EuPA Open Proteomics 8 (2015) 85–93

Contents lists available at ScienceDirect

EuPA Open Proteomics

journal homepage: www.elsevier.com/locate/euprot

The proteomic landscape of glioma stem-like cells

Cheryl F. Lichti[a,b,*], Norelle C. Wildburger[a,c], Alexander S. Shavkunov[a], Ekaterina Mostovenko[a], Huiling Liu[a], Erik P. Sulman[d], Carol L. Nilsson[a,b]

[a] Department of Pharmacology & Toxicology, University of Texas Medical Branch, 301 University Blvd, Galveston, TX 77555-0617, United States
[b] UTMB Cancer Center, University of Texas Medical Branch, 301 University Blvd, Galveston, TX 77555-1074, United States
[c] Neuroscience Graduate Program, Graduate School of Biomedical Sciences, University of Texas Medical Branch, 301 University Blvd, Galveston, TX 77555-1074, United States
[d] Department of Radiation Oncology, The University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, United States

ARTICLE INFO

Article history:
Received 27 February 2015
Received in revised form 26 June 2015
Accepted 29 June 2015
Available online 17 July 2015

Keywords:
Glioblastoma
Glioma stem cells
Label-free quantitative proteomics
Ingenuity Pathway Analysis
DAVID
Unsupervised hierarchical clustering

ABSTRACT

Glioma stem-like cells (GSCs) are hypothesized to provide a repository of cells in tumors that can self-replicate and are radio- and chemo-resistant. GSC lines, representing several glioma subtypes, have been isolated and characterized at the transcript level. We sought to characterize 35 GSC lines at the protein level using label-free quantitative proteomics. Resulting relative fold changes were used to drive unsupervised hierarchical clustering for the purpose of classifying the cell lines based on proteomic profiles. Bioinformatics analysis identified synoviolin, serine/arginine-rich splicing factor 2, symplekin, and IL-5 as molecules of interest in progression and/or treatment of glioma.

1. Introduction

Primary brain tumors comprise 3% of all cancer diagnoses, and of these, GBM [World Health Organization (WHO) grade IV astrocytoma] makes up over half (52%) of all cases [1,2]. Despite an aggressive therapeutic regimen including surgical resection followed by some combination of chemotherapy and radiotherapy, the disease is ultimately fatal, with median survival only slightly over a year after diagnosis [2–6]. A deeper molecular-level understanding of the origins of GBM is critical in the quest to discover new therapeutic targets for this disease that few survive.

One factor that contributes to poor clinical outcome is the presence of a small subpopulation (<1%) of cells within the tumor which are both radio- and chemotherapy resistant [7–11]. These cells, termed glioma stem cells or glioma stem-like cells (GSCs) are postulated to provide a repository of cells for tumor recurrence [7–11]. GBM has been classified into several subgroups based upon patterns of gene expression [12–16], and this same classification scheme may be applied to GSCs. According to The Cancer Genome Atlas (TCGA), the GBM subgroups are classical, proneural, and mesenchymal [12–16]. Tumors themselves are not homogeneous; a recent study of biopsy samples revealed that cells from different regions of the same tumor show different molecular phenotypes [17]. Therefore, it is critical to understand GSCs at a molecular level in order to design an effective treatment regimen for GBM.

The purpose of our study was to perform a proteomic comparison of 35 GSCs, derived from patient tumors, in order to gain a deeper understanding of protein-level changes associated with GSCs and to identify potential therapeutic targets. The origin of GSCs has yet to be definitively determined, although glioma may originate from neural stem cells [18–22], glial cells [23], oligodendrocyte precursor cells [24], neurons [25] or astrocytes [25,26]. Given the difficulty in identifying an appropriate control for GSCs, each cell line was quantified relative to a mixed control sample containing an equal amount of protein from each cell line. The Catalog of Somatic Mutations in Cancer (COSMIC) database [27–31] was queried in order to compare our results with genome-level studies of patient tumor samples, especially for those proteins with no previous association to GBM. A combination of the Database for Annotation, Visualization and Integrated Discovery (DAVID) webtool [32,33] and Ingenuity Pathway Analysis (IPA) was used to determine affected pathways for each cell line, and IPA was used to determine predicted upstream regulators. Unsupervised hierarchical clustering of proteins and of upstream regulators was used to determine similar

* Corresponding author. Department of Pharmacology & Toxicology, University of Texas Medical Branch, 301 University Blvd, Galveston, TX 77555-0617, United States.
E-mail address: cflichti@utmb.edu (C.F. Lichti).

http://dx.doi.org/10.1016/j.euprot.2015.06.008

behavior between the cell lines. We identified symplekin (SYMPK) as a protein whose expression is significantly changed across several of our cell lines [34]. In addition, we identified novel putative upstream regulators interleukin-5 (IL5) and synoviolin (SYVN1). From a cluster of proteins which demonstrated lower expression in mesenchymal stem cells, we identified serine/arginine rich splicing factor 2 (SRSF2) as an upstream regulator.

2. Materials and methods

2.1. Chemicals and Reagents

LC-MS grade acetonitrile and water were purchased from J.T. Baker (Philipsburg, NJ). Formic acid and RIPA buffer were purchased from Pierce (Rockford, IL). Iodoacetamide (IAA), dithiothreitol (DTT), triethylammonium bicarbonate (TEAB) were obtained from Sigma–Aldrich (St. Louis, MO). Sequencing grade trypsin was purchased from Promega (Madison, WI), and PMSF from CalBiochem (Darmstadt, Germany). All chemicals were used without further purification.

2.2. Cell culture conditions

Isolation of GSCs from patient tumors was performed as previously described [19] in accordance with the institutional review board of The University of Texas M.D. Anderson Cancer Center, and are named in the order they were acquired. GSCs were cultured according to previously published methods [19,35]. Upon dissociation of cells, GSCs were enriched using CD133 via flow cytometry. CD133+ cells are grown in serum-free medium as neurospheres as previously described [21,22]. All cell lines were tested to exclude the presence of *Mycoplasma* infection.

2.3. Proteomic analysis of GSCs

Sample preparation and nanoLC–MS/MS analysis of GSCs was performed as previously described [34]. Briefly, protein (100 μg) isolated from 2×10^6 cells was reduced (TCEP) and alkylated (iodoacetamide). After precipitation using four volumes (440 μL) of ice cold acetone for 2 h at −20 °C, protein was resuspended in 8 M urea (12.5 μL) and digested with trypsin (10 μg in 87.5 μL of TEAB buffer) for 24 h at 37 °C.

Chromatographic separation and mass spectrometric analysis was performed with a nanoLC chromatography system (Easy-nLC 1000, Thermo Scientific), coupled on-line to a hybrid linear iontrap–Orbitrap mass spectrometer (Orbitrap Elite, Thermo Scientific) through a Nano-Flex II nanospray ion source (Thermo Scientific) as previously described [34]. Briefly, samples were analyzed in groups of block-randomized triplicates [36], with three GSCs and M37 (mixed control sample consisting of equal protein from 37 cell lines) in each group. Each block consisted of one technical replicate of each GSC and the M37; the run order for each block was randomized so that data acquisition was not performed in the same order for any of the three blocks. A total of three blocks was acquired for each group of samples, resulting in acquisition of three data files for each GSC and M37. Peptides (1 μg cell protein digest) were separated by gradient elution using a C_{18} column (10 cm × 75 μm ID,15 μm tip, ProteoPep II, 5 μm, 300 Å, New Objective) using a 4 h gradient. Mobile phases were 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (ACN; B). All nanoLC–MS/MS data were acquired using XCalibur, version 2.7 SP1 (Thermo Fisher Scientific) using a Top 10 HCD method as previously described [34]. The data files have been deposited into ProteomeXchange repository (PXD001890) [37–40].

2.4. Bioinformatic analysis

Data files were analyzed as previously described [34]. Instrument .raw files for each experimental block were imported into Progenesis LC–MS software (version 18.214.1528, Nonlinear Dynamics) for *m/z* and retention time alignment. This process combines observations for all samples in the block as single measurements for each peptide feature, which allows the best peptide spectrum match for a particular peptide feature to be projected onto all runs within the experimental block. Next, the top 5 spectra for each feature were exported as a combined .mgf file for database searching in PEAKS [41–43] (version 6, Bioinformatics Solutions Inc., Waterloo, ON) and Mascot (version 2.3.02, Matrix Science). Database searches were performed as previously described [34] (10 ppm parent ion tolerance, 0.025 Da fragment ion tolerance, fixed carbamidomethyl cysteine, variable oxidation (methionine), deamidation (asparagine, glutamine) and phosphorylation (serine, threonine, tyrosine), with a maximum of three post-translational modifications per peptide; trypsin with two missed cleavages). Peptide-spectrum matches were then exported from PEAKS as a .xml file and re-imported into Progenesis LC–MS in order to assign peptide-spectrum matches to features. After filtering to remove peptide-spectrum matches below 95% peptide probability (as calculated in PEAKS, using the Peptide Prophet algorithm [44]), manual conflict resolution was performed by removing lower scoring peptide spectrum matches in order to ensure that a single unique peptide sequence was assigned to each feature. Feature intensities were normalized using the default normalization algorithm in Progenesis LC-MS (http://www.nonlinear.com/progenesis/qi-for-proteomics/v2.0/faq/how-normalisation-works.aspx), and normalized peptide intensity data was exported and filtered to remove non-unique peptides, methionine-containing peptides [45], and all modified peptides except those containing cysteine carbamidomethylation. Peptide intensities were imported into DanteR (version 0.1.1) [46,47] for protein quantification as previously described [34]. Briefly, intensities for peptides with the same sequence were combined into a single entry by summation, in order to correct for MS1-level misalignment and to fold together measurements representing multiple charge states of the same peptide. The resulting peptide intensities were $\log_2$-transformed and combined to generate protein abundances (RRollup) without considering proteins with a single peptide assignment. Default settings were used: 50% minimum presence of at least one peptide, minimum dataset presence 3, *p*-value cutoff of 0.05 for Grubbs' Test, minimum of 5 peptides for Grubbs' Test. A one-way ANOVA was performed for each experimental block, relative to M37, to obtain estimated fold changes, and *p*-values were corrected for multiple testing [48]. In order to compare results across experimental blocks, estimated protein fold changes from all experimental blocks were standardized to correct for analytical differences, and the resulting standardized fold changes were imported into Ingenuity Pathway Analysis. Resulting z-scores for upstream regulators and biological and disease functions for all cell lines were exported and collated. After filtering to remove entries with no missing values, unsupervised hierarchical clustering of proteins was performed using a Euclidean distance metric and a Ward linkage metric in Mass Profiler Pro (Agilent, Santa Clara, CA). After filtering to remove entries with <20% missing values, unsupervised hierarchical clustering of upstream regulators was performed using a Euclidean distance metric and an average linkage metric in DanteR.

For additional analysis of biological function, Gene Ontology (GO) analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) webtool [32,33]. For each cell line, the proteins were separated into lists of proteins which were increased and decreased relative to mixed control.