Journal of proteome research

Article

Integrated Chromosome 19 Transcriptomic and Proteomic Datasets Derived from Glioma Cancer Stem Cell Lines

Cheryl F. Lichti, Huiling Liu, Alexander S. Shavkunov, Ekaterina Mostovenko, Erik P. Sulman, Ravesanker Ezhilarasan, Qianghu Wang, Roger Kroes, Joseph R. Moskal, David Fenyö, Betül Akgöl Oksuz, Charles A Conrad, Frederick F. Lang, Frode S. Berven, Ákos Végvári, Melinda Rezeli, György Marko-Varga, Sophia Hober, and Carol L. Nilsson

J. Proteome Res., **Just Accepted Manuscript** • Publication Date (Web): 24 Nov 2013

ACS Publications
High quality. High impact.

Journal of Proteome Research is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Integrated Chromosome 19 Transcriptomic and Proteomic Datasets Derived from Glioma Cancer Stem Cell Lines

Cheryl F. Lichti[1], Huiling Liu[1], Alexander S. Shavkunov[1], Ekaterina Mostovenko[1], Erik P. Sulman[2], Ravesanker Ezhilarasan[2], Qianghu Wang[3], Roger A. Kroes[4], Joseph C. Moskal[4], David Fenyö[5], Betül Akgöl Oksuz[5], Charles A. Conrad[6], Frederick F. Lang[7], Frode S. Berven[8], Ákos Végvári[9], Melinda Rezeli[9], György Marko-Varga[9], Sophia Hober[10], and Carol L. Nilsson*[1]

[1]Department of Pharmacology and Toxicology, UTMB Cancer Center, University of Texas Medical Branch, Galveston, Texas 77555, United States

[2]Department of Radiation Oncology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas, United States

[3]Department of Bioinformatics, The University of Texas M.D. Anderson Cancer Center, Houston, Texas, United States

[4]The Falk Center for Molecular Therapeutics, McCormick School of Engineering and Applied Sciences, Northwestern University, 1801 Maple Street, Evanston, Illinois 60201, United States

[5]Department of Biochemistry, New York University Langone Medical Center New York, New York 10016, United States

[6]Department of Neuro-Oncology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas, United States

[7]Department of Neurosurgery, The University of Texas M.D. Anderson Cancer Center, Houston, Texas, United States

[8]Department of Biomedicine, University of Bergen, 5009 Bergen, Norway

[9]Clinical Protein Science & Imaging, Biomedical Center, Department of Measurement Technology and Industrial Engineering, Lund University, 221 84 Lund, Sweden

[10]School of Biotechnology, Department of Proteomics, Royal Institute of Technology, 106 91 Stockholm, Sweden

*To whom correspondence should be addressed: Department of Pharmacology and Toxicology, UTMB Cancer Center, University of Texas Medical Branch, Galveston, Texas 77555-1074, United States, Telephone: +1-409-747-1840. FAX: +1-409-772-9648. E-mail: carol.nilsson@utmb.edu.

Running title: Chromosome 19 Transcripts and Proteins in Glioma Stem Cells

Abstract

One sub-project within the global Chromosome 19 Consortium is to define chromosome 19 gene and protein expression in glioma-derived cancer stem cells (GSCs). Chromosome 19 is notoriously linked to glioma by 1p/19q co-deletions and clinical tests are established to detect that specific aberration. GSCs are tumor-initiating cells and are hypothesized to provide a repository of cells in tumors that can self-replicate and be refractory to radiation and chemotherapeutic agents developed for the treatment of tumors. In this pilot study, we performed RNA-Seq, label-free quantitative protein measurements in six GSC lines, and targeted transcriptomic analysis using a chromosome 19-specific microarray in an additional six GSC lines. The data have been deposited to the ProteomeXchange with identifier PXD000563. Here, we present insights into differences in GSC gene and protein expression, including the identification of proteins listed as having no or low evidence at the protein level in the Human Protein Atlas, as correlated to chromosome 19 and GSC subtype. Furthermore, the upregulation of proteins downstream of adenovirus-associated viral integration site 1 (AAVS1) in GSC11 in response to oncolytic adenovirus treatment was demonstrated. Taken together, our results may indicate new roles for chromosome 19, beyond the 1p/19q co-deletion, in the future of personalized medicine for glioma patients.

Keywords: Chromosome-Centric Human Proteome Project, proteins, mRNA, RNA-Seq, mass spectrometry, bioinformatics, glioma, glioma stem cells, cancer proteomics, chromosome 19, oncolytic virus, neurocan core protein, symplekin

Introduction

The objective of the Chromosome-Centric Human Proteome Project (C-HPP) is to map all proteins encoded by the human chromosomal complement and identify compelling correlates of protein biological functions and their role in disease[1]. The C-HPP has joined forces with the Encyclopedia of DNA Elements (ENCODE) project to achieve these goals[2]. Importantly, the ENCODE consortium has provided an initial "parts list" of the human genome[3], and continues to refine the first draft, published in 2012. The ENCODE project is highly synergistic with the goals of the C-HPP, in which global research efforts are focused on the identification and characterization of missing proteins, those that lack any credible mass spectrometric or antibody detection. Because not all human proteins are expressed in all tissues, a variety of normal and diseased tissues are currently under scrutiny by consortium members. The Chromosome 19 Consortium is an international, multicenter, multi-investigator group that develops complementary analytical platforms and integrates results derived from evidence of chromosome 19 activity in human tissues[4]. Several disease studies are underway within the Consortium, including the role of chromosome 19 in neurodegeneration, lung cancer, prostate cancer, and glioma.

The inherent genomic instability in gliomas results in chromosomal duplications, amplifications of specific genes, and activating mutations[5]. Chromosome 19 is linked to glioma by 1p/19q co-deletions, which are a positive prognostic indicator: 123 months mean survival versus 16 months in patients with tumors that are 1p/19q intact[6]. Because tumors with the co-deletion respond favorably to temozolomide, clinical testing is recommended.[7] Further, upregulation of a novel, previously uncharacterized

chromosome 19 protein, ER membrane protein complex subunit 10 (EMC10), located in a genomic region implicated in many cancers (19q13.33) was found to suppress glioma growth. However, amplification and overexpression of Rhophilin-2 (RHPN2), another chromosome 19 protein, was recently linked to dramatically decreased survival in glioma patients.[8]

In many cancer types, small subsets of tumor-initiating and treatment-resistant cells have been identified. In the cancer stem cell (CSC) hypothesis, stem cells have the capacity to make new tumors and produce progeny cells of many different types. Cancer stem-like cells have been described for leukemias[9] as well as solid tumors, including glioma[10]. Traditional chemotherapeutic regimens developed to debulk tumors have little effect on CSCs, and radiation therapy is equally inefficient[11].

The Cancer Genome Atlas Research Network has defined molecular-genetic subtypes of gliomas: mesenchymal, classical, neural and proneural[12]. Mesenchymal subtypes are characterized by neurofibromin (NF1, chromosome 17) loss, phosphatase and tensin homolog (PTEN, chromosome 10) loss or mutation, and inactivating cellular tumor antigen p53 (TP53, chromosome 17) mutations. Classical subtypes are typified by EGFR (chromosome 7) amplification, overexpression or mutation and PTEN loss or mutation. Proneural (PN1) subtypes are characterized by platelet-derived growth factor receptor alpha (PDGRFA, chromosome 4) amplification, mutations of isocitrate dehydrogenase (IDH1, chromosome 2) and phosphatidylinositol-3-kinase (PI3K, protein group), and expression of pro-neuronal markers such as OLIG2 (chromosome 21). Finally, neural (PN2) subtypes carry EGFR amplification or overexpression and express neuronal markers. While tumor classifications based on this system differ in their

response to treatments and partly guide patient treatment plans, median survival rates do not differ greatly between patients with gliomas characterized by different molecular-genetic subtypes.

Expanding our understanding of the biological drivers of treatment resistance in GSCs could serve to identify new therapeutic targets. In this study, we applied a global, integrated transcriptomic-proteomic workflow (**Figure 1**) in the analysis of six GSC lines and a targeted C19 transcription analysis[4] of twelve GSC lines, derived from four different GSC subtypes. We compared protein and mRNA expression in GSCs to their current evidence status in the Human Protein Atlas[13] and neXtProt[14]; several of the proteins identified were previously listed as having no- to low evidence of expression. We report our findings derived from analysis of six GSC lines and the inter-cell line differences in chromosome 19 expression at the level of transcription and protein expression and discuss them in the context of GSC subtypes. Further, proteomic studies of GSC responses to a therapeutic oncolytic adenovirus, Delta-24-RGD, revealed differential upregulation of proteins located downstream from AAVS1 (19q13.4) in one of four GSC lines.

Experimental Section

Cell culture conditions

Isolation of GSCs (GSC2, 11, 13, 17, 23, 8-11) from patient tumors was performed as previously described [15] in accordance with the institutional review board of The University of Texas M.D. Anderson Cancer Center, and are named in the order that they were acquired. GSCs were cultured according to a published method[15,16]. All cell

lines were tested to exclude the presence of *Mycoplasma* infection. Downstream transcriptomic and proteomic analyses were performed on identical cell culture batches in order to reduce the influence of batch variance in the comparative assays.

Oncolytic virus treatment

GSCs (GSC2, 11, 13, and 23) were cultured as described in the previous section, then dissociated and plated in 12 well plates (2×10^4 cells per well) and immediately infected with Delta-24-RGD at a multiplicity of infection of 10. Cells were harvested for proteomic analysis at 24 hrs and 48 hrs after infection or control treatment.

Transcriptomic Analysis of 6 Glioma Stem Cell Lines

Paired-end sequencing assays were performed on 37 glioma stem cell lines (GSC) using Illumina HiSeq platform. Only data from cell lines matching validated proteomic datasets (six as of the date of manuscript submission) were considered in this study. Each GSC line generated about 50 million paired-ends, each end was 75 bp in size. The average phred quality scores (APQS)[17] of each specimen ranged from 35.58 to 36.10. In order to generate more stable transcriptome mapping results, a trimming procedure was performed by using phred quality score<7. After trimming, the APQSs ranged from 37.01 to 37.23. Then Burroughs-Wheeler alignment[18], Samtools[19], and Genome Analysis Toolkit[20] were used to map short reads to the human transcriptome and RPKM values were generated for each of the 135,994 transcripts of 21,165 protein coding genes in Ensembl database (release version: Ensembl 64).

Targeted Transcriptomic Analysis of C19 Transcripts

A custom targeted oligonucleotide microarray platform was used to examine the expression of 1,382 chromosome 19-specific transcripts[4,21]. A full description of our targeted transcriptome profiling method was recently published[4]. Importantly, the quality of this platform has been rigorously evaluated in terms of dynamic range, discrimination power, accuracy, reproducibility and specificity. The ability to reliably measure even low levels of statistically significant differential gene expression stems from coupling (a) stringently designed and quality controlled chip manufacturing and transcript labeling protocols, (b) rigorous data analysis algorithms, and (c) flexible ontological and interactome analyses capable of demonstrating significant correlations between the expression of specific gene sets.

In these studies, the relative quantitation of individual transcript abundance in 12 GSC lines was compared to that derived from human neural stem cells (hNSCs, a kind gift from Dr. Ping Wu, UTMB Dept. of Neuroscience). Three each mesenchymal (GSC2, 17, 37), classical (10-6, 11, 47), PN1 (8-11, 23, 46) and PN2 (13, 34, 35) were included in the study. Briefly, total RNA extracted and purified from 12 defined glioma-derived stem (GSC) cell lines and normal hNSCs was used as the substrate for RNA amplification and labeling. We employed a universal reference design[22] and comprehensive statistical analysis platforms to facilitate the acquisition of expression profiles. Following hybridization and sequential high-stringency washes, individual Cy3 and Cy5 fluorescence hybridization to each spot on the microarray was quantitated by a high resolution confocal laser scanner. For each time point, RNA samples from each of

lines were analyzed in triplicate. As each transcript-specific oligonucleotide was also spotted in triplicate on the array, there were a total of 27 individual expression measurements per gene in each experimental group. Statistically significant differentially expressed genes were identified using the permutation-based Significance Analysis of Microarrays algorithm (SAM software package, v4.0, Stanford University, Palo Alto, CA)[23]. In our analyses, appropriately normalized data were analyzed using two-class, unpaired analysis on a minimum of 5000 permutations and was performed by comparing expression data derived from the different GSC lines versus hNSCs. In order to maximize the information derived from the subsequent GoMiner-based ontological analyses, the cutoff for significance in these experiments was set at a false discovery rate (FDR) of approximately 10%.

Genes identified as differentially expressed by SAM analysis were examined for their biological association to the gene ontology (GO) categories as defined by the GO Consortium. This provides both additional statistical stringency to the identified genes and identifies groups of related genes or “gene families” on chromosome 19 which were modulated in the various lines. Analyses were performed using the ontological mapping software GoMiner[24]. This software calculated the enrichment or depletion of individual ontological categories with genes that had changed expression and identified cellular pathways potentially relevant to GSC development. Pathways within three independent functional hierarchies, namely biological process, molecular function, and cellular component, were queried.

Proteomic Analysis of GSCs

$2x10^6$ cells were lysed with RIPA buffer (Thermo Fisher Scientific, Rockford, IL, 25mM TrisHCl pH 7.6, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS) mixed with Halt protease inhibitor EDTA-free, Halt phosphatase inhibitor cocktail and Pierce universal nuclease (Thermo Fisher Scientific, Rockford, IL).The protein concentration was determined by BCA Protein Assay Kit (Pierce), and the resulting protein (100 µg total protein) was reduced and alkylated. Five µL of 200 mM tris (2-carboxyethyl) phosphine (TCEP) buffered with triethylammonium bicarbonate (TEAB) was added to each sample (final TCEP concentration was 10 mM) and incubated at 55 °C for 1 h. Five µL of 375 mM iodoacetamide (buffered with TEAB) was added and incubated in the dark for 30 min. Proteins were precipitated in four volumes (440 µL) of ice cold acetone for 2 h at -20 °C. Samples were centrifuged at 10,000× *g* for 30 min (4 °C) after which the supernatants were removed and discarded. Pellets were air dried and resuspended in 12.5 µL of 8 M urea. Trypsin (10 µg in 87.5 µL of TEAB buffer) was added, and the samples were incubated for 24 h at 37 °C. An external standard comprised of proteins from all cell lines ("M37") was used for relative quantitation. Block randomization (random.org) was employed and M37 was included in each block. Samples were analyzed in triplicate.

Chromatographic separation and mass spectrometric analysis was performed with a nano-LC chromatography system (Easy-nLC 1000, Thermo Scientific), coupled on-line to a hybrid linear ion trap-Orbitrap mass spectrometer (Orbitrap Elite, Thermo Scientific) through a Nano-Flex II nanospray ion source (Thermo Scientific)[25]. Mobile phases were 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (ACN, B). After equilibrating the column in 95% solvent A and 5% solvent B, the samples (5 µL

in 5% v/v ACN/0.1% (v/v) formic acid in water, corresponding to 1 µg cell protein digest) were injected onto a trap column (C_{18}, 100 µm ID×2 cm) and subsequently eluted (250 nL/min) by gradient elution onto a C_{18} column (10 cm x 75 µm ID,15 µm tip, ProteoPep II, 5µm, 300 Å, , New Objective). The gradient was as follows: isocratic at 5% B, 0-8 min; 5% to 30% B, 8-188 min; 30% to 95% B, 188-220 min; and isocratic at 95% B, 220-240 min. Total run time, including column equilibration, sample loading, and analysis was 260 min.

All LC-MS/MS data were acquired using XCalibur, version 2.7 SP1 (Thermo Fisher Scientific). The survey scans (*m/z* 350-1650) (MS) were acquired in the Orbitrap at 60,000 resolution (at *m/z* = 400) in profile mode, followed by top 10 Higher Energy Collisional Dissociation (HCD) fragmentation centroid MS/MS spectra, acquired at 15K resolution in data-dependent analyses (DDA) mode. The automatic gain control targets for the Orbitrap were 1 x 10^6 for the MS scans and 5 x 10^4 for MS/MS scans. The maximum injection times for the MS1 and MS/MS scans in the Orbitrap both 200 ms. For MS/MS acquisition, the following settings were used: parent threshold = 10,000; isolation width = 4.0 Da; normalized collision energy = 30%; activation time = 100 ms. Monoisotopic precursor selection, charge state screening, and charge state rejection were enabled, with rejection of singly charged and unassigned charge states. Dynamic exclusion was used to remove selected precursor ions (−/+10 ppm) for 90 s after MS/MS acquisition. A repeat count of 1 and a maximum exclusion list size of 500 was used. The following ion source parameters were used: capillary temperature 275 °C, source voltage 2.2 kV, source current 100 uA, and S-lens RF level 40%.

Proteomic Data Analysis

MS files (.raw) were imported into Progenesis LC-MS (version 18.214.1528, Nonlinear Dynamics) for *m/z* and retention time alignment. The top 5 spectra for each feature were exported (charge deconvolution, top 1000 peaks) as a combined .mgf file for database searching in PEAKS[26] (version 6, Bioinformatics Solutions Inc., Waterloo, ON) against the UniprotKB/Swissprot-Human database (July 2013 version, 20,264 proteins), appended with the cRAP contaminant database. PEAKS DB and Mascot (version 2.3.02, Matrix Science) searches were performed with a parent ion tolerance of 10 ppm, fragment ion tolerance of 0.025 Da, fixed carbamidomethyl cysteine, and variable modifications of oxidation (M), phosphorylation (STY), and deamidation (NQ). Trypsin was specified as the enzyme, allowing for 2 missed cleavages and a maximum of 3 PTMs per peptide. An additional search for unexpected modifications was performed with the entire Unimod database. Finally, homology searching was performed using the SPIDER algorithm[27] to identify peptides resulting from nonspecific cleavages or amino acid substitutions. Mascot and PEAKS SPIDER searches were combined (inChorus), using a 1% false discovery rate cutoff for both search engines. The resulting peptide-spectrum matches (95% peptide probability) were imported into Progenesis LC-MS. Conflict resolution was performed manually to ensure that a single peptide sequence was assigned to each feature by removing lower scoring peptides. The resulting normalized peptide intensity data were exported, and the peptide list was filtered to remove non-unique peptides, methionine-containing peptides, and all modified peptides except cysteine carbamidomethylation. For quantification, the filtered list of peptide intensities was imported into DanteR (version 0.1.1)[28], and intensities for peptides of the

same sequence were combined to form a single entry. The resulting peptide intensities were $\log_2$ transformed and combined to protein abundances (RRollup) using the default settings, excluding one-hit wonders (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). The resulting proteins were quantified by 1-way ANOVA relative to M37; *p*-value adjustment for multiple testing was performed according to Benjamini and Hochberg[29]. Chromosome 19 proteins were then subjected to further bioinformatic analysis as described in the following section.

For identification of chromosome 19 proteins, peptides modified by acetylation, phosphorylation, methionine oxidation, and N-terminal pyroglutamate from glutamine were considered in addition to peptides considered for quantification. BLAST searches were conducted in those instances where protein assignments were based on one or two peptides in order to confirm uniqueness. The list of all identified proteins, along with number of peptides supporting protein assignments, can be found in **S Table 2**. The data have been deposited to the ProteomeXchange via the PRIDE partner repository with identifier PXD000563.

Bioinformatic Analysis of Transcriptomic and Proteomic Data

For comparison of protein expression for cell lines measured in separate analytical sets, unsupervised hierarchical clustering and PCA analysis was performed using fold changes relative to M37 for all proteins measured in the six cell lines. Fold change values were standardized to Z-scores, imported into DanteR, and mean centered to 0. Hierarchical clustering was performed for both proteomic and

transcriptomic data by use of a Euclidean distance metric, an average agglomeration method, and no row scaling, with the results visualized as heat maps.

Ingenuity Pathway Analysis

Normalized quantitative data sets were analyzed by use of Ingenuity Pathway Analysis (Ingenuity Systems, version 16542223, build 220217, 22 June 2013, (www.ingenuity.com). The data set contained protein identifiers, fold changes relative to M37 and ANOVA *p*-values. The protein networks were created using proteins with *p*-value <0.05 and molecular interactions described in the scientific literature. Networks represent a highly interconnected set of proteins derived from the input data set.

Results and Discussion

Chromosome 19 Targeted Transcriptomics

Of the 1,382 chromosome 19 genes analyzed, between 70-75% were expressed in each of the cell lines. Utilizing the reference design, a key element of our transcriptomic platform, affords us the ability to customize our analyses to identify statistically significantly differential chromosome 19-specific gene expression patterns between the phenotypically diverse stem cell subtypes. These tailored analyses allow us potential insights into important questions such as; (i) what are the differences between each of the GSC subtypes and normal human neural stem cells (e.g., transcripts associated with either “stemness” or frank differences between neural- and glial-derived cells), and (ii) what are the differences between each of the GSC subtypes

(i.e., genes potentially involved in glioma stem cell progression). For this communication, we will primarily focus on the former.

The identity of the differentially expressed transcripts for PN1, PN2, classical, and mesenchymal GSC subtypes as compared to hNSCs (at a 10% False Discovery Rate; FDR) are listed in **Supplementary Table 1**. A significant proportion (~20%) of the transcripts were differentially expressed in the proneural (PN1 and PN2) and classical subtypes when compared to those expressed in hNSCs. The identity of these transcripts, as well as the gene families represented by these genes were largely similar. In stark contrast, the mesenchymal GSCs were clearly differentiated from the proneural and classical GSCs at the level of both proportion of differentially expressed transcripts identified (>30%, **Table 1**), as well as the identity of those transcripts. GoMiner-based ontological analyses of the mesenchymal (M) GSCs compared to both hNSCs and the neural (PN1) GSCs (**Table 2**) demonstrated enrichment in several molecular pathways with established relevance to stem cell genesis and progression. Of particular interest, increases in the expression of specific genes related to carbohydrate binding were significantly associated with the mesenchymal phenotype when compared to both hNSCs and the less invasive proneural GSCs. The expression of cell surface glycoconjugates are well established to play a major role in the modulation of the invasive phenotype[30]. There are also several unique molecular pathways identified in this analysis. For example, despite the large number genes encoding transcription factors present on chromosome 19 (the large number can skew significance values), there is still a very significant modulation of their expression in the mesenchymal GSCs.

Chromosome 19 transcripts potentially encoding candidate proteins as yet unidentified (open reading frames; ORFs) were also queried in these analyses. Of particular interest, of the 43 ORFs represented on the arrays, 31 (72%) were expressed in the GSC lines. Of these, 17 (55%) were differentially expressed in 2 or more of the subtypes. The remaining transcripts were present in the only one of the subtypes. Our integrated platform will not only facilitate the identification of these novel proteins, but help elucidate their role in the development of highly malignant glial tumors.

Lastly, the expression of 12,042 genes of 481 GBM samples and 10 solid tissue normal samples was detected using Affymetrix human genome U133A platform, downloaded from TCGA data portal (https://tcga-data.nci.nih.gov/tcga/). Among these genes, 762 genes are located on chromosome 19. 144, 155, 83 and 99 samples of the 481 GBM tumors were classified into classical, mesenchymal, neural, proneural subtypes respectively[31]. 30%~40% of the genes were differentially expressed in proneural, classical, mesenchymal and classical subtypes when compared to those expressed in normal samples. Mesenchymal subtypes were differentially expressed from the proneural subtype, which is highly consistent with the finding in GSCs.

Identification of No- to Medium-Evidence Chromosome 19 Proteins

We detected 213 proteins and quantified (2 unique peptides/protein) 184 chromosome 19 proteins at high confidence expressed in six GSC lines (**Supplementary Tables 2 and 3**) and queried the Human Protein Atlas (HPA, proteinatlas.org) and neXtProt (www.nextprot.org) to examine evidence for existence of these proteins. For all identified proteins, evidence of existence, according to HPA and

neXtProt, and peptide counts can be found in S Table 2 and in the following discussion. Several identified proteins had been seen previously only at the RNA level or were of low confidence (**Table 3**). Of the verified protein hits, protein tweety homolog 1 (TTYH1) and neurocan (NCAN) are low confidence entries. Mucin-16, ribonucleoprotein PTB-binding 1 and E3 ubiquitin-protein ligase (MUC16, RAVER1 and UHRF1, respectively) have no record in HPA, but do show protein evidence in neXtProt. Two proteins, ceramide synthase 1 (CERS1) and long-chain fatty acid transport protein 1 (SLC27A1) have only RNA evidence in HPA, but both show evidence at the protein level in Proteomics DB (proteomicsdb.org).

Neurocan (NCAN, core protein of an extracellular matrix protein) is an intriguing finding, because it is a brain-specific chondroitin sulfate proteoglycan that is believed to mediate neuronal adhesion and neurite growth during development. At the mRNA level, increased levels of NCAN have been seen in non-small-cell lung cancer brain metastases and in astrocytoma tissues, linking it to tumor invasivity[4]. Our original list of proteins included RUXGL_HUMAN. However, upon performing a BLAST search, we found that the peptides assigned to RUXGL were also shared by RUXG_HUMAN. This illustrates the caution that must be exercised when validating the identification of newly identified proteins.

A large number of proteins (87) display medium evidence for existence (**Supplementary Table 2**). Within the chromosome 19 proteomic data, we found expression of zinc-finger proteins listed as having medium levels of protein evidence. Certain zinc finger proteins play a central role in aberrant signaling processes in glioma[32]. ZNF428 (C19orf37) has few findings in the literature. Interestingly, this protein

was relatively decreased in GSC2, 11 and 13, a distribution which does not coincide with the current classification of GSCs.

Five ORFs in the proteomic dataset were also detected at the transcript level [C19orf10, C19orf43, C19orf1 (TOMM40), C19orf5 (MAP1S), and C19orf37 (ZNF428)]. C19orf10, a high evidence protein, encodes a stromal-derived growth factor, previously detected in bone marrow and synovial fluid[33]. Its expression in GSCs has not been described previously, and its concentration was within a two-fold range across the cell lines. It may be of interest to glioma pathology, because it promotes lymphoid cell proliferation. C19orf43 is a medium-level evidence 18 kDa protein to which no biological function has yet been ascribed. This protein was decreased ~20-fold in GSC11 relative to M37, whereas another protein, C19orf1 (TOMM40) was highly increased in this cell line. It encodes an import protein integral to the outer mitochondrial membrane.

Quantitative Proteomics

For the six cell lines, 161 chromosome 19 proteins were quantified relative to M37 in at least one cell line (**Supplementary Table 3**); volcano plots illustrating fold changes are shown in **Figure 2**. In three of the six cell lines (GSC2, **Figure 2A**; GSC11, **Figure 2C**; and GSC13, **Figure 2D**), the protein demonstrating the highest relative fold change is symplekin (SYMPK), a component of the mRNA polyadenylation machinery that has been associated with tumorigenicity in colon[34] and lung[35] cancers. Our findings represent the first report of SYMPK in gliomas, making it a target for our follow-up studies. GSC2, GSC11, and GSC13 also demonstrated decreased levels of ZNF428 relative to M37. Along with GSC23, these same cell lines showed significantly

decreased levels of far upstream element-binding protein 2 (KHSRP), a key regulator of miRNAs in the DNA damage response[36].

In order to compare the six GSC cell lines to one another, we performed principal component analysis (PCA, **Figure 3**) and hierarchical clustering analysis (**Figure 4B**) at the protein level. Hierarchical clustering was also performed at the transcript level (**Figure 4A**). In the 3D PCA plot, three cell lines (GSC23, 11, and 8-11) were the most proximal, indicating a high degree of similarity in overall protein expression patterns. The other cell lines (GSC2, 13 and 17) were distinctly different from each other and from the three other cell lines. In the heatmaps for both the proteomic and transcriptomic data (**Figure 4**), the two mesenchymal cell lines (GSC2 and GSC17) clustered together, as did the (pro-)neural cell lines. Taken together, these results indicate that the various GSC cell line classes show similar expression patterns for chromosome 19 proteins.

Ingenuity Pathway Analysis

We examined putative functional roles of chromosome 19 proteins expressed in GSC lines by studying them in the context of Ingenuity Pathway Analysis (**Figure 5**). Notably, there are several chromosome 19 proteins linked in GSC2 to the RB tumor suppressor (Rb) in the extended network. This pathway is universally disrupted in glioma[37].

Upregulation of Chromosome 19 Proteins in Response to Oncolytic Adenovirus Treatment

We studied the effect of oncolytic adenovirus therapy on four GSC lines (GSC2, GSC11, GSC13, and GSC23). Delta-24-RGD is a replication-competent adenovirus that targets the RB pathway in gliomas[38]. Integration of adenovirus requires two *trans*-acting viral proteins (Rep68 and Rep78) and *cis*-acting DNA elements that display Rep binding sites[39]. The integration may occur in the AAVS1 site on chromosome 19q13.4, in exon1 of the protein phosphatase 1 regulatory subunit 12C PPP1R12C)[40]. We found that proteins derived from genes downstream of the AAVS1 site were upregulated in GSC11 but not three other GSC lines, at 24 hrs and 48 hrs after infection with Delta-24-RGD virus (**Figure 6**). These results indicate that the patient from whom GSC11 was derived had a previous CNS infection by adenovirus. It is not yet known if previous adenovirus infection may influence the efficacy of Delta-24-RGD in larger patient groups. These unexpected results serve to demonstrate how new biological insights can be derived by examining protein expression in the context of chromosomal localization of the encoding elements.

Conclusion

We have defined patterns of chromosome 19 mRNA and protein expression in several GSC lines. The cells provided a source of chromosome 19 mRNA and proteins that demonstrated expression of previously uncharacterized gene products. Our results underscore the importance of studying a variety of healthy and diseased tissues in the context of the C-HPP. Furthermore, we detected differential regulation of chromosome 19 activity in the context of histological sub-types and in response to treatment with oncolytic adenovirus. Our results have expanded the knowledge of the role of

chromosome 19 beyond the well-known impact of 1p/19q co-deletion. Our findings may have relevance for selection of GSC lines for testing responses to pre-clinical and clinical compounds.

Acknowledgements

Support from the Cancer Prevention and Research Institute of Texas (CLN) and the University of Texas Medical Branch (CLN) is gratefully acknowledged, as are the Biomolecular Resource Facility Mass Spectrometry Lab (CFL) and the Department of Pharmacology & Toxicology (CFL). The DanTE software was written by Tom Taverner (t.taverner@gmail.com) and Ashoka Polpitiya for the U.S. Department of Energy (PNNL, Richland, WA, USA, http://omics.pnl.gov/software). Dr. Ping Wu (UTMB, Department of Neuroscience & Cell Biology) is gratefully acknowledged for the gift of hNSCs.

Figure legends

Tables.

Table 1. Numbers of differentially expressed chromosome 19 transcripts by molecular-histological subclasses compared to hNSCs.

Table 2. GoMiner-based ontological analyses of the mesenchymal (M) GSCs compared to hNSCs and the neural (PN1) GSCs.

Table 3. Proteins identified in six GSC lines that are described as having low or no protein evidence in the Human Protein Atlas version 11.0, with additional evidence from neXtProt.

Figures.

Figure 1. Workflow for integrated proteomic and transcriptomic analysis of GSC cell lines. Glioma stem cell lines, derived from patient tumor samples, were analyzed by three approaches: targeted chromosome 19 microarray (1), quantitative proteomics (2), and RNA-Seq (3). Identification and quantification were performed at the transcript (4) and protein (5) levels, and a custom protein database was generated from the RNA-Seq data (6). Comparisons were made between transcript and protein data (7), and the custom protein database was used to search for proteins (8). Validated protein identifications were queried against protein databases in order to determine levels of protein evidence (9), and quantitative proteomic data were used to generate networks in Ingenuity Pathway Analysis (10).

Figure 2. Volcano plots (*p* value vs $\log_2$ fold change) for GSC2 (**A**), GSC8-11 (**B**), GSC11 (C), GSC13 (D), GSC17 (E), and GSC23 (F) relative to M37. Proteins that are differentially expressed relative to M37 include peptidyl-prolyl cis-trans isomerase NIMA-interacting 1 (PIN1), nuclear factor 1 X-type (NFIX1), E3 ubiquitin-protein ligase UHRF1, epidermal growth factor receptor substrate 15-like 1 (EPS15L1), lysophospholipid acyltransferase 7 (MBOAT7), calpain small subunit 1 (CAPNS1), 40S ribosomal protein S28 (RPS28), ribonuclease H2 subunit A(RNASEH2A), isochorismatase domain-containing protein 2, mitochondrial (ISOC2), immunity-related GTPase family Q protein (IRGQ), and DNA repair protein XRCC1. A complete list of measured fold changes and *p*-values can be found in **Supplementary Table 3**.

Figure 3. Three-dimensional principal component analysis (PCA) plot, demonstrating variation between GSC2 (red), GSC8-11 (yellow), GSC11 (green), GSC13 (cyan), GSC17 (blue), and GSC23 (pink) based upon fold change relative to M37 for proteins identified in all 6 cell lines.

Figure 4. Heat map of C19 transcripts (**A**) and proteins (**B**) for six GSC cell lines. Chromosome 19 heat map for six GSC cell lines. Unsupervised hierarchical clustering was performed using $\log_2$ fold changes relative to hNSCs (**A**) and Z-scaled fold changes relative to M37 (**B**) for proteins identified in all 6 GSC cell lines.

Figure 5. Network generated by use of Ingenuity Pathway Analysis. Gene identifiers for all chromosome 19 proteins identified in GSC2 sample were uploaded and those proteins with a *p*-value< 0.05 were considered. Fifteen proteins were associated to proteins in the retinoblastoma group (Rb), a pathway that is universally disrupted in glioma. The associated proteins included retinoblastoma-like protein 1 (RBL1), retinoblastoma-associated protein (RB1), transcription factors AP1 (JUN) and activator BRG1 (SMARCA4), and several proteins involved in cell cycle control.

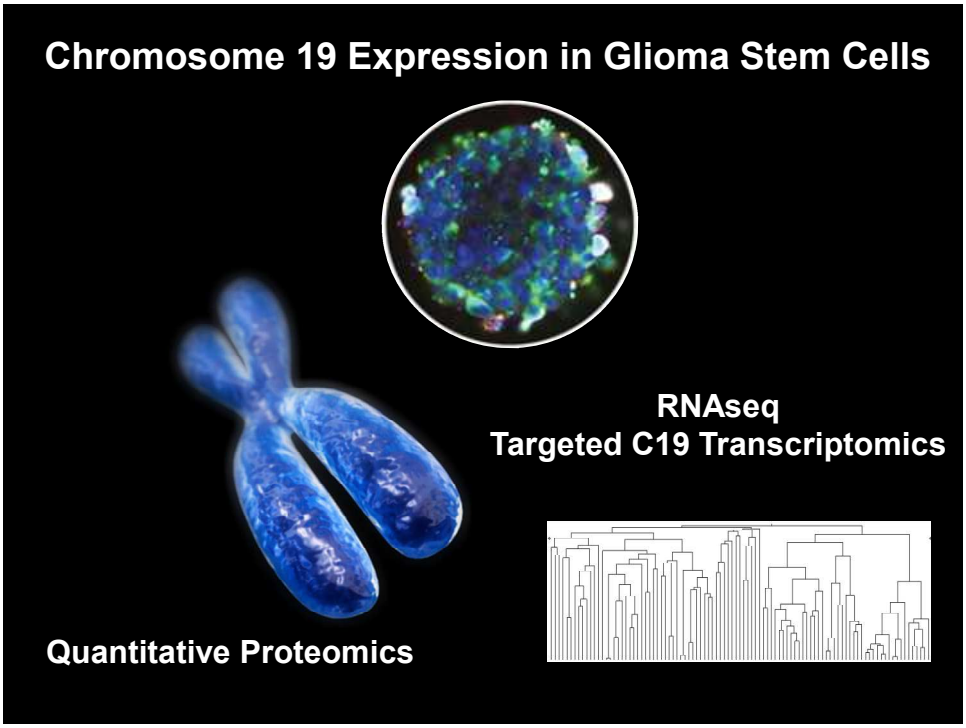
Figure 6. Upregulation of Chromosome 19q13 proteins in GSC11 in response to treatment with oncolytic virus. Fold changes at 24h (black bars) and 48h (red bars) are calculated relative to control. The most dramatically increased protein was rRNA 2'-O-methyltransferase fibrillarin (FBL), a protein involved in rRNA pre-processing. Other proteins more highly expressed at 24h include SUMO-activating enzyme subunit 2 (UBA2) and 26S protease regulatory subunit 6B (PSMC4).

References

(1) Paik, Y.-K.; Omenn, G. S.; Uhlen, M.; Hanash, S.; Marko-Varga, G.; Aebersold, R.; Bairoch, A.; Yamamoto, T.; Legrain, P.; Lee, H.-J.; Na, K.; Jeong, S.-K.; He, F.; Binz, P.-A.; Nishimura, T.; Keown, P.; Baker, M. S.; Yoo, J. S.; Garin, J.; Archakov, A.; Bergeron, J.; Salekdeh, G. H.; Hancock, W. S. *J. Proteome Res.* **2012**, *11*, 2005-2013.
(2) Paik, Y.-K.; Hancock, W. S. *Nat. Biotechnol.* **2012**, *30*, 1065-1067.
(3) Consortium, E. P. *Nature* **2012**, *489*, 57-74.
(4) Nilsson, C. L.; Berven, F.; Selheim, F.; Liu, H.; Moskal, J. R.; Kroes, R. A.; Sulman, E. P.; Conrad, C. A.; Lang, F. F.; Andrén, P. E.; Nilsson, A.; Carlsohn, E.; Lilja, H.; Malm, J.; Fenyö, D.; Subramaniyam, D.; Wang, X.; Gonzales-Gonzales, M.; Dasilva, N.; Diez, P.; Fuentes, M.; Végvári, Á.; Sjödin, K.; Welinder, C.; Laurell, T.; Fehniger, T. E.; Lindberg, H.; Rezeli, M.; Edula, G.; Hober, S.; Marko-Varga, G. *J. Proteome Res.* **2012**, *12*, 135-150.
(5) Vranova, V.; Necesalova, E.; Kuglik, P. *Oncol Rep.* **2007**, *17*, 457-464; Ruano, Y.; Mollejo, M.; Ribalta, T.; Fiano, C.; Camacho, F.; Gomez, E.; de Lope, A.; Hernandez-Moneo, J.-L.; Martinez, P.; Melendez, B. *Mol. Cancer* **2006**, *5*, 39; Hui, A. B.; Lo, K. W.; Yin, X. L.; Poon, W. S.; Ng, H. K. *Lab. Invest.* **2001**, *81*, 717-723.
(6) Ino, Y.; Zlatescu, M. C.; Sasaki, H.; Macdonald, D. R.; Stemmer-Rachamimov, A. O.; Jhung, S.; Ramsay, D. A.; von Deimling, A.; Louis, D. N.; Cairncross, J. G. *J. Neurosurg.* **2000**, *92*, 983-990.
(7) Yip, S.; Iafrate, A. J.; Louis, D. N. *J. Neuropathol. Exp. Neurol.* **2008**, *67*, 1-15 10.1097/nen.0b013e31815f65fb.
(8) Danussi, C.; Akavia, U. D.; Niola, F.; Jovic, A.; Lasorella, A.; Pe'er, D.; Iavarone, A. *Cancer Res.* **2013**.
(9) Bhatia, M.; Wang, J. C. Y.; Kapp, U.; Bonnet, D.; Dick, J. E. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 5320-5325; Bonnet, D.; Dick, J. E. *Nat. Med.* **1997**, *3*, 730-737; Lapidot, T.; Sirard, C.; Vormoor, J.; Murdoch, B.; Hoang, T.; Caceres-Cortes, J.; Minden, M.; Paterson, B.; Caligiuri, M. A.; Dick, J. E. *Nature* **1994**, *367*, 645-648.
(10) Singh, S. K.; Hawkins, C.; Clarke, I. D.; Squire, J. A.; Bayani, J.; Hide, T.; Henkelman, R. M.; Cusimano, M. D.; Dirks, P. B. *Nature* **2004**, *432*, 396-401.
(11) Sulman, E.; Aldape, K.; Colman, H. *Curr.t Probl. Cancer* **2008**, *32*, 124-142.
(12) TCGA. *Nature* **2008**, *455*, 1061-1068; Network, T. C. G. A. R. *Nature* **2013**, *494*, 506-506.
(13) Fagerberg, L.; Oksvold, P.; Skogs, M.; Älgenäs, C.; Lundberg, E.; Pontén, F.; Sivertsson, Å.; Odeberg, J.; Klevebring, D.; Kampf, C.; Asplund, A.; Sjöstedt, E.; Al-Khalili Szigyarto, C.; Edqvist, P.-H.; Olsson, I.; Rydberg, U.; Hudson, P.; Ottosson Takanen, J.; Berling, H.; Björling, L.; Tegel, H.; Rockberg, J.; Nilsson, P.; Navani, S.; Jirström, K.; Mulder, J.; Schwenk, J. M.; Zwahlen, M.; Hober, S.; Forsberg, M.; von Feilitzen, K.; Uhlén, M. *J. Proteome Res.* **2012**, *12*, 2439-2448.
(14) Gaudet, P.; Argoud-Puy, G.; Cusin, I.; Duek, P.; Evalet, O.; Gateau, A.; Gleizes, A.; Pereira, M.; Zahn-Zabal, M.; Zwahlen, C.; Bairoch, A.; Lane, L. *Journal of proteome research* **2013**, *12*, 293-8; Lane, L.; Argoud-Puy, G.; Britan, A.; Cusin, I.; Duek, P. D.; Evalet, O.; Gateau, A.; Gaudet, P.; Gleizes, A.; Masselot, A.; Zwahlen, C.; Bairoch, A. *Nucleic acids research* **2012**, *40*, D76-83.
(15) Galli, R.; Binda, E.; Orfanelli, U.; Cipelletti, B.; Gritti, A.; De Vitis, S.; Fiocco, R.; Foroni, C.; Dimeco, F.; Vescovi, A. *Cancer Res.* **2004**, *64*, 7011-7021.
(16) Jiang, H.; Gomez-Manzano, C.; Aoki, H.; Alonso, M. M.; Kondo, S.; McCormick, F.; Xu, J.; Kondo, Y.; Bekele, B. N.; Colman, H.; Lang, F. F.; Fueyo, J. *J. Natl. Cancer Inst.* **2007**, *99*, 1410-1414.
(17) Ewing, B.; Hillier, L.; Wendl, M. C.; Green, P. *Genome Res.* **1998**, *8*, 175-185.
(18) Li, H.; Durbin, R. *Bioinformatics* **2009**, *25*, 1754-1760.
(19) Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; Subgroup, G. P. D. P. *Bioinformatics* **2009**, *25*, 2078-2079.
(20) McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; DePristo, M. A. *Genome Res.* **2010**, *20*, 1297-1303.
(21) Kroes, R. A.; Dawson, G.; Moskal, J. R. *J. Neurochem.* **2007**, *103*, 14-24.

(22) Churchill, G. A. *Nat. Genet.* **2002**, *32*, S490-495.
(23) Tusher, V. G.; Tibshirani, R.; Chu, G. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 5116-5121.
(24) Zeeberg, B.; Feng, W.; Wang, G.; Wang, M.; Fojo, A.; Sunshine, M.; Narasimhan, S.; Kane, D.; Reinhold, W.; Lababidi, S.; Bussey, K.; Riss, J.; Barrett, J.; Weinstein, J. *Genome Biol.* **2003**, *4*, R28.
(25) Michalski, A.; Damoc, E.; Lange, O.; Denisov, E.; Nolting, D.; Müller, M.; Viner, R.; Schwartz, J.; Remes, P.; Belford, M.; Dunyach, J.-J.; Cox, J.; Horning, S.; Mann, M.; Makarov, A. *Mol. Cell. Proteomics* **2012**, *11*.
(26) Zhang, J.; Xin, L.; Shan, B.; Chen, W.; Xie, M.; Yuen, D.; Zhang, W.; Zhang, Z.; Lajoie, G. A.; Ma, B. *Mol. Cell. Proteomics* **2012**, *11*; Han, X.; He, L.; Xin, L.; Shan, B.; Ma, B. *J. Proteome Res.* **2011**, *10*, 2930-2936.
(27) Han, Y.; Ma, B.; Zhang, K. *J. Bioinform. Comput. Biol.* **2005**, *3*, 697-716.
(28) Karpievitch, Y.; Stanley, J.; Taverner, T.; Huang, J.; Adkins, J. N.; Ansong, C.; Heffron, F.; Metz, T. O.; Qian, W.-J.; Yoon, H.; Smith, R. D.; Dabney, A. R. *Bioinformatics* **2009**, *25*, 2028-2034; Polpitiya, A. D.; Qian, W.-J.; Jaitly, N.; Petyuk, V. A.; Adkins, J. N.; Camp, D. G.; Anderson, G. A.; Smith, R. D. *Bioinformatics* **2008**, *24*, 1556-1558.
(29) Benjamini, Y.; Hochberg, Y. *J. Royal Stat. Soc. Series B* **1995**, *57*, 289-300.
(30) Moskal, J. R.; Kroes, R. A.; Dawson, G. *Expert Revi. Neurother.* **2009**, *9*, 1529-1545.
(31) Verhaak, R. G. W.; Hoadley, K. A.; Purdom, E.; Wang, V.; Qi, Y.; Wilkerson, M. D.; Miller, C. R.; Ding, L.; Golub, T.; Mesirov, J. P.; Alexe, G.; Lawrence, M.; O'Kelly, M.; Tamayo, P.; Weir, B. A.; Gabriel, S.; Winckler, W.; Gupta, S.; Jakkula, L.; Feiler, H. S.; Hodgson, J. G.; James, C. D.; Sarkaria, J. N.; Brennan, C.; Kahn, A.; Spellman, P. T.; Wilson, R. K.; Speed, T. P.; Gray, J. W.; Meyerson, M.; Getz, G.; Perou, C. M.; Hayes, D. N. *Cancer Cell* **2010**, *17*, 98-110.
(32) Santoni, M.; Buerattini, L.; Nabissi, M.; Morelli, M. B.; Berardi, R.; Santoni, G.; Cascinu, S. *Curr. Protein Pept. Sci.* **2013**, *14*, 133-140.
(33) Weiler, T.; Du, Q.; Krokhin, O.; Ens, W.; Standing, K.; El-Gabalawy, H.; Wilkins, J. *Arthritis Res. Ther.* **2007**, *9*, R30; Tulin, E. E.; Onoda, N.; Nakata, Y.; Maeda, M.; Hasegawa, M.; Nomura, H.; Kitamura, T. *J. Immunol.* **2001**, *167*, 6338-6347.
(34) Buchert, M.; Papin, M.; Bonnans, C.; Darido, C.; Raye, W. S.; Garambois, V.; Pélegrin, A.; Bourgaux, J.-F.; Pannequin, J.; Joubert, D.; Hollande, F. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 2628-2633.
(35) Cappell, K. M.; Larson, B.; Sciaky, N.; Whitehurst, A. W. *Mol. Cell. Biol.* **2010**, *30*, 5135-5144.
(36) Zhang, X.; Wan, G.; Berger, F. G.; He, X.; Lu, X. *Mol. Cell* **2011**, *41*, 371-383.
(37) Goldhoff, P.; Clarke, J.; Smirnov, I.; Berger, M. S.; Prados, M. D.; James, C. D.; Perry, A.; Phillips, J. J. *J. Neuropathol. Exp. Neurol.* **2012**, *71*, 83-89.
(38) Jiang, H.; Gomez-Manzano, C.; Lang, F. F.; Alemany, R.; Fueyo, J. *Curr. Gene Ther.* **2009**, *9*, 422-427.
(39) Janovitz, T.; Klein, I. A.; Oliveira, T.; Mukherjee, P.; Nussenzweig, M. C.; Sadelain, M.; Falck-Pedersen, E. *J. Virol.* **2013**, *87*, 8559-8568; Surosky, R. T.; Urabe, M.; Godwin, S. G.; McQuiston, S. A.; Kurtzman, G. J.; Ozawa, K.; Natsoulis, G. *J. Virol.* **1997**, *71*, 7951-9; Urabe, M.; Kogure, K.; Kume, A.; Sato, Y.; Tobita, K.; Ozawa, K. *J. Gen. Virol.* **2003**, *84*, 2127-2132.
(40) Linden, R. M.; Ward, P.; Giraud, C.; Winocour, E.; Berns, K. I. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 11288-11294; Samulski, R. J.; Zhu, X.; Xiao, X.; Brook, J. D.; Housman, D. E.; Epstein, N.; Hunter, L. A. *EMBO J.* **1991**, *10*, 5071-5078.

TOC Graphic


Chromosome 19 Expression in Glioma Stem Cells
RNAseq
Targeted C19 Transcriptomics
Quantitative Proteomics

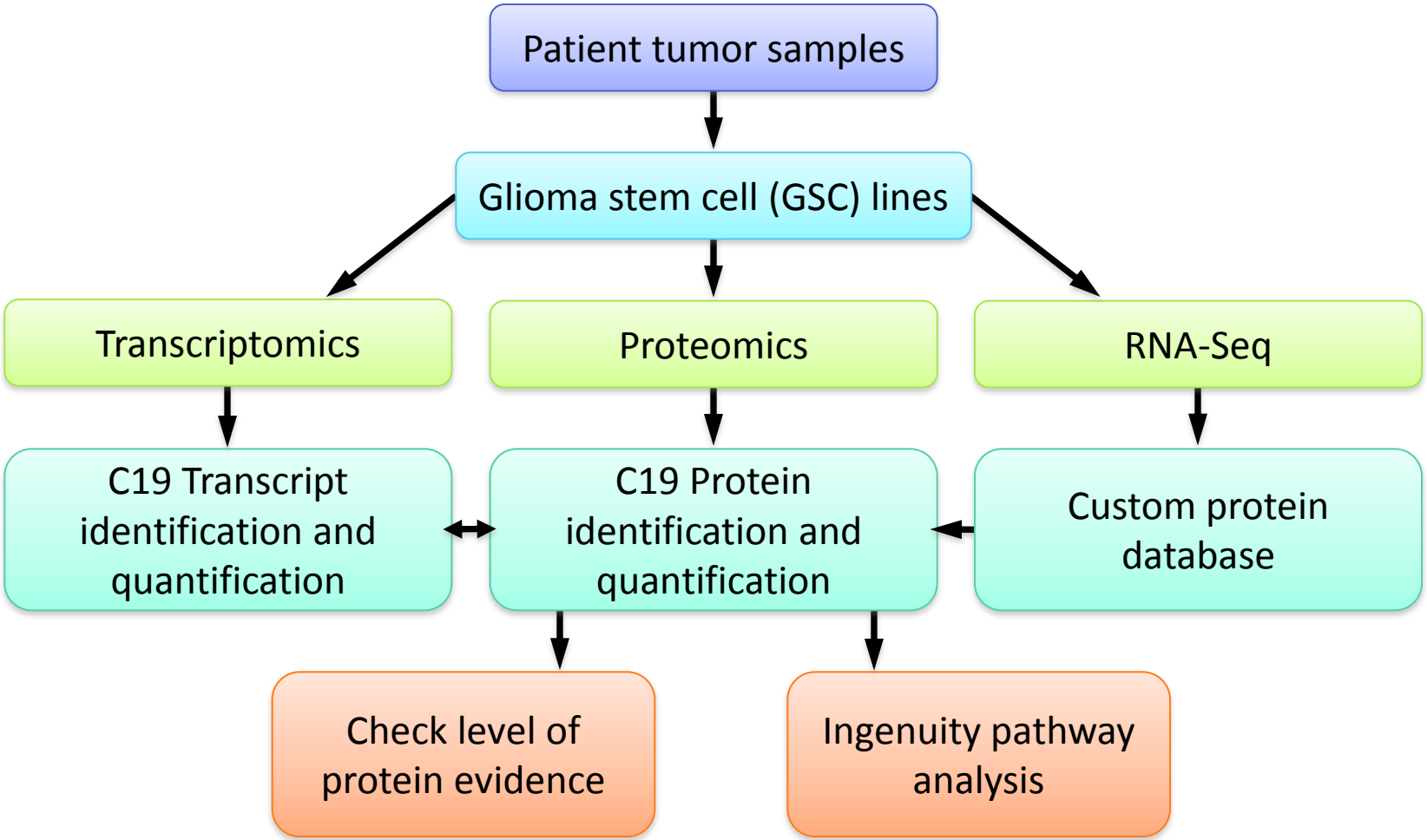

Patient tumor samples
Glioma stem cell (GSC) lines
Transcriptomics
Proteomics
RNA-Seq
C19 Transcript identification and quantification
C19 Protein identification and quantification
Custom protein database
Check level of protein evidence
Ingenuity pathway analysis

Figure 2A (GSC2)

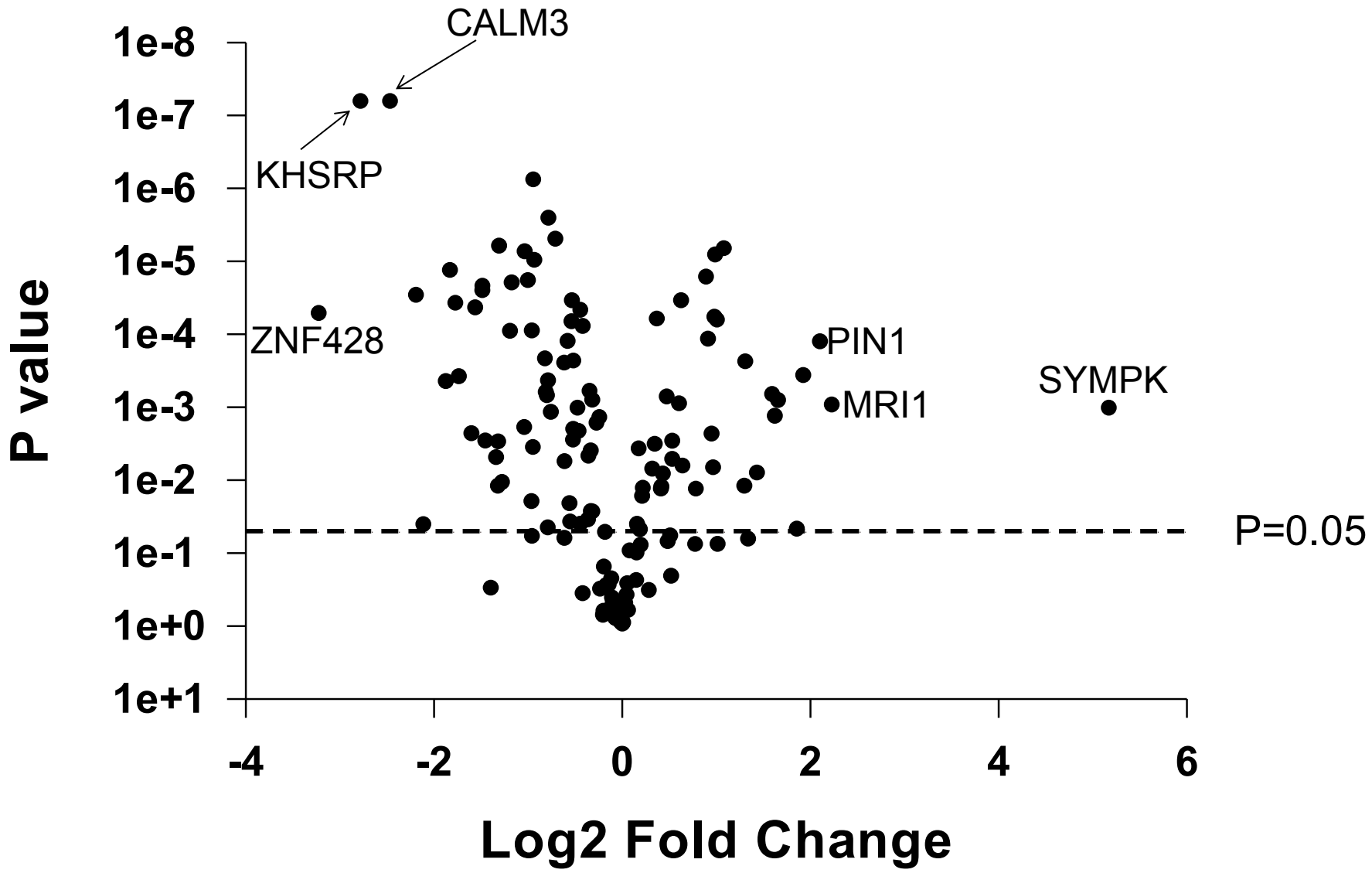


Figure 2B (GSC8-11)

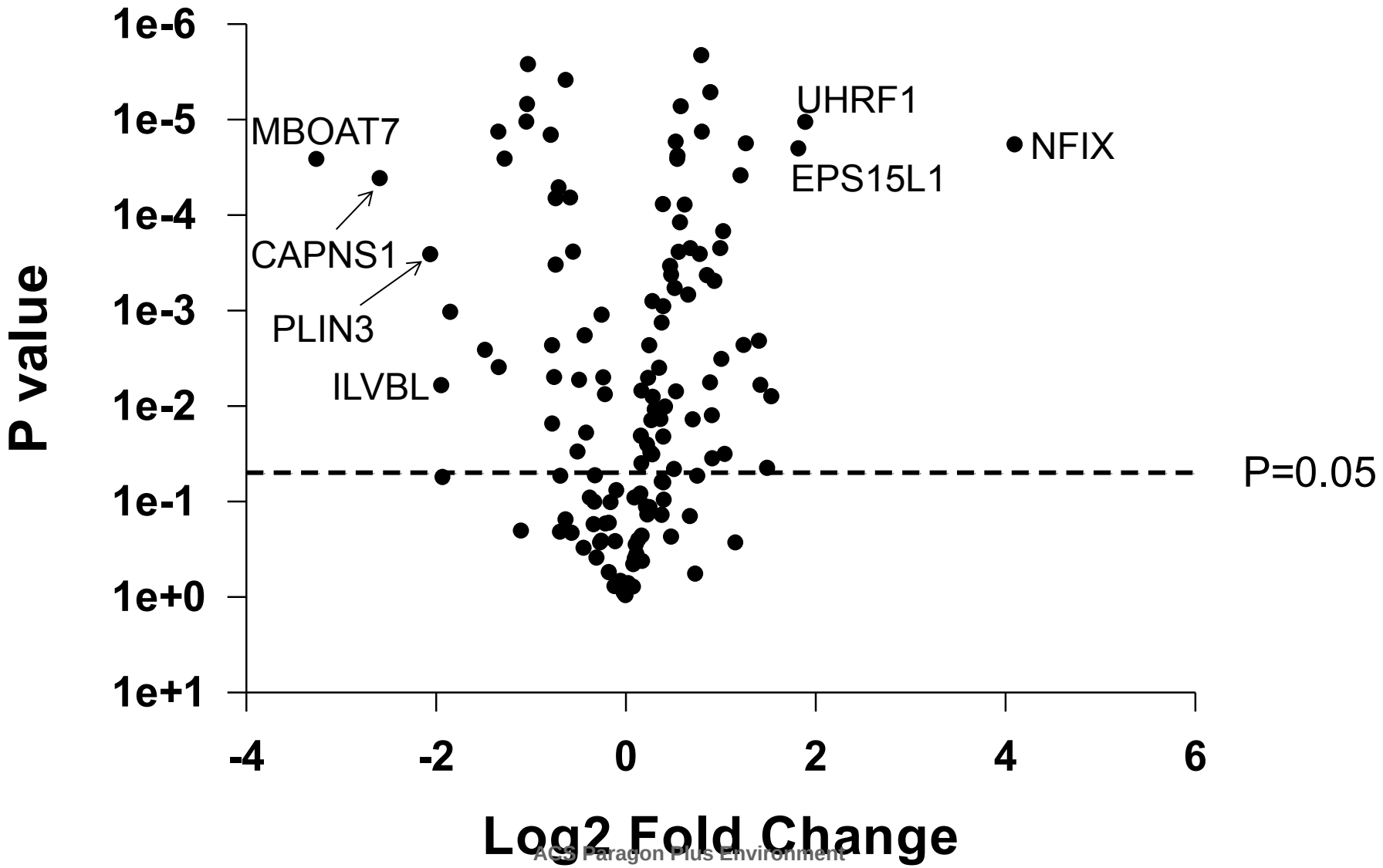


Figure 2C (GSC11)

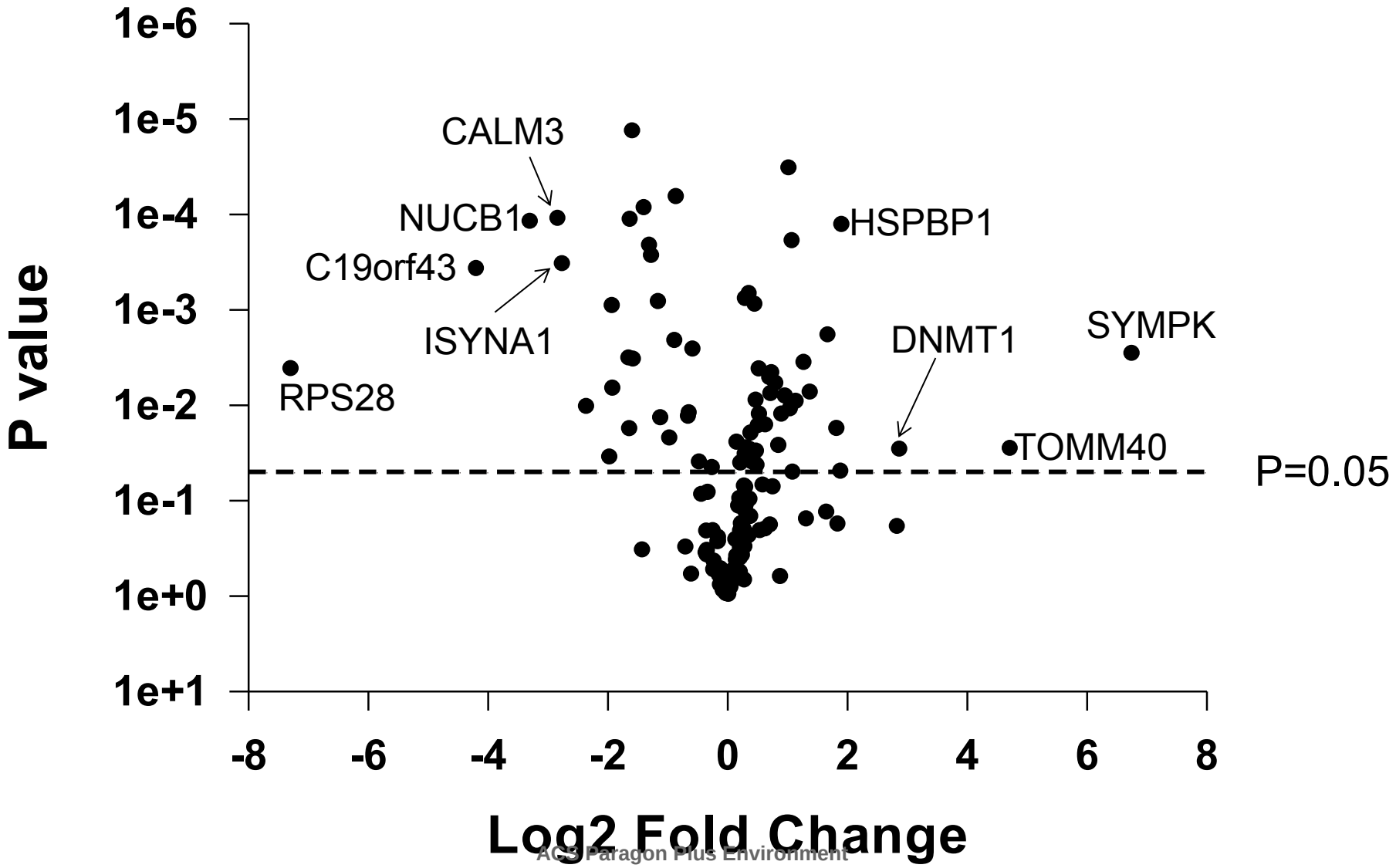


Figure 2D (GSC13)

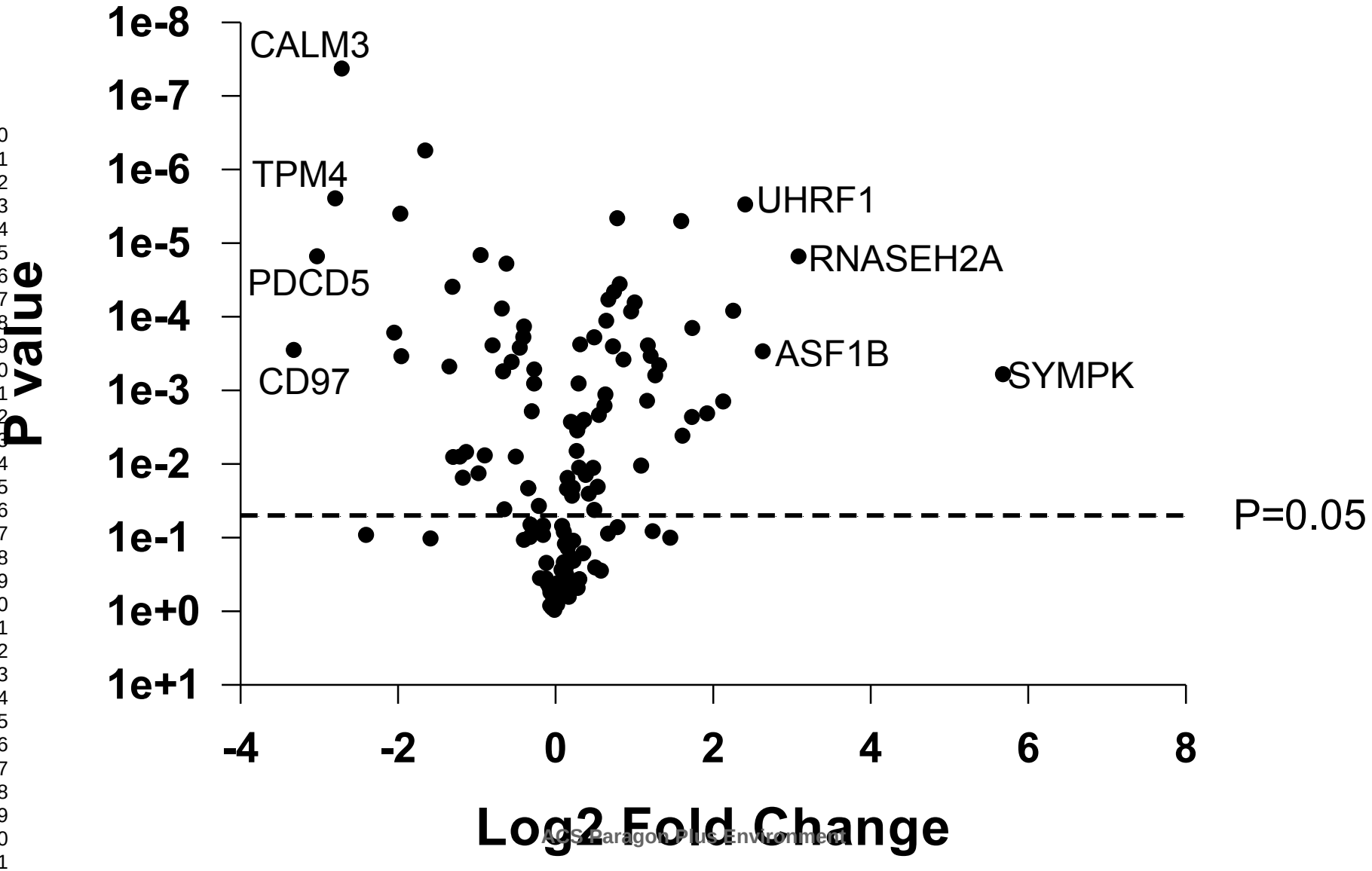


Figure 2E (GSC17)

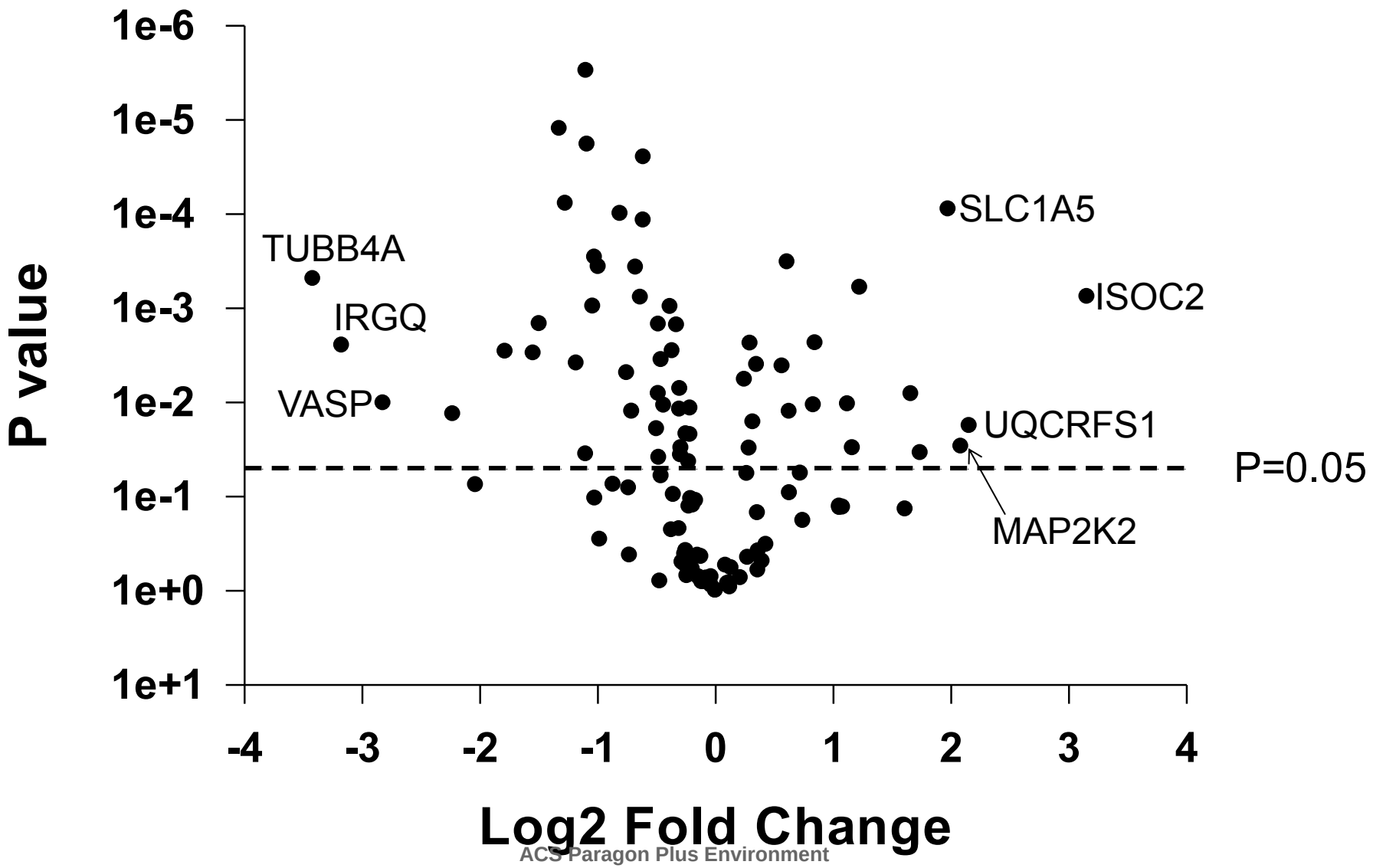


Figure 2F (GSC23)

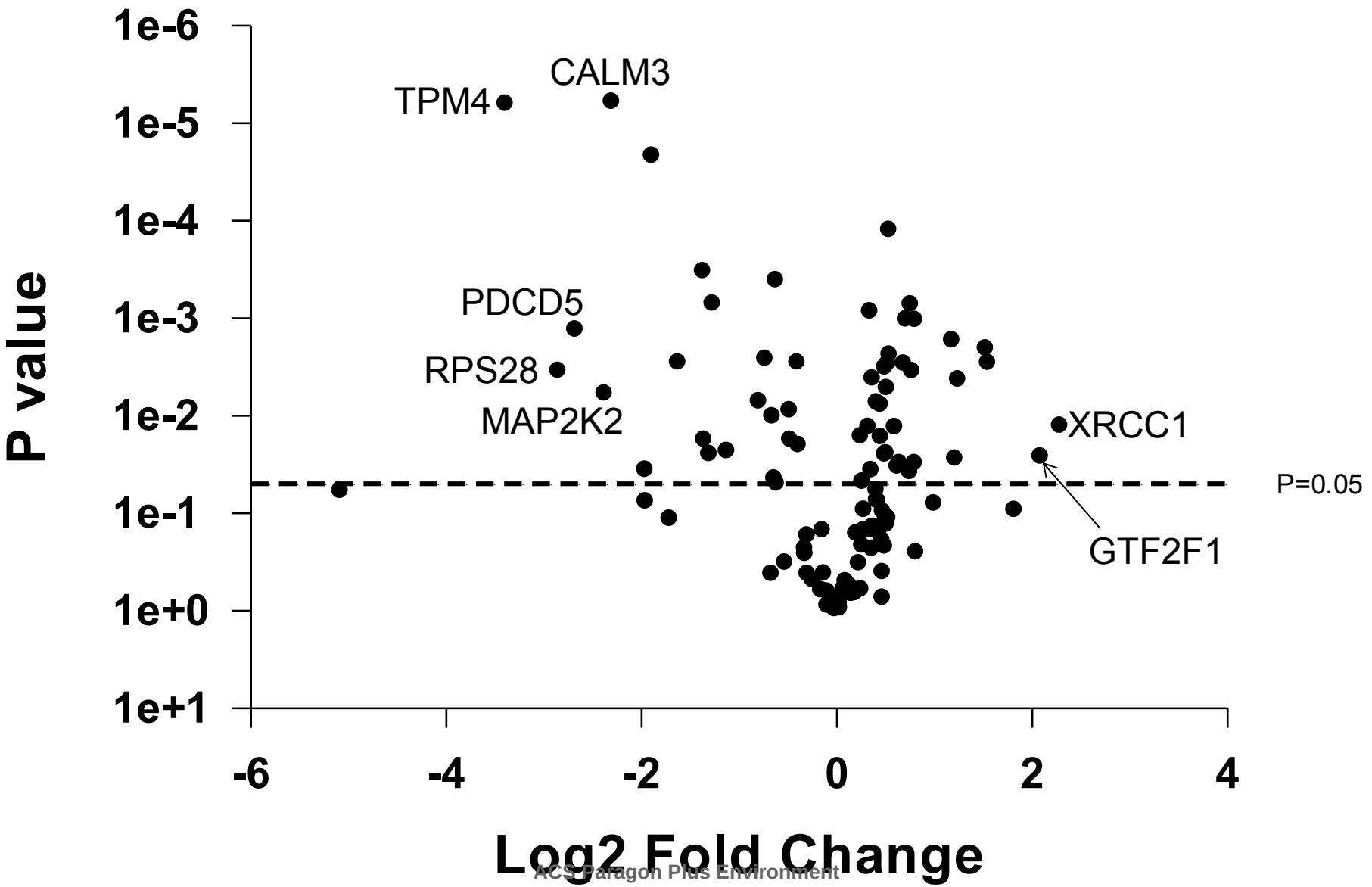


Figure 3

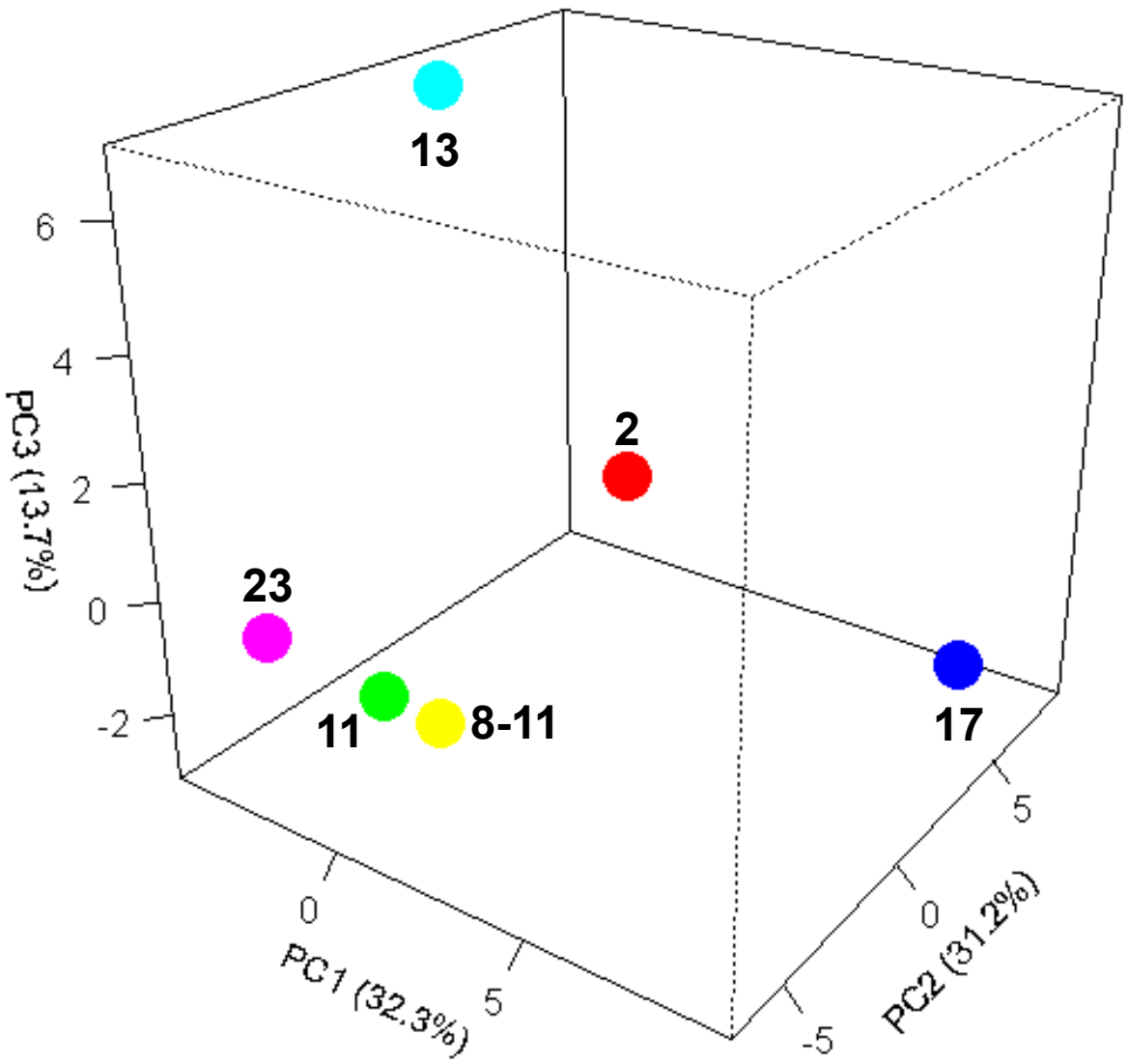


Figure 4A

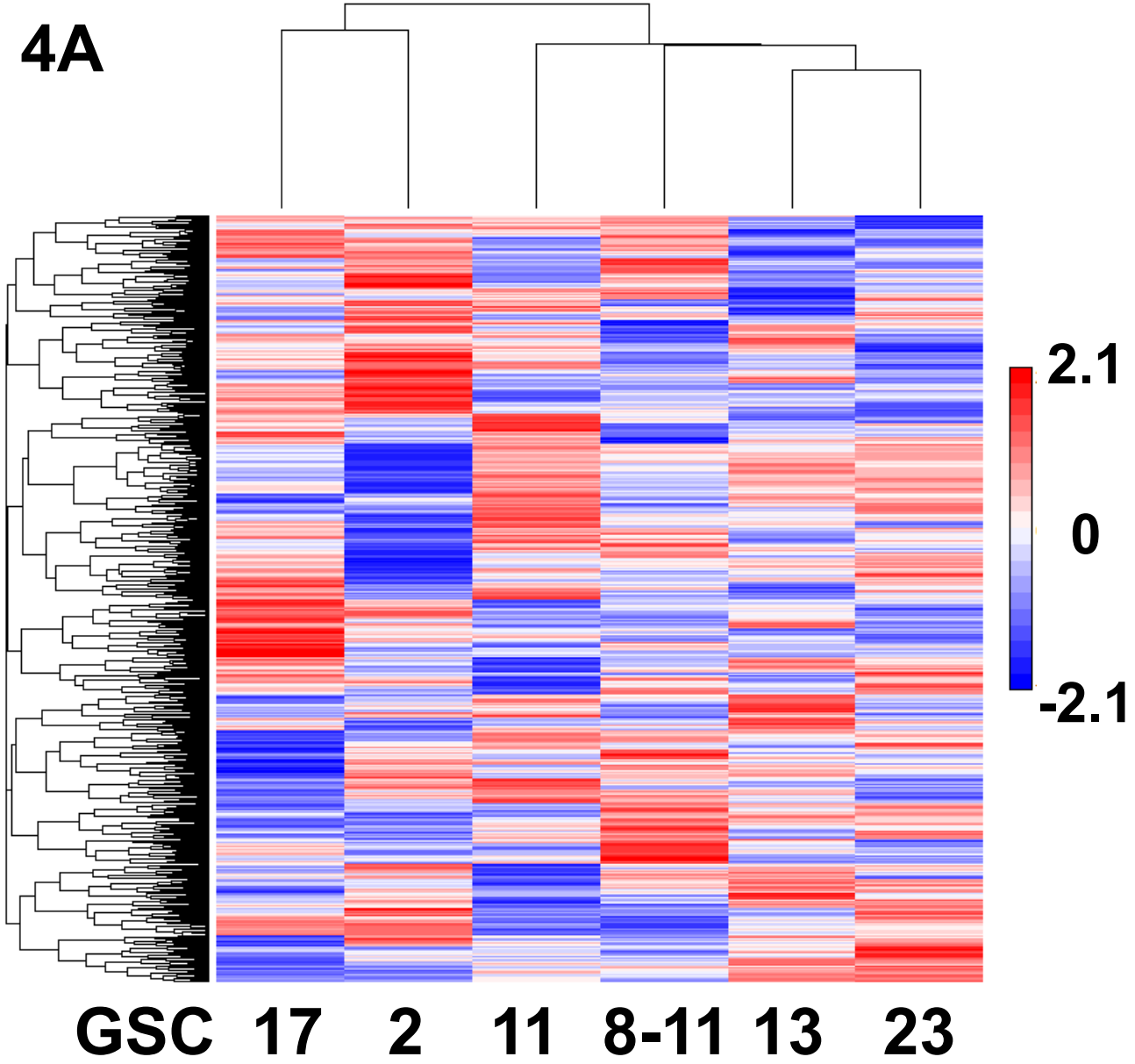
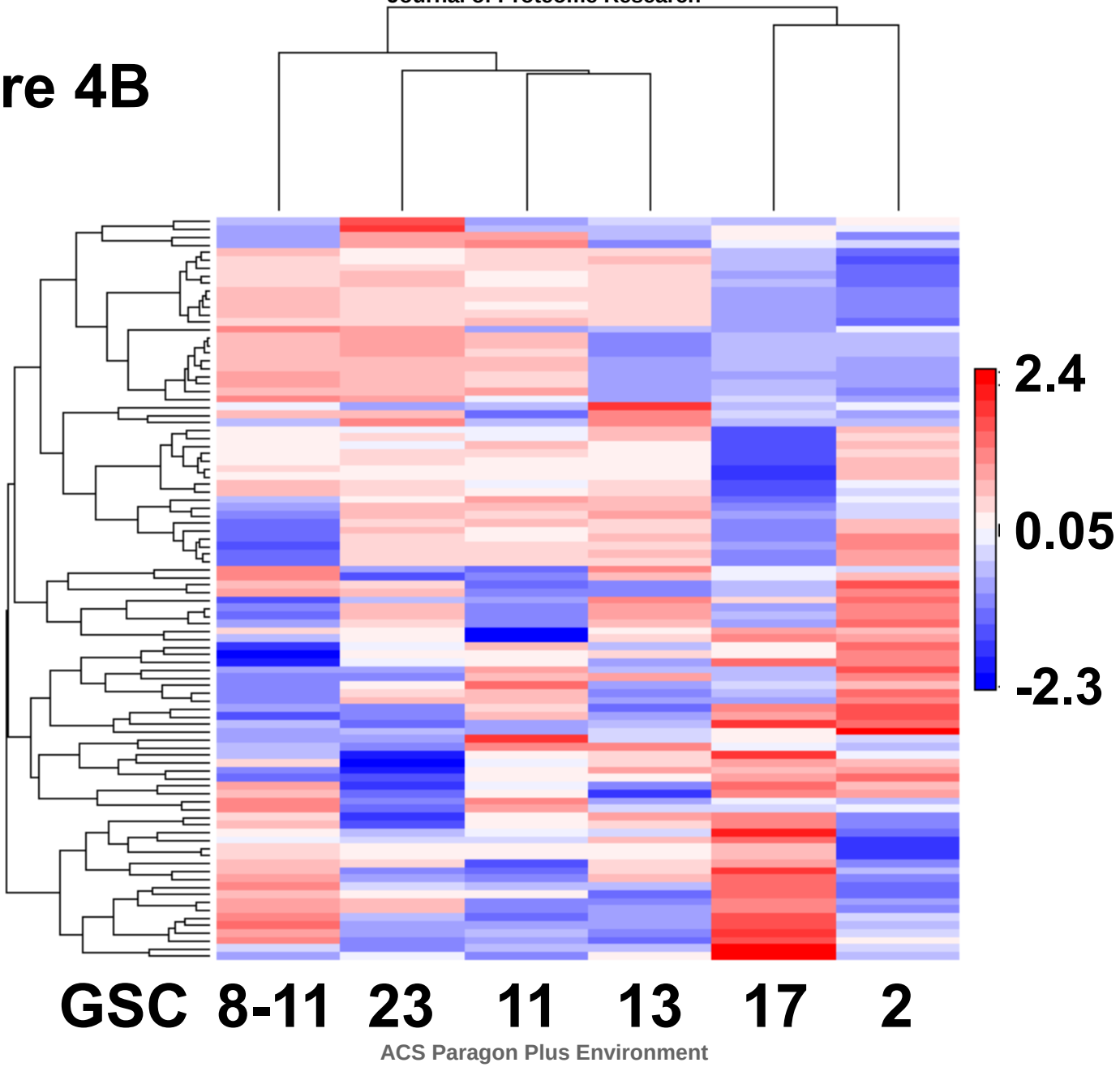


Figure 4B



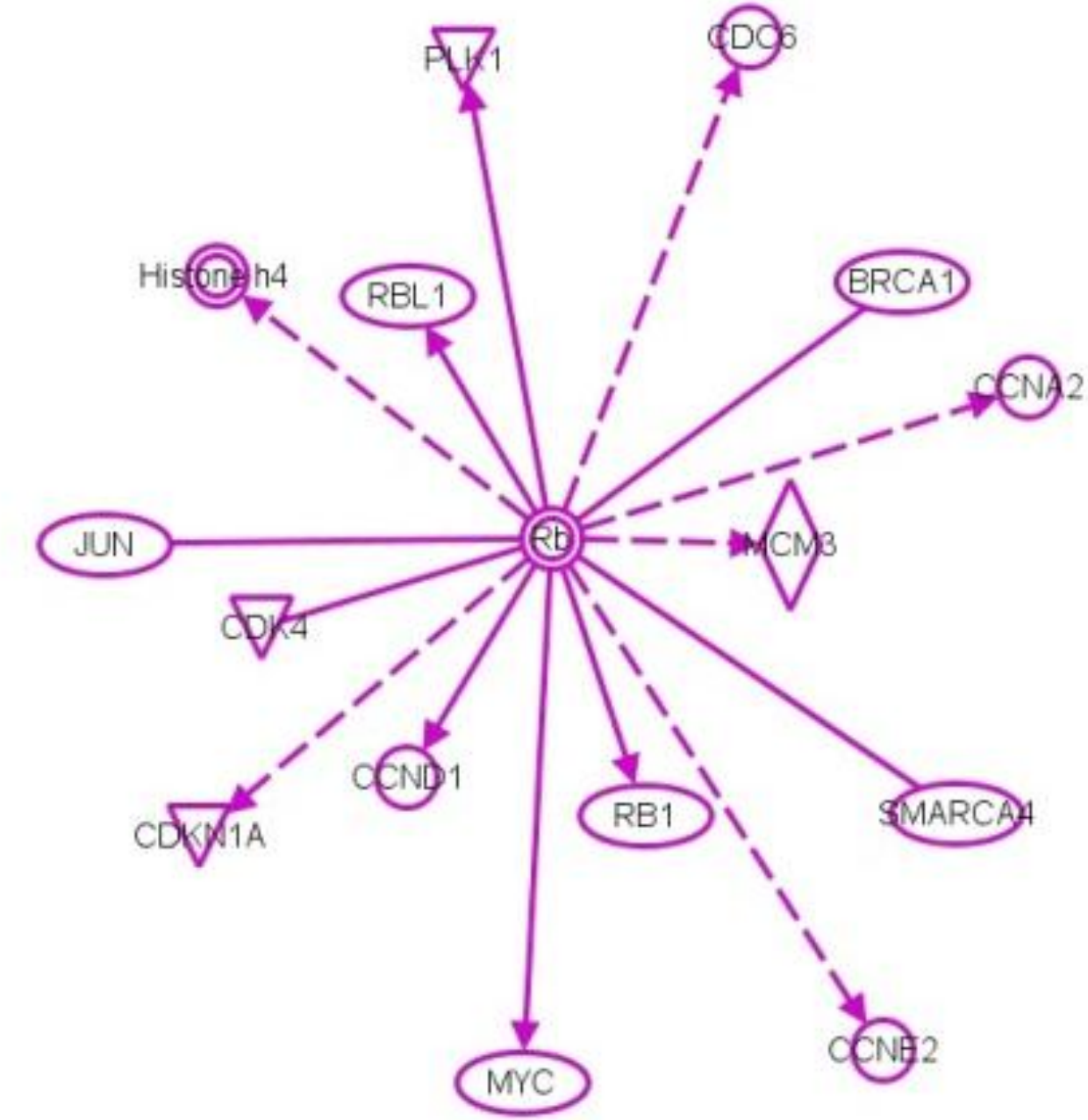
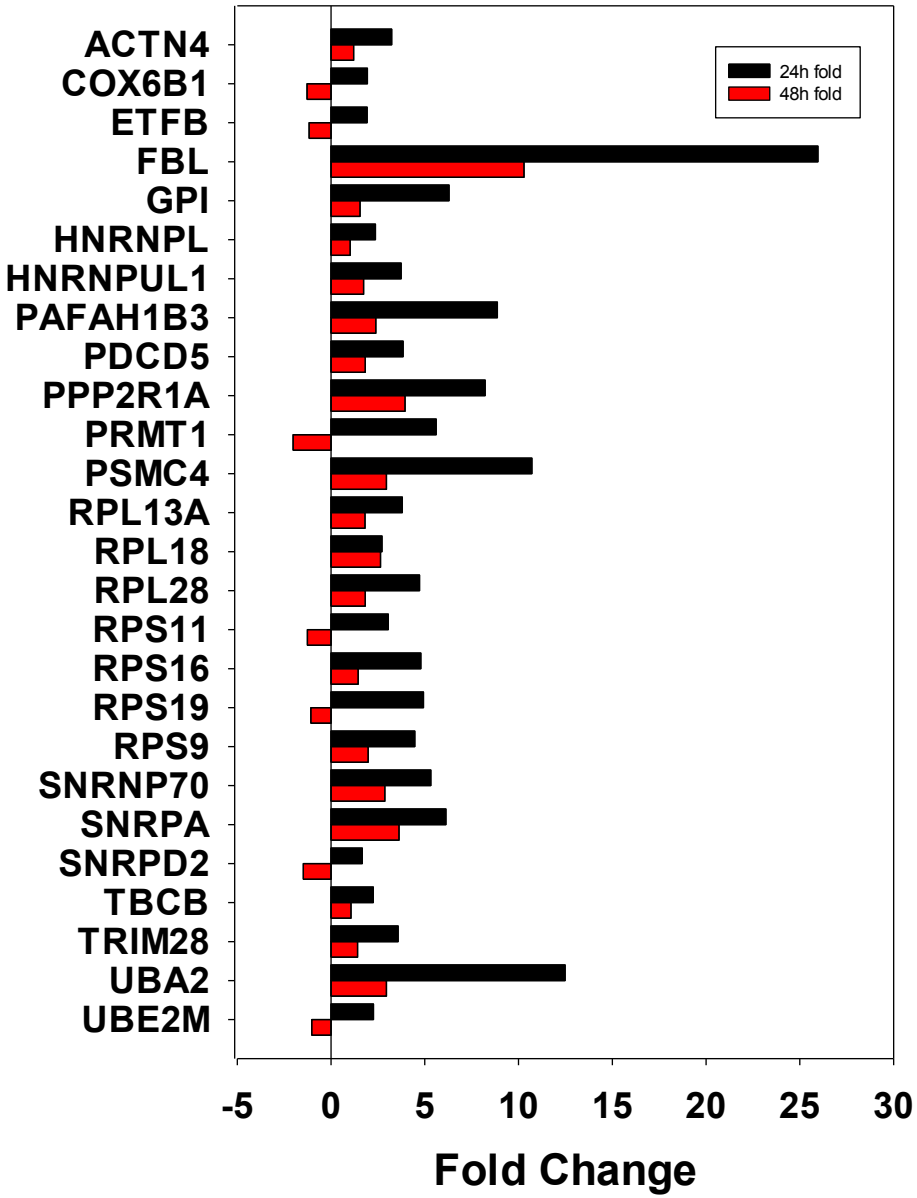

PLK1
CDC6
Histone h4
RBL1
BRCA1
CCNA2
JUN
Rb
MCM3
CDK4
CCND1
CDKN1A
RB1
SMARCA4
MYC
CCNE2

Figure 6


ACTN4
COX6B1
ETFB
FBL
GPI
HNRNPL
HNRNPUL1
PAFAH1B3
PDCD5
PPP2R1A
PRMT1
PSMC4
RPL13A
RPL18
RPL28
RPS11
RPS16
RPS19
RPS9
SNRNP70
SNRPA
SNRPD2
TBCB
TRIM28
UBA2
UBE2M
24h fold
48h fold
-5
0
5
10
15
20
25
30
Fold Change

	P1 *vs.* hNSC	P2 *vs.* hNSC	C *vs.* hNSC	M *vs.* hNSC	M *vs.* P1
Total # differentially expressed	361	247	272	441	287
# Upregulated	173	127	136	197	103
# Downregulated	188	120	136	244	184

Mesenchymal *vs.* hNSC

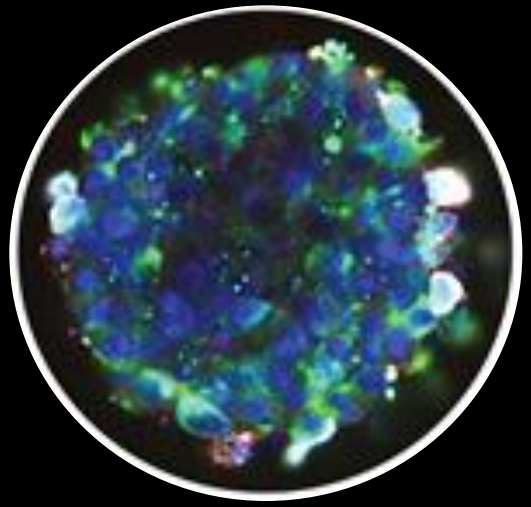
GO ID	Term	Total on array	Downregulated	Upregulated	P-Value (Down)	P-Value (Up)
30246	carbohydrate binding	43	5	16	0.905	**0.0001**
22843	voltage-gated cation channel activity	14	3	5	0.4659	**0.036**
4518	nuclease activity	10	1	4	0.8598	**0.0402**
3700	sequence-specific DNA binding transcription factor activity	84	22	17	**0.0301**	0.0695
8135	translation factor activity, nucleic acid binding	5	4	0	**0.0042**	1
47485	protein N-terminus binding	5	3	0	**0.0419**	1

Mesenchymal *vs.* Proneural 1

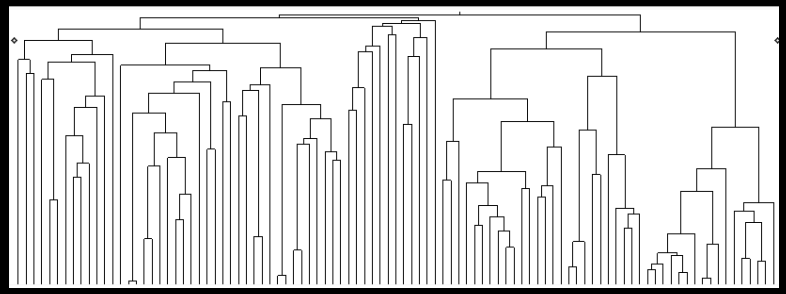
GO ID	Term	Total on array	Downregulated	Upregulated	P-Value (Down)	P-Value (Up)
5515	protein binding	433	59	42	0.4795	**0.0347**
30246	carbohydrate binding	43	4	7	0.8525	**0.0413**
5509	calcium ion binding	32	2	6	0.9435	**0.0311**
15276	ligand-gated ion channel activity	6	1	3	0.5805	**0.0074**
4540	ribonuclease activity	4	1	2	0.4393	**0.0317**
47485	protein N-terminus binding	5	3	0	**0.0195**	1
42393	histone binding	4	3	0	**0.0086**	1
4697	protein kinase C activity	3	2	0	**0.0492**	1

Accession	Name	Gene Name	Description	HPA Evidence Summary	neXtProt Evidence Level
Q8WXI7	MUC16_HUMAN	MUC16	Mucin-16	No Record in HPA	protein level
Q8IY67	RAVR1_HUMAN	RAVER1	Ribonucleoprotein PTB-binding 1	No Record in HPA	protein level
Q96T88	UHRF1_HUMAN	UHRF1	E3 ubiquitin-protein ligase UHRF1	No Record in HPA	protein level
Q6PCB7	S27A1_HUMAN	SLC27A1	Long-chain fatty acid transport protein 1	Only RNA	protein level
P27544	CERS1_HUMAN	CERS1	Ceramide synthase 1	Only RNA	protein level
Q9H313	TTYH1_HUMAN	TTYH1	Protein tweety homolog 1	Low	protein level
O14594	NCAN_HUMAN	NCAN	Neurocan core protein	Low	protein level

Chromosome 19 Expression in Glioma Stem Cells



RNAseq
Targeted C19 Transcriptomics



Quantitative Proteomics