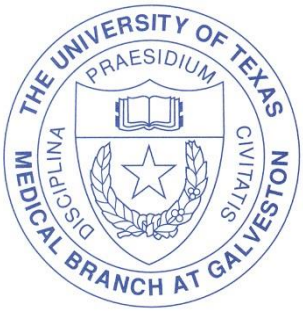
Report on research program (summer 2015)

Research Development Project on Glioma Stem Cells by Quantitative Mass Spectrometry

K. Barbara Sahlin

September, 2015

LUND UNIVERSITY



CANCER PREVENTION & RESEARCH INSTITUTE OF TEXAS

Content

1. Objective of the Study

The objective of the study was to analyze the protein expression in glioma stem cells (GSCs) of glioblastoma multiforme (GBM) from biobanked materials. Our goal was to validate and quantify expression of proteins containing single amino acid variants (SAVs) that were previously identified in a collaboration between University of Texas Medical Branch (UTMB) and Lund University (LU) (1).

2. Background

GSCs are multipotent and invasive cells derived from GBM, the most common type of primary brain cancer. Fewer than 10% of patients survive five years after diagnosis due to GSCs ability to cause tumor recurrence. GSCs are aggressive and resistant to radiation and chemotherapies. Our hypothesis is that GSCs harbor protein SAVs due to germline or somatic mutations, where proteins that may have cell protective properties are transformed into promoters of genetic instability, invasivity, or epigenetic or metabolic control (1).

Proteomics is the study of the proteome, the protein complement of the genome and genomics. These "omics" terms symbolize the redefinition of how we think about biology and the workings of living systems. There are four essential tools in proteomics: analytical protein-separation technology, mass spectrometry (MS), software matching MS data, and specific protein sequences in databases (2).

Mass spectrometry (MS) is defined as an instrumental method for identifying the chemical constitution of proteins by means of the separation of gaseous ions according to their differing mass and charge. The unit is called mass-to-charge, *m/z* (3,4). There are several different types of mass spectrometers available, for example the quadrupole orbitrap hybrid, Q Exactive, which is used in the present study. In quantitative MS, synthetic heavy labeled peptides are spiked into biological samples in a defined quantity. These peptides are isotopes of the endogenous peptides and, therefore, maintain the same chemical properties. The synthetic heavy labeled peptides provide an internal standard and enable quantification (5).

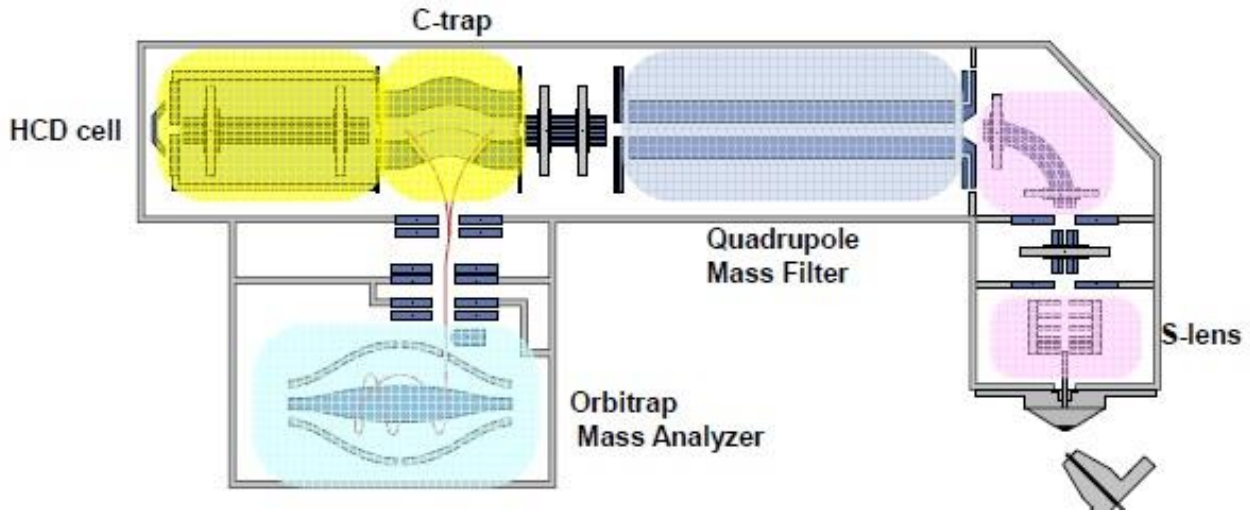


Figure 1. A representation of how the Q Exactive acquires data. In the bottom right corner there is a steady flow of prepared sample from the nano-Liquid Chromatography needle, which is electrosprayed as it exits the column. The peptides become gas phase ions and enter the Q Exactive. The S-lens (pink color) focuses the sample and provides transmission. The Quadrupole Mass Filter (gray color) filters the peptides of interest. The C-trap (yellow color) is a curved linear trap which collects the fragments prior to detection in the orbitrap (blue color). HCD cell (yellow color) stands for Higher-energy Collisional Dissociation and can be utilized during data acquisition for MS/MS data. In such a case, the ionized peptides enter the HCD cell from the C-trap and return to the C-trap as smaller fragments to be transported to the Orbitrap for high resolution detection (9).

The Q Exactive provides high mass resolution, accurate mass determination, and fast cycle time in data collection. The Q Exactive is comprised of four main functional parts: 1) ion source, 2) ion optics, 3) ion selection in the quadrupole and 4) ion detection in the oribtrap. The instrument can provide fragment ion data, MS/MS, which allows us to sequence peptides (Fig. 1).

There is currently a great need to further understand the pathophysiology in the recurrence of GBM to identify and develop new drug targets. This study aims to quantify protein SAVs that have previously been identified in GSCs as germline or a result from somatic mutation (1). We hypothesize that some of these mutations are involved in the disease progression of GBM, "driver" SAVs.

3. Methods

Previously, proteins were extracted from biobanked GSC lines from MD Anderson Cancer Center in Houston, Texas (6). The study was approved by the Institutional Review Board of The University of Texas MD Anderson Cancer Center, and informed consent was obtained from all subjects (N=36). Proteins were extracted from all cell lines and digested into peptides with trypsin.

Samples were further prepared by reduction and alkylation, then spiked with known amounts of synthetic heavy labeled peptides (Thermo Fisher Scientific) for quantitative mass spectrometry (MS) analysis (7). Peptide mixtures were separated

using a nano-Liquid Chromatrography II system (Easy nLC, Thermo Fisher Scientific) with a 300-µl/min flow rate on a C_{18} column, where Solvent A consisted of 0.1% formic acid in water and Solvent B of 0.1% formic acid in acetonitrile. The peptides were eluted with a 45-min linear gradient starting with 10% B to 35% B, followed by a 5-min linear gradient to 90% B and a column wash at 90% B for 5 min (8). Analyses of SAV peptides were scheduled (scheduled reaction monitoring) and acquired in a Q Exactive MS (Thermo Fisher Scientific). By use of a proprietary database that includes all protein variants (derived from Ensembl), we identified hundreds of variant peptides in our samples (1).

Method setup and data for the mass spectrometry analysis was performed by use of Skyline software (10). We screened all cell lines that harbored the corresponding variant RNA transcript with parallel reaction monitoring (PRM) in a Q Exactive MS. RPKM (quantitative mRNA data) was calculated in 46 cell lines for the corresponding SAV transcript in all GSCs in the MD Anderson Cancer Center biobank (1).

4. Results

We validated and quantified mutant proteoforms in GSCs within the present study. A total of 189 peptides were quantified in our study, however, data from some of the peptides were excluded due to the peptides behaving poorly during data acquisition. For example, peptides containing the amino acid methionine are easily oxidized

which change how the peptide behaves during the nLC separation. Some of the GSC samples were excluded entirely due to problems during the data acquisition. The final number of GSCs that provided adequate data was 29 (Fig. 2).

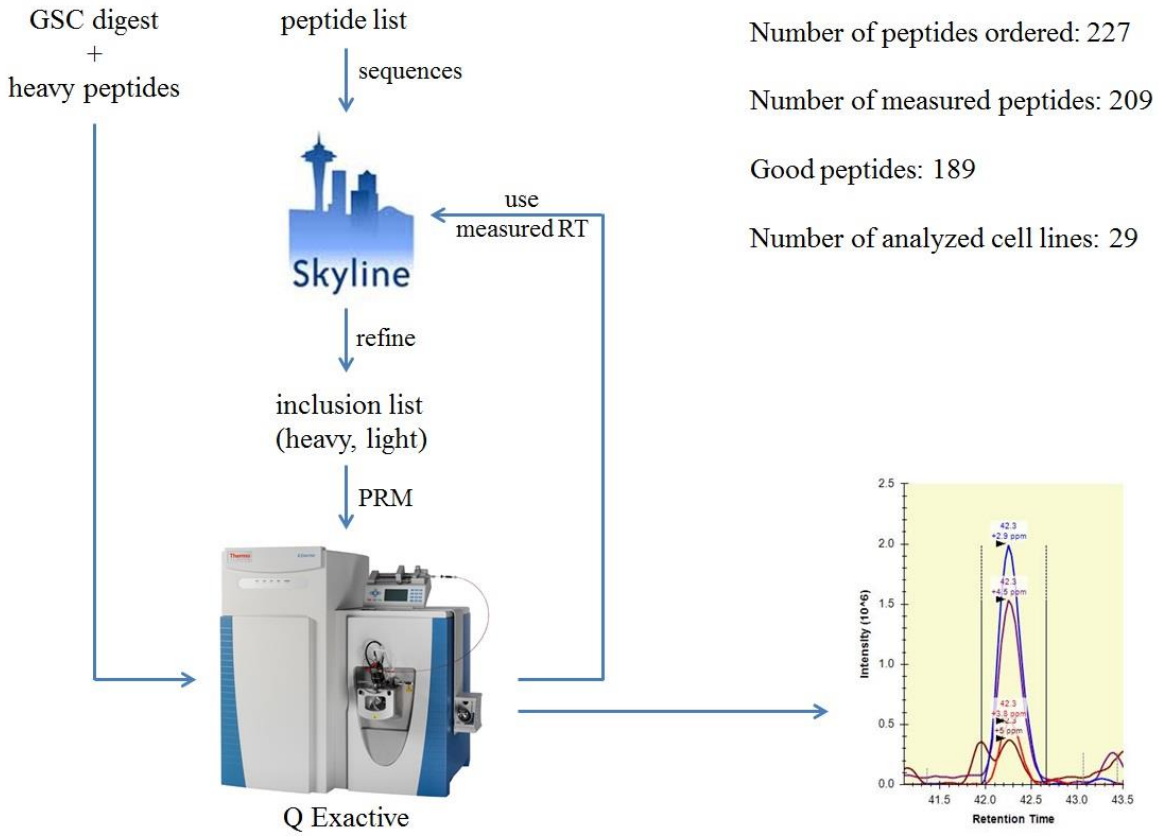


Figure 2. Method utilized at Lund University. The initial cell line count was 36, however, seven were excluded due to problems in the data acquisitions. The peptide list was created based on previous studies (1), which was the base for the method on the data acquisition on the Q Exactive. The peptides ordered were synthetic heavy labeled peptides that were of interest in the GSC samples. The MS provided *m/z* and intensity for the peptides in the sample.

Branched chain amino acid transferase 2 (BCAT2) T186R was the first novel target we identified in GSCs (1). BCAT2 T186R is a germline variant with prevalence of 25% in

patient tumor-derived GSCs, compared to 9.6% of the general population (1000 Genomes Data). This mitochondrial enzyme metabolizes branched chain amino acids (BCAAs), which are easily absorbed through the blood-brain barrier, and produces glutamate and alpha-ketoacids (11). Glutamate is neurotoxic and allows expansion of the brain tumor. Alpha-ketoacids are carbon skeletons for metabolism and fuel tumor growth (Fig. 3).

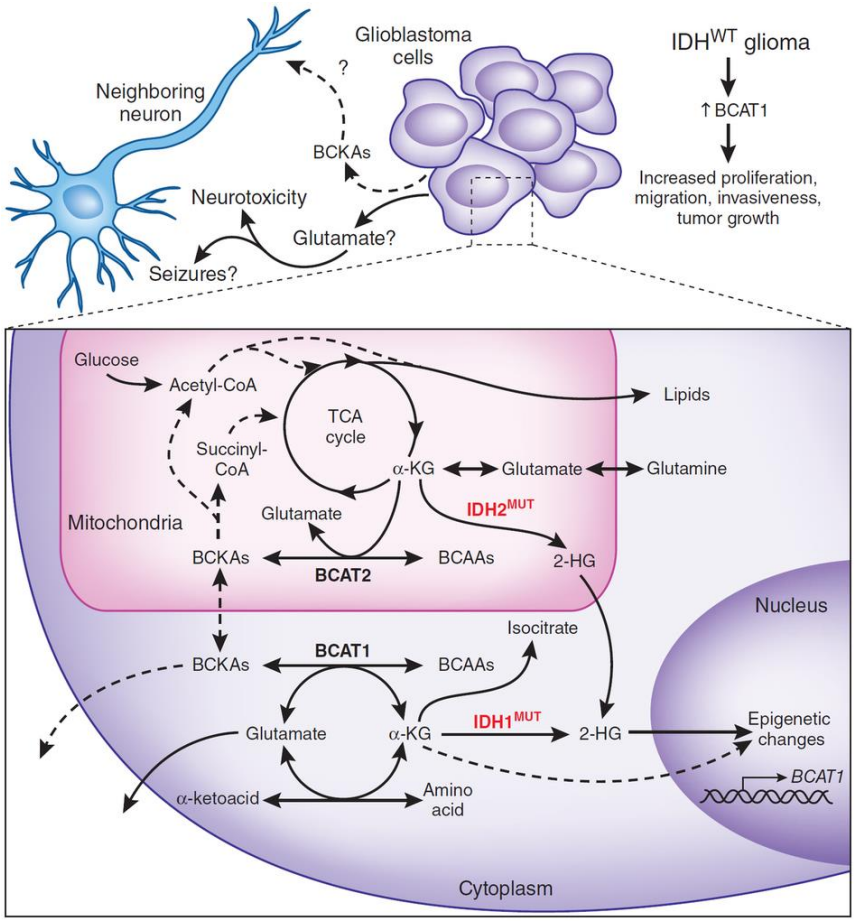


Figure 3. Overview of metabolic enzymes implicated in GBM. Branched chain amino acid transferases (BCATs) transfer branched chain amino acids (BCAA) nitrogen to α-ketoglutarate (α-KG) to produce glutamate and the corresponding keto-acid, which can be further metabolized to the tricarboxylic acid cycle (TCA) intermediates. Glutamate may contribute to the neurotoxicity observed in GBM. Mutant isocitrate dehydrogenase (IDH) proteins consume α-KG to produce 2-hydroxyglutarate (2-HG) and influence chromatin modifications (11). BCAT2 has not previously been implicated in GBM pathology. Figure is taken from reference 11.

GSCs that express the BCAT2 variant, therefore, may have a growth advantage in the brain environment.

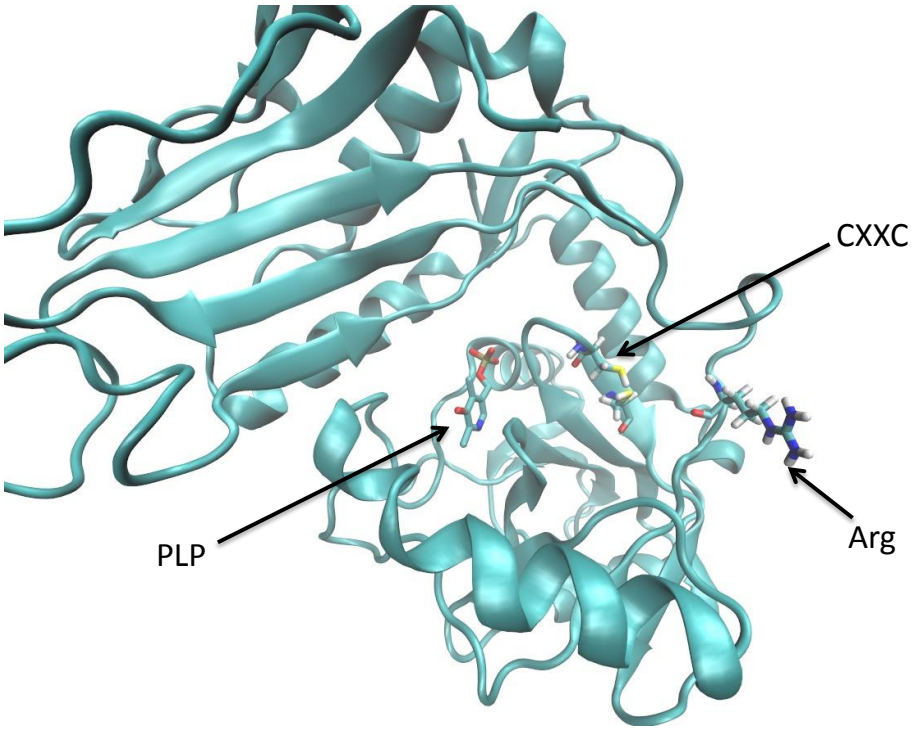


Figure 4. Crystal structure of oxidized human mitochondrial branched chain aminotransferase 2 monomer with computationally generated T186R mutation. It has been demonstrated that the CXXC center (C321, C324) contributes to enzyme activity through reversible disulfide bond formation that prevents substrate from correctly orienting with the pyridoxal phosphate (PLP) co-factor. The Thr186Arg substitution is located along a flexible loop in close proximity to the CXXC center and thus may interact with the Cys residues, influencing the reversibility of the oxidation reaction and thereby enzyme activity (1). Figure is taken from reference 1.

From the original 227 SAVs, we identified a subset of 73 variants and their paired non-variant peptides for parallel reaction monitoring (PRM), in a multiplexed approach. Ten SAVs are expressed in 20 or more cell lines. Some variants have a known correlation to GBM, such as Nestin V130A (Fig. 5).

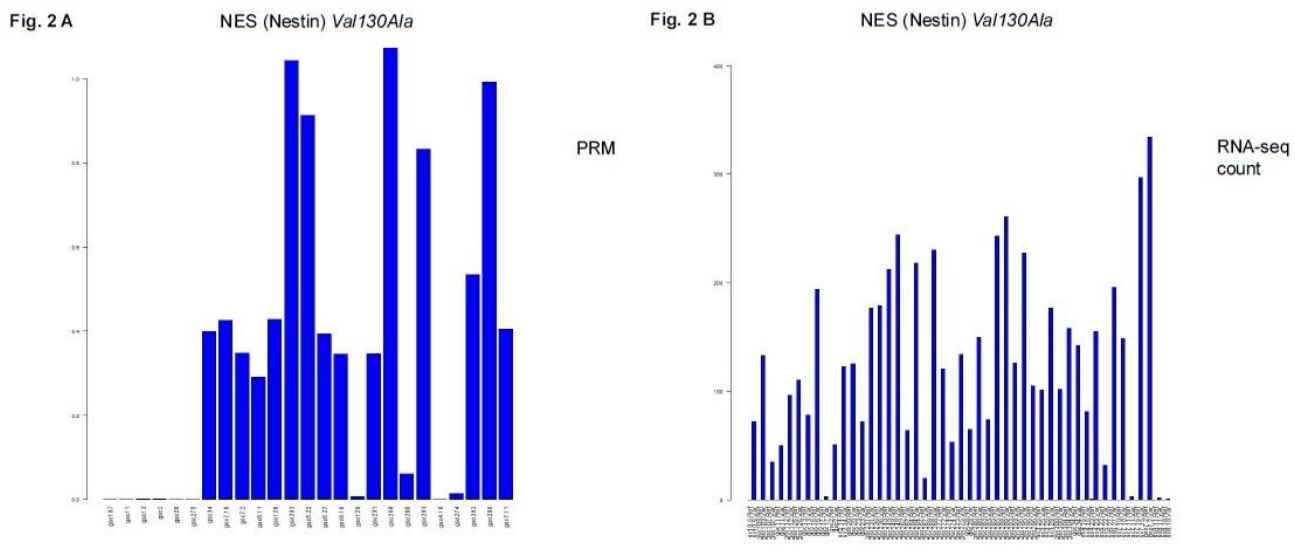


Figure 5. Comparison of quantitative proteomics and RPKM (reads per kilobase of transcript per million reads mapped). The x axis is GSC line and the y axis is arbitrary units based on counts. **A.** Quantitative data derived from PRM analysis of a Nestin SAV peptide. The variant was reliably detected in most GSC lines. **B.** RPKM (quantitative mRNA data) for the corresponding SAV transcript in all GSCs in the MDACC biobank (n=46).

5. Conclusions

We could conclude that the PRM methodology was successful in the validation and quantification of SAVs in GSCs. We suspect that other SAVs, along with BCAT2 T186R and Nestin V130A, may be oncogenic "drivers", and that they also may be expressed in brain metastatic tumors of non-neuronal origin, such as malignant melanoma. In the future, we will investigate SAV expression in six brain-metastatic and six non-brain metastatic melanoma cell lines derived from patients.

If primary and secondary (non-GBM) brain tumors share pathological similarities, new treatments could be developed to target both tumor types of the brain. Tumors

metastatic to the brain also carry high patient mortality and are more frequent than primary brain tumors. This comparison has not previously been performed. The identification of new potential therapeutic targets could improve treatments in these underserved patient categories.

References

1. Lichti CF, Mostovenko E, Wadsworth PA, Lynch GC, Pettitt BM, Sulman EP, et al. Systematic identification of single amino acid variants in glioma stem-cell-derived chromosome 19 proteins. J Proteome Res. 2015,14(2):778-86.

2. Liebler D. Introduction to proteomics. Totowa: Humana Press Inc; 2002.

3. Marshall AG, Hendrickson CL. High-resolution mass spectrometers. Annu Rev Anal Chem. 2008;1:579-99.

4. Marko-Varga G. Genomics and Proteomics for Clinical Discovery and Development. Dordrecht, Heidelberg, New York, London: Springer; 2014.

5. Picotti P, Aebersold R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nature Methods. 2012,9(6):555-566.

6. Galli R, Binda E, Orfanelli U, Cipelletti B, Gritti A, De Vitis S, et al. Isolation and Characterization of Tumorigenic, Stem-like Neural Precursors from Human Glioblastoma. Cancer Res. 2004,64(19):7011−7021.

7. Lichti CF, Liu H, Shavkunov AS, Mostovenko E, Sulman EP, Ezhilarasan R, et al. Integrated Chromosome 19 Transcriptomic and Proteomic Data Sets Derived from Glioma Stem-Cell Lines. J Proteome Res. 2014, 13(1):191-9.

8. Végvári A, Sjödin K, Rezeli M, Malm J, Lilja H, Laurell T, Marko-Varga G. Identification of a novel proteoform of prostate specific antigen (SNP-L132I) in clinical samples by multiple reaction monitoring. Mol Cell Proteomics. 2013 Oct;12(10):2761-73.

9. Kellmann M, Crone C, Xuan Y, Moehring T. Q Exactive Training Guide. Thermo Fisher Scientific; 2013.

10. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. 2010 Apr 1;26(7):966-8.

11. Tönjes M, Barbus S, Park YJ, Wang W, Schlotter M, Lindroth AM, et al. BCAT1 promotes cell proliferation through amino acid catabolism in gliomas carrying wild-type IDH1. Nat Med. 2013 Jul;19(7):901-8.