Tirosh I, Venteicher AS, Hebert C, Escalante LE, Patel AP, Yizhak K, Fisher JM, Rodman C, Mount C, Filbin MG, *et al* (2016) Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* 539: 309–313
- Tixier F, Um H, Bermudez D, Iyer A, Apte A, Graham MS, Nevel KS, Deasy JO, Young RJ & Veeraraghavan H (2019) Preoperative MRI-radiomics features improve prediction of survival in glioblastoma patients over MGMT methylation status alone. *Oncotarget* 10: 660–672
- Touat M, Li YY, Boynton AN, Spurr LF, Iorgulescu JB, Bohrson CL, Cortes-Ciriano I, Birzu C, Geduldig JE, Pelton K, *et al* (2020) Mechanisms and therapeutic implications of hypermutation in gliomas. *Nature* 580: 517–523
- Turner KM, Deshpande V, Beyter D, Koga T, Ruser J, Lee C, Li B, Arden K, Ren B, Nathanson DA, *et al* (2017) Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* 543: 122–125
- Valtorta S, Dico A Lo, Raccagni I, Gaglio D, Belloli S, Politi LS, Martelli C, Diceglie C, Bonanomi M, Ercoli G, *et al* (2017) Metformin and temozolomide, a synergic option to overcome resistance in glioblastoma multiforme models. *Oncotarget* 8: 113090–113104
- Venkataramani V, Tanev DI, Strahle C, Studier-Fischer A, Fankhauser L, Kessler T, Körber C, Kardorff M, Ratliff M, Xie R, *et al* (2019) Glutamatergic synaptic input to glioma cells drives brain tumour progression. *Nature* 573: 532–538
- Venkataramani V, Yang Y, Schubert MC, Reyhan E, Tetzlaff SK, Wißmann N, Botz M, Soyka SJ, Beretta CA, Pramatarov RL, *et al* (2022) Glioblastoma hijacks neuronal mechanisms for brain invasion. *Cell* 185: 2899–2917.e31
- Venkatesh HS, Morishita W, Geraghty AC, Silverbush D, Gillespie SM, Arzt M, Tam LT, Espenel C, Ponnuswami A, Ni L, *et al* (2019) Electrical and synaptic integration of glioma into neural circuits. *Nature* 573: 539–545
- Venneti S & Thompson CB (2017) Metabolic Reprogramming in Brain Tumors

- Venteicher AS, Tirosh I, Hebert C, Yizhak K, Neftel C, Filbin MG, Hovestadt V, Escalante LE, Shaw ML, Rodman C, *et al* (2017) Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* (80-) 355
- Verduin M, Compter I, Steijvers D, Postma AA, Eekers DBP, Anten MM, Ackermans L, Ter Laan M, Leijenaar RTH, Van De Weijer T, *et al* (2018) Noninvasive Glioblastoma Testing: Multimodal Approach to Monitoring and Predicting Treatment Response. *Dis Markers* 2018
- Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, *et al* (2010) Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 17: 98–110
- Vescovi AL, Galli R & Reynolds BA (2006) Brain tumour stem cells. *Nat Rev Cancer* 6: 425–436
- Vinci M, Burford A, Molinari V, Kessler K, Popov S, Clarke M, Taylor KR, Pemberton HN, Lord CJ, Gutteridge A, *et al* (2018) Functional diversity and cooperativity between subclonal populations of pediatric glioblastoma and diffuse intrinsic pontine glioma cells. *Nat Med* 24: 1204–1215
- Visigalli I, Moresco RM, Belloli S, Politi LS, Gritti A, Ungaro D, Matarrese M, Turolla E, Falini A, Scotti G, *et al* (2009) Monitoring disease evolution and treatment response in lysosomal disorders by the peripheral benzodiazepine receptor ligand PK11195. *Neurobiol Dis* 34: 51–62
- Vitiello GAF, Losi Guembarovski R, Amarante MK, Ceribelli JR, Carmelo ECB & Watanabe MAE (2018) Interleukin 7 receptor alpha Thr244Ile genetic polymorphism is associated with susceptibility and prognostic markers in breast cancer subgroups. *Cytokine* 103: 121–126
- Wakimoto H, Kesari S, Farrell CJ, Curry WT, Zaupa C, Aghi M, Kuroda T, Stemmer-Rachamimov A, Shah K, Liu TC, *et al* (2009) Human glioblastoma-derived cancer stem cells: Establishment of invasive glioma models and treatment with oncolytic herpes simplex virus vectors. *Cancer Res* 69: 3472–3481
- Wan F, Zhang S, Xie R, Gao B, Campos B, Herold-Mende C & Lei T (2010) The utility and limitations of neurosphere assay, cd133 immunophenotyping and side population assay in glioma stem cell research. *Brain Pathol* 20: 877–889
- Wang J, Cazzato E, Ladewig E, Frattini V, Rosenbloom DIS, Zairis S, Abate F, Liu Z, Elliott O, Shin YJ, *et al* (2016) Clonal evolution of glioblastoma under therapy. *Nat Genet* 48: 768–776
- Wang Q, Hu B, Hu X, Kim H, Squatrito M, Scarpace L, deCarvalho AC, Lyu S, Li P, Li

- Y, *et al* (2017) Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer Cell* 32: 42-56.e6
- Weller M (2018) Where does O6-methylguanine DNA methyltransferase promoter methylation assessment place temozolomide in the future standards of care for glioblastoma? *Cancer* 124: 1316–1318
- Weller M, van den Bent M, Tonn JC, Stupp R, Preusser M, Cohen-Jonathan-Moyal E, Henriksson R, Rhun E Le, Balana C, Chinot O, *et al* (2017) European Association for Neuro-Oncology (EANO) guideline on the diagnosis and treatment of adult astrocytic and oligodendroglial gliomas. *Lancet Oncol* 18: e315–e329
- Weller M, Tabatabai G, Kästner B, Felsberg J, Steinbach JP, Wick A, Schnell O, Hau P, Herrlinger U, Sabel MC, *et al* (2015) MGMT promoter methylation is a strong prognostic biomarker for benefit from dose-intensified temozolomide rechallenge in progressive Glioblastoma: The DIRECTOR Trial. *Clin Cancer Res* 21: 2057–2064
- Wen PY, Weller M, Lee EQ, Alexander BM, Barnholtz-Sloan JS, Barthel FP, Batchelor TT, Bindra RS, Chang SM, Antonio Chiocca E, *et al* (2020) Glioblastoma in adults: A Society for Neuro-Oncology (SNO) and European Society of Neuro-Oncology (EANO) consensus review on current management and future directions. *Neuro Oncol* 22: 1073–1113
- Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, Matese JC, Perou CM, Hurt MM, Brown PO, *et al* (2002) Identification of Genes Periodically Expressed in the Human Cell Cycle and Their Expression in Tumors. *Mol Biol Cell* 13: 1977–2000
- Wick A, Kessler T, Elia AEH, Winkler F, Batchelor TT, Platten M & Wick W (2018) Glioblastoma in elderly patients: Solid conclusions built on shifting sand? *Neuro Oncol* 20: 174–183
- Wick W, Gorlia T, Bady P, Platten M, Van Den Bent MJ, Taphoorn MJB, Steuve J, Brandes AA, Hamou MF, Wick A, *et al* (2016) Phase II study of radiotherapy and temsirolimus versus radiochemotherapy with temozolomide in patients with newly diagnosed glioblastoma without MGMT promoter hypermethylation (EORTC 26082). *Clin Cancer Res* 22: 4797–4806
- Wick W, Gorlia T, Bendszus M, Taphoorn M, Sahm F, Harting I, Brandes AA, Taal W, Domont J, Idbaih A, *et al* (2017) Lomustine and Bevacizumab in Progressive Glioblastoma. *N Engl J Med* 377: 1954–1963
- De Witt Hamer PC, Robles SG, Zwinderman AH, Duffau H & Berger MS (2012)

- Impact of intraoperative stimulation brain mapping on glioma surgery outcome: A meta-analysis. *J Clin Oncol* 30: 2559–2565
- Wong AJ, Ruppert JM, Bigner SH, Grzeschik CH, Humphrey PA, Bigner DS & Vogelstein B (1992) Structural alterations of the epidermal growth factor receptor gene in human gliomas. *Proc Natl Acad Sci U S A* 89: 2965–2969
- Wood MD, Reis GF, Reuss DE & Phillips JJ (2016) Protein analysis of glioblastoma primary and posttreatment pairs suggests a mesenchymal shift at recurrence. *J Neuropathol Exp Neurol* 75: 925–935
- Woroniecka K, Chongsathidkiet P, Rhodin K, Kemeny H, Dechant C, Harrison Farber S, Elsamadicy AA, Cui X, Koyama S, Jackson C, et al (2018) T-cell exhaustion signatures vary with tumor type and are severe in glioblastoma. *Clin Cancer Res* 24: 4175–4186
- Xi Y Bin, Guo F, Xu ZL, Li C, Wei W, Tian P, Liu TT, Liu L, Chen G, Ye J, et al (2018) Radiomics signature: A potential biomarker for the prediction of MGMT promoter methylation in glioblastoma. *J Magn Reson Imaging* 47: 1380–1387
- Xie Y, Bergström T, Jiang Y, Johansson P, Marinescu VD, Lindberg N, Segerman A, Wicher G, Niklasson M, Baskaran S, et al (2015) The Human Glioblastoma Cell Culture Resource: Validated Cell Models Representing All Molecular Subtypes. *EBioMedicine* 2: 1351–1363
- Xue C, Zhang B, Deng J, Liu X, Li S & Zhou J (2021) Differentiating Giant Cell Glioblastoma from Classic Glioblastoma With Diffusion-Weighted Imaging. *World Neurosurg* 146: e473–e478
- Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, Kos I, Batinic-Haberle I, Jones S, Riggins GJ, et al (2009) IDH1 and IDH2 Mutations in Gliomas. *N Engl J Med* 360: 765–773
- Yoshida GJ (2020) Applications of patient-derived tumor xenograft models and tumor organoids. *J Hematol Oncol* 13: 1–16
- Youngblood MW, Stupp R & Sonabend AM (2021) Role of Resection in Glioblastoma Management. *Neurosurg Clin N Am* 32: 9–22
- Zhang H, Schneider T, Wheeler-Kingshott CA & Alexander DC (2012) NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage* 61: 1000–1016
- Zhang L, Yu H, Yuan Y, Yu JS, Lou Z, Xue Y & Liu Y (2020) The necessity for standardization of glioma stem cell culture: A systematic review. *Stem Cell Res Ther* 11: 1–7
- Zhao J, Li J bin, Wang J yan, Wang Y liang, Liu D wei, Li X bei, Song Y kun, Tian Y su, Yan X, Li Z hao, et al (2018) Quantitative analysis of neurite orientation

dispersion and density imaging in grading gliomas and detecting IDH-1 gene mutation status. *NeuroImage Clin* 19: 174–181

Zhou M, Scott J, Chaudhury B, Hall L, Goldgof D, Yeom KW, Iv M, Ou Y, Kalpathy-Cramer J, Napel S, *et al* (2018) Radiomics in Brain Tumor: Image Assessment, Quantitative Feature Descriptors, and Machine-Learning Approaches. *Am J Neuroradiol* 39: 208–216

A handwritten signature in black ink, appearing to read 'Hall', written in a cursive style.

9 APPENDIX: R Scripts

1_Raw_counts_prep.R

```
#load libraries
library(dplyr)
library(stringr)
library(readxl)

#download TCGA GBM mRNA gene expression data from
https://gdac.broadinstitute.org
#select mRNASeq_genes_raw_counts
#read database
raw_counts <-
read.delim('./A_gdac.broadinstitute.org_GBM.mRNAseq_Preprocess.Level_3.20160
12800.0.0/GBM.uncv2.mRNAseq_raw_counts.txt')
#check dimensions for consistency
dim(raw_counts)

#look for names of columns to select lines you need
write.csv(colnames(raw_counts),'samples.txt')

#read file with correspondence sample-subtype
sample_subtype <- read.delim('sample_subtypes.txt', header = F)

#backup di raw_counts
raw_counts_backup <- raw_counts

#get entries for gene.id
gene_id <- as.vector(raw_counts$HYBRIDIZATION.R)

#keep only gene.id (remove portion of the entry with '|' and after it)
gene_id_correct <- t(as.data.frame(strsplit(gene_id, '|')))
gene_id <- as.data.frame(gene_id_correct[,1])

#substitute gene.id column with correct gene.id
raw_counts$HYBRIDIZATION.R <- gene_id$`gene_id_correct[, 1]`
```

```

#name it accordingly
names(raw_counts)[names(raw_counts) == 'HYBRIDIZATION.R'] <- 'Gene.ID'

#save the database
write.csv(raw_counts,'raw_counts_dataset.txt')

```

2a_Subsetting_raw_counts.R

```

library(dplyr)
library(stringr)

#Get raw_counts_db by importing it, or by result of script 1
#Get sample subtype by importing it, or by result of script 1
raw_counts_backup <- raw_counts

#summary of samples
table(sample_subtype$V2)

#Remove recurrent samples
sample_subtype <- sample_subtype[-c(4,7,13,15,17),]

#select samples of each subtype
MES <- filter(sample_subtype, sample_subtype$V2 == 'MES')
PN <- filter(sample_subtype, sample_subtype$V2 == 'PN')
CL <- filter(sample_subtype, sample_subtype$V2 == 'CL')
Healthy <- filter(sample_subtype, sample_subtype$V2 == 'Healthy')

#remove gene_id '?'
raw_counts <- filter(raw_counts, Gene.ID != '?')

#assign names to rows
row.names(raw_counts) <- raw_counts$Gene.ID
#duplicate entry for SLC35E2 --> identify them
duplicate <- base::grep('SLC35E2', raw_counts$Gene.ID)
#mean exp for columns in the upper row
mean_value <- colMeans(raw_counts[duplicate,2:length(raw_counts)])

```

```

raw_counts[duplicate[1],2:length(raw_counts)] <- mean_value
#eliminate lower row of the duplicate
raw_counts <- raw_counts[-duplicate[2],]

#assign names to rows
row.names(raw_counts) <- raw_counts$Gene.ID

#remove duplicate Gene.ID column
raw_counts <- raw_counts[,-1]

#select cases to extract for each DEG by binding rows
MES_Healthy <- bind_rows(MES,Healthy)
PN_Healthy <- bind_rows(PN,Healthy)
CL_Healthy <- bind_rows(CL,Healthy)
MES_PN <- bind_rows(MES,PN)
CL_PN <- bind_rows(CL,PN)
MES_CL <- bind_rows(MES,CL)

#subset expression_db with only the cases needed for each DEG
Cts_DEG_MES_Healthy <- select(raw_counts, MES_Healthy$V1)
Cts_DEG_PN_Healthy <- select(raw_counts, PN_Healthy$V1)
Cts_DEG_CL_Healthy <- select(raw_counts, CL_Healthy$V1)
Cts_DEG_MES_PN <- select(raw_counts, MES_PN$V1)
Cts_DEG_CL_PN <- select(raw_counts, CL_PN$V1)
Cts_DEG_MES_CL <- select(raw_counts, MES_CL$V1)

```

3a_DGE.R

```
library(DESeq2)
library(dplyr)
library(pheatmap)
library(xlsx)

#check class of entries for samples
class(Cts_DEG_MES_Healthy$TCGA.02.0055.01)

#transform into numeric
#length(colnames(cts))
#i <- 1
#for (i in 1:length(colnames(cts))) {
#  cts[,i] <- as.integer(cts[,i])
#}

#CLASSICAL

#transform into matrix
cts <- as.matrix(Cts_DEG_CL_Healthy)

#round values to nearest integer
i <- 1
j <- 1
for (i in 1:nrow(cts)) {
  for (j in 1:ncol(cts)) {
    cts[i,j] <- round(cts[i,j])
  }
}

#assign condition to samples
coldata <- CL_Healthy

#adjust column names
names(coldata) <- c('sample','subtype')
```

```

#make as factor, relevel on Healthy subtype
coldata$subtype <- factor(coldata$subtype, levels = c('Healthy', 'CL'))

#check if correct
table(coldata$subtype)
class(coldata$subtype)
base::colnames(cts) == coldata$sample

#construct a DESeqDataSet
dds_CL_Healthy <- DESeqDataSetFromMatrix(countData = cts,
                                         colData = coldata,
                                         design = ~ subtype)

#Prefilter genes with counts < number of samples (arbitrarily chosen)
keep <- rowSums(counts(dds_CL_Healthy)) >= 9
dds_CL_Healthy <- dds_CL_Healthy[keep,]

#Differential Gene Expression Analysis
dds_CL_Healthy <- DESeq(dds_CL_Healthy)
dds_CL_Healthy

#results function automatically performs independent filtering based on the mean
#of normalized counts for each gene, optimizing the number of genes which will
#have an adjusted p value below a given FDR cutoff, alpha. By default the
argument
#alpha is set to 0.1
res_CL_Healthy <- results(dds_CL_Healthy, alpha = 0.05)
summary(res_CL_Healthy)
res_CL_Healthy
sum(res_CL_Healthy$padj<0.05, na.rm = T)

#Order according to decreasing Log2FC
resOrdered_LFC <- as.data.frame(res_CL_Healthy)
resOrdered_LFC <- arrange(resOrdered_LFC, desc(log2FoldChange))
summary(resOrdered_LFC)

```

```

resOrdered_LFC

#select only log2FC and adjusted p-value
resOrdered_LFC <- resOrdered_LFC[,c(2,6)]

#write xlsx ranked DEG list (including also non-significant p-values) for GSEA
#keep only Gene.ID, Log2FC, adjusted p-values
write.xlsx(resOrdered_LFC, 'DEG_TCGA_CL_vs_Healthy_no_recurrent.xlsx',
sheetName = "Sheet1",
          col.names = TRUE, row.names = TRUE, append = FALSE)

#PRONEURAL

#transform into matrix
cts <- as.matrix(Cts_DEG_PN_Healthy)

#round values to nearest integer
i <- 1
j <- 1
for (i in 1:nrow(cts)) {
  for (j in 1:ncol(cts)) {
    cts[i,j] <- round(cts[i,j])
  }
}

#assign condition to samples
coldata <- PN_Healthy

#adjust column names
names(coldata) <- c('sample','subtype')

#make as factor, relevel on Healthy subtype
coldata$subtype <- factor(coldata$subtype, levels = c('Healthy', 'PN'))

#check if correct

```

```

table(coldata$subtype)
class(coldata$subtype)
base::colnames(cts) == coldata$sample

#construct a DESeqDataSet
dds_PN_Healthy <- DESeqDataSetFromMatrix(countData = cts,
                                         colData = coldata,
                                         design = ~ subtype)

#Pre-filter genes with counts < number of samples (arbitrarily chosen)
keep <- rowSums(counts(dds_PN_Healthy)) >= 11
dds_PN_Healthy <- dds_PN_Healthy[keep,]

#Differential Gene Expression Analysis
dds_PN_Healthy <- DESeq(dds_PN_Healthy)
dds_PN_Healthy

#results function automatically performs independent filtering based on the mean
#of normalized counts for each gene, optimizing the number of genes which will
#have an adjusted p value below a given FDR cutoff, alpha. By default the
argument
#alpha is set to 0.1
res_PN_Healthy <- results(dds_PN_Healthy, alpha = 0.05)
summary(res_PN_Healthy)
res_PN_Healthy
sum(res_PN_Healthy$padj<0.05, na.rm = T)

#Order according to decreasing Log2FC
resOrdered_LFC <- as.data.frame(res_PN_Healthy)
resOrdered_LFC <- arrange(resOrdered_LFC, desc(log2FoldChange))
summary(resOrdered_LFC)
resOrdered_LFC

#select only log2FC and adjusted p-value
resOrdered_LFC <- resOrdered_LFC[,c(2,6)]

```



```

#write xlsx ranked DEG list (including also non-significant p-values) for GSEA
#keep only Gene.ID, Log2FC, adjusted p-values
write.xlsx(resOrdered_LFC, 'DEG_TCGA_PN_vs_Healthy_no_recurrent.xlsx',
sheetName = "Sheet1",
          col.names = TRUE, row.names = TRUE, append = FALSE)

```

```

#MESENCHYMAL

```

```

#transform into matrix
cts <- as.matrix(Cts_DEG_MES_Healthy)

```

```

#round values to nearest integer

```

```

i <- 1
j <- 1
for (i in 1:nrow(cts)) {
  for (j in 1:ncol(cts)) {
    cts[i,j] <- round(cts[i,j])
  }
}

```

```

#assign condition to samples
coldata <- MES_Healthy

```

```

#adjust column names
names(coldata) <- c('sample','subtype')

```

```

#make as factor, relevel on Healthy subtype
coldata$subtype <- factor(coldata$subtype, levels = c('Healthy', 'MES'))

```

```

#check if correct
table(coldata$subtype)
class(coldata$subtype)
base::colnames(cts) == coldata$sample

```

```

#construct a DESeqDataSet
dds_MES_Healthy <- DESeqDataSetFromMatrix(countData = cts,
                                         colData = coldata,
                                         design = ~ subtype)

#Pre-filter genes with counts < number of samples (arbitrarily chosen)
keep <- rowSums(counts(dds_MES_Healthy)) >= 14
dds_MES_Healthy <- dds_MES_Healthy[keep,]

#Differential Gene Expression Analysis
dds_MES_Healthy <- DESeq(dds_MES_Healthy)
dds_MES_Healthy

#results function automatically performs independent filtering based on the mean
#of normalized counts for each gene, optimizing the number of genes which will
#have an adjusted p value below a given FDR cutoff, alpha. By default the
argument
#alpha is set to 0.1
res_MES_Healthy <- results(dds_MES_Healthy, alpha = 0.05)
summary(res_MES_Healthy)
res_MES_Healthy
sum(res_MES_Healthy$padj<0.05, na.rm = T)

#Order according to decreasing Log2FC
resOrdered_LFC <- as.data.frame(res_MES_Healthy)
resOrdered_LFC <- arrange(resOrdered_LFC, desc(log2FoldChange))
summary(resOrdered_LFC)
resOrdered_LFC

#select only log2FC and adjusted p-value
resOrdered_LFC <- resOrdered_LFC[,c(2,6)]

#write xlsx ranked DEG list (including also non-significant p-values) for GSEA
#keep only Gene.ID, Log2FC, adjusted p-values

```

```

write.xlsx(resOrdered_LFC, 'DEG_TCGA_MES_vs_Healthy_no_recurrent.xlsx',
sheetName = "Sheet1",
          col.names = TRUE, row.names = TRUE, append = FALSE)

```

```

#MESENCHYMAL vs PRONEURAL

```

```

#transform into matrix

```

```

cts <- as.matrix(Cts_DEG_MES_PN)

```

```

#round values to nearest integer

```

```

i <- 1

```

```

j <- 1

```

```

for (i in 1:nrow(cts)) {

```

```

  for (j in 1:ncol(cts)) {

```

```

    cts[i,j] <- round(cts[i,j])

```

```

  }

```

```

}

```

```

#assign condition to samples

```

```

coldata <- MES_PN

```

```

#adjust column names

```

```

names(coldata) <- c('sample','subtype')

```

```

#make as factor, relevel on Healthy subtype

```

```

coldata$subtype <- factor(coldata$subtype, levels = c('PN', 'MES'))

```

```

#check if correct

```

```

table(coldata$subtype)

```

```

class(coldata$subtype)

```

```

base::colnames(cts) == coldata$sample

```

```

#construct a DESeqDataSet

```

```

dds_MES_PN <- DESeqDataSetFromMatrix(countData = cts,
                                     colData = coldata,

```

```

design = ~ subtype)

#Prefilter genes with counts < number of samples (arbitrarily chosen)
keep <- rowSums(counts(dds_MES_PN)) >= 15
dds_MES_PN <- dds_MES_PN[keep,]

#Differential Gene Expression Analysis
dds_MES_PN <- DESeq(dds_MES_PN)
dds_MES_PN

#results function automatically performs independent filtering based on the mean
#of normalized counts for each gene, optimizing the number of genes which will
#have an adjusted p value below a given FDR cutoff, alpha. By default the
argument
#alpha is set to 0.1
res_MES_PN <- results(dds_MES_PN, alpha = 0.05)
summary(res_MES_PN)
res_MES_PN
sum(res_MES_PN$padj<0.05, na.rm = T)

#Order according to decreasing Log2FC
resOrdered_LFC <- as.data.frame(res_MES_PN)
resOrdered_LFC <- arrange(resOrdered_LFC, desc(log2FoldChange))
summary(resOrdered_LFC)
resOrdered_LFC

#select only log2FC and adjusted p-value
resOrdered_LFC <- resOrdered_LFC[,c(2,6)]

#write xlsx ranked DEG list (including also non-significant p-values) for GSEA
#keep only Gene.ID, Log2FC, adjusted p-values
write.xlsx(resOrdered_LFC, 'DEG_TCGA_MES_vs_PN_no_recurrent.xlsx',
sheetName = "Sheet1",
          col.names = TRUE, row.names = TRUE, append = FALSE)

```

```

#CLASSICAL vs PRONEURAL

#transform into matrix
cts <- as.matrix(Cts_DEG_CL_PN)

#round values to nearest integer
i <- 1
j <- 1
for (i in 1:nrow(cts)) {
  for (j in 1:ncol(cts)) {
    cts[i,j] <- round(cts[i,j])
  }
}

#assign condition to samples
coldata <- CL_PN

#adjust column names
names(coldata) <- c('sample','subtype')

#make as factor, relevel on Healthy subtype
coldata$subtype <- factor(coldata$subtype, levels = c('PN', 'CL'))

#check if correct
table(coldata$subtype)
class(coldata$subtype)
base::colnames(cts) == coldata$sample

#construct a DESeqDataSet
dds_CL_PN <- DESeqDataSetFromMatrix(countData = cts,
                                     colData = coldata,
                                     design = ~ subtype)

#Prefilter genes with counts < number of samples (arbitrarily chosen)

```

```

keep <- rowSums(counts(dds_CL_PN)) >= 10
dds_CL_PN <- dds_CL_PN[keep,]

#Differential Gene Expression Analysis
dds_CL_PN <- DESeq(dds_CL_PN)
dds_CL_PN

#results function automatically performs independent filtering based on the mean
#of normalized counts for each gene, optimizing the number of genes which will
#have an adjusted p value below a given FDR cutoff, alpha. By default the
argument
#alpha is set to 0.1
res_CL_PN <- results(dds_CL_PN, alpha = 0.05)
summary(res_CL_PN)
res_CL_PN
sum(res_CL_PN$padj<0.05, na.rm = T)

#Order according to decreasing Log2FC
resOrdered_LFC <- as.data.frame(res_CL_PN)
resOrdered_LFC <- arrange(resOrdered_LFC, desc(log2FoldChange))
summary(resOrdered_LFC)
resOrdered_LFC

#select only log2FC and adjusted p-value
resOrdered_LFC <- resOrdered_LFC[,c(2,6)]

#write xlsx ranked DEG list (including also non-significant p-values) for GSEA
#keep only Gene.ID, Log2FC, adjusted p-values
write.xlsx(resOrdered_LFC, 'DEG_TCGA_CL_vs_PN_no_recurrent.xlsx', sheetName
= "Sheet1",
          col.names = TRUE, row.names = TRUE, append = FALSE)

#MESENCHYMAL vs CLASSICAL

#transform into matrix
cts <- as.matrix(Cts_DEG_MES_CL)

```

```

#round values to nearest integer
i <- 1
j <- 1
for (i in 1:nrow(cts)) {
  for (j in 1:ncol(cts)) {
    cts[i,j] <- round(cts[i,j])
  }
}

#assign condition to samples
coldata <- MES_CL

#adjust column names
names(coldata) <- c('sample','subtype')

#make as factor, relevel on Healthy subtype
coldata$subtype <- factor(coldata$subtype, levels = c('CL', 'MES'))

#check if correct
table(coldata$subtype)
class(coldata$subtype)
base::colnames(cts) == coldata$sample

#construct a DESeqDataSet
dds_MES_CL <- DESeqDataSetFromMatrix(countData = cts,
                                     colData = coldata,
                                     design = ~ subtype)

#Prefilter genes with counts < number of samples (arbitrarily chosen)
keep <- rowSums(counts(dds_MES_CL)) >= 13
dds_MES_CL <- dds_MES_CL[keep,]

#Differential Gene Expression Analysis
dds_MES_CL <- DESeq(dds_MES_CL)

```

```
dds_MES_CL
```

```
#results function automatically performs independent filtering based on the mean  
#of normalized counts for each gene, optimizing the number of genes which will  
#have an adjusted p value below a given FDR cutoff, alpha. By default the  
argument
```

```
#alpha is set to 0.1
```

```
res_MES_CL <- results(dds_MES_CL, alpha = 0.05)
```

```
summary(res_MES_CL)
```

```
res_MES_CL
```

```
sum(res_MES_CL$padj<0.05, na.rm = T)
```

```
#Order according to decreasing Log2FC
```

```
resOrdered_LFC <- as.data.frame(res_MES_CL)
```

```
resOrdered_LFC <- arrange(resOrdered_LFC, desc(log2FoldChange))
```

```
summary(resOrdered_LFC)
```

```
resOrdered_LFC
```

```
#select only log2FC and adjusted p-value
```

```
resOrdered_LFC <- resOrdered_LFC[,c(2,6)]
```

```
#write xlsx ranked DEG list (including also non-significant p-values) for GSEA
```

```
#keep only Gene.ID, Log2FC, adjusted p-values
```

```
write.xlsx(resOrdered_LFC, 'DEG_TCGA_MES_vs_CL_no_recurrent.xlsx',
```

```
sheetName = "Sheet1",
```

```
col.names = TRUE, row.names = TRUE, append = FALSE)
```

4_xlsx_to_rnk_no_NA.R

```
library(readxl)
```

```
gene_list <- read_xlsx('./2_Datasets/No_recurrent/DEG_TCGA_XXX.xlsx',
```

```
col_names = T)
```

```
gene_list <- as.data.frame(gene_list)
```

```
#filter for adjusted p-values NA
```



```

sum(is.na(gene_list$padj))
keep_rows <- !is.na(gene_list$padj)
gene_list <- gene_list[keep_rows,]

#keep only LogFC and arrange it in descending order
gene_list <- gene_list[,c(1,2)]
gene_list <- arrange(gene_list, desc(log2FoldChange))

write.table(gene_list, file =
'./preranked_DEG_TCGA_MES_vs_PN_no_recurrent.rnk', sep = '\t',
          quote = F, col.names = T, row.names = F)

```

0_Getting_preranked_no_NA.R

```

library(readxl)
library(dplyr)

gene_list <- read_xls('./XXXXXX.xls', col_names = T)
gene_list <- as.data.frame(gene_list)

#filter for adjusted p-values NA
sum(is.na(gene_list$padj))
keep_rows <- !is.na(gene_list$padj)
gene_list <- gene_list[keep_rows,]

#keep only LogFC and arrange it in descending order
gene_list <- gene_list[,c(1,2)]

#to invert order of comparison
#gene_list$log2FoldChange <- gene_list$log2FoldChange * (-1)

gene_list <- arrange(gene_list, desc(log2FoldChange))

write.table(gene_list, file = './preranked_DEG_XXXXXXX.rnk', sep = '\t',
          quote = F, col.names = T, row.names = F)

```

0_Signatures_to_gmx.R

```
library(readxl)

Sig <- read_xls('./XXXXXXXXXXXXX.xls', col_names = T)
Sig <- as.data.frame(Sig)

write.table(Sig, file = './XXXXX_signature.gmx', sep = '\t',
            quote = F, col.names = T, row.names = F)
```

QuantoPlot.R

```
library(xlsx)
library(tidyverse)
library(ggplot2)

#read quantification xlsx
quant <- read.xlsx('./1_Quantifications_xlsx/26_Test_IL7R.xlsx', sheetIndex = 1)
quant$Sample <- as.factor(quant$Sample)

#perform relative expression to housekeeping protein
quant$Relative_Expression <- quant$IL7R/quant$GAPDH

#add sample subgroup
subgroup <- c('MES', 'MES', 'MES', 'PN', 'PN', 'PN', 'CTRL', 'CTRL')
quant$Subgroup <- factor(subgroup, levels = c('CTRL', 'PN', 'MES'))

#plot relative expression
quantoplot <- ggplot(quant, aes(Sample, Relative_Expression)) +
  geom_col(aes(fill = Subgroup)) + xlab('Sample') + ylab('IL7R/GAPDH') +
  scale_fill_manual(values = c('dark green', 'purple', 'red'))

quantoplot

#boxplot relative expression relative to subgroup

quantobox <- ggplot(quant, aes(Subgroup, Relative_Expression)) +
```

```

    geom_boxplot(aes(fill = Subgroup)) + xlab('Subgroup') + ylab('IL7R/GAPDH')
+
    scale_fill_manual(values = c('dark green', 'purple', 'red')) +
    theme(legend.position = 'none')

```

quantobox

```

#look for statistical differences
#subset dataset
quant_stat <- filter(quant, quant$Subgroup != 'CTRL')
#perform t.test or Wilcox.test
t.test(quant_stat$Relative_Expression~quant_stat$Subgroup)

```

Getting_signature.R

```

library(tidyverse)
library(readxl)
setwd('./1-DGE/')

##PN vs MES GLANCE
Sig <- read_xls('./DGE_GLANCE_PN_vs_MES_CSC.xls')
Sig <- as.data.frame(Sig)

##Ordina variabili per Adjusted p-Value e controlla non ci siano significatività
##SeqC NA o 0 e tieni solo i +1 (significativamente up) e -1
##(significativamente down)
Sig <- arrange(Sig, Sig$padj)
table(Sig$significance_SEQC)
Sig_rank <- filter(Sig, Sig$significance_SEQC!=0)

##Seleziona solo le colonne che ti interessano (ID, Adj_pVal, LogFC)
Sig_rank <- Sig_rank[,c(1,2,4)]

##Seleziona solo le entries upregolate in PN (essendo log2, scelgo soglia per
##livelli almeno 10 volte -> log2(10) = 3.32, adj_pVal<=0.05)
Sig_PN <- filter(Sig_rank, Sig_rank$log2FoldChange>=(3.32))

```

```

Sig_PN <- filter(Sig_PN, Sig_PN$padj<=0.05)
nrow(Sig_PN)

##Seleziona solo le entries upregolate in MES (essendo log2, scelgo soglia per
##livelli almeno 10 volte -> log2(10) = 3.32, adj_pVal<=0.05)
Sig_MES <- filter(Sig_rank, Sig_rank$log2FoldChange<=(-3.32))
Sig_MES <- filter(Sig_MES, Sig_MES$padj<=0.05)
nrow(Sig_MES)

##Salva file geni per firma
write.table(Sig_PN, file = './PN_signature_logFC3.32.xls', sep = '\t',
            quote = F, col.names = T, row.names = F)
write.table(Sig_MES, file = './MES_signature_logFC3.32.xls', sep = '\t',
            quote = F, col.names = T, row.names = F)

##MC,MB (MES) vs PN GLANCE
Sig <- read_xls('./DGE_GLANCE_MC-MB_vs_PN_CSC.xls')
Sig <- as.data.frame(Sig)

##Ordina variabili per Adjusted p-Value e controlla non ci siano significatività
##SeqC NA o 0 e tieni solo i +1 (significativamente up) e -1
##(significativamente down)
Sig <- arrange(Sig, Sig$padj)
table(Sig$significance_SEQC)
Sig_rank <- filter(Sig, Sig$significance_SEQC!=0)

##Seleziona solo le colonne che ti interessano (ID, Adj_pVal, LogFC)
Sig_rank <- Sig_rank[,c(1,2,4)]

##Seleziona solo le entries upregolate in PN (essendo log2, scelgo soglia per
##livelli almeno 10 volte -> log2(10) = 3.32, adj_pVal<=0.005)
Sig_PN <- filter(Sig_rank, Sig_rank$log2FoldChange<=(-3.32))
Sig_PN <- filter(Sig_PN, Sig_PN$padj<=0.005)
nrow(Sig_PN)

```

```

##Seleziona solo le entries upregolate in MES (essendo log2, scelgo soglia per
##livelli almeno 10 volte -> log2(10) = 3.32, adj_pVal<=0.005)
Sig_MES <- filter(Sig_rank, Sig_rank$log2FoldChange>=(3.32))
Sig_MES <- filter(Sig_MES, Sig_MES$padj<=0.005)
nrow(Sig_MES)

##Salva file geni per firma
write.table(Sig_PN, file = './PN_vs_MC-MB_signature_logFC3.32.xls', sep = '\t',
            quote = F, col.names = T, row.names = F)
write.table(Sig_MES, file = './MC-MB_signature_logFC3.32.xls', sep = '\t',
            quote = F, col.names = T, row.names = F)

##RNA-Seq Unsupervised Clustering A (MES) vs B (non-MES/other MES) GLANCE
Sig <- read_xls('./DGE_GLANCE_RNASeq_UC_A_vs_B.xls')
Sig <- as.data.frame(Sig)

##Ordina variabili per Adjusted p-Value e controlla non ci siano significatività
##SeqC NA o 0 e tieni solo i +1 (significativamente up) e -1
##(significativamente down)
Sig <- arrange(Sig, Sig$padj)
table(Sig$significance_SEQC)
Sig_rank <- filter(Sig, Sig$significance_SEQC!=0)

##Seleziona solo le colonne che ti interessano (ID, Adj_pVal, LogFC)
Sig_rank <- Sig_rank[,c(1,2,4)]

##Seleziona solo le entries upregolate in B (essendo log2, scelgo soglia per
##livelli almeno 10 volte -> log2(50) = 5.64, adj_pVal<=0.005)
Sig_PN <- filter(Sig_rank, Sig_rank$log2FoldChange<=(-5.64))
Sig_PN <- filter(Sig_PN, Sig_PN$padj<=0.005)
nrow(Sig_PN)

##Seleziona solo le entries upregolate in A (essendo log2, scelgo soglia per
##livelli almeno 10 volte -> log2(50) = 5.64, adj_pVal<=0.005)

```

```

Sig_MES <- filter(Sig_rank, Sig_rank$log2FoldChange>=(5.64))
Sig_MES <- filter(Sig_MES, Sig_MES$padj<=0.005)
nrow(Sig_MES)

##Salva file geni per firma
write.table(Sig_PN, file = './B(other)_vs_A(MES)_signature_logFC5.64.xls',
            sep = '\t', quote = F, col.names = T, row.names = F)
write.table(Sig_MES, file = './A(MES)_signature_logFC5.64.xls', sep = '\t',
            quote = F, col.names = T, row.names = F)

```

Upregulated_genes.R

```

getwd()
setwd('./Desktop/Lab/4- RNASeq/GLANCE - RNASeq/Upregulated for Enrichr/')

library(readxl)
library(dplyr)

DEG <- read_xls('./XXXXXX.xls')
DEG <- as.data.frame(DEG)
DEG <- filter(DEG, padj < 0.05)
DEG <- DEG[,c(1,2,4)]
DEG_up <- filter(DEG, log2FoldChange > 0)
DEG_down <- filter(DEG, log2FoldChange < 0)

write.table(DEG_up, file = './PN_vs_MES_upregulated.xls', sep = '\t',
            quote = F, col.names = T, row.names = F)
write.table(DEG_down, file = './MES_vs_PN_upregulated.xls', sep = '\t',
            quote = F, col.names = T, row.names = F)

```

3_Differential_metabolites_MannWhitney.R

```
library(dplyr)
library(readxl)
library(xlsx)
library(omu)
library(pheatmap)
library(RColorBrewer)
library(seriation)
library(dendextend)

#read database
metabo_raw <- read_xlsx('./1_Db/SRSI-01-20MD DATA TABLES.XLSX',
                      sheet = 'Mass_extracted-norm Data')
metabo_raw <- as.data.frame(metabo_raw)

rownames(metabo_raw) <- metabo_raw$PARENT_SAMPLE_NAME
metabo_raw <- metabo_raw[,-1]

#most of the data are between 0 and 1; therefore, apply transformation by
multiplying
# for 1000 and then round to nearest integer
i <- 1
j <- 1
for (i in 1:nrow(metabo_raw)) {
  for (j in 1:ncol(metabo_raw)) {
    metabo_raw[i,j] <- round(metabo_raw[i,j]*1000)
  }
}

i <- 1
for (i in 1:ncol(metabo_raw)) {
  metabo_raw[,i] <- as.integer(metabo_raw[,i])
}
```

```

#read file for correspondence to samples
metabo_metadata <- read_xlsx('./1_Db/SRSI-01-20MD DATA TABLES.XLSX',
                             sheet = 'Sample Meta Data')
metabo_metadata <- as.data.frame(metabo_metadata)

#read file for correspondence to metabolites
metabo_ID <- read_xlsx('./1_Db/SRSI-01-20MD DATA TABLES.XLSX',
                       sheet = 'Chemical Annotation')
metabo_ID <- as.data.frame(metabo_ID)

#prepare for sample and metabolite name substitution
metabo_samples <- metabo_metadata[,c(1,9)]
rownames(metabo_samples) <- metabo_samples$PARENT_SAMPLE_NAME

metabo_metabolites <- metabo_ID[,c(1,11,16,15)]
rownames(metabo_metabolites) <- metabo_metabolites$CHEM_ID

#substitute with corresponding names
i <- 1
for (i in (1:nrow(metabo_raw))) {
  rownames(metabo_raw)[i] <- ifelse(rownames(metabo_raw)[i] %in%
rownames(metabo_samples)[i],
                                metabo_samples[i,2], rownames(metabo_raw)[i])
}

#transpose dataset, so that metabolites are in rows and samples in columns
metabo_raw <- t(metabo_raw)
metabo_raw <- as.data.frame(metabo_raw)
colnames(metabo_raw) <- gsub(' ','_',colnames(metabo_raw))

#insert columns for metabolite name and KEGG and reorder
metabo_raw$Metabolite <- NA
metabo_raw$KEGG <- NA
metabo_raw$HMDB <- NA
metabo_raw <- metabo_raw[,c(37:39,1:36)]

```



```

#insert corresponding metabolite name, KEGG ID and HMDB ID
i <- 1
for (i in (1:nrow(metabo_raw))) {
  metabo_raw[i,1] <- ifelse(rownames(metabo_raw)[i] %in%
rownames(metabo_metabolites)[i],
  metabo_metabolites[i,2])
}

i <- 1
for (i in (1:nrow(metabo_raw))) {
  metabo_raw[i,2] <- ifelse(rownames(metabo_raw)[i] %in%
rownames(metabo_metabolites)[i],
  metabo_metabolites[i,3])
}

i <- 1
for (i in (1:nrow(metabo_raw))) {
  metabo_raw[i,3] <- ifelse(rownames(metabo_raw)[i] %in%
rownames(metabo_metabolites)[i],
  metabo_metabolites[i,4])
}

i <- 1
for (i in (1:nrow(metabo_raw))) {rownames(metabo_raw)[i] <- i}

#save database as csv to reimport it with omu::read.metabo (as dataframe and
cpd)
write.csv(metabo_raw, file = './1_Db/metabo_counts.csv', row.names = FALSE)

#reimport db with read.metabo
metabo_raw <- read_metabo('./1_Db/metabo_counts.csv')
metabo_raw$Metabolite <- as.factor(metabo_raw$Metabolite)
metabo_raw$KEGG <- as.factor(metabo_raw$KEGG)

```

```

#create annotation metadata
subtypes <- c('PN', 'MES', 'HN', 'HW', 'MYC')
metabo_samples$Group <- 'CTRL'

i <- 1
j <- 1
for (i in (1:nrow(metabo_samples))) {
  for (j in (1:length(subtypes))) {
    metabo_samples$Group[i] <- ifelse(grepl(subtypes[j],
metabo_samples$CLIENT_SAMPLE_ID[i]),
                                     subtypes[j], metabo_samples$Group[i])
  }
}

sample_metadata <- metabo_samples[,c(2,3)]
colnames(sample_metadata) <- c('Sample','Group')

sample_metadata$Sample <- gsub(' ','_',sample_metadata$Sample)

sample_metadata$Group <- as.factor(sample_metadata$Group)
sample_metadata$Sample <- as.factor(sample_metadata$Sample)

#keep only samples you need for differential metabolite enrichment
tokeep <- c('Metabolite', 'KEGG', 'HMDB', 'PN_0605_n1', 'PN_0605_n2',
'PN_0605_n3',
          'PN_0512_n1', 'PN_0512_n2', 'PN_0512_n3', 'PN_0801_n1', 'PN_0801_n2',
          'MES_U87_n1', 'MES_1312_n1',
'MES_1312_n2','MES_1312_n3','MES_1312_n4',
          'MES_ZL_n1', 'MES_ZL_n2')

sample_metadata <- filter(sample_metadata, sample_metadata$Sample %in%
tokeep)

metabo_raw <- metabo_raw[,tokeep]

```

```

i <- 1
for (i in (1:nrow(sample_metadata))) {rownames(sample_metadata)[i] <- i}

#perform differential metabolite enrichment (Welch and Student's t-test gives error
#for too low variance in group, try with non-parametric Mann Whitney U test)

metabo_stats <- omu_summary(count_data = metabo_raw, metadata =
sample_metadata,
                           numerator = 'MES', denominator = 'PN',
                           response_variable = 'Metabolite', Factor = 'Group',
                           log_transform = TRUE, p_adjust = 'BH', test_type = 'mwu')

metabo_res <- metabo_stats[,c(1,13,14,6,7,9,11,12,15:29)]

#write csv file
write.csv(metabo_res, file = './1_Db/Metabo_diff_MES_vs_PN_all', row.names =
FALSE)

#read csv file
metabo_res_01 <- read.csv('./1_Db/Metabo_diff_MES_vs_PN_all')

#write xlsx file
write.xlsx(metabo_res_01, file = './1_Db/Metabo_diff_MES_vs_PN.xlsx', row.names
= FALSE)

#keep only metabolites with pval < 0.1
metabo_res_01 <- filter(metabo_res_01, metabo_res_01$pval<0.1)

#write xlsx file
write.xlsx(metabo_res_01, file = './1_Db/Metabo_diff_MES_vs_PN_signif<0.1.xlsx',
row.names = FALSE)

```

5_Get_significant_sig_for_DrugRepo.R

```
#Classical vs Healthy
res_CL_Healthy_df <- as.data.frame(res_CL_Healthy)
res_CL_Healthy_DrugRepo <- tibble::rownames_to_column(res_CL_Healthy_df)
res_CL_Healthy_DrugRepo <- filter(res_CL_Healthy_DrugRepo, padj <= 0.05)
res_CL_Healthy_DrugRepo <- res_CL_Healthy_DrugRepo[,c(1,3)]
base::colnames(res_CL_Healthy_DrugRepo) <- c('Genes','log2FoldChange')
write.table(res_CL_Healthy_DrugRepo, 'CL_Healthy_sig.txt', row.names = F, sep =
'\t')

#PN vs Healthy
res_PN_Healthy_df <- as.data.frame(res_PN_Healthy)
res_PN_Healthy_DrugRepo <- tibble::rownames_to_column(res_PN_Healthy_df)
res_PN_Healthy_DrugRepo <- filter(res_PN_Healthy_DrugRepo, padj <= 0.05)
res_PN_Healthy_DrugRepo <- res_PN_Healthy_DrugRepo[,c(1,3)]
base::colnames(res_PN_Healthy_DrugRepo) <- c('Genes','log2FoldChange')
write.table(res_PN_Healthy_DrugRepo, 'PN_Healthy_sig.txt', row.names = F, sep =
'\t')

#MES vs Healthy
res_MES_Healthy_df <- as.data.frame(res_MES_Healthy)
res_MES_Healthy_DrugRepo <- tibble::rownames_to_column(res_MES_Healthy_df)
res_MES_Healthy_DrugRepo <- filter(res_MES_Healthy_DrugRepo, padj <= 0.05)
res_MES_Healthy_DrugRepo <- res_MES_Healthy_DrugRepo[,c(1,3)]
base::colnames(res_MES_Healthy_DrugRepo) <- c('Genes','log2FoldChange')
write.table(res_MES_Healthy_DrugRepo, 'MES_Healthy_sig.txt', row.names = F,
sep = '\t')

#MES vs PN
res_MES_PN_df <- as.data.frame(res_MES_PN)
res_MES_PN_DrugRepo <- tibble::rownames_to_column(res_MES_PN_df)
res_MES_PN_DrugRepo <- filter(res_MES_PN_DrugRepo, padj <= 0.05)
res_MES_PN_DrugRepo <- res_MES_PN_DrugRepo[,c(1,3)]
base::colnames(res_MES_PN_DrugRepo) <- c('Genes','log2FoldChange')
write.table(res_MES_PN_DrugRepo, 'MES_PN_sig.txt', row.names = F, sep = '\t')
```

6b_Get_Genes_for_CMAP.R

```
library(dplyr)
```

```
#read L1000 genes
```

```
L1000 <- read.delim('./4_Genes_for_CMAP/L1000.txt')
```

```
L1000 <- L1000$pr_gene_symbol
```

```
#Up CL vs Healthy
```

```
top_up_CL <- as.data.frame(res_CL_Healthy)
```

```
top_up_CL <- tibble::rownames_to_column(top_up_CL)
```

```
top_up_CL <- filter(top_up_CL, padj <= 0.05)
```

```
top_up_CL <- filter(top_up_CL, log2FoldChange > 0)
```

```
top_up_CL <- arrange(top_up_CL, padj)
```

```
tokeep <- top_up_CL$rowname %in% L1000
```

```
top_up_CL <- top_up_CL[tokeep,c(1,3)]
```

```
#if more than 150
```

```
top_up_CL <- top_up_CL[c(1:150),]
```

```
base::colnames(top_up_CL) <- c('Genes','log2FoldChange')
```

```
write.table(top_up_CL, 'CL_Healthy_UP_150_L1000.txt', row.names = F, sep = '\t')
```

```
#Down CL vs Healthy
```

```
top_down_CL <- as.data.frame(res_CL_Healthy)
```

```
top_down_CL <- tibble::rownames_to_column(top_down_CL)
```

```
top_down_CL <- filter(top_down_CL, padj <= 0.05)
```

```
top_down_CL <- filter(top_down_CL, log2FoldChange < 0)
```

```
top_down_CL <- arrange(top_down_CL, padj)
```

```
tokeep <- top_down_CL$rowname %in% L1000
```

```
top_down_CL <- top_down_CL[tokeep,c(1,3)]
```

```
#if more than 150
```

```
top_down_CL <- top_down_CL[c(1:150),]
```

```
base::colnames(top_down_CL) <- c('Genes','log2FoldChange')
```

```
write.table(top_down_CL, 'CL_Healthy_DOWN_150_L1000.txt', row.names = F, sep = '\t')
```

```
#Up MES vs Healthy
```

```
top_up_MES <- as.data.frame(res_MES_Healthy)
```

```

top_up_MES <- tibble::rownames_to_column(top_up_MES)
top_up_MES <- filter(top_up_MES, padj <= 0.05)
top_up_MES <- filter(top_up_MES, log2FoldChange > 0)
top_up_MES <- arrange(top_up_MES, padj)
tokeep <- top_up_MES$rowname %in% L1000
top_up_MES <- top_up_MES[tokeep,c(1,3)]
#if more than 150
top_up_MES <- top_up_MES[c(1:150),]
base::colnames(top_up_MES) <- c('Genes','log2FoldChange')
write.table(top_up_MES, 'MES_Healthy_UP_150_L1000.txt', row.names = F, sep =
'\t')

```

```
#Down MES vs Healthy
```

```

top_down_MES <- as.data.frame(res_MES_Healthy)
top_down_MES <- tibble::rownames_to_column(top_down_MES)
top_down_MES <- filter(top_down_MES, padj <= 0.05)
top_down_MES <- filter(top_down_MES, log2FoldChange < 0)
top_down_MES <- arrange(top_down_MES, padj)
tokeep <- top_down_MES$rowname %in% L1000
top_down_MES <- top_down_MES[tokeep,c(1,3)]
#if more than 150
top_down_MES <- top_down_MES[c(1:150),]
base::colnames(top_down_MES) <- c('Genes','log2FoldChange')
write.table(top_down_MES, 'MES_Healthy_DOWN_150_L1000.txt', row.names = F,
sep = '\t')

```

```
#Up PN vs Healthy
```

```

top_up_PN <- as.data.frame(res_PN_Healthy)
top_up_PN <- tibble::rownames_to_column(top_up_PN)
top_up_PN <- filter(top_up_PN, padj <= 0.05)
top_up_PN <- filter(top_up_PN, log2FoldChange > 0)
top_up_PN <- arrange(top_up_PN, padj)
tokeep <- top_up_PN$rowname %in% L1000
top_up_PN <- top_up_PN[tokeep,c(1,3)]
#if more than 150
top_up_PN <- top_up_PN[c(1:150),]

```

```
base::colnames(top_up_PN) <- c('Genes','log2FoldChange')
write.table(top_up_PN, 'PN_Healthy_UP_150_L1000.txt', row.names = F, sep =
'\t')
```

```
#Down PN vs Healthy
```

```
top_down_PN <- as.data.frame(res_PN_Healthy)
top_down_PN <- tibble::rownames_to_column(top_down_PN)
top_down_PN <- filter(top_down_PN, padj <= 0.05)
top_down_PN <- filter(top_down_PN, log2FoldChange < 0)
top_down_PN <- arrange(top_down_PN, padj)
tokeep <- top_down_PN$rowname %in% L1000
top_down_PN <- top_down_PN[tokeep,c(1,3)]
#if more than 150
top_down_PN <- top_down_PN[c(1:150),]
base::colnames(top_down_PN) <- c('Genes','log2FoldChange')
write.table(top_down_PN, 'PN_Healthy_DOWN_150_L1000.txt', row.names = F,
sep = '\t')
```

```
#Up MES vs PN
```

```
top_up_MES_PN <- as.data.frame(res_MES_PN)
top_up_MES_PN <- tibble::rownames_to_column(top_up_MES_PN)
top_up_MES_PN <- filter(top_up_MES_PN, padj <= 0.05)
top_up_MES_PN <- filter(top_up_MES_PN, log2FoldChange > 0)
top_up_MES_PN <- arrange(top_up_MES_PN, padj)
tokeep <- top_up_MES_PN$rowname %in% L1000
top_up_MES_PN <- top_up_MES_PN[tokeep,c(1,3)]
#if more than 150
top_up_MES_PN <- top_up_MES_PN[c(1:150),]
base::colnames(top_up_MES_PN) <- c('Genes','log2FoldChange')
write.table(top_up_MES_PN, 'MES_PN_UP_150_L1000.txt', row.names = F, sep =
'\t')
```

```
#Down MES vs PN
```

```
top_down_MES_PN <- as.data.frame(res_MES_PN)
top_down_MES_PN <- tibble::rownames_to_column(top_down_MES_PN)
top_down_MES_PN <- filter(top_down_MES_PN, padj <= 0.05)
```

```
top_down_MES_PN <- filter(top_down_MES_PN, log2FoldChange < 0)
top_down_MES_PN <- arrange(top_down_MES_PN, padj)
tokeep <- top_down_MES_PN$rowname %in% L1000
top_down_MES_PN <- top_down_MES_PN[tokeep,c(1,3)]
#if more than 150
top_down_MES_PN <- top_down_MES_PN[c(1:150),]
base::colnames(top_down_MES_PN) <- c('Genes','log2FoldChange')
write.table(top_down_MES_PN, 'MES_PN_DOWN_150_1000.txt', row.names = F,
sep = '\t')
```