Spatial reference and alternative proteome analysis of glioblastoma reveals molecular signatures and associates survival with specific markers

Marie Duhamel (marie.duhamel@univ-lille.fr)
Univ.Lille, Inserm, CHU Lille, U1192, Laboratoire Protéomique, Réponse Inflammatoire et Spectrométrie de Masse (PRISM), F-59000 Lille, France https://orcid.org/0000-0002-4006-5605

Lauranne Drelich
Univ.Lille, Inserm, CHU Lille, U1192, Laboratoire Protéomique, Réponse Inflammatoire et Spectrométrie de Masse (PRISM), F-59000 Lille, France

Maxence Wisztorski
Universite Lille, U1192 Inserm https://orcid.org/0000-0003-1320-075X

Soulaimane Aboulouard
University of Lille https://orcid.org/0000-0002-2045-4785

Jean-pascal Gimeno
Univ.Lille, Inserm, CHU Lille, U1192, Laboratoire Protéomique, Réponse Inflammatoire et Spectrométrie de Masse (PRISM), F-59000 Lille

Nina Ogrinc
Univ. Lille, Inserm, CHU Lille, U1192 - Protéomique Réponse Inflammatoire Spectrométrie de Masse - PRISM https://orcid.org/0000-0002-0773-0095

Patrick Devos
Univ. Lille, CHU Lille, ULR 2694 - METRICS : Évaluation des technologies de santé et des pratiques médicales, F-59000 Lille https://orcid.org/0000-0001-7803-9552

Tristan Cardon
Universite Lille, U1192 Inserm https://orcid.org/0000-0003-1751-0528

Michael Weller
University Hospital Zurich https://orcid.org/0000-0002-1748-174X

Fabienne ESCANDE
CHRU de Lille

Fahed Zairi
CHU Lille, Service de neurochirurgie, F-59000 Lille

Claude-Alain Maurage
Univ.Lille, Inserm, CHU Lille, U1172 - LilNCog - Lille Neuroscience & Cognition, F-59000 Lille

Emilie Le Rhun

Univ.Lille, Inserm, CHU Lille, U1192, Laboratoire Protéomique, Réponse Inflammatoire et Spectrométrie de Masse (PRISM), F-59000 Lille

Isabelle Fournier

Unversité de Lille, INserm, CHU Lille https://orcid.org/0000-0003-1096-5044

Michel salzet

Unversité de Lille, INserm, CHU Lille https://orcid.org/0000-0003-4318-0817

Article

Keywords: glioblastoma, mass spectrometry imaging, spatially-resolved mass spectrometry, 57 prognosis, SpiderMass technology, signaling pathways, alternative proteins

Posted Date: April 26th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1572039/v1

Version of Record: A version of this preprint was published at Nature Communications on November 4th, 2022. See the published version at https://doi.org/10.1038/s41467-022-34208-6.

Spatial reference and alternative proteome analysis of glioblastoma reveals molecular signatures and associates survival with specific markers

Marie Duhamel[1 §], Lauranne Drelich[1 §], Maxence Wisztorski [1§], Soulaimane Aboulouard[1], Jean-Pascal Gimeno[1], Nina Ogrinc[1], Patrick Devos[2], Tristan Cardon[1], Michael Weller[5], Fabienne Escande[3], Fahed Zairi[4], Claude-Alain Maurage[3], Emilie Le Rhun[1,4,5*], Isabelle Fournier[1*]and Michel Salzet[1 *]

[1]Univ.Lille, Inserm, CHU Lille, U1192, Laboratoire Protéomique, Réponse Inflammatoire et Spectrométrie de Masse (PRISM), F-59000 Lille, France

[2] Univ. Lille, CHU Lille, ULR 2694 - METRICS : Évaluation des technologies de santé et des pratiques médicales, F-59000 Lille, France

[3] CHU Lille, Service de biochimie et biologie moléculaire, CHU Lille, F-59000 Lille France

[4] CHU Lille, Service de neurochirurgie, F-59000 Lille, France

[5]Department of Neurology & Clinical Neuroscience Center, University Hospital and University of Zurich, Zurich, Switzerland

[§]These authors share first authorship

Corresponding authors:

* To whom correspondence should be addressed. Michel Salzet, Isabelle Fournier, Emilie Le Rhun, Phone: +33 320 43 41 94; Fax: +33 320 43 40 54; Email: michel.salzet@univ-lille.fr , isabelle.fournier@univ-lille.fr, emilie.lerhun@usz.ch

SUMMARY

Molecular heterogeneity is a key feature of glioblastoma pathology impeding patient's stratification and leading to high discrepancies between patients mean survivals. We performed a spatial proteomics analysis on a cohort of 96 glioblastoma patients with survival varying from few months to more than 4 years. 46 tumors were analyzed by spatially-resolved high resolution mass spectrometry proteomics. Integrative analysis of protein expression and clinical information allowed us to identify three molecular regions associated with immune, neurogenesis and tumorigenesis signatures. Several of these molecular signatures can be enriched within the same tumor sample leading to high intra-tumoral heterogeneity. Nevertheless, a set of proteins was found statistically significant based on patient's survival times, 10 of which stem from alternative AltORF or non-coding RNA. From these proteins, 5 were selected as survival markers. Classification of patients based on the expression of these 5 proteins leads to a clear difference in survival. The expression of these 5 proteins was validated by immunofluorescence on an external cohort of 50 glioblastoma patients, with a similar correlation with their survival.

Taken together, our work has enabled the characterization of new molecular regions within glioblastoma tissues based on protein expression which can help to guide glioblastoma prognosis and to improve the current glioblastoma classification.

Key Words: glioblastoma, mass spectrometry imaging, spatially-resolved mass spectrometry, prognosis, SpiderMass technology, signaling pathways, alternative proteins

Significance:

Glioblastoma are very heterogeneous tumors with survival times usually inferior to 20 months. We conducted a spatial proteomics analysis of glioblastoma to stratify glioblastoma based on their molecular signatures. Three molecular signatures were identified across tissues *i.e.* neurogenesis, immune and RNA processing and metabolism signatures. We showed that several of these signatures can be enriched within the same tumor sample, preventing to classify glioblastoma based on them and demonstrating high intra-tumoral heterogeneity. We correlated these results with the TCGA data. Despite a high heterogeneity, we nevertheless identified 5 specific prognostic proteins with differential expression according to the survival length of patients which were validated on an external glioblastoma cohort. These markers can help to stratify glioblastoma patients into homogeneous subgroups.

Highlights

- A novel stratification of glioblastoma based on mass spectrometry proteomics has been established.
- Three tumor regions with different molecular features were identified.
- A single tumor can be represented by more than one molecular region.
- 5 prognosis markers associated with either long or short survival were validated on an external cohort of glioblastoma patients.
- This new classification may improve prognosis.

Abbreviations

A: astrocytoma

ACN: acetonitrile

ATRX: alpha-thalassemia/mental retardation syndrome X-linked

CDKN2A: cyclin-dependent kinase inhibitor 2A

CGH: comparative genomic hybridization

DNA: deoxyribonucleic acid

EGFR: epidermal growth factor receptor

F: female

FDR: false discovery rate

FFPE: formalin-fixed paraffin-embedded

gCIMP: CpG island methylator phenotype

HCD: Higher energy Collision Dissociation

HES: Hematoxylin Eosin Safran

IDH: Isocitrate dehydrogenase

LC: Liquid Chromatography

H3F3A: H3 Histone, Family 3A

LESA: Liquid Extraction Surface Analysis

LFQ: Label-Free Quantification

M: male

MALDI: Matrix-Assisted Laser Desorption/Ionization

MALDI MSI: MALDI Mass Spectrometry Imaging

TOF: Time-Of-Flight

MeOH: Methanol

MGMT: O^6-methylguanine-DNA methyltransferase

MRI: Magnetic Resonance Imaging

MSI: Mass Spectrometry Imaging

O: oligodendroglioma

PSM: peptide spectrum matches

PTEN: phosphatase and tensin homolog deleted on chromosome 10

ROI: Region of interest

RNA: Ribonucleic acid

SNEA: Subnetwork Enrichment Analysis

TERT: telomerase reverse transcriptase

TFA: Trifluoroacetic acid

TP53: tumor protein p53

WHO: World Health Organization

Introduction

Glioblastoma represents the main malignant primary brain tumor (Ostrom et al., 2021). The prognosis is poor with a median survival estimated at 16 months in clinical studies (Chinot et al., 2014; Gilbert et al., 2014; Stupp et al., 2009; Stupp et al., 2017; Weathers and Gilbert, 2014; Weller et al., 2021) and around 12 months in contemporary population-based studies(Gramatzki et al., 2018). Approximately 5% of patients survive more than 5 years (Ostrom et al., 2021). Favorable therapy-independent prognostic factors include younger age and higher neurological performance status at diagnosis. Furthermore, low postoperative residual tumor volume has been associated with improved outcome. In a cohort of 232 patients with centrally confirmed glioblastoma who survived at least 5 years, the median age at diagnosis was 52 years (range 21-77 years) and most patients had a gross total resection initially (Weller et al., 2019).

Morphological criteria for the diagnosis of glioblastoma according to the World Health Organization (WHO) central nervous system tumor classification of 2021 (Louis et al., 2021). include mitotic activity, anaplastic nuclear features, microvascular proliferation and necrosis. Morphological variants include giant cell glioblastoma, gliosarcoma and epithelioid glioblastoma. Isocitrate dehydrogenase (IDH) 1 or 2 mutations now exclude the diagnosis of glioblastoma. Tumors with morphological features of glioblastoma which exhibit IDH mutations are now referred to as Astrocytoma, IDH-mutant, WHO grade 4 (Brat et al., 2020). Conversely, IDH wildtype tumors that do not fulfill morphological WHO grade 4 criteria are still diagnosed as glioblastoma if they exhibit at least one of the following alterations: EGFR amplification, a +7/-10 genotype or TERT promoter mutation (Brat et al., 2020). Standard treatment of glioblastoma includes maximum safe resection followed by radiotherapy with concomitant and maintenance temozolomide (Weller et al., 2021).

Efforts to further subclassify glioblastoma have been restricted to the genomic, transcriptomic and epigenetic levels. In 2008, the Cancer Genome Atlas (TCGA) group delineated three main signaling pathways affected by genetic alterations in glioblastoma, receptor tyrosine kinase/RAS/PI3K, p53 and RB (Brennan et al., 2013). Genome methylation profiling in adult patients with IDH wildtype glioblastoma allowed the definition of three epigenetic subtypes, (i) receptor tyrosine kinase (RTK) I often with PDGFR amplification, (ii) RTK II or classical often with EGFR amplification, CDKN2A/B deletion, and PTEN mutation, and (iii) mesenchymal (Sturm et al., 2012). Any clinical relevance of the methylation classes in glioblastoma remains controversial. The DNA methylation-based classification of CNS tumors has meanwhile evolved to a comprehensive machine-learning approach (Capper et al., 2018) that has shaped the new WHO classification (Louis et al. 2021), resulting also in the delineation of further rare methylation classes of glioblastoma. Prior to the introduction of methylation

profiling, a classification based on transcriptional profiling revealed four subtypes of glioblastoma: proneural, neural, classic and mesenchymal (Verhaak et al., 2010a). The neural subtype is no longer maintained since it may reflect contamination by normal brain tissue, but it has become apparent that transcriptomic profiles are less homogeneous and stable than genome or methylome classifiers. Despite these efforts, these approached have found limited clinical application and only a few biomarkers are being used in clinic. Proteomic approaches have been less frequently explored, although they can identify and quantify the final product of altered genomics and transcriptomics and may better characterize the activation of specific pathways (Deighton et al., 2010; Dilillo et al., 2017,Goplen, 2010 #14255; Kalinina et al., 2011). Proteomic analyses of gliomas have been performed to identify proteomic differences between grades and genomic alterations (Djuric et al., 2019). More recently, proteogenomic approaches have been used to stratify glioblastoma patients demonstrating a stronger association of protein expression with patient survival compared to RNA transcripts (Yanovich-Arad et al., 2020). Another study has performed a multi-omics strategy to investigate glioblastoma biology (Wang et al., 2021). However, glioblastoma are highly heterogeneous tumors and a spatially-resolved proteomics approach may bring new insights in glioblastoma biology to improve their stratification. The determination of specific proteomic signatures could help to improve the distinction between the different glioblastoma subtypes and to guidec management.

In the current study, we present a spatially-resolved proteomic approach to characterize glioblastoma. We analyzed a cohort of 96 glioblastoma patients of varying survival. A spatially resolved proteomic approach guided by mass spectrometry imaging enabled us to stratify patients into three molecular groups. Our strategy provides new insights into intertumoral and intratumoral heterogeneities by considering the glioblastoma microenvironment which is of prime importance in tumor development. Based on our proteomic study, 5 prognostic protein markers were identified. The expression of these 5 proteins are indicators of short and long survival and can therefore help to stratify patients. We validated our results on an external cohort of 50 glioblastoma patients by immunofluorescence. Altogether, these results highlight the potential of spatially resolved proteomics to decipher glioblastoma molecular heterogeneity and to identify markers associated with survival.

Materials & Methods

Patient samples and consent

Tumors from 96 patients were included in the study**.** 46 patients with newly diagnosed glioblastoma were prospectively enrolled between September 2014 and November 2018 at Lille University Hospital, France. Patients were adult, had no medical history of other cancers or previous cancer treatment, no known genetic disease potentially leading to cancer and no neurodegenerative disease. Tumors samples were processed within 2 hours after sample extraction in the surgery room to limit the risk of degradation of proteins. Approval was obtained from the research ethics committee (ID-RCB 2014-A00185-42) before initiation of the study. The study adhered to the principles of the Declaration of Helsinki and the Guidelines for Good Clinical Practice and is registered at NCT02473484. Informed consent was obtained from patients. For the validation cohort used for IF analysis, 50 formalin-fixed paraffin-embedded (FFPE) glioblastoma tissues were obtained from the Pathology department of Lille Hospital, France. IDH mutant tumors were excluded from the study.

Deoxyribonucleic acid (DNA) extraction and quantification

Molecular analyses were performed on DNA extracted FFPE tissues. The following tests were performed: Comparative genomic hybridization (CGH) array and assessment of O^6-methylguanine-DNA methyltransferase (MGMT) promoter methylation status. All tissues used for DNA extraction were histologically evaluated to determine the tumor cell content. Analyses were performed on all tissue samples. Samples with less than 40% of tumor cells content were considered as not interpretable when no molecular abnormalities were found. DNA extraction from FFPE was performed using the kit QIAamp DNA FFPE Tissue (Qiagen). CGH profiles were determined using a SurePrint G3 Human CGH Microarray Kit, 8x60K (Aligent) and the CytoGenomics v2.7 software. The limit of resolution was 1 Mb. Presence of 1p/19q codeletion, gain of chromosome 7, loss of chromosome 10, EGFR amplification and homozygous deletion of the cyclin-dependent kinase inhibitor 2A (CDKN2A) gene was systematically evaluated. The MGMT promoter methylation status (CpGs 74-78) was determined after bisulfite treatment by pyrosequencing on a PyroMark Q96 with kit MGMT PyroMark (Qiagen). The presence of a methylation was score positive when a minimum of 8% of methylation was observed.

MALDI mass spectrometry imaging (MALDI MSI)

A Leica CM1510S cryostat (Leica Microsystems, Nanterre, France) was used to cut twelve micrometer sections in order to perform the MALDI MSI analysis (Fournier et al., 2003; Lemaire

et al., 2007; Lemaire et al., 2006a; Lemaire et al., 2006b; Wisztorski et al., 2016). These tissue sections were deposited on ITO-coated glass slides (LaserBio Labs, Valbonne, France) and vacuum-dried during 15 min. Tissue sections were then soaked in different solutions to remove abundant lipids: (1) 1 min in 70% ethanol, (2) 1 min in 100% ethanol, (3) 1 min in acetone and (4) 30 s in chloroform with concomitant drying between washings. An electrospray nebulizer connected to a syringe pump (flow rate 180 nL/min) was used to uniformly spray a trypsin solution (60 µg/mL in NH4HCO3 50 mM) on the tissue surface for 15 min. ImagePrep (Bruker Daltonics, Bremen, Germany) was used as an incubation chamber by microspraying water heated to 37 °C for 2 h (60 cycles with 2 s spraying, 180 s incubation and 60 s drying using the nitrogen flow). For optimal digestion, a constant humidity atmosphere was maintained inside the spray chamber by filling a small container with 95°C water. After digestion, HCCA/ANI (Lemaire et al., 2006) a solid ionic matrix was deposited using ImagePrep. Briefly, 36 µL of aniline were added to 5 mL of a solution of 10 mg/mL HCCA dissolved in ACN/0.1% TFA aqueous (7:3, v/v). A real-time control of the deposition is performed by monitoring scattered light to obtain a uniform layer of matrix. MALDI MSI experiments were done on an Ultraflex II MALDI-TOF/TOF instrument (Bruker) with a smartbeam II solid state laser. Mass spectra were acquired in positive reflector mode between 800–4000 m/z range. Recorded spectra were averaged from 400 laser shots per pixel acquired at 200Hz laser repletion rate and. with a 70 µm spatial resolution raster.

MALDI MSI data processing and analysis

The MALDI-MSI data were analyzed using SCiLS Lab software (SCiLS Lab 2019, SCiLS GmbH). Common processing methods for MALDI MSI were applied with a baseline removal using a convolution method and data were normalized using Total Ion Count (TIC) method (Klein et al., 2014; Trede et al., 2012). Then, the resulting pre-processing data were clustered to obtain a spatial segmentation using the bisecting k means algorithm (Alexandrov et al., 2010). Different spatial segmentations were performed. First, an individual segmentation was applied to each tissue separately. Then, the data from all tissues were clustered together to obtain a global segmentation. Briefly, the spatial segmentation consists of grouping all spectra according to their similarity using a clustering algorithm and all pixels of a same cluster are colour coded. To limit the pixel-to-pixel variability, edge-preserving image denoising was applied. Note that a color is arbitrary assigned to a cluster and that several disconnected regions can have the same color, i.e. the same molecular content. The results of segmentation are represented on a dendrogram resulting from a hierarchical clustering. The branches of the dendrogram were defined based on a distance calculation between each cluster. The selection of different branches of the dendrogram will give a segmentation map where regions of distinct

molecular composition were differentially color-coded. The individual segmentation provides information concerning the heterogeneity of the tissue section and the global segmentation is used to group patients with a similar molecular signature. For comparison, global segmentation was also performed using the Ward clustering method with IMAGEREVEAL MS Ver.1.1 (Shimadzu). The global spatial segmentation allowed to determine regions of interest (ROIs) which were then subjected to on-tissue microdigestion followed by microextraction for protein identifications.

SpiderMass analyses

The global design of the instrument setup has been described (Saudemont et al., 2018). Briefly, the system is composed of three parts including a laser system for micro-sampling of tissues set remotely, a transfer line allowing for transfer of the micro-sampled material to the third part, which is the mass spectrometer itself (Fatou et al., 2016). The first part is composed of a tunable wavelength OPO which is tunable from 2.8 µm to 3.1 µm (Radiant version 1.0.1, OPOTEK Inc., Carlsbad, CA, USA) pumped by a pulsed Nd:YAG laser (pulse duration: 5 ns, λ=1064 nm, Quantel, Les Ulis, France). A biocompatible laser fiber (450 µm inner diameter; length of 1 m; Infrared Fiber Systems, Silver Spring, CO, USA) is connected to the laser system output and a handpiece including a 4 cm focusing lens is attached to the end of the laser fiber. The handpiece with a 4 cm focusing lens allows the user to hold the system and screen the surface of raw tissues at a resolution of 400 µm. In these experiments the irradiation time was fixed to 10 sec at 4 mJ/pulse laser energy corresponding to a laser fluence of ~3 J/cm2. The laser energy was measured at the focal point of the focusing lens using a power meter (ThorLabs, Maisons-Laffitte, France). The second part of the system corresponds to a 3-meter length transfer line made from a Tygon ND 100-65 tubing (2.4 mm inner diameter, 4 mm outer diameter, Akron, Ohio, USA). The transfer line is attached on one side onto the laser hand piece at the end of the laser fiber and on its other side directly connected to the mass spectrometer (Xevo, Waters, Manchester, United Kingdom) from which the conventional electrospray source was removed and replaced by an atmospheric pressure interface (Fatou et al., 2016). Each acquisition was accompanied by a 150 µL/min isopropanol infusion. Spectral acquisition was performed both in positive and negative ion resolution mode with a scan time of 1 sec. Prior to SpiderMass analysis, the samples were taken out of the -20°C freezer and thawed to RT for 30 s. The spectral acquisition sequence was composed of 2 or 3 acquisitions using 1-sec irradiation periods. The ROI were selected using the morphological controls and acquired peptide MALDI-MSI data prior to each SpiderMass acquisition to ensure that each acquisition was performed on the same histological area (Ogrinc et al., 2019).

Classification model construction

For data analysis, all raw data files produced with the SpiderMass instrument were imported into the Abstract Model Builder (AMX v.0.9.2092.0) software. After importation, spectra were first pre-processed. The pre-processing steps include background subtraction, total ion count normalization, lockmass correction and re-binning to a 0.1 or 0.2 Da window. All processed MS spectra obtained from the 30 histologically validated samples were then used to build a principal component analysis and linear discriminant analysis (PCA-LDA) classification model (Ogrinc et al., 2019). The first step consisted of PCA to reduce data multidimensionality by generating features that explain most of the variance observed. These features were then subjected to supervised analysis using LDA by setting the classes that the model will be based upon. LDA attempts to classify the sample spectra and assess the model by cross validation. Cross-validation was carried out by either using the “20% out” or the “leave one patient out” methods. For the first method, 20% of MS spectra are randomly taken from the total spectra and the model is reconstructed from the remaining 80%. The remaining 20% of spectra are used to interrogate the reconstructed model. The permutation is automatically reiterated for 5 cycles before reporting the cross-validation results. For the second method, the spectra are grouped by patient and left out one by one; at each step the model without the patient is interrogated against this model.

Spatially-resolved proteomics

On-tissue digestion

A total of 122 ROIs were selected from MALDI-MSI. Spatially resolved microproteomics was performed on the predefined ROIs according to the previously published protocol (Quanico et al., 2013). Briefly, tissue sections of 20 µm thickness were cut and subjected to different washes to remove lipids. Then, on-tissue digestion is performed using a LysC-trypsin solution (40 µg/mL in Tris-HCl 50 mM, pH 8.0). This solution was deposited using a piezoelectric microspotter (CHIP-1000, Shimadzu, CO, Kyoto, Japan) on each ROIs with a total area of 1 mm^2 (4×4 spots of 200 µm. Enzyme droplet was maintained for a total of 2 h digestion. After enzyme deposition 0.1% TFA was spotted for 25 cycles with 100 pL on each spot/cycle.

Microextraction by liquid microjunction

After tissue microdigestion, the triptic peptides were extracted using an automated platform, the TriVersa Nanomate platform (Advion Biosciences Inc., Ithaca, NY, USA) with Liquid Extraction Surface Analysis (LESA) option (Quanico et al., 2013). Briefly, a volume of solvent was aspirated onto a tip and dispensed onto the digested region. The droplet formed was maintained between the tip and the tissue and then aspirated after 15 s. The recovery solution

is finally pooled in a low binding tube. Three extractions steps were performed per region using different solutions: (1) 0.1% TFA, (2) ACN/0.1% TFA (8:2, v/v), and (3) MeOH: 0.1% TFA (7:3, v/v). Two extraction cycles per point were performed to increase the amount of material collected.

NanoLC-MS & MS/MS analysis

Prior to MS analysis, the reconstituted samples were desalted using C-18 Ziptip (Millipore, Saint-Quentin-en-Yvelines, France), eluted with 80% ACN and vacuum-dried. The dried samples were resuspended in 0.1% FA aqueous/ACN (98:2, v/v). Peptides separation was performed by reverse phase chromatography, using a NanoAcquity UPLC system (Waters) coupled to a Q-Exactive Orbitrap mass spectrometer (Thermo Scientific) via a nano electrospray source. A pre-concentration column (nanoAcquity Symmetry C18, 5 µm, 180 µm x 20 mm) and an analytical column (nanoAcquity BEH C18, 1.7 µm, 75 µm x 250 mm) were used. A 2 h linear gradient of acetonitrile in 0.1% formic acid (5%-35%) was applied, at the flow rate of 300 nl/min. For MS and MS/MS analysis, a data dependent mode was defined to analyze the 10 most intense ions of MS analysis (Top 10). The MS analysis was performed with an m/z mass range between 300 to 1600, a resolution of 70,000 FWHM, an AGC of 3e6 ions and a maximum injection time of 120 ms. The MS/MS analysis was performed with an m/z mass range between 200 to 2000, an AGC of 5e4 ions, a maximum injection time of 60 ms and the resolution was set at 17,500 FWHM. To avoid any batch effect during the analysis, the extractions were chosen at random to create analysis sequences.

Data analysis

All MS data were searched with MaxQuant software (Cox and Mann, 2008; Tyanova et al., 2015) (Version 1.5.3.30) using Andromeda search engine (Cox et al., 2011) against the complete proteome for *Homo sapiens* (UniProt, release July 2018, 20 412 entries). Trypsin was selected as enzyme and two missed cleavages were allowed, with N-terminal acetylation and methionine oxidation as variable modifications. The mass accuracies were set to 6 ppm and 20 ppm respectively for MS and MS/MS spectra. False discovery rate (FDR) at the peptide spectrum matches (PSM) and protein levels was estimated using a decoy version of the previously defined databases (reverse construction) and set to 1%. A minimum of 2 peptides with at least one unique is necessary to complete the identification of a protein. The MaxLFQ algorithm (Cox et al., 2014) was used to performed label-free quantification of the proteins. The resulting file was analyzed using Perseus software (version 1.6.0.7). First, hits from the reverse database, proteins with only modified peptides and potential contaminants were removed. Statistical analyses were performed using ANOVA with a truncation value based on

"Benjamini Hocheberg FDR" of 5%. Three categorical annotation groups were used for the ANOVA, i.e. (1) the color group based on the three colors from Scils global segmentation of the 46 samples (Red; Yellow and Blue), (2) the patient groups which are determined by the main color present in each tumor sample (Groups A, B, C) and (3) the patients` survival time (patients with an OS > to the third quartile, patients with an OS between the first and the third quartile and patients with an OS < to the first quartile). Proteins significantly different were selected and normalized by a Z-score with matrix access by rows. For representation, a hierarchical clustering was performed using the Euclidean parameter for the distance calculation, and the average option for linkage in the rows and columns of the trees with a maximum of 300 clusters.

System biology analyses

An annotation analysis of gene ontology terms for the identified proteins were performed using PANTHER Classification System (version 14.1, http://www.pantherdb.org), FunRich (Version 3.1.3) (Pathan et al., 2017) and the STRING database (version 11.0, www.string-db.org) (Szklarczyk et al., 2019). Potential interaction network was then loaded into Cytoscape 3.7.2 with relative expression data using Idmapper (Otasek et al., 2019). The Reactome FI plugging was used to select a subnetwork of gene ontology terms and NCI database-associated disease-specific proteins. The relationships between the differentially expressed proteins among all conditions were also depicted based on the Ariadne ResNet database (Yuryev et al., 2009) using Elseviers' Pathway Studio (version 11.0, Elsevier). The subnetwork Enrichment Analysis (SNEA) algorithm was used to detect the statistically significant altered biological pathways in which the identified proteins are involved.

Human Pathology Atlas

The glioma data contained in the Human pathology atlas (Uhlen et al., 2017a) were used. Based on TCGA transcriptomics and antibody-based protein data from 153 patients, this database identified 268 potentially prognostic genes (201 unfavorable and 67 favorable prognoses). These data were compared to the proteins identified in our study.

Alternative protein identification

RAW data obtained by nanoLC-MS/MS analysis were analyzed using Proteome Discoverer V2.3 (Thermo Scientific) LFQ quantification with the following parameters: trypsin as enzyme, 2 missed cleavages, methionine oxidation as variable modification and carbamidomethylation of cysteines as static modification, Precursor Mass Tolerance: 10 ppm and fragment mass tolerance: 0.6 Da. The validation was performed using Percolator with a FDR set to 0.001%.

A consensus workflow was then applied for the statistical arrangement, using the high confidence protein identification. The protein database was uploaded from Openprot (https://openprot.org/) and included RefProt, novel isoforms and AltProts predicted from both Ensembl and RefSeq annotations (GRCh38.83, GRCh38.p7) (Brunet et al., 2021; Delcourt et al., 2018; Vanderperre et al., 2013) for a total of 658263 entries. The identified abundance was extracted to PD2.3 and loaded in Perseus to performed statistical analysis and graphical representation.

Statistical analyses

Descriptive analyses were performed on clinical data. Patients were divided into 3 groups according to the quartiles of the overall survival (<Q1, Q1-Q3, > Q3). The Cox model was used to determine which proteins were most associated with overall survival. Stepwise analysis and bootstrap methods (500 samples) were used to guarantee the robustness of the results. The proteins selected after this step were used to carry out a hierarchical classification (Euclidean distance and Ward's method) on the 46 patients to determine if there were any subgroups (clusters). Finally, the clinical variables were analyzed according to the different clusters in order to provide a clinical description of the clusters obtained. Statistical analyses were performed using the SAS Software, V9.4.

Confirmatory immunohistochemistry analyses

Survival group validation was performed using antibodies directed against ALCAM, RPS14, ANXA11, PPP1R12A. The tissues were incubated with a primary antibody at 4°C overnight, followed by application of a secondary antibody (alexa fluor conjugated antibody, 1/1 000 dilution) for 1 hour at RT. For the validation cohort, dewaxing and antigen retrieval with citrate buffer were first performed before the incubation with the antibodies. We used the following primary antibodies: ALCAM (R&D Systems; 1/40 dilution), RPS14 (Invitrogen, 1/100 dilution), ANXA11 (OriGene, 1/100 dilution), PPP1R12A (Invitrogen, 1/250 dilution). All slides were imaged on the Zeiss LSM700 confocal microscope. Three to four pictures were taken for each tumor section. Processing of the images and fluorescence intensity quantification was performed using ImageJ software.

Results

Clinical characteristics

Fifty glioblastoma samples from a prospective cohort were collected (**Supp. Figure 1A**). Four tumors with an IDH1 mutation were excluded from the study (**Suppl. Figure 1A**, tumor samples with a star). Among the remaining 46 patients (**Table 1**), thirty-one (67%) were male, the median age at diagnosis was 60 (interquartile range (IQR), 51-66), the median Karnofsky performance status at diagnosis was 90 (80-90). Twenty-six (57%) patients had a gross total resection. A methylation of the MGMT promoter was noted in 15 tumors (33%), an EGFR amplification was noted in 24 cases (52%) and a homozygous CDKN2A deletion in 34 cases (74%). A standard treatment was initiated in 42 patients (91%). At the time of the analysis, 38 patients (83%) had progressed. After a median follow-up of 19.4 months (IQR 13.5-32), 43 patients (93%) had died. The median overall survival was 19.4 months. The pathologist (CAM) defined regions of interest for each tumor sample: tumor, necrosis, and endothelial proliferation, after hematoxylin eosin safran (HES) staining **(Supp. Figure 1B**).

MALDI MSI allows patient grouping based on molecular features

Considering the heterogeneity of glioblastoma, we conducted spatially resolved proteomic studies guided by mass spectrometry imaging (MSI) (**Figure 1A**). A comparison between the pathologist`s annotations and the MSI molecular images showed discrepancies for many samples (**Figure 1B**, **Supp. Data 1**). A global clustering was then performed by subjecting spectra from all tissue samples to spatial segmentation (**Figure 1C**). Three main regions were identified *i.e.* red (region A), yellow (region B) and blue (region C) areas according to the segmentation map (**Figure 1C**). Each colored region shared common molecular characteristics, meaning that the spectra in each of these areas were similar. Some specific ions can be attributed to each region: m/z 967,621 and 1492,916 were specifically present in the region A, m/z 1914,591, 2375,074 and 2376,274 were specific to the region B and m/z 1473,312, 2045,815, 2046,615 and 2237,849 were specific to the region C. Images of some group-specific ions are shown in **Figure S1C and D.** The Ward clustering method using IMAGEREVEAL MS Ver.1.1 software confirmed the segmentation of the 46 tumors into 3 groups with similar specific ions (**Supp. Figure 1D and E**).

In order to validate the classification obtained by MALDI MSI, we analysed 30 samples by SpiderMass technology (Ogrinc et al., 2019; Saudemont et al., 2018). Following the acquisition of the MS spectra in positive ion mode, a PCA analysis of the generated spectra acquired from 30 tumor tissues was performed. The features of the PCA were subjected to a supervised analysis using linear discriminant analysis (LDA) (Balog et al., 2013; Schafer et al.,

2009) which yielded 3 groups (**Figure 1Da**). According to **Figure 1Da**, LDA 1 discriminated region A from region C and the LDA 2 separated region B from regions A and C. The LDA analysis of the SpiderMass data therefore allowed the samples to be grouped in the same way as the MALDI-MSI classification. Some examples of discriminant ions (*m/z*) between the three regions, corresponding to lipids, are presented as their normalized intensities in **Figure 1Db**. The most discriminating peaks for group A in LD2+ correspond to *m/z* 746.75, and 810.65; for group B in LD2- correspond to *m/z* 718.55, 724.65, 744.55, 751.55, 778.55, 862.65 and 890.65; for group C in LD1- correspond to *m/z* 725.4, 754.6, 788.65, and 936.85. To consolidate the classification, validation was performed using either 20% randomly patients taken out or the one-patient-out method (**Inset table in Figure 1D**). Excellent cross validation results were obtained using 20% randomly patient taken out method with 100% and 91.85% correct classification rates with and without outliers respectively and good classification using the one-patient-out method with 92.92% and 77.78% including or not outliers respectively (**Inset table in Figure 1D**). These results of outliers and misclassifications (mainly group B) reflect the fact that each group is not represented by only one colored region.

Identification of specific signaling pathway signatures for each group

In order to understand the molecular differences between the three regions, spatially-resolved tissue proteomic was undertaken on the 46 tissue samples (Wisztorski et al., 2016). On each tissue, 2 to 5 specific micro extraction points were selected according to the molecular regions identified by spatial segmentation of MALDI MSI data (**Supp. Data 1**) in order to analyse the tumor heterogeneity and micro-environment presenting with specific protein signatures in each group. This resulted in a total of 135 micro-extraction points. Each extraction point was associated with one of the three regions identified by MALDI-MSI (red-A, yellow-B and blue-C). In all tumor samples, 28 extractions were performed in the red region (A), 20 in the yellow region (B) and 87 in the blue region (C) (**Supp. Table 1, Supp. Figure 1B**). From these shotgun proteomic experiments, a total of 4936 proteins were identified (**Supp. Data 2**).

First, we measured the correlation between all the extraction points from the 46 glioblastoma samples by a Pearson correlation analysis. This analysis allows the grouping of the samples according to their similarities without bias. Hierarchical clustering of all the samples based on the correlation coefficients between them reveals a grouping of the samples according to the MSI identified colored regions (**Figure 2A**). The first cluster regroups mainly samples belonging to the red region. The second cluster contains only samples belonging to the yellow region while the third cluster is more represented by samples extracted from the blue region. With this analysis, we confirmed the heterogeneity of glioblastoma tissues and validated again the MALDI-MSI segmentation.

To better understand the differences between each identified group, ANOVA tests with a Benjamini Hochberg FDR of 0.05 was performed. A total of 1183 proteins showed a significant difference in expression between the three groups (**Figure 2B, Supp. Data 3**). Two main branches were identified in the heatmap. The first branch was composed of 100% of samples extracted from the yellow region (region B). The second branch separates group A (red region) from group C (blue region). This branch is then separated into two sub-branches with the first one corresponding to region A and regrouping 79.2% of the samples extracted from the red region and 20.8% of the samples extracted from the blue region. The second sub-branch corresponds to the largest cluster, group C and contains 82.8% of the samples extracted from the blue region, 9% of the samples extracted from the red region and 8% of the samples extracted from the yellow region. We confirmed that each sample from the same colored region has the same proteomic profile (**Supp. Data 3**). Three specific clusters of over-expressed proteins for each region were identified using the heatmap (**Figure 2B**) *i.e.* cluster 1 corresponds to proteins overexpressed in group B; cluster 2, to proteins overexpressed in group A and cluster 3, to proteins overexpressed in group C. The lists of overexpressed proteins per group are presented in **Supplementary Data 3**.

In group A (mainly represented in cluster 2), the proteins are associated with neuro-developmental genes, that are characteristic of neuronal/glial lineages or progenitor cells. Most proteins were related to neurogenesis and axon guidance (dihydropyrimidinase-related protein 1 (CRMP1), misshapen-like kinase 1 (MINK1), neuromodulin (GAP43), dihydropyrimidinase-related protein 5 (DPYSL5), dihydropyrimidinase-related protein 4 (DPYSL4), microtubule-associated protein tau (MAPT), kinesin-like protein KIF2A (KIF2A), neurofilament heavy polypeptide (NEFH), unconventional myosin-XVIIIa (MYO18A), MAGUK p55 subfamily member 2 (MPP2), alpha-internexin (INA), CLIP-associating protein 2 (CLASP2) (**Supp. Data 4**). Using the functional enrichments analysis tool of String database, the most representative Reactome pathway was devoted to axon guidance. Nine of the 16 proteins identified in this pathway are involved in neuron development projection, morphogenesis, and guidance (**Supp. Figure 2Aa**). System biology analyses using SNEA and Cytoscape confirmed that the proteins in group A (Cluster 2) are involved in neurite outgrowth, synaptogenesis, synaptic vesicle transport and neurotransmission (**Figures 2C**). Interestingly, among the identified proteins some are known to be involved in tumorigenesis like mitogen-activated protein kinase 3 (MAPK3), protein kinase C alpha type (PRKCA) and some were already identified in glioblastoma *e.g.* CRMP1, DPYSL2 (i.e. CRMP2), (Jovčevska et al., 2017) DPYSL5 (i.e. CRMP5) (Moutal et al., 2015), GAP43 (Gutmann et al., 2002; Huang et al., 2003; Voigt et al., 2017), as well as Tau protein encoded by MAPT in low-grade glioma (Zaman et al., 2019).

Proteins overexpressed in group B (mainly represented in cluster 1) were linked to microglial activation and more generally immune system activation. Indeed, among the proteins identified, 10 proteins are linked to the immune response such as complement C1q subcomponent subunit C and B (C1QC and C1QB), complement factor H (CFH), haptoglobin (HP), kininogen 1 (KNG1), histidine-rich glycoprotein (HRG), transthyretin (TTR), grancalcin (GCA), proteins S100-A9 (S100A9) & S100-A12 (S100A12), erythrocyte band 7 integral membrane protein (STOM) and galectin-3-binding protein (LGALS3). Immunoglobulin heavy and light chains (IGHG2; IGKC; IGHG1; IGLC6; IGHM and IGHA1) and macrophage markers, macrophage-capping protein (CAPG) were also detected (**Supp. Data 4**). Moreover, some proteins are related to iron transporters like ceruloplasmin (CP), serotransferrin (TF), hemopexin (HPX) and haptoglobin, and other proteins are associated to coagulation *e.g.* transthyretin, kininogen-1 (KNG1), plasminogen (PLG). Most of these proteins are known to be present in human plasma (Uhlen et al., 2017b). These results are in accordance with histological annotations reflecting that most of the extraction points belonging to region B present intense proliferation of capillary endothelial cells with inflammation and hemorrhage (**Supp. Figure 1B**). The cytoscape and SNEA analysis (**Figure 2D**) confirmed that most of the proteins are involved in the complements and coagulation cascades, inflammation, ischemia, vascularization, wood healing, and cancer. The same pathways were found in Reactome (**Supp. Figure 2Ab**). Some of these proteins have already been identified in the TCGA glioma database (see below) and are mostly associated with unfavorable prognosis, *e.g,* Grancalcin and CAPG (**Supp. Figure 2B**). These results are in accordance with histological annotations reflecting that most of the extraction points of the region B are in areas of intense proliferation of capillary endothelial cells with inflammation and hemorrhage **(Supp. Figure 1B).**

The overexpressed proteins in the group C (mainly represented in cluster 3) are mainly involved in tumor growth (Hepatoma-derived growth factor (HDGF), Developmentally regulated GTP-binding protein 2 (DRG2)), but also in virus infection (Eukaryotic translation initiation factor 3 subunit L (EIF3L), Double-stranded RNA-binding protein Staufen homolog 1 (STAU1) and Interferon-induced double-stranded RNA-activated protein kinase (EIF2AK2)) (**Supp. Data 4**). KEGGS analyses confirmed a network of proteins involved in Epstein Barr virus infection (**Supp. Figure 2Ac**). Cytoscape pathway analyses established that this group is linked to viral infection and antiviral immune response (**Figure 2D**). System biology analyses confirmed the involvement of proteins in virus infection (transfection, reproduction) and transcriptomic modification at the RNA level (RNA splicing, metabolism, replication) (**Figure 2D**). Some other markers of the group C are known to be bad prognosis indicators such as EIF2AK2 and ZC3HAV1.

Identification of alternative proteins

Using the OpenProt alternative proteins database (Brunet et al., 2021), 257 AltProts were identified in our glioblastoma cohort and 170 were quantified. After ANOVA tests with a p-value of 0.05, 58 were differentially expressed between the three regions (**Figure 2F**). In region A, four AltProts are over-expressed coming from ncRNA, IP_2390879 issued from LOC107985743, IP_244732 from KIFC3, involved in cell adhesion, IP_672223 from GBP1P1 and IP_710015 from LRRC37A9P (**Supp. Table 2**). In region B, we found a cluster of nine over-expressed AltProts. Five are transcribed from ncRNA, two are located at the 5'UTR of mRNA, one at the 3'UTR and one results from a frame shift in the CDS **(Supp. Table 2**). In region C, 45 AltProts are over-expressed: 24 from ncRNA, six from the 5'UTR, 10 from the 3'UTR and five result from the frame shift in the CDS (**Supp. Table 2**). Taken together, we identified several AltProts issued from ncRNA (~57%) which is in line with our previous work on a glioma cell line (NCH82) (Cardon et al., 2020b and (Cardon et al., 2021)).

Correlation between TCGA and proteomic data

We then compared our almost 5000 identified proteins to the TCGA database, on which 682 genes show an elevated expression in glioma; 282 of these 682 genes were found in our samples (**Supp. Table 3**). Of these 682 genes, 268 genes are suggested as prognostic indicators based on transcriptomic data from 153 patients; 201 genes are associated with an unfavourable prognosis and 67 genes are associated with a favourable prognosis. In our proteomic data, we found 12 proteins associated with an unfavourable prognosis: 7 proteins are over-expressed in region A (CEND1, DMTN, PAK1, MAP2K1, THY1, VSNL1 and FN3KRP), 2 proteins are over-expressed in region B (AEBP1 and PDIA4) and 3 proteins are over-expressed in region C (POR, ERLIN2 and DBNL) (**Table 2**). We also found 9 proteins associated with a favourable prognosis: 7 proteins are over-expressed in region A (GLUD1, GDI2, SARS, SEPT2, PHGDH, KPNA3 and ARHGEF7), and 2 proteins are overexpressed in region C (PABPC1 and RBBP4) (**Table 3**).

Integrating proteomics and survival data

Overall survival was associated with MGMT status (**Supp. Figure 3Ac**) and KPS (**Supp. Figure 3Ab**) but not with the extent of resection (**Supp. Figure 3Ad**). In order to find new prognostic proteins from our proteomic data, we performed an ANOVA test on the entire proteomic dataset (n=46 patients) according to OS. The cohort was divided arbitrarily into 3 groups: 11 patients (25%) with OS > to the third quartile, 23 patients (50%) with an OS between the first and the third quartiles and 12 patients (25%) with an OS < to the first quartile were included in this analysis. 114 reference proteins and 10 AltProt showed significance between

these 3 groups of patients defined by their OS (**Supp. Data 5**). Then, using a Cox model, 28 proteins were significant with a p value<0.01 (**Table 4**). After a step-by-step analysis and a bootstrap procedure, 5 proteins remained highly significantly correlated with survival: ALCAM, RPS14, ANXA11, PPP1R12A and the AltProt IP_652563 (**Figure 3A**). Based on the expression of these 5 proteins, 2 clusters of patients were identified (respectively cluster 1 and cluster 2) (**Figure 3B**). The OS of the patients from the 2 clusters differed significantly (**Figure 3C**). The expression of the 28 proteins was compared between patients of clusters 1 and 2 (**Figure 3D and 3E**). 14 proteins are overexpressed in cluster 2 and associated with a poor prognosis (ANXA6, RPL11, HMGA1, IGHM, EIF3C, TUBA1A, GPHN, ANXA11, AP1G1, CDC42, PDCD6, IGHV3, IP_652563 and ALCAM). 14 proteins are overexpressed in cluster 1 and associated with a better prognosis (FXR1, RPS20, CALM3, S100B, CPNE6, RPS14, PPP1R12A, MTDH, WIBG, ACIN1, LASP1, THRAP3, PML, CDC5L).

Among the 5 proteins highly correlated with the survival based on the bootstrap procedure, IP_652563 is an AltProt issued from an ncRNA. This ncRNA is transcribed from the ENSG00000206028 gene which is expressed in glioma cell lines (Expression Atlas). This AltProt is a poor prognosis indicator whose expression is high in tumors of cluster 2. ALCAM and ANXA11 are the two other bad prognosis markers overexpressed in cluster 2. PPP1R12A and RPS14 are good prognosis markers overexpressed in cluster 1 (**Figure 3A**). We confirmed the overexpression of these 5 markers in either cluster 1 or 2 based on the LFQ proteomic values (**Figure 3F**). We further validated the expression of 4 of the 5 prognosis markers (ALCAM, RPS14, ANXA11 and PPP1R12A) by immunohistochemistry in the two clusters of patients. Representative images are presented in **Figure 4A**. For the AltProt, we could not perform this validation due to lack of antibodies. We confirmed in patients from cluster 2 a higher expression of ANXA11, which correlates well with the proteomic data (**Figure 4B**). ALCAM was found to be higher expressed by proteomics in cluster 2. Although not significant, a slight increase of fluorescence was observed in tumors of cluster 2 as well. The expression of ALCAM is associated with blood vessels as shown in **Figure 4A** and is known to participate in immune cell infiltration. Even though no difference in fluorescence was measured, blood vessels appeared to show different morphologies between patients of cluster 1 and 2 as shown in **Figure 4A.** RPS14 and PPP1R12A are expressed at higher levels in the tumors of cluster 1 (longer OS), which was also found by proteomics (**Figure 4B**). In order to confirm the power of these 5 markers to predict survival in glioblastoma, we validated their expression by immunohistochemistry on an independent cohort of 50 patients (**Figure 4C**). Patients were grouped according to their survival time: 13 patients had a low survival (less than 1 year), 25 patients had an intermediate survival (between 1 year and 2 years) and 12 patients had a high survival (more than 2 years). RPS14 and PPP1R12A were expressed at higher levels in the

tumors of patients with longer survival compared to tumors of patients with a low and intermediate survivals. ANXA11 was expressed at higher levels in tumors of patients with a low and intermediate survivals compared to patients with longer survival. No statistical differences of expression were observed for ALCAM expression, as observed for the first cohort of patients. These results confirm the validity of the identified prognostic markers, except for ALCAM.

Discussion

In this work, we investigated the biology and heterogeneity of glioblastoma by a proteomic approach at a low spatial resolution to capture the tumor microenvironment. A non-targeted MALDI-MSI analysis followed by spatial segmentation using different algorithms allowed to highlight molecular heterogeneity among these tumors. We validated these observations with SpiderMass technology with a 93% good classification. Three sub regions were identified (A- Red, B-Yellow, and C-Blue regions). To decode the biological pathways involved in these three regions, we performed a spatially resolved proteomic analysis that confirmed the data. Molecular signatures of different tumor subtypes were identified among the groups. From these data, we derived three molecular signatures. Region A is enriched in genes related to neurotransmission and synaptogenesis. Proteins overexpressed in region B are associated with immune infiltration while in region C, we mainly identified proteins involved in RNA processing and metabolism.

Region A is associated with neuro-developmental genes, characteristic of neuronal/glial lineages or neural progenitor cells (NPC) (**Figure 2B**). These included nervous system development markers (like CRMP family, GAP43, MAPT), oligodendrocyte development and differentiation markers (like ABI1, ASPA, CNP, CNTNAP1), stem and progenitor cell signatures (like TRIM2). The NPC-like state is correlated with markers for immature neurons (beta-3-tubulin), markers for mature neurons (NeuN) and markers indicative for synapses (synaptophysin, SV2A) (Beier et al., 2018). In our data, we found Stathmin 1, NEFH, NEFM and NEFL (Neftel et al., 2019) which are also markers of the NPC-like state of the GSC. Region B is enriched in proteins linked to immune status with macrophages infiltration, (**Figure 2C**) such as complement factors, immunoglobulin heavy and light chains (IGHG2; IGKC; IGHM; IGHG1; IGLC6 and IGHA1), macrophage markers (CAPG) and coagulation cascade proteins (HP, KNG1, HRG, TTR, GCA, S100A9, STOM). In a study of (Cheng et al., 2016), eight immune related genes (FOXO3, IL6, IL10, ZBTB16, CCL18, AIMP1, FCGR2B, and MMP9) were identified and used as unfavorable prognostic markers in glioblastoma. High-risk patients exhibited an enhanced intensity of local immune response compared to low-risk ones. From the 8 signature genes, AIMP2 was identified in region B, too.

GSC markers but with a “stem-to-invasion” path were also identified in region B. CD44, NES and VIM, enriched in region B, are markers of the mesenchymal like state.

The presence of class I self-antigen HLA proteins (HLA-A3 and HLA-B07) in group B is interesting since a positive correlation between HLA expression and some cancers has been demonstrated, such as cervical or nasopharyngeal carcinomas (Machulla et al., 2001). In a previous study based on HLA antigen frequencies in patients with glioma, patients positive for HLA-A*25 had a 3.0-fold increased risk of glioma ($p = 0.04$) and patients positive for HLA-B*27, a 2.7-fold risk ($p = 0.03$), compared with the control population. In contrast, the relationship between HLA-B*07 expression and higher risk to develop a glioma is very rare (Tang et al., 2005), as well as for HLA-A*3 (Zhang et al., 2007). Taken together, these data confirmed that there is interpatient molecular heterogeneity that may be related to tumor phenotype and cellular plasticity (Neftel et al., 2019) but not directly with transcriptional classification of glioblastoma (proneural, neural, classic and mesenchymal) (Verhaak et al., 2010b). Finally, systemic biology analyses revealed that group C is linked to an anti-viral immune response and viral infection, in addition to RNA processing. Recent studies have reported a link between glioblastoma and perinatal viral exposure (Akhtar et al., 2018; Dickinson et al., 2002; Limam et al., 2019; Strojnik et al., 2017). Further Epstein-Barr virus has been implicated in glioblastoma etiology (Zavala-Vega et al., 2019). Moreover, some studies have also reported that cytomegalovirus (CMV) promotes murine glioblastoma growth via pericyte recruitment and angiogenesis (Krenzlin et al., 2019). In human, CMV nucleic acids and proteins have been observed within glioblastoma tumor tissue (Rahman et al., 2019), although the link between glioblastoma and CMV remains very controversial (Baumgarten et al., 2014).

The comparison with the TCGA specific glioma gene signature showed that 21 of them were associated with survival among the 3 groups identified in our study. Most of the proteins were identified in group A and are related to nervous system development, neuron differentiation axon guidance, 3 proteins were identified in group B and are linked to cytokine secretion and 5 both in groups A & C related to Notch signalling. Notch signaling is an evolutionarily conserved pathway that regulates important biological processes, such as cell proliferation, apoptosis, migration, self-renewal, and differentiation. Growing evidence reveals that Notch signaling is highly active in glioma stem cells, in which it suppresses differentiation and maintains stem-like properties, contributing to glioblastoma tumorigenesis and conventional-treatment resistance (Bazzoni and Bentivegna, 2019)

Taken together, we have revealed three main molecular regions in glioblastomas. Each region has a distinct molecular pattern, reflecting a specific molecular phenotype of the tumors. These different groups may be explained by an early differentiation due to the presence, in

primary tumours, of subpopulation of cells with a distinct functional profile as well as the existence of cells with a high invasive potency. A recent study (Pang et al., 2019) proposed that glioblastoma stem cells (GSCs) acquire a high invasive activity through a mechanism called the 'stem-to-invasion path' and that long noncoding RNAs are one of the key factors. It has been demonstrated that these non-coding genomic regions can result in the synthesis of proteins, so called alternative proteins, forming an unexplored ghost proteome with unknown function in cancer (Delcourt et al., 2018). 170 alternative proteins (AltProts) were found significantly variable in the three groups identified above. Although the function of these AltProts remains poorly understood, they can have a role in regulation of transcription and can also be present in extracellular vesicles (Murgoci et al., 2020). Finally, more than 50% of the AltProts identified in the present study come from the translation of ncRNAs transcribed from pseudogenes. Seven AltProts have been identified in common with our previous study on the NCH82 glioma cell line, (IP_2323408 and IP_261897 described as an ncRNA and IP_755940, IP_593099, IP_774693, IP_572422 and IP_671464 from non-coding regions of mRNA. These last five AtlProts are pseudogenes for: HNRNPA1P30, TUBB2BP1, TUBAP2, TUBBP1, and TPI1P1 respectively. These pseudogenes, for which no protein has been observed yet, express their transcripts in glioma cell lines (Expression Atlas) (Petryszak et al., 2016). Interestingly, the last one IP_079312, from the mRNA encoding EDARADD was correlated with a low survival rate in ovarian cancer patients (Cardon et al., 2020b). Recently it has been demonstrated that pseudogenes can also be used as signatures for glioma prognosis. 6 pseudogenes (SP3P, ANXA2P3, PTTG3P, LPAL2, CLCA3P, and TDH) were reported to be associated with overall survival in glioma (Gao et al., 2015). Nine other pseudogenes (TP73-AS1, AC078883.3, RP11-944L7.4, HAR1B, RP4-635E18.7, HOTAIR, SAPCD1-AS1, AC104653.1, and RP5-1172N10.2.) constitute a set of prognosis markers to predict survival of patients with glioma (Lei et al., 2018). All these results provide novel insights into the biological role of pseudogenes in cancer and especially in glioma. Additionally, the novel identified AltProts translated from ncRNAs add additional information to the already known pseudogenes in glioma.

In another study, in which we studied interaction partners of AltProts in NCH82 cells (Cardon et al., 2020a), we identified five significantly different AltProts. One of them has been identified as overexpressed in region B: IP_156671 which originates from the 3'UTR of the transcript coding for SLC13A1. The four others are overexpressed in group C: IP_261897 coming from an ncRNA, IP_063564, IP_256988 both issued from the 3'UTR region of the CLDN19 and TBX21 genes respectively and IP_073718 originating from a shift in the reading frame of the CCDC181 gene.

Finally, the present study allowed us to highlight the presence of 5 new prognostic proteins for glioblastoma: PPP1R12A and RPS14 are favourable prognostic markers while ALCAM, ANXA11, and AltProt IP_652563 are unfavourable prognostic markers. The expression of these markers was validated on an external cohort of patients. These proteins were already identified as prognostic markers in lung and renal cancers (Human Protein Atlas).

In conclusion, we present here a spatial proteomic characterization in clinical samples of glioblastoma. The proteomic signatures we identified demonstrate the intratumoral molecular heterogeneity of glioblastoma tumors. While in previous studies, these signatures have been shown to be associated with survival (Yanovich-Arad et al., 2021), we showed that several of these signatures can be detected in a single tumor preventing their use as prognostic indicators. Despite this high heterogeneity, we have shown that some common markers could be identified for tumors of patients with inferior survival and inversely for tumors of patients with a longer survival, with a validation on an external cohort of patients. In addition, our dataset can serve as a starting point to guide the development of new personalized therapeutic strategies and better treatment decisions.

Acknowledgments

This research was supported by grants from the Ministère de L'Education Nationale, de L'Enseignement Supérieur et de la Recherche, ANR (IF), SIRIC ONCOLille (MS), Grant INCa-DGOS-Inserm 6041aa (IF, MS), and INSERM, Ligue contre le Cancer (EL) .

AUTHORS CONTRIBUTION

Conceptualization : MS, IF, ELR, Methodology : MD, MWi, LD, MS, IF, JP, SA, NO, Software : LD, MWi, TC, FZ; Validation: MD, LD, CAM, FE; Formal Analysis : MS, MD, LD, TC, MWi, PD; Investigation : MD, MWi, LD, ELR, IF, MS; Data curation : MWi, MS, LD, MD, Writing : MS, MD, ELR, MWe, TC, LD. Original Draft : MS, LD, MD, MWi; Supervision, Project : MWi, ELR, IF, MS; Administration : ELR, MS, IF; Funding Acquisition IF, MS; ELR

Declaration of Interests

Dr Le Rhun has received honoraria for lectures or advisory board from Adastra, Bayer, Janssen, Leo Pharma, Pierre Fabre, and Seattle Genetics.

Dr. Weller has received research grants from Apogenix, Merck, Sharp & Dohme, Merck (EMD), Philogen and Quercis, and honoraria for lectures or advisory board participation or consulting from Adastra, Bayer, Bristol Meyer Squibb, Medac, Merck, Sharp & Dohme, Merck (EMD), Nerviano Medical Sciences, Novartis, Orbus, Philogen and y-Mabs.

The other authors declare no competing interests.

Supplementary Figures

Supp. Figure 1. A. Scanned pictures after hematoxylin-eosin staining of the 46 glioblastoma and **B.** anathomo-pathologist annotations. **C.** MALDI MSI images of characteristic m/z ions for each region. **D.** Ward clustering method gives 3 main branches with same characteristic ions. **E**. Principal component analysis (PCA) of each individual spectra reveals separation between the three regions.

Supp. Figure 2. **A. a)** Analysis of proteins overexpressed in group A shows an involvement in axon guidance. **b)** Proteins overexpressed in group B and mainly involved in complements, coagulation cascade and inflammation **c)** Analysis of overexpressed proteins in group C shows a network of proteins involved in Epstein barr infection. **B.** Correlation between CAPG expression (a) and Grancalcin (b) and glioma patient survival according to the TGCA data. Patients were divided based on level of expression into “low” or “high”.

Supp Figure 3. Global survival curve of all patients according to the Karnofsky indice (b), MGMT statut (c) and resection quality (d) 0-1 = Total resection, subtotal 2-3 = partial resection biopsies

DATA AVAILIBILITY

Proteomic datasets including MaxQuant files and annotated MS/MS datasets were uploaded to the ProteomeXchange Consortium via the PRIDE database, and then assigned the dataset identifier PXD016165".

References

Akhtar, S., Vranic, S., Cyprian, F. S., and Al Moustafa, A.-E. (2018). Epstein–Barr virus in gliomas: cause, association, or artifact? Frontiers in oncology *8*, 123.

Alexandrov, T., Becker, M., Deininger, S. O., Ernst, G., Wehder, L., Grasmair, M., von Eggeling, F., Thiele, H., and Maass, P. (2010). Spatial segmentation of imaging mass spectrometry data with edge-preserving image denoising and clustering. Journal of proteome research *9*, 6535-6546.

Balog, J., Sasi-Szabo, L., Kinross, J., Lewis, M. R., Muirhead, L. J., Veselkov, K., Mirnezami, R., Dezso, B., Damjanovich, L., Darzi, A*., et al.* (2013). Intraoperative tissue identification using rapid evaporative ionization mass spectrometry. Sci Transl Med *5*, 194ra193.

Baumgarten, P., Michaelis, M., Rothweiler, F., Starzetz, T., Rabenau, H. F., Berger, A., Jennewein, L., Braczynski, A. K., Franz, K., and Seifert, V. (2014). Human cytomegalovirus infection in tumor cells of the nervous system is not detectable with standardized pathologico-virological diagnostics. Neuro-oncology *16*, 1469-1477.

Bazzoni, R., and Bentivegna, A. (2019). Role of notch signaling pathway in glioblastoma pathogenesis. Cancers *11*, 292.

Beier, C. P., Rasmussen, T., Dahlrot, R. H., Tenstad, H. B., Aarø, J. S., Sørensen, M. F., Heimisdóttir, S. B., Sørensen, M. D., Svenningsen, P., and Riemenschneider, M. J. (2018). Aberrant neuronal differentiation is common in glioma but is associated neither with epileptic seizures nor with better survival. Scientific reports *8*, 1-12.

Brat, D. J., Aldape, K., Colman, H., Figrarella-Branger, D., Fuller, G. N., Giannini, C., Holland, E. C., Jenkins, R. B., Kleinschmidt-DeMasters, B., and Komori, T. (2020). cIMPACT-NOW update 5: recommended grading criteria and terminologies for IDH-mutant astrocytomas. Acta neuropathologica *139*, 603-608.

Brennan, C. W., Verhaak, R. G., McKenna, A., Campos, B., Noushmehr, H., Salama, S. R., Zheng, S., Chakravarty, D., Sanborn, J. Z., and Berman, S. H. (2013). The somatic genomic landscape of glioblastoma. Cell *155*, 462-477.

Brunet, M. A., Lucier, J.-F., Levesque, M., Leblanc, S., Jacques, J.-F., Al-Saedi, H. R., Guilloy, N., Grenier, F., Avino, M., and Fournier, I. (2021). OpenProt 2021: deeper functional annotation of the coding potential of eukaryotic genomes. Nucleic acids research *49*, D380-D388.

Capper, D., Jones, D. T., Sill, M., Hovestadt, V., Schrimpf, D., Sturm, D., Koelsche, C., Sahm, F., Chavez, L., and Reuss, D. E. (2018). DNA methylation-based classification of central nervous system tumours. Nature *555*, 469-474.

Cardon, T., Fournier, I., and Salzet, M. (2021). Unveiling a Ghost Proteome in the Glioblastoma Non-Coding RNAs. Frontiers in Cell and Developmental Biology *9*.

Cheng, W., Ren, X., Zhang, C., Cai, J., Liu, Y., Han, S., and Wu, A. (2016). Bioinformatic profiling identifies an immune-related risk signature for glioblastoma. Neurology *86*, 2226-2234.

Chinot, O. L., Wick, W., Mason, W., Henriksson, R., Saran, F., Nishikawa, R., Carpentier, A. F., 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.

Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Molecular & cellular proteomics : MCP *13*, 2513-2526.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol *26*, 1367-1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. Journal of proteome research *10*, 1794-1805.

Deighton, R. F., McGregor, R., Kemp, J., McCulloch, J., and Whittle, I. R. (2010). Glioma pathophysiology: insights emerging from proteomics. Brain Pathology *20*, 691-703.

884 Delcourt, V., Staskevicius, A., Salzet, M., Fournier, I., and Roucou, X. (2018). Small Proteins Encoded by
885 Unannotated ORFs are Rising Stars of the Proteome, Confirming Shortcomings in Genome
886 Annotations and Current Vision of an mRNA. Proteomics *18*, e1700058.
887 Dickinson, H., Nyari, T., and Parker, L. (2002). Childhood solid tumours in relation to infections in the
888 community in Cumbria during pregnancy and around the time of birth. British journal of cancer *87*,
889 746-750.
890 Dilillo, M., Ait-Belkacem, R., Esteve, C., Pellegrini, D., Nicolardi, S., Costa, M., Vannini, E., De Graaf, E.,
891 Caleo, M., and McDonnell, L. (2017). Ultra-high mass resolution MALDI imaging mass spectrometry
892 of proteins and metabolites in a mouse model of glioblastoma. Scientific reports *7*, 1-11.
893 Djuric, U., Lam, K. B., Kao, J., Batruch, I., Jevtic, S., Papaioannou, M.-D., and Diamandis, P. (2019).
894 Defining Protein Pattern Differences Among Molecular Subtypes of Diffuse Gliomas Using Mass
895 Spectrometry*[S]. Molecular & Cellular Proteomics *18*, 2029-2043.
896 Fatou, B., Saudemont, P., Leblanc, E., Vinatier, D., Mesdag, V., Wisztorski, M., Focsa, C., Salzet, M.,
897 Ziskind, M., and Fournier, I. (2016). In vivo Real-Time Mass Spectrometry for Guided Surgery
898 Application. Scientific reports *6*, 25919.
899 Fournier, I., Day, R., and Salzet, M. (2003). Direct analysis of neuropeptides by in situ MALDI-TOF mass
900 spectrometry in the rat brain. Neuro Endocrinol Lett *24*, 9-14.
901 Gao, K.-M., Chen, X.-c., Zhang, J.-x., Wang, Y., Yan, W., and You, Y.-P. (2015). A pseudogene-signature
902 in glioma predicts survival. Journal of experimental & clinical cancer research *34*, 1-7.
903 Gilbert, M. R., Dignam, J. J., Armstrong, T. S., Wefel, J. S., Blumenthal, D. T., Vogelbaum, M. A., Colman,
904 H., Chakravarti, A., Pugh, S., and Won, M. (2014). A randomized trial of bevacizumab for newly
905 diagnosed glioblastoma. New England Journal of Medicine *370*, 699-708.
906 Gramatzki, D., Roth, P., Rushing, E., Weller, J., Andratschke, N., Hofer, S., Korol, D., Regli, L., Pangalu,
907 A., and Pless, M. (2018). Bevacizumab may improve quality of life, but not overall survival in
908 glioblastoma: an epidemiological study. Annals of Oncology *29*, 1431-1436.
909 Gutmann, D. H., Huang, Z. Y., Hedrick, N. M., Ding, H., Guha, A., and Watson, M. A. (2002). Mouse
910 glioma gene expression profiling identifies novel human glioma-associated genes. Annals of
911 Neurology: Official Journal of the American Neurological Association and the Child Neurology
912 Society *51*, 393-405.
913 Huang, Z.-y., Wu, Y., Burke, S. P., and Gutmann, D. H. (2003). The 43,000 growth-associated protein
914 functions as a negative growth regulator in glioma. Cancer research *63*, 2933-2939.
915 Jovčevska, I., Zupanec, N., Urlep, Ž., Vranič, A., Matos, B., Stokin, C. L., Muyldermans, S., Myers, M. P.,
916 Buzdin, A. A., and Petrov, I. (2017). Differentially expressed proteins in glioblastoma multiforme
917 identified with a nanobody-based anti-proteome approach and confirmed by OncoFinder as
918 possible tumor-class predictive biomarker candidates. Oncotarget *8*, 44141.
919 Kalinina, J., Peng, J., Ritchie, J. C., and Van Meir, E. G. (2011). Proteomics of gliomas: initial biomarker
920 discovery and evolution of technology. Neuro-oncology *13*, 926-942.
921 Klein, O., Strohschein, K., Nebrich, G., Oetjen, J., Trede, D., Thiele, H., Alexandrov, T., Giavalisco, P.,
922 Duda, G. N., and von Roth, P. (2014). MALDI imaging mass spectrometry: Discrimination of
923 pathophysiological regions in traumatized skeletal muscle by characteristic peptide signatures.
924 Proteomics *14*, 2249-2260.
925 Krenzlin, H., Behera, P., Lorenz, V., Passaro, C., Zdioruk, M., Nowicki, M. O., Grauwet, K., Zhang, H.,
926 Skubal, M., and Ito, H. J. T. J. o. c. i. (2019). Cytomegalovirus promotes murine glioblastoma growth
927 via pericyte recruitment and angiogenesis. *129*.
928 Lei, B., Yu, L., Jung, T. A., Deng, Y., Xiang, W., Liu, Y., and Qi, S. (2018). Prospective series of nine long
929 noncoding RNAs associated with survival of patients with glioblastoma. Journal of Neurological
930 Surgery Part A: Central European Neurosurgery *79*, 471-478.
931 Lemaire, R., Menguellet, S. A., Stauber, J., Marchaudon, V., Lucot, J. P., Collinet, P., Farine, M. O.,
932 Vinatier, D., Day, R., Ducoroy, P*., et al.* (2007). Specific MALDI imaging and profiling for biomarker
933 hunting and validation: fragment of the 11S proteasome activator complex, Reg alpha fragment, is
934 a new potential ovary cancer biomarker. Journal of proteome research *6*, 4127-4134.

Lemaire, R., Tabet, J. C., Ducoroy, P., Hendra, J. B., Salzet, M., and Fournier, I. (2006a). Solid ionic matrixes for direct tissue analysis and MALDI imaging. Anal Chem *78*, 809-819.
Lemaire, R., Wisztorski, M., Desmons, A., Tabet, J. C., Day, R., Salzet, M., and Fournier, I. (2006b). MALDI-MS direct tissue analysis of proteins: Improving signal sensitivity using organic treatments. Anal Chem *78*, 7145-7153.
Limam, S., Missaoui, N., Mestiri, S., Yacoubi, M., Krifa, H., Selmi, B., and Mokni, M. (2019). Epstein-Barr virus infection in gliomas. Current research in translational medicine *67*, 129-133.
Louis, D. N., Perry, A., Wesseling, P., Brat, D. J., Cree, I. A., Figarella-Branger, D., Hawkins, C., Ng, H., Pfister, S. M., and Reifenberger, G. (2021). The 2021 WHO classification of tumors of the central nervous system: a summary. Neuro-oncology *23*, 1231-1251.
Machulla, H. K., Steinborn, F., Schaaf, A., Heidecke, V., and Rainov, N. G. (2001). Brain glioma and human leukocyte antigens (HLA)–is there an association. Journal of neuro-oncology *52*, 253-261.
Moutal, A., Honnorat, J., Massoma, P., Désormeaux, P., Bertrand, C., Malleval, C., Watrin, C., Chounlamountri, N., Mayeur, M.-E., and Besançon, R. (2015). CRMP5 controls glioblastoma cell proliferation and survival through notch-dependent signaling. Cancer research *75*, 3519-3528.
Murgoci, A.-N., Cardon, T., Aboulouard, S., Duhamel, M., Fournier, I., Cizkova, D., and Salzet, M. (2020). Reference and ghost proteins identification in rat C6 glioma extracellular vesicles. Iscience *23*, 101045.
Neftel, C., Laffy, J., Filbin, M. G., Hara, T., Shore, M. E., Rahme, G. J., Richman, A. R., Silverbush, D., Shaw, M. L., and Hebert, C. M. (2019). An integrative model of cellular states, plasticity, and genetics for glioblastoma. Cell *178*, 835-849. e821.
Ogrinc, N., Saudemont, P., Balog, J., Robin, Y. M., Gimeno, J. P., Pascal, Q., Tierny, D., Takats, Z., Salzet, M., and Fournier, I. (2019). Water-assisted laser desorption/ionization mass spectrometry for minimally invasive in vivo and real-time surface analysis using SpiderMass. Nature protocols *14*, 3162-3182.
Ostrom, Q. T., Cioffi, G., Waite, K., Kruchko, C., and Barnholtz-Sloan, J. S. (2021). CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2014–2018. Neuro-oncology *23*, iii1-iii105.
Otasek, D., Morris, J. H., Bouças, J., Pico, A. R., and Demchak, B. (2019). Cytoscape automation: empowering workflow-based network analysis. Genome biology *20*, 1-15.
Pang, B., Xu, J., Hu, J., Guo, F., Wan, L., Cheng, M., and Pang, L. (2019). Single-cell RNA-seq reveals the invasive trajectory and molecular cascades underlying glioblastoma progression. Molecular oncology *13*, 2588-2603.
Pathan, M., Keerthikumar, S., Chisanga, D., Alessandro, R., Ang, C. S., Askenase, P., Batagov, A. O., Benito-Martin, A., Camussi, G., Clayton, A*., et al.* (2017). A novel community driven software for functional enrichment analysis of extracellular vesicles data. J Extracell Vesicles *6*, 1321455.
Quanico, J., Franck, J., Dauly, C., Strupat, K., Dupuy, J., Day, R., Salzet, M., Fournier, I., and Wisztorski, M. (2013). Development of liquid microjunction extraction strategy for improving protein identification from tissue sections. Journal of proteomics *79*, 200-218.
Rahman, M., Dastmalchi, F., Karachi, A., and Mitchell, D. (2019). The role of CMV in glioblastoma and implications for immunotherapeutic strategies. Oncoimmunology *8*, e1514921.
Saudemont, P., Quanico, J., Robin, Y. M., Baud, A., Balog, J., Fatou, B., Tierny, D., Pascal, Q., Minier, K., Pottier, M*., et al.* (2018). Real-Time Molecular Diagnosis of Tumors Using Water-Assisted Laser Desorption/Ionization Mass Spectrometry Technology. Cancer Cell *34*, 840-851 e844.
Schafer, K. C., Denes, J., Albrecht, K., Szaniszlo, T., Balog, J., Skoumal, R., Katona, M., Toth, M., Balogh, L., and Takats, Z. (2009). In vivo, in situ tissue analysis using rapid evaporative ionization mass spectrometry. Angew Chem Int Ed Engl *48*, 8240-8242.
Strojnik, T., Duh, D., and Lah, T. T. (2017). Prevalence of neurotropic viruses in malignant glioma and their onco-modulatory potential. in vivo *31*, 221-229.
Stupp, R., Hegi, M. E., Mason, W. P., van den Bent, M. J., Taphoorn, M. J., Janzer, R. C., Ludwin, S. K., 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. The Lancet Oncology *10*, 459-466.
Stupp, R., Taillibert, S., Kanner, A., Read, W., Steinberg, D., Lhermitte, B., Toms, S., Idbaih, A., Ahluwalia, M. S., Fink, K*., et al.* (2017). Effect of Tumor-Treating Fields Plus Maintenance Temozolomide vs Maintenance Temozolomide Alone on Survival in Patients With Glioblastoma: A Randomized Clinical Trial. JAMA *318*, 2306-2316.
Sturm, D., Witt, H., Hovestadt, V., Khuong-Quang, D.-A., Jones, D. T., Konermann, C., Pfaff, E., Tönjes, M., Sill, M., and Bender, S. (2012). Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. Cancer cell *22*, 425-437.
Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., and Bork, P. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic acids research *47*, D607-D613.
Tang, J., Shao, W., Dorak, M. T., Li, Y., Miike, R., Lobashevsky, E., Wiencke, J. K., Wrensch, M., Kaslow, R. A., and Cobbs, C. S. (2005). Positive and negative associations of human leukocyte antigen variants with the onset and prognosis of adult glioblastoma multiforme. Cancer Epidemiology and Prevention Biomarkers *14*, 2040-2044.
Trede, D., Kobarg, J. H., Oetjen, J., Thiele, H., Maass, P., and Alexandrov, T. (2012). On the importance of mathematical methods for analysis of MALDI-imaging mass spectrometry data. Journal of Integrative Bioinformatics (JIB) *9*, 1-11.
Tyanova, S., Temu, T., Carlson, A., Sinitcyn, P., Mann, M., and Cox, J. (2015). Visualization of LC-MS/MS proteomics data in MaxQuant. Proteomics *15*, 1453-1456.
Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., Benfeitas, R., Arif, M., Liu, Z., and Edfors, F. (2017a). A pathology atlas of the human cancer transcriptome. Science *357*, eaan2507.
Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., Benfeitas, R., Arif, M., Liu, Z., and Edfors, F. J. S. (2017b). A pathology atlas of the human cancer transcriptome. *357*, eaan2507.
Vanderperre, B., Lucier, J. F., Bissonnette, C., Motard, J., Tremblay, G., Vanderperre, S., Wisztorski, M., Salzet, M., Boisvert, F. M., and Roucou, X. (2013). Direct detection of alternative open reading frames translation products in human significantly expands the proteome. PLoS One *8*, e70698.
Verhaak, R. G., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., Miller, C. R., Ding, L., Golub, T., and Mesirov, J. P. (2010a). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer cell *17*, 98-110.
Verhaak, R. G., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., Miller, C. R., Ding, L., Golub, T., and Mesirov, J. P. J. C. c. (2010b). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *17*, 98-110.
Voigt, A., Nowick, K., and Almaas, E. (2017). A composite network of conserved and tissue specific gene interactions reveals possible genetic interactions in glioma. PLoS computational biology *13*, e1005739.
Wang, L.-B., Karpova, A., Gritsenko, M. A., Kyle, J. E., Cao, S., Li, Y., Rykunov, D., Colaprico, A., Rothstein, J. H., and Hong, R. (2021). Proteogenomic and metabolomic characterization of human glioblastoma. Cancer Cell *39*, 509-528. e520.
Weathers, S. P., and Gilbert, M. R. (2014). Advances in treating glioblastoma. F1000Prime Rep *6*, 46.
Weller, M., Reifenberger, G., Le Rhun, E., Clarke, J. L., Soffietti, R., Wick, A., Chinot, O. L., Ducray, F., Hau, P., and McDonald, K. L. (2019). Molecular genetic, host-derived and clinical determinants of long-term survival in glioblastoma: First results from the ETERNITY study (EORTC 1419). In, (American Society of Clinical Oncology).
Weller, M., van den Bent, M., Preusser, M., Le Rhun, E., Tonn, J. C., Minniti, G., Bendszus, M., Balana, C., Chinot, O., and Dirven, L. (2021). EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood. Nature reviews Clinical oncology *18*, 170-186.

Wisztorski, M., Desmons, A., Quanico, J., Fatou, B., Gimeno, J. P., Franck, J., Salzet, M., and Fournier, I. (2016). Spatially-resolved protein surface microsampling from tissue sections using liquid extraction surface analysis. Proteomics *16*, 1622-1632.

Yanovich-Arad, G., Ofek, P., Yeini, E., Mardamshina, M., Danilevsky, A., Shomron, N., Grossman, R., Satchi-Fainaro, R., and Geiger, T. (2021). Proteogenomics of glioblastoma associates molecular patterns with survival. Cell Reports *34*, 108787.

Yuryev, A., Kotelnikova, E., and Daraselia, N. J. E. o. o. d. d. (2009). Ariadne's ChemEffect and Pathway Studio knowledge base. *4*, 1307-1318.

Zaman, S., Chobrutskiy, B. I., Sikaria, D., and Blanck, G. J. O. r. (2019). MAPT (Tau) expression is a biomarker for an increased rate of survival for low-grade glioma. *41*, 1359-1366.

Zavala-Vega, S., Palma-Lara, I., Ortega-Soto, E., Trejo-Solis, C., de Arellano, I. T.-R., Ucharima-Corona, L. E., Garcia-Chacón, G., Ochoa, S. A., Xicohtencatl-Cortes, J., and Cruz-Córdova, A. J. C. R. i. O. (2019). Role of Epstein-Barr Virus in Glioblastoma. *24*.

Zhang, J. G., Eguchi, J., Kruse, C. A., Gomez, G. G., Fakhrai, H., Schroter, S., Ma, W., Hoa, N., Minev, B., and Delgado, C. J. C. C. R. (2007). Antigenic profiling of glioma cells to generate allogeneic vaccines or dendritic cell–based therapeutics. *13*, 566-575.

1055 Tables

1056 **Table 1. Clinical characteristics**

	Total population (n=46)
Sex	
female, n (%)	15 (33)
male, n (%)	31 (67)
Age at diagnosis (years)	
median (IQR)	60 (51-66)
Karnofsky performance status at diagnosis	
median (IQ)	90 (80-90)
0-80, n (%)	14 (30)
90-100, n (%)	32 (70)
Main location of the tumor	
frontal, n (%)	11 (24)
occipital, n (%)	3 (6)
parietal, n (%)	12 (26)
temporal, n (%)	20 (43)
Extent of surgical resection	
Gross total, n (%)	26 (57)
partial, n (%)	19 (41)
biopsy, n (%)	1 (2)
MGMT promoter methylation status	
not methylated, n (%)	31 (67)
methylated, n (%)	15 (33)
EGFR amplification	
no, n (%)	22 (48)
yes, n (%)	24 (52)
Chromosome 7 gain combined with chromosome 10 loss (+7/-10)	
no, n (%)	12 (26)
yes, n (%)	34 (74)
EGFR amplification combined with 7 gain / 10 loss	
EGFR amplification or gain 7 / lost 10	41 (89)
EGFR amplification without gain 7 / lost 10	7 (15)
EGFR amplification and gain 7 / lost 10	17 (37)
gain 7 / lost 10 without EGFR amplification	17 (37)
Homozygous CDKN2A deletion	
no, n (%)	18 (39)
yes, n (%)	28 (61)
Median follow-up (months) median (IQR)	19.4 (13.5-32.0)
Initial treatment	
RT/TMZ followed by 6 cycles of TMZ, n (%)	18 (39)
RT/TMZ followed bythen more than 6 cycles of months TMZ, n (%)	4 (9)
RT/TMZ followed by less than 6 cycles of TMZ, n (%)	20 (43)
other treatment*, n (%)	2 (4)
clinical study, n (%)	1 (2)
no treatment, n (%)	1 (2)
Progression	
yes, n (%)	38 (83)
no, n (%)	3 (6)
unknown, n (%)	5 (11)

Progression-free survival (months)	
median (IQR)	10.6 (7.1-16.3)
Treatment at first progression (n=38)	
yes, n (%)	33* (87)
no, n (%)	5 (13)
Death	
yes, n (%)	43 (93)
no, n (%)	3 (7)
Survival from surgery (months)	
median (IQR)	19.4 (13.5-32.0)
Survival	
upper IQR, n (%)	12 (26)
intermediate IQR, n (%)	23 (50)
lower IQR, n (%)	11 (24)

* One patient: RT only, one patient: 6 cycles TMZ then SRT

Abbreviations:

EGFR, epidermal growth factor receptor

IQR: interquartile range

MGMT: O^6-methylguanine DNA methyltransferase

RT: radiotherapy

SRT: stereotactic radiotherapy

TMZ: temozolomide

Table 2: Proteins associated with unfavorable prognostic in glioma and identified in regions A, B and C

Uniprot	Gene description	Gene	Region
Q8N111	Cell cycle exit and neuronal differentiation 1	CEND1	A
Q08495	Dematin actin binding protein	DMTN	A
Q13153	P21 (RAC1) activated kinase 1	PAK1	A
Q02750	Mitogen-activated protein kinase kinase 1	MAP2K1	A
P04216	Thy-1 cell surface antigen	THY1	A
Q9HA64	Ketosamine-3-kinase	FN3KRP	A
P62760	Visinin like 1	VSNL1	A
Q8IUX7	AE binding protein 1	AEBP1	B
P13667	Protein disulfide isomerase family A member 4	PDIA4	B
P16435	Cytochrome p450 oxidoreductase	POR	C
Q9UJU6	Drebrin-like protein	DBNL	C
O94905	Erlin-2	ERLIN2	C

Table 3: Proteins associated with favorable prognostic in glioma and identified in regions A, B and C

Uniprot	Gene description	Gene	Group
P00367	Glutamate dehydrogenase 1	GLUD1	A
Q14155	Rho guanine nucleotide exchange factor 7	ARHGEF7	A
P50395	GDP dissociation inhibitor 2	GDI2	A
P49591	Seryl-tRNA synthetase	SARS	A
Q15019	Septin-2	SEPT2	A
O43175	D-3-phosphoglycerate dehydrogenase	PHGDH	A
O00505	Importin subunit alpha-4	KPNA3	A
P11940	Poly(A) binding protein cytoplasmic 1	PABPC1	C

Table 4. Proteins associated with survival after Cox model p = 0.01

Parameter	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
IP_652563*	0.28507	0.07527	14.3427	0.0002	1.33	1.147	1.541
FXR1	-1.25301	0.36448	11.8185	0.0006	0.286	0.14	0.584
RPS20	-0.7618	0.23552	10.4623	0.0012	0.467	0.294	0.741
ANXA6	0.51404	0.15933	10.4094	0.0013	1.672	1.224	2.285
ALCAM*	0.56831	0.17706	10.3015	0.0013	1.765	1.248	2.498
RPL11	-0.78558	0.25342	9.6093	0.0019	0.456	0.277	0.749
CALM3	0.25136	0.08171	9.4639	0.0021	1.286	1.096	1.509
HMGA1	-0.34607	0.11316	9.352	0.0022	0.707	0.567	0.883
S100B	0.24907	0.0818	9.2715	0.0023	1.283	1.093	1.506
IGHM	0.32975	0.10848	9.2399	0.0024	1.391	1.124	1.72
EIF3C	-1.04772	0.34798	9.0653	0.0026	0.351	0.177	0.694
CPNE6	0.33732	0.11439	8.6952	0.0032	1.401	1.12	1.753
TUBA1A	0.44037	0.15337	8.244	0.0041	1.553	1.15	2.098
RPS14*	-0.64519	0.22592	8.1556	0.0043	0.525	0.337	0.817
GPHN	0.44548	0.15631	8.1221	0.0044	1.561	1.149	2.121
ANXA11*	0.27504	0.09713	8.0179	0.0046	1.317	1.088	1.593
PPP1R12A*	-1.23054	0.43941	7.8424	0.0051	0.292	0.123	0.691
AP1G1	0.83198	0.29958	7.7128	0.0055	2.298	1.277	4.134
MTDH	-0.63924	0.2339	7.4688	0.0063	0.528	0.334	0.835
WIBG	-0.58575	0.21444	7.4613	0.0063	0.557	0.366	0.848
ACIN1	-0.58334	0.21379	7.4451	0.0064	0.558	0.367	0.848
LASP1	-0.7371	0.27361	7.2578	0.0071	0.478	0.28	0.818
THRAP3	-0.49936	0.18628	7.1865	0.0073	0.607	0.421	0.874
CDC42	0.51331	0.1933	7.0515	0.0079	1.671	1.144	2.44
PDCD6	0.3683	0.13911	7.0097	0.0081	1.445	1.1	1.898
PML	-0.37898	0.14353	6.9716	0.0083	0.685	0.517	0.907
IGHV3_20	0.24137	0.09238	6.8269	0.009	1.273	1.062	1.526
CDC5L	-0.52553	0.20365	6.6596	0.0099	0.591	0.397	0.881

* proteins that remained significantly correlated to survival after step by step and bootstrap analyses.

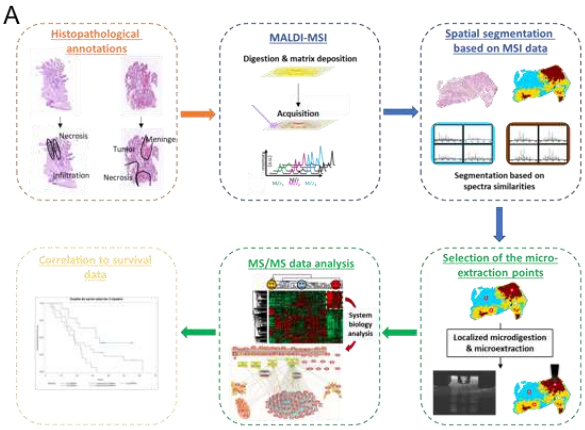
Figure 1: Histological, MALDI MSI and SpiderMass data

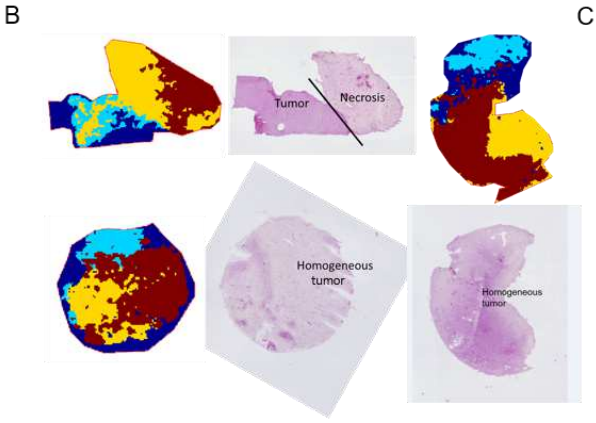
A. General workflow of the MALDI-MS imaging combined with microproteomics used for glioblastoma inter- and intra-tumor heterogeneities characterization.

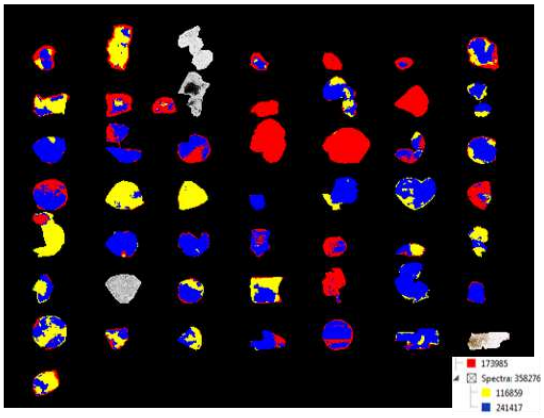
B. Representative annotated histopathology images of three glioblastoma samples and their corresponding segmentation map obtained from MALDI-MSI data. Colors represent molecularly different regions. Note that for 2 different tissues, similar colors are not equivalent to similar molecular groups. The segmentation map shows different clusters for each case and non-observable with HES coloration.

C. Global segmentation maps of all tissues together after MALDI-MSI analysis. Colors represent molecularly different regions as shown in the corresponding dendrogram. The segmentation map gives 3 main clusters. The 4 tumors which are not segmented correspond to the IDH mutant tumors, which were excluded from the analysis.

D. The built PCA-LDA classification model based on 3 glioma groups; Group A (red), Group B (yellow), Group C (blue). a) LDA representation of the 3-class PCA-LDA (right). The table (right) represents the “20% out” and “leave-one-patient-out” cross-validation results of the built classification model. b) LD2 loading spectra (top) indicate the discrimination between Group A (red) and Group B (yellow). The 10 most discriminatory lipid peaks are indicated by the blue dash line. LD1 loading spectra (bottom) indicate the discrimination between Group A (red) and Group C (blue). The 10 most discriminatory lipid peaks are indicated by the blue dash line.

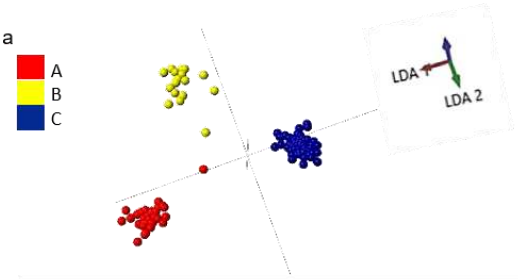






D

a



Model type	Acquisition methods	n classes	n tested spectra	n passes	n failures	N outliers	Correct classification rate (%)	
							Excluding outliers	Including outliers
Cross Validation by 20% randomly taken								
Glioma	Positive	3	135	124	0	11	100%	91.85%
Cross Validation by 1 patient out method								
Glioma	Positive	3	135	105	8	22	92.92%	77.78%

b

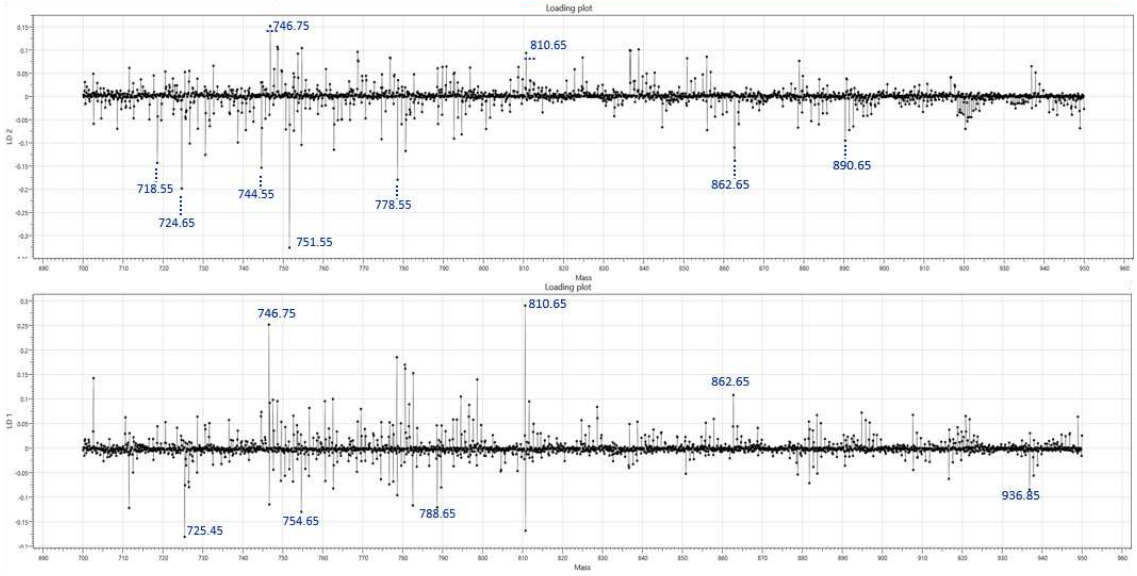


Figure 2. Shotgun microproteomics analysis

A. Matrix correlation map between all microextraction points from the 46 tumors. Correlation coefficients are calculated between each sample and are represented on a heatmap.

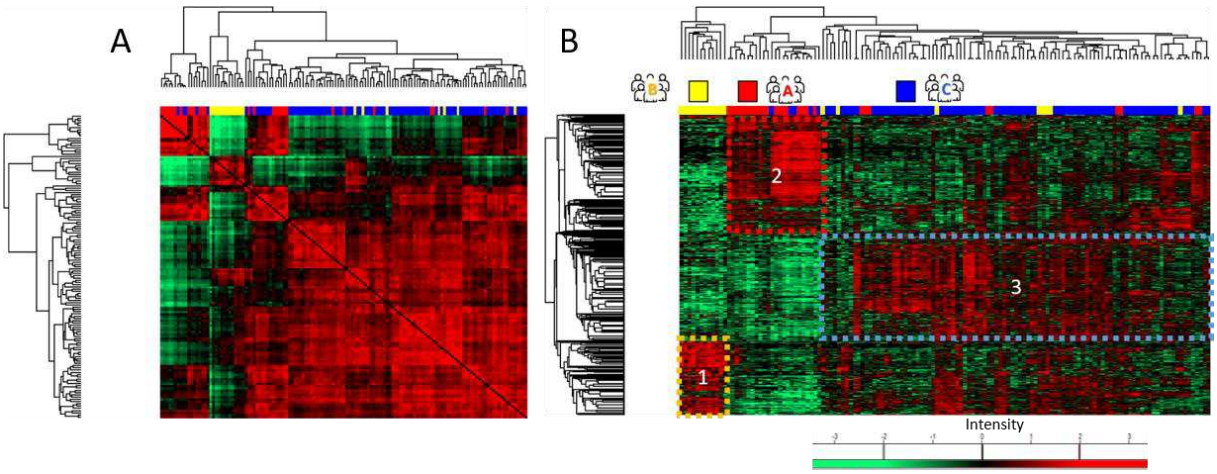
B. Heatmap of proteins with different regulation profiles as determined after label free quantification in the three groups highlighting the presence of 3 clusters. Shotgun proteomics was performed after on-tissue trypsin digestion followed by microextraction at the spots determined from MALDI MSI data.

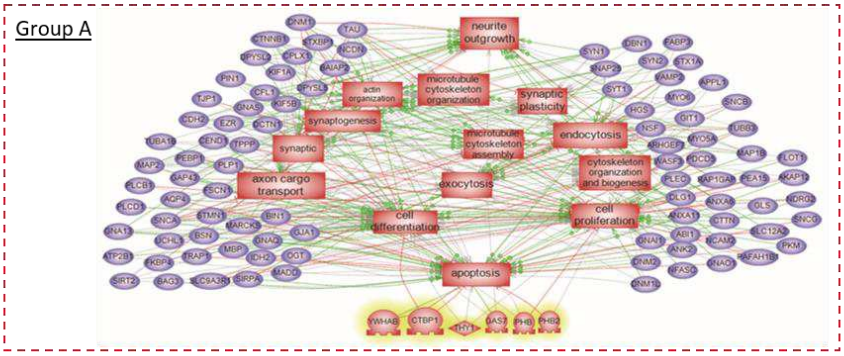
C. Pathway analysis of proteins overexpressed in group A reveals that a large majority of protein is involved in (a) neurogenesis, brain development, synaptogenesis and cytoskeleton organization.

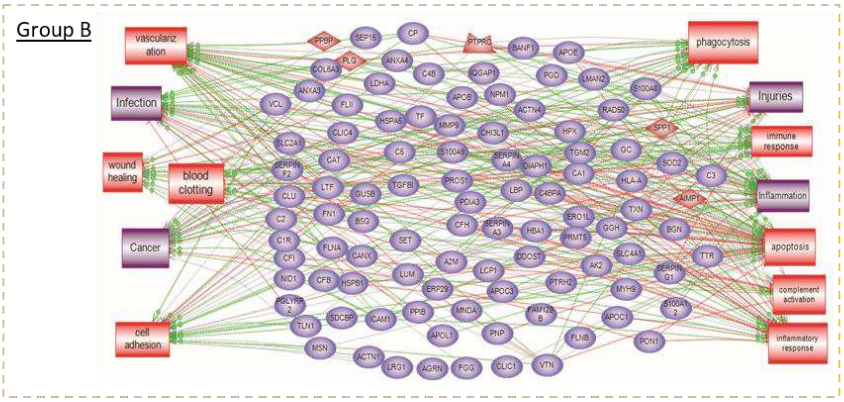
D. Pathway analysis of proteins overexpressed in group B reveals that majority of proteins are involved in injuries, inflammation and more generally immune system response and vascularization.

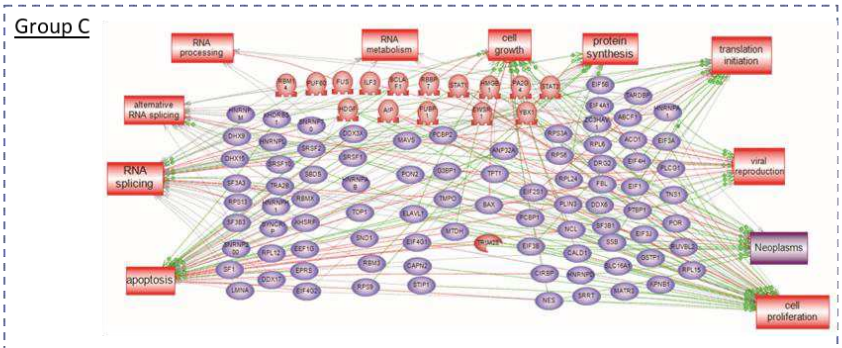
E. Pathway analysis of proteins overexpressed in group C shows implication in cell proliferation, neoplastic processes, RNA metabolism and processing and viral reproduction.

F. Heatmap of alternative proteins with different regulation profiles as determined after label free quantification in the three regions highlighting the presence of 3 clusters.


A
B
2
3
1
Intensity


C
Group A
neurite outgrowth
microtubule cytoskeleton organization
synaptic plasticity
synaptogenesis
endocytosis
microtubule cytoskeleton assembly
synaptic
axon cargo transport
exocytosis
cytoskeleton organization and biogenesis
cell differentiation
cell proliferation
apoptosis


D
Group B
vascularization
Infection
wound healing
blood clotting
Cancer
cell adhesion
phagocytosis
Injuries
immune response
Inflammation
apoptosis
complement activation
inflammatory response


E
Group C
RNA processing
RNA metabolism
cell growth
protein synthesis
translation initiation
alternative RNA splicing
RNA splicing
viral reproduction
Neoplasms
apoptosis
cell proliferation

F

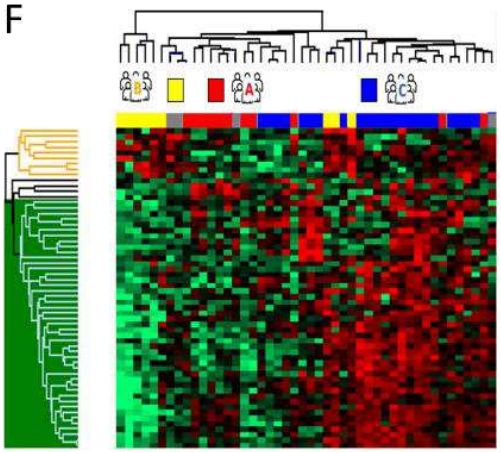


Figure 3. Proteomic and survival analysis

A,B. Analysis of maximum likelihood estimates of the 5 proteins significantly correlated with survival (ANXA11, RPS14, ALCAM, PPP1R12A and AltProt IP_652563) identified after a step by step analysis and bootstrap procedure and B. patient clustering based on these proteins

C. Overall survival of the 46 patients according to the expression of the 5 prognostic markers. Two clusters of patients were identified with a clear difference in their survival. Cluster 1 has longer survival than cluster 2.

D. Heatmap of the 28 proteins significant in the Cox model (p=0.01) between the 2 groups of patients defined by their OS (left).

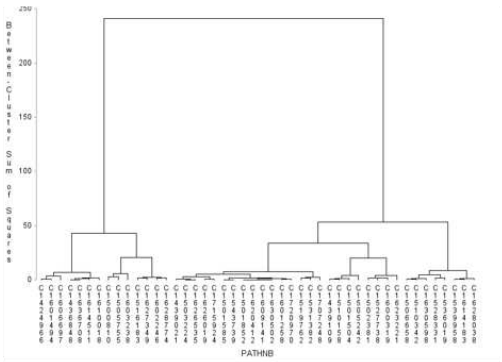
E. Boxplots of the 28 prognosis proteins significant after applying the Cox model. Their LFQ values were compared between patients of cluster 1 (long survival) and cluster 2 (short survival).

F. Boxplots of the 5 prognostic markers identified after a step by step analysis and bootstrap procedure. Their LFQ values were compared between patients of cluster 1 (long survival) and cluster 2 (short survival).

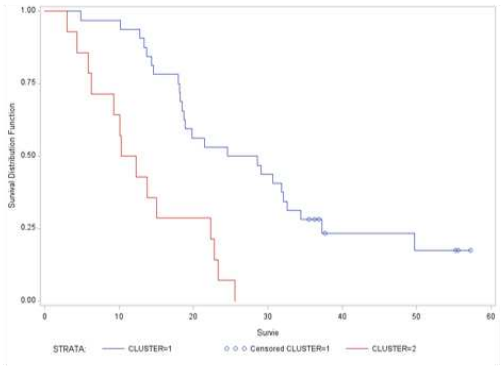
A

Analysis of Maximum Likelihood Estimates								
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
ALCAM	1	0,692	0,188	13,552	0,000	1,997	1,382	2,886
RPS14	1	1,476	0,536	7,583	0,006	4,377	1,530	12,520
ANXA11	1	0,463	0,157	8,714	0,003	1,589	1,168	2,162
PPP1R12A	1	-2,056	0,675	9,286	0,002	0,128	0,034	0,480
IP_652563	1	0,268	0,080	11,262	0,001	1,308	1,118	1,529

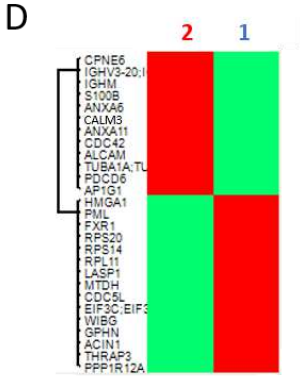
B



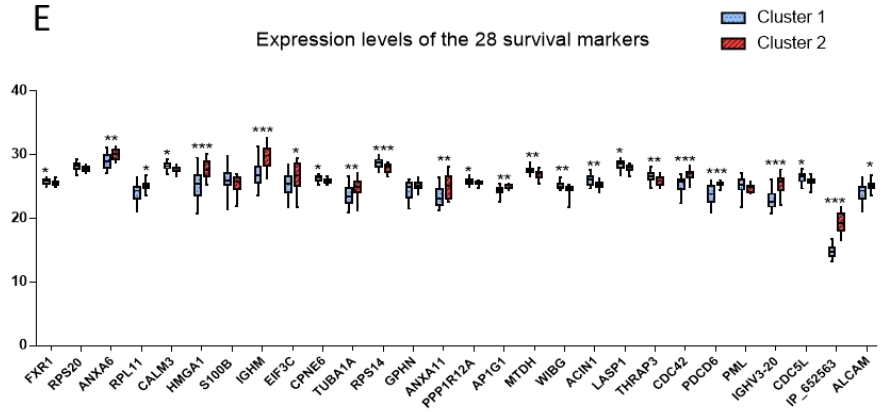
C



D



E



F

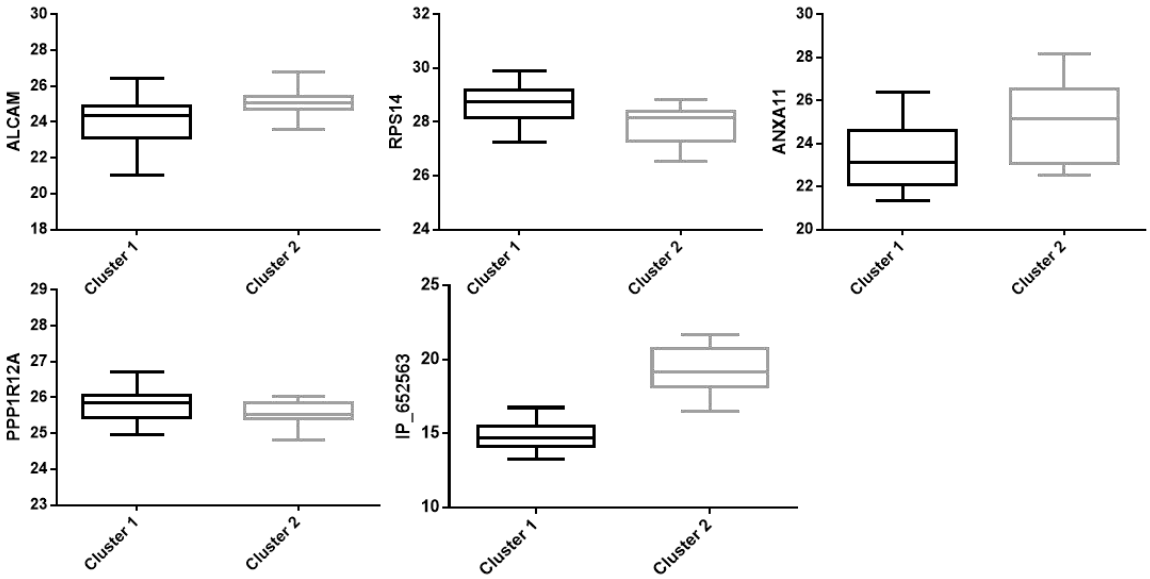
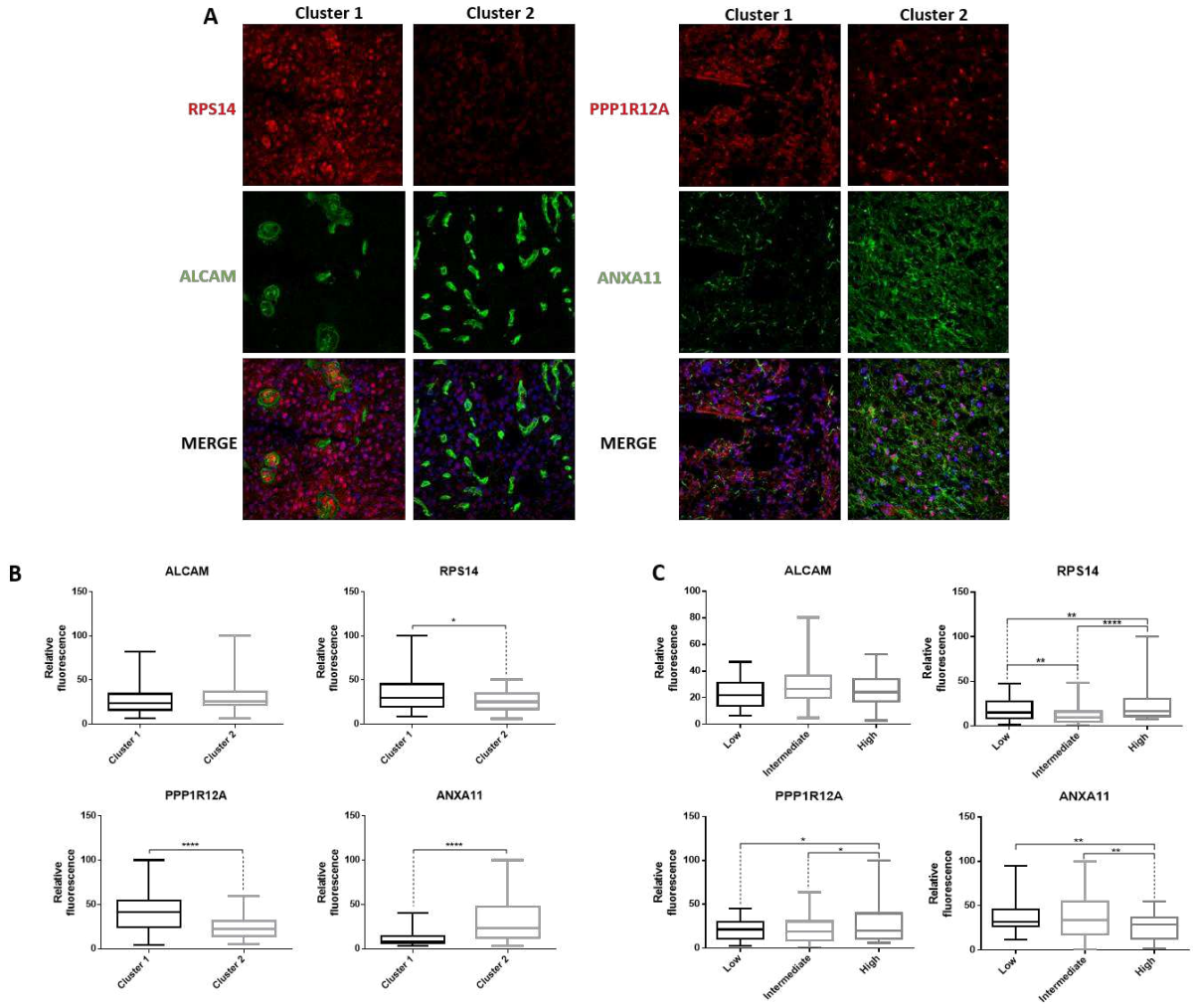


Figure 4. Validation immunohistochemistry of the panel of survival markers identified.

A. Representative fluorescence images of the 4 proteins in the two OS clusters of patients. ANXA11 and ALCAM are associated to a bad prognosis while PPP1R12A and RPS14 are related to a good prognosis. Images were acquired with a confocal microscope at 40x magnification.

B. Quantification of fluorescence intensities of the 4 proteins in the two OS clusters. Images taken from 14 tumors of cluster 1 and 9 tumors of cluster 2 were quantified. For each tumor, 3 to 4 images were acquired and quantified. Significant differences were identified using unpaired t test with **** $p<0.0001$; *** $p<0.001$; ** $p<0,01$ and * $p<0.05$.

C. Quantification of fluorescence intensities of the 4 proteins in an external cohort of glioblastoma patients (50 patients). Patients were classified according to their survival times (low, intermediate and high). The fluorescence intensities of images taken from 50 tumors were quantified. For each tumor, 3 to 4 images were acquired and quantified. Significant differences were identified using unpaired t test with **** $p<0.0001$; *** $p<0.001$; ** $p<0,01$ and * $p<0.05$.


A
Cluster 1
Cluster 2
RPS14
ALCAM
MERGE
PPP1R12A
ANXA11
B
C
Relative fluorescence
Low
Intermediate
High

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supp.Data5OSversuscluster.xlsx
- Supp.Data4Subnetworkperregion.xlsx
- Supp.Data3clusterGROUPRougeAJauneBBleuC.xlsx
- Supp.FiguresandTables.pdf
- Supp.Data2Totalmatrix.xlsx
- Supp.data1SegmentationindividualIDHwt.pdf