Q-Cell Glioblastoma Resource: Proteomics Analysis Reveals Unique Cell-States Are Maintained in 3D Culture

Rochelle C. J. D'Souza [1], **Carolin Offenhäuser** [1], **Jasmin Straube** [2], **Ulrich Baumgartner** [1], **Anja Kordowski** [1], **Yuchen Li** [1], **Brett W. Stringer** [1], **Hamish Alexander** [1,3], **Zarnie Lwin** [1,3], **Po-Ling Inglis** [1,3], **Rosalind L. Jeffree** [1,3], **Terrance G. Johns** [4], **Andrew W. Boyd** [1,5] and **Bryan W. Day** [1,5,6,*]

[1] Cell and Molecular Biology Department, QIMR Berghofer Medical Research Institute, Sid Faithfull Brain Cancer Laboratory, Brisbane, QLD 4006, Australia; Rochelle.D'Souza@qimrberghofer.edu.au (R.C.J.D.); Carolin.Offenhauser@qimrberghofer.edu.au (C.O.); Ulrich.Baumgartner@qimrberghofer.edu.au (U.B.); Anja.Kordowski@qimrberghofer.edu.au (A.K.); Michelle.Li@qimrberghofer.edu.au (Y.L.); Brett.W.Stringer@gmail.com (B.W.S.); Hamish.Alexander@health.qld.gov.au (H.A.); Zarnie.Lwin@health.qld.gov.au (Z.L.); Po-Ling.Inglis@health.qld.gov.au (P.-L.I.); Lindy.Jeffree@health.qld.gov.au (R.L.J.); Andrew.Boyd@qimrberghofer.edu.au (A.W.B.)
[2] Immunology Department, QIMR Berghofer Medical Research Institute, Gordon and Jessie Gilmour Leukaemia Research Laboratory, Brisbane, QLD 4006, Australia; Jasmin.Straube@qimrberghofer.edu.au
[3] Royal Brisbane and Women's Hospital, Brisbane, QLD 4006, Australia
[4] Telethon Kids Institute, Perth, WA 6009, Australia; Terrance.Johns@telethonkids.org.au
[5] School of Biomedical Sciences, The University of Queensland, Brisbane 4072, Australia
[6] School of Biomedical Sciences, Faculty of Health, Queensland University of Technology, Brisbane 4059, Australia
* Correspondence: Bryan.Day@qimrberghofer.edu.au; Tel.: +61-7-3845-3885

check for updates

Abstract: Glioblastoma (GBM) is a treatment-refractory central nervous system (CNS) tumour, and better therapies to treat this aggressive disease are urgently needed. Primary GBM models that represent the true disease state are essential to better understand disease biology and for accurate preclinical therapy assessment. We have previously presented a comprehensive transcriptome characterisation of a panel (n = 12) of primary GBM models (Q-Cell). We have now generated a systematic, quantitative, and deep proteome abundance atlas of the Q-Cell models grown in 3D culture, representing 6167 human proteins. A recent study has highlighted the degree of functional heterogeneity that coexists within individual GBM tumours, describing four cellular states (MES-like, NPC-like, OPC-like and AC-like). We performed comparative proteomic analysis, confirming a good representation of each of the four cell-states across the 13 models examined. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis identified upregulation of a number of GBM-associated cancer pathway proteins. Bioinformatics analysis, using the OncoKB database, identified a number of functional actionable targets that were either uniquely or ubiquitously expressed across the panel. This study provides an in-depth proteomic analysis of the GBM Q-Cell resource, which should prove a valuable functional dataset for future biological and preclinical investigations.

Keywords: glioblastoma (GBM); proteomics; GBM cell-states; recurrence; tumour heterogeneity; therapeutic targets

1. Introduction

Glioblastoma (GBM) is the most common and aggressive form of adult brain cancer, and patients have not experienced significant increases in overall survival for several decades [1–3]. This lack of progress is, in part, attributable to the significant heterogeneity that exists within these tumours [4–8]. Recent seminal studies by Suva and colleagues have discovered that multiple dynamic cell-states (MES-like, NPC-like, OPC-like and AC-like) can reside within a single tumour entity, highlighting the degree of heterogeneity and plasticity present within GBM [9]. Therefore, therapeutic strategies with the ability to target each functional state have the potential to produce more durable clinical responses. Another significant factor preventing successful disease translation has been the lack of faithful models to drive meaningful neuro-oncology discoveries. To address this problem, we embarked upon the generation and characterisation of a panel of clinically relevant primary in vitro and in vivo GBM models (termed Q-Cell) [10,11]. Our previous Q-Cell characterisation studies have primarily focused on transcriptome expression analysis using microarray and RNA Sequencing (RNA-Seq) approaches. This study sought to expand upon our previous transcriptome findings by better understanding the proteomic profile of these models. Relative protein abundances of 6167 human proteins were quantitated using mass spectrometry (MS) and their expression patterns analysed bioinformatically. Principal component analysis, conducted on independent Q-Cell replicates, revealed consistent protein expression within a given model. These data indicated a high degree of fidelity at the protein level. All Q-Cell GBM models have been generated from IDH1 WT (wild-type) disease specimens. Not surprisingly, given this IDH1 wild-type (WT) status, we observed a high degree of correlation when comparing total protein expression between each model. Correlation at the protein level was most pronounced in a pair-matched pre- and post-treatment GBM specimen (SB2/SB2b). We also sought to analyse the relevance of the recently described GBM cell-states in our Q-Cell resource. Proteomics analysis revealed a good representation of each of the four cell-states across the 13 models examined. We also characterised a number of hybrids (n = 5) that displayed significant heterogeneity with elevated protein expression of multiple cell-state markers within a given model. This finding was consistent with the Suva study, where the single-cell RNA-Seq data showed a significant number of tumour cells in transition states [9]. Despite a small sample size, detailed analysis of the single pair-matched model SB2/SB2b highlighted a transition at recurrence to a more MES-like state. This is consistent with the findings of others, indicating an increase in mesenchymal disease when GBM tumours recur post-therapy [4,12]. Furthermore, of the 257 cell-state-characterisation genes assessed, we detected 153 (57%) at the protein level. This finding indicated reasonably high levels of mRNA translation, providing significant scope for future biological and preclinical therapy studies aimed at targeting multiple GBM cell-states.

Unsupervised hierarchical clustering identified that eight of the models tended to express structural proteins associated with a broad mesenchymal-like signature, while the remaining five models expressed proteins associated with a more neuronal-like GBM signature. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis identified upregulation of a number of GBM-associated cancer pathway proteins across the 13 models, as expected. We also undertook bioinformatics analysis, using the OncoKB database, to identify functional actionable targets that were either uniquely or ubiquitously expressed across the models.

The present effort has sought to provide a detailed proteomics analysis and build upon our previously described transcriptomic analysis of the Q-Cell GBM resource. Findings of this study have increased our understanding of GBM tumour heterogeneity and cell-states at the protein level and further validated our previous mRNA datasets. Given the highly functional nature of proteomics analysis, we hope this freely available GBM resource and combined datasets will provide a highly relevant functional platform to drive meaningful biological and translational neuro-oncology discoveries into the future.

2. Materials and Methods

2.1. Cell Culture

We have developed a characterised GBM patient-derived cell line resource (Q-Cell) [10–13], in which lines are maintained as glioma neural stem cell (GNS) cultures [14] or as 3D neurosphere cultures [15] using StemPro NSC SFM (Invitrogen, Carlsbad, CA, USA) or KnockOut™ DMEM (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) as per the manufacturer's guidelines. Characterisation data is freely available from https://www.qimrberghofer.edu.au/q-cell/. All tissues were collected following ethical approval from the Royal Brisbane Women's Hospital and QIMR Berghofer Human Research Ethics Committees. Ethical approval number: P3420, HREC/17/QRBW/577 Novel Therapies for Brain Cancer. In order to maintain pluripotency, KnockOut™ DMEM (Gibco) media were supplemented with GlutaMAX™ Supplement (Gibco), StemPro™ Neural Supplement (Gibco), Recombinant Human EGF (Gibco), Recombinant Human FGFb (Gibco), and Penicillin/Streptomycin (Gibco). Cells were cultured on flasks coated with Basement Membrane Matrigel®Matrix (Corning, New York, NY, USA). Passaging the cells was done by detaching the cells from the flask surface using Accutase®solution (Sigma-Aldrich, St. Louis, MO, USA). Glioma cells were cultured as glioma neural stem (GNS) cultures as outlined in detail in [14] or as tumourspheres using StemPro® NSC SFM.

2.2. Proteome Sample Preparation

GBM cells (5×10^6) were seeded in T75 flasks with StemPro® NSC SFM (Thermo Fisher Scientific, Waltham, MA, USA). Cell lysates were collected in biological triplicates 72 hours postseeding. Cells were lysed in buffer containing 2% SDS (Sigma-Aldrich), EDTA-free protease inhibitor (Roche, Basel, Switzerland) and 100 mM tris-HCl (pH 7.6) [16]. The lysates were sonicated and incubated for 5 min at 95 °C, protein concentration was quantified, and samples were processed by an in-solution digest protocol. Briefly, 100 μg of cell lysate was precipitated overnight with 5 volumes of ice-cold acetone. The precipitate was collected by centrifugation and resuspended in 8 M urea (Sigma-Aldrich). The samples were reduced with dithiothreitol (Sigma-Aldrich) for 1 h at room temperature, alkylated with iodoacetamide (Sigma-Aldrich) in the dark, diluted to a final concentration of 2 M urea, 10% acetonitrile (Merck, Kenilworth, NJ, USA) and 1 μg of trypsin (Promega, Madison, WI, USA) and digested to peptides overnight at 37 °C. The peptides were desalted on C18 stage tips as described [17] and resuspended in 0.1% formic acid.

2.3. Liquid Chromatography (LC) and Mass Spectrometry (MS) Analysis

Peptides were injected under trapping conditions using a Waters (Milford, MA, USA) NanoAcquity system interfaced to a linear triple quadrupole (LTQ)-Orbitrap elite mass spectrometer (Thermo Fisher Scientific). Acidified digested samples were loaded onto a Waters 2G-V/M C18 Symmetry trap (5 μm particle size, 180 μm × 20 mm) at 5 μL/min in 99% solvent A (0.1% (v/v) aqueous formic acid) and 1% solvent B (100% (v/v) acetonitrile containing 0.1% (v/v) formic acid) for 3 minutes, and the peptides were subsequently separated inline using a pre-equilibrated analytical column (Waters C18 BEH 130 Å, 1.7 μm particle size, 75 μm × 200 μm) at a flow rate of 0.3 μL/min and temperature of 35 °C. The gradient ran at 300 nL/min from 2% B (0.1% formic acid in acetonitrile) to 40% B over 190 min, rose to 95% B to wash the column, then re-equilibrated at 2% B for the next injection. Eluate was delivered into the mass spectrometer via a Nanospray Flex Ion Source (Thermo Fisher Scientific) containing a 10 μm P200P coated silica emitter (New Objective, Woburn, MA, USA). Typical spray voltage was 1.8 kV with no sheath, sweep, or auxiliary gases; the heated capillary temperature was set to 285 °C. The LTQ-Orbitrap Elite mass spectrometer (Thermo Fisher Scientific) was controlled using Xcalibur 2.2 SP1.48 software (Thermo Fisher Scientific) and operated in a data-dependent acquisition mode to automatically switch between Orbitrap-MS and collision induced dissociation (CID)- based ion trap-MS/MS. The survey full scan mass spectra (from m/z 380 to 1700) were acquired in the Orbitrap with a resolving power of 120,000 after accumulating ions to an automatic gain control (AGC) target value of 1.0×10^6 charges in

the LTQ. MS/MS spectra were concurrently acquired in the LTQ on the 20 most intense ions from the survey scan, using an AGC target value of 1.0×10^4. Charge state filtering (unassigned precursors and singly charged ions were not selected for fragmentation) and dynamic exclusion (repeat count 1, repeat duration 30 s, exclusion list size 500, exclusion duration 90 s) were used. CID fragmentation conditions in the LTQ were 35% normalized collision energy, activation q of 0.25, 10 ms activation time, and minimum ion selection intensity of 5000 counts. Maximum ion injection times were 200 ms and 50 ms for survey full scans and MS/MS scans, respectively.

2.4. Data Processing and Analysis

Spectra were analyzed using MaxQuant [18] version 1.5.3.30, and Andromeda [19]. The MS2 spectra were searched against the UniProt FASTA database version 31/01/2013. Enzyme specificity was set to trypsin, allowing for cleavage N-terminal to proline and between aspartic acid and proline. The search included carbamidomethylation of cysteine as a fixed modification and N-acetylation and oxidation of methionine as variable modifications. Up to two missed cleavages were allowed for protease digestion, and enzyme specificity was set to trypsin, defined as C-terminal to arginine and lysine excluding proline. The *identify* module in MaxQuant was used to filter (1% false identification rate (FDR)) identifications at the peptide and protein level. The identity of precursor peptides present in MS1, but not selected for fragmentation and identification by MS2 in a given run, was obtained by transferring peptide identifications based on accurate mass and retention times across liquid chromatography–mass spectrometry (LC–MS) runs where possible using MaxQuant [20]. Protein identifications were collapsed to the minimal number that contained the set of identified peptides. Proteome quantification was performed in MaxQuant using the extracted ion chromatography (XIC)-based label-free quantification (LFQ) algorithm [21]. In MaxQuant, a quantification event was reported only when isotope pattern could be detected and was consistent in terms of charge state of peptide. For quantification, intensities were determined as the intensity maximum over the retention time profile. Intensities of different isotopic peaks in an isotope pattern were summed up for further analysis. All RAW files and protein-based quantification results are available for download from the Q-Cell website at https://www.qimrberghofer.edu.au/q-cell/.

2.5. Bioinformatics and Statistical Analysis

Bioinformatics analyses were performed using Perseus in MaxQuant [22]. Proteins identified on the basis of at least one unique peptide were used for all subsequent analyses. We selected the normalized abundances of proteins that were quantified in duplicates from at least one cell line. For comparing differences between all cell lines, biological triplicates were grouped by cell line, and the analysis of variance (ANOVA) was performed. We used the ANOVA method with largest power, permutation-based FDR of 0.05, and at least 250 repetitions for truncation. A two-sided student's t-test was used to perform the comparison between two cell lines SB2 and SB2b employing a *p*-value cutoff of 0.05.

Unsupervised hierarchical clustering was performed on the z-score-transformed protein intensities or the z-scored log2 normalised mRNA read counts. Proteins were annotated with GO (Gene Ontology) biological process terms [23] and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway terms [24]. Statistical analysis of the enrichment of proteins with functional annotations was evaluated using a Fisher's exact test (*p*-value <0.05). An enrichment score of greater than 1 indicates enrichment, and a score from 0 to 1 indicates negative enrichment. For comparison of mRNA and protein expression of the GBM-associated genes, the normalized protein intensities and normalized mRNA read counts were matched by gene name, and Pearson's coefficient of correlation was calculated.

2.6. Cell-State ssGSEA

NPC1, NPC2, MES1, MES2, APC, and OPC marker genes were extracted from [9]. NPC1 and NPC2 were combined into a single gene set called NP, and MES1 and MES2 were combined into

MES. z-score-scaled protein intensities of the models were assessed for NPC, MES, APC, and OPC enrichment using the R GSVA library (v1.32) with the ssGSEA algorithm as developed by Barbie et al. (2009) [25]. Principal component analysis was performed on scaled and centered ssGSEA enrichment scores. Replicates were summarised using an ellipse with the Euclidean distance from the centre and radius of 1. Heat maps were generated on the average ssGSEA enrichment scores of the cell line replicates. Hierarchical clustering of models and marker gene sets was performed on the 1-Spearman correlation dissimilarity matrix using the ward.D2 algorithm. A similar analysis was performed on the log2-normalised mRNA read counts from the previous transcriptomics study.

2.7. Cytoscape-Based Data Visualisation

Proteins identified from the list of metamodule genes describing the six different cell-states (MES1-, MES2-, NPC1-, NPC2-, OPC-, and AC-like) were visualised as nodes and grouped by the respective cell-states. z-score-scaled protein intensity values were plotted using continuous colour mapping of the nodes in Cytoscape version 3.7.2 [26]. For enrichment analysis, the top 100 significantly regulated proteins with the highest fold change absolute value were analysed for enrichment of gene ontology (GO) molecular functions, GO biological processes, and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways using the ClueGO App version 2.5.5 [27] in Cytoscape version 3.7.2 [26]. Analysis parameters were set to include GO terms (level 3 to 8), and KEGG pathways for which at least 5 proteins and 4% of associated proteins per term/pathway were present in the list of regulated proteins. A two-sided hypergeometric test with Benjamini–Hochberg correction was performed to identify significantly enriched/depleted terms (p-value <0.05) against a custom reference set consisting of all proteins identified in the SB2 and SB2b MS datasets. The kappa score threshold for clustering of related terms was set to 0.4.

3. Results

3.1. Q-Cell Proteomics Analysis

To accurately quantify proteome expression in the Q-Cell GBM panel, we performed label-free quantitative MS analysis of protein lysates from GBM cells grown in 3D neurosphere suspension cultures [15]. Primary models were initially grown as adherent GNS cultures on matrigel under serum-free conditions [14]. In order to avoid confounding our results with high abundance proteins present in the basement membrane of matrigel, cells were grown in 3D suspension cultures as neurospheres for 72 hours before lysis. We have previously shown that GBM cells either grown as neurospheres or on a substrate of matrigel tend to display equivalent characteristics [28]. Cell lysates were subjected to a defined proteomics workflow (Figure 1A). Peptides were analysed with 4 hour gradients on an LTQ-Orbitrap Elite mass spectrometer with CID fragmentation. Combined analysis of the spectra from all samples resulted in the identification of 6180 proteins at FDR of 1%. Using the label-free quantification algorithm in MaxQuant [21], we measured the abundance of 6167 proteins in at least one pairwise comparison (Table S1). Analysis of triplicate samples showed excellent reproducibility when compared using principal component analysis of all quantified proteins (Figure 1B). Next, we compared the expression of the proteome by calculating the Pearson correlation coefficient (r) between each model. As expected, the highest correlation (r = 0.97) was observed between the pair-matched primary and recurrent model (SB2 and SB2b), while the least similar model identified was PB1 (r = 0.82) (Figure 1C).

We next measured the number of proteins identified in each of the triplicate runs from each Q-Cell model. Proteomics analysis identified an average of 2200 common proteins between each model. PB1 and SJH1 expressed the largest number of unique proteins (145 and 85 respectively) and therefore correlated the least to the remaining 11 models (Figure 1D and Table S1). Some of the uniquely expressed genes include *CD44* (quantified in FPW1), *ERBB2* (quantified in RN1), and *OLIG2* (quantified in PB1) and are genes with key roles in brain cancer. *CDKN1A* and *CDKN2A* were highly

expressed in RKI1 and detected in JK2 and MMK1; these cell lines are the only ones in our panel which do not have deletion of the coding gene at the genome level.
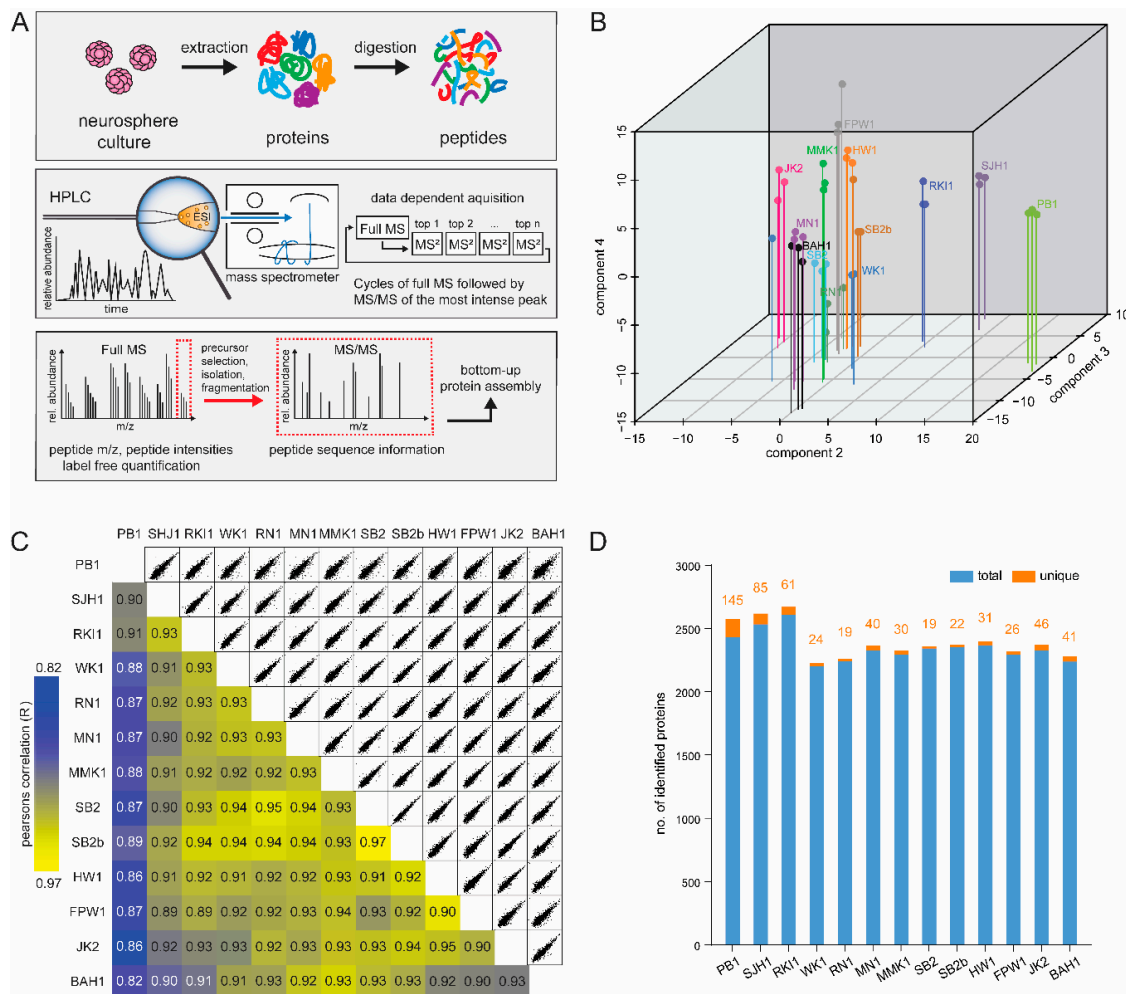
Figure 1. Proteome mapping of 13 Q-cell cell lines. (**A**) Graphical illustration of the workflow for glioblastoma (GBM) cell line proteome analysis. GBM primary cell lines were cultured as neurospheres, lysed in SDS-based buffer, trypsin digested by in-solution digestion in biological triplicate. LC–MS/MS with 4 hour runs and CID fragmentation were performed in the LTQ-Orbitrap Elite. (**B**) 3D-PCA plot. The proteome of thirteen cell lines measured in triplicates segregated into major cell types showing well-correlated triplicates within a cell line. (**C**) The matrix of 78 correlation plots revealed very high correlations between protein intensities in triplicates (Pearson correlation coefficient 0.82–0.97 between cell types). The colour code follows the indicated values of the correlation coefficient. (**D**) A bar chart showing the number of total proteins identified in each of the cell lines with false identification rate (FDR) of 1%.

3.2. GBM Cell-State Analysis

As outlined above, four dynamic cell-states which functionally drive intratumoural heterogeneity within GBM have recently been described [9]. To better understand the contribution of these GBM cell-states within our Q-Cell resource, we firstly analysed 257 unique genes, separating tumours into six metamodules (MES1-, MES2-, NPC1-, NPC2-, OPC-, and AC-like) as per Suva and colleagues [9] encompassing each of the identified four cell-states. We next matched gene expression to the 6172 identified proteins from our MS analysis. Identified proteins corresponded to 38/50 and 29/50 genes from MES1 and MES2, 24/50 and 28/50 genes from NPC1 and NPC2, 26/39 genes from

OPC and 30/50 genes from AC-like metamodules respectively (Table S1). A total of 153 proteins were identified from the corresponding 257 unique cell-state genes outlined by Suva and colleagues.

To identify the contribution of each cell-state in the Q-Cell panel, z-score-scaled protein intensities were assessed for enrichment of the four states using an ssGSEA algorithm [25]. We thus obtained a cell-state score, which was used to predict the predominant cell-state of each model (Figure 2A and Figure S1). Four of the models showed an MES-like state, while two of each model showed an NPC-like and AC-like profile. Four models showed an enrichment of more than one cell-state, termed hybrids. This finding was consistent with Suva and colleagues, highlighting a significant number of GBM cells present in transition states [9]. Principal component analysis showed an enrichment of the predominant cell-state when compared to the rest (Figure 2B and Figure S2). JK2 and MN1 displayed a clear MES-like phenotype matching a total of 39 proteins from the Suva study, with 25 of these proteins showing very high expression (Figure 3 and Table S1). Similarly, HW1 and FPW1 showed a strong AC-like phenotype with high expression of 15 matched proteins, and RKI1 and WK1 displayed the strongest NPC-like phenotype, while BAH1 showed an OPC-like phenotype (Figure 3). Our analysis identified four of the Q-cell lines, PB1, SHJ1, RN1, and MMK1, to be intermediate hybrids, showing a strong ssGSEA score of more than one cell-state, which is concordant with the results from Suva and colleagues (Figure 3). In addition, we similarly generated cell-state scores for 12 of these cell lines previously analysed by transcriptomics using 252 genes that were identified (Figure 2C and Table S2) [11]. The proteomics-based and transcriptomics-based cell-state phenotypes for our cell lines were remarkably similar, except for FPW1, which switched from being AC-like to OPC-like.

Our panel of 13 primary GBM cell lines consisted of one pair-matched primary and one recurrent tumour, designated SB2 and SB2b; respectively. SB2 showed an OPC-like state, while SB2b was enriched for the MES-like cell-state with increased expression of marker proteins SERPINE1, GSN, NNMT, SOD2, PLP2, and PRDX6 (Figure 4A,B) with an overall increase of the ssGSEA MES score. We performed a student's t-test (two-sided, *p*-value cut off >0.05) to identify the significantly regulated subset of proteins between the two cell lines (Table S3) and then performed enrichment analysis to look for GO molecular functions and KEGG pathways that are among the top 100 most regulated genes in the recurrent SB2b model. The analysis revealed an enrichment of DNA recombination and replication, ECM organisation and cell–substrate junctions, antigen presentation, morphogenesis and organ development, and calcium homeostasis (Figure 3C and Table S4).
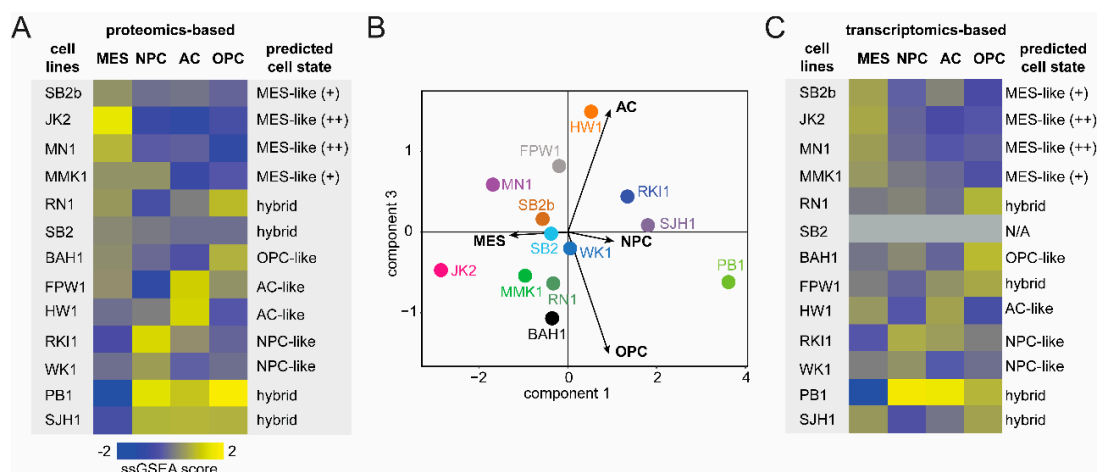

Figure 2. ssGSEA based cell-state scores (**A**) Heatmap of the z-scored ssGSEA scores of the thirteen cell line proteomes with predicted cell-states on the right. (**B**) PCA plot showing the distribution of the cell lines across the four cell-states. (**C**) Heatmap of the z-scored ssGSEA scores of the twelve cell line transcriptomes with predicted cell-states on the right. (Note: SB2 was not analysed by transcriptomics). Abbreviations: MES-like, mesenchymal-like; AC-like, astrocyte-like; OPC-like, oligodendrocyte progenitor cell-like; NPC-like, neural progenitor cell-like.
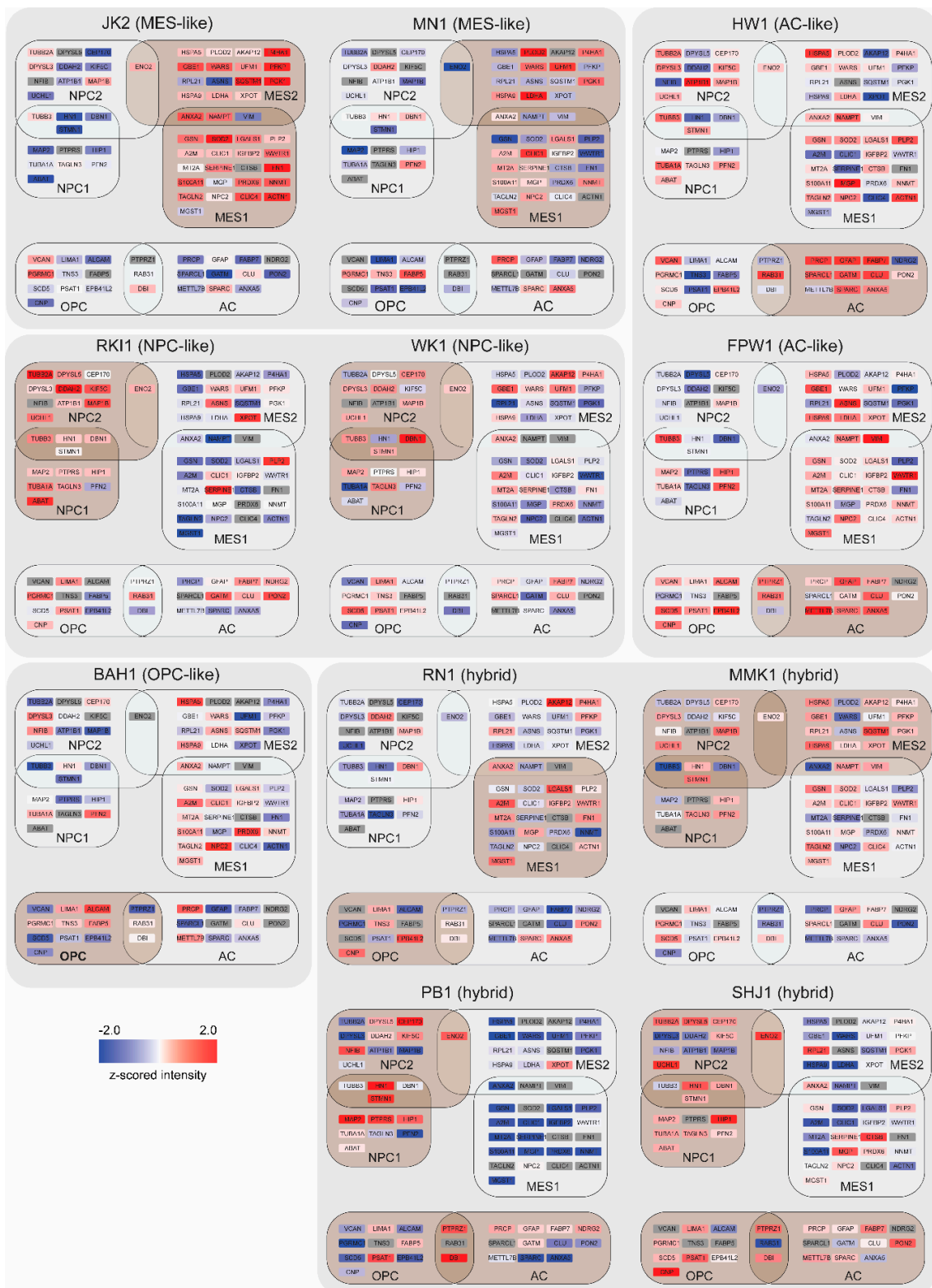
Figure 3. Depiction of cell-states in 11 GBM cell lines. z-scored expression values of genes that contribute to MES1-like, MES2-like, NPC1-like, NPC2-like, and AC-like metamodules (gene list from [9]). Brown boxes indicate predominant cell-state(s) in each cell line. Colour key: red, high expression; blue, low expression; grey, not detected.
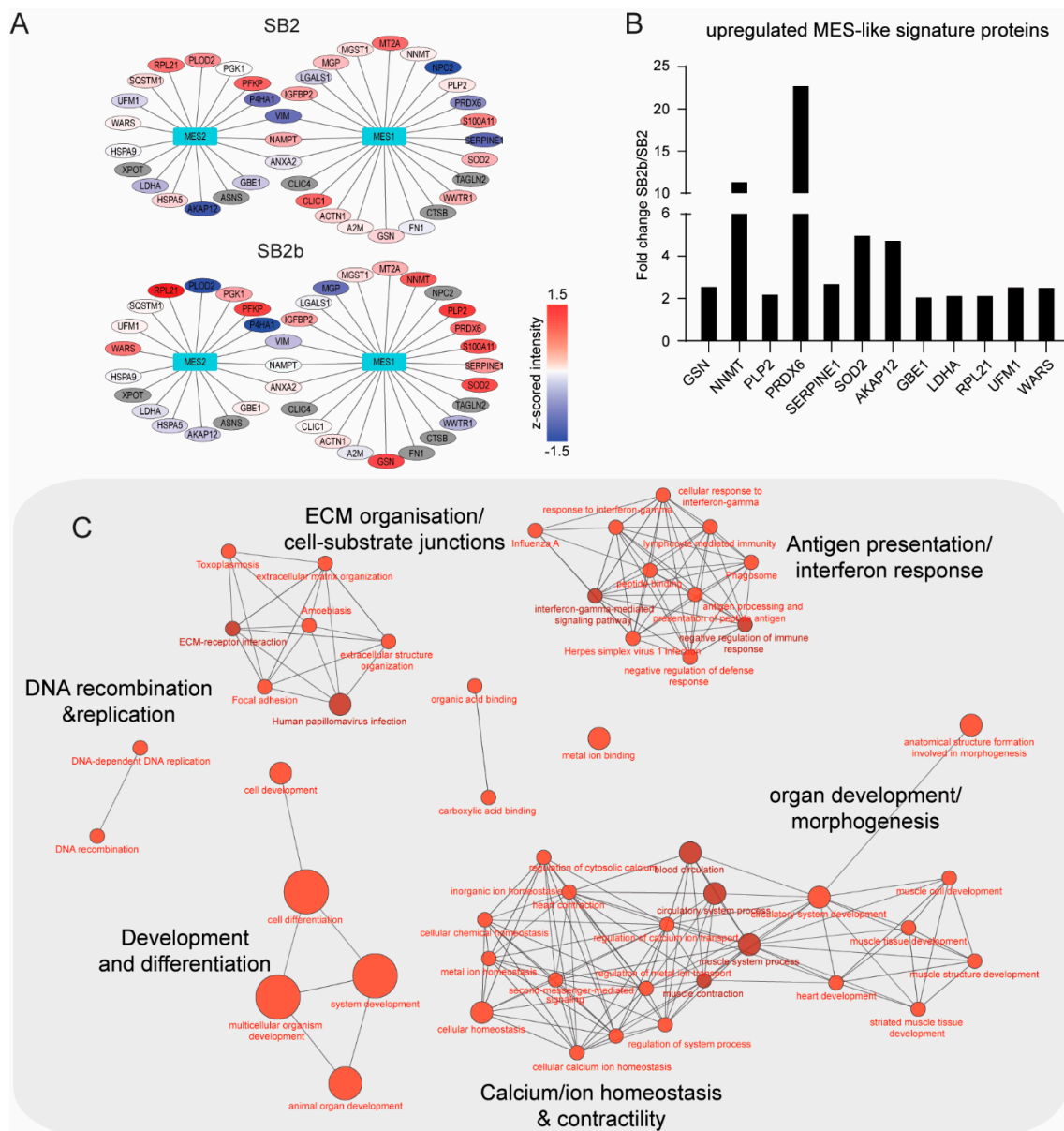
Figure 4. Changing cell-states in a pair-matched recurrent model. (**A**) z-scored expression values of genes that contribute to MES1-like and MES2-like metamodules (gene list from [9]). Colour key: red, high expression; blue, low expression; grey, no expression. (**B**) A bar graph of MES-1-like and MES-2-like genes with >2-fold increase in expression in the recurrent model (SB2b) when compared with the primary model (SB2). (**C**) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and gene ontology (GO) molecular function terms enriched in the top 100 most regulated proteins in the recurrent SB2b model when compared to the primary SB2B model. Node size reflects the number of regulated proteins associated with a particular term/pathway. Node colour intensity increases with increasing significance of term enrichment.

3.3. Bioinformatics and Gene-Ontology-Based Protein Analysis

To gain a broader understanding of the differences in biological processes and signalling pathways between Q-Cell models, we performed an analysis of variance (ANOVA) at a 1% permutation-based FDR cut-off (Table S1). Amongst the top ten regulated proteins were GBM-associated structural and signalling members, including glial fibrillary acidic protein (GFAP), fibronectin (FN1), and Notch1 [29–31]. PB1 protein expression correlated the least with the rest of the Q-Cell models (Figure 1) and displays a

number of unique morphological characteristics, including high growth and small cell morphology in culture (data not shown). The highest exclusively expressed proteins, identified in PB1, were ubiquitin-conjugating enzyme-2 (UBE2) and SKI. Interestingly, the SKI protein has been associated with high cellular tumours and functions to negatively regulate transforming growth factor beta (TGF-beta) by direct interaction with Smads [32,33].

We next performed a Fisher's exact test to evaluate the enrichment of KEGG pathways in the ANOVA-positive subset of proteins. Analysis revealed an enrichment of the KEGG term "Pathways in Cancer", that consists of proteins mediating extracellular matrix (ECM), focal adhesion, and other associated pathways (Table S5). We extracted all identified member proteins belonging to this pathway and performed unsupervised hierarchical clustering. Results showed a segregation of eight models with higher expression of EMT/-mesenchymal-related proteins BAH1, MMK1, MN1, RN1, SB2b, SB2, WK1, and JK2, termed mesenchymal protein cluster 1 and cluster 2 (Figure 5).
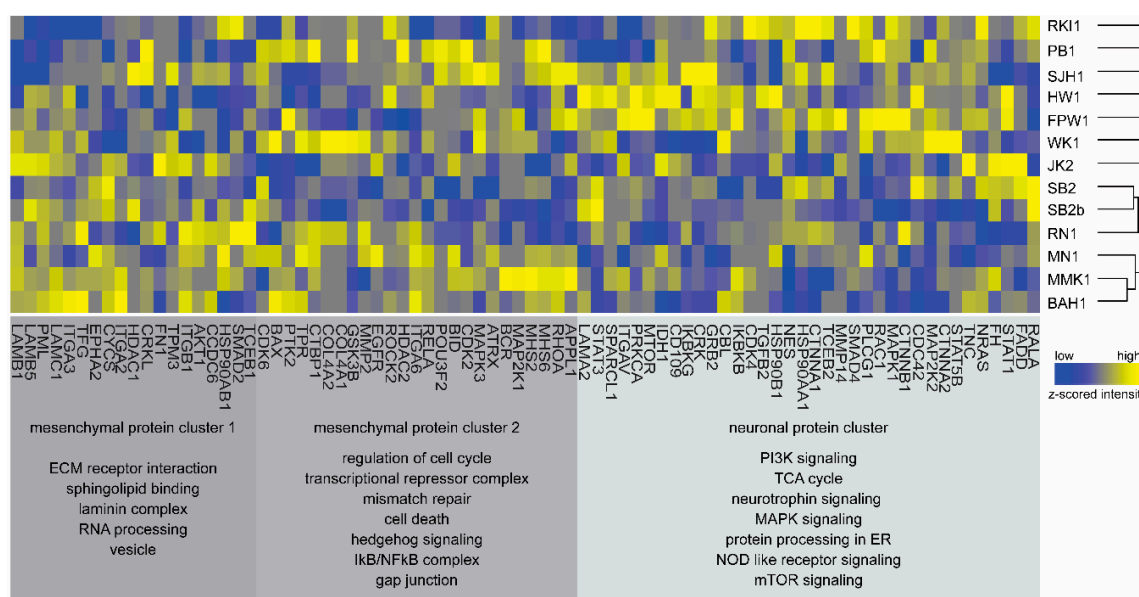
Figure 5. Functional analysis of the expressed proteome. Heatmap of proteins (z-scored intensities) belonging to the KEGG cancer pathway. The proteins segregate into three clusters: mesenchymal cluster 1, mesenchymal cluster 2, and neuronal cluster listing the pathways that proteins represent. Dendrogram depicts the similarity in protein profiles among the 13 cell lines. Colour legend: blue, low expression; and yellow, high expression.

Mesenchymal cluster 1 and cluster 2 consist of structural proteins known to be over-expressed in predominantly mesenchymal-like disease, including laminins, integrins, and collagens, as well as proteins involved in GAP junctions and cell cycle [34]. The remaining Q-Cell panel displayed strong expression of proteins correlated with genes identified from the TCGA GBM neuronal signature and was therefore termed neuronal protein cluster. Models in this cluster showed higher expression of proteins associated with PI3K and mTOR signalling pathways, TCA cycle, NFKB, and MAPK signalling. CDKN2A was found expressed only in MMK1 and RKI1, in concordance with the observation that these were the only lines that did not have a homozygous deletion on the gene level in our characterisation data [10,11]. All individual proteins and their z-scored expression values are provided in Table S6. We compared the proteome-based expression of proteins shown in Figure 6 to the transcriptomics-based RNA-Seq expression values from our previous study [11]. The Pearson's coefficient of correlation was low to moderate between mRNA and protein levels for most cell lines and ranged between 0.18 and 0.56, with the exception of PB1, which showed a high correlation of 0.78 (Figure S3). While genes such as *EGFR*, *CDKN2A*, *PRKCA*, and *HSP90B1* showed good correlation between mRNA and protein; others such as *lamininB1*, *NRAS*, and *tenascin* showed a poor correlation. A caveat to our data was that

some key GBM-associated genes such as *PTEN*, *MYC* and *MET* were not identified in our proteomics analysis. In addition, we also performed hierarchical clustering of mRNA expression values and found that the Q-Cell panel segregated into two major clusters similar to our proteomics data (Figure S4).

To identify potentially actionable targets that could be either uniquely or ubiquitously expressed in the Q-Cell panel, we matched the OncoKB annotated genes to our dataset. OncoKB is a comprehensive precision oncology database that offers evidence-based drug information on FDA-approved therapies and other investigational agents. We identified 298/642 OncoKB annotated genes in our dataset and extracted the list of proteins identified that corresponded to genes that could be targeted with FDA-approved drugs (Figure 6). Previous therapy studies using the Q-Cell resource have validated our proteomic findings. Johns and colleagues showed that BAH1, which expresses high-level EGFR, effectively responded to gefitinib and panitumumab [35].

	RKI1	PB1	SJH1	HW1	FPW1	WK1	JK2	SB2	SB2b	RN1	MN1	MMK1	BAH1
SMARCB1 (tazemetostat)	✓	✓	✓										
EZH2 (tazemetostat)		✓											
CDKN2A (abemaciclib, palbociclib)	✓												
CDK4 (abemaciclib, palbociclib)	✓	✓				✓						✓	
AKT1 (AZD5363)				✓		✓		✓	✓	✓			
MTOR (everolimus, temsirolimus)	✓		✓	✓	✓		✓	✓	✓			✓	
TSC2 (everolimus)		✓	✓	✓						✓	✓		
ERBB2 (neratinib)										✓			
PDGFRA (imatinib)			✓							✓	✓		
EGFR (gefitinib, panitumumab)					✓		✓				✓		✓
NRAS (cetuximab, panitumumab)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
RAF1 (cobimetinib)		✓											
MAP2K1 (cobimetinib, trametinib)	✓	✓											
MAP2K2 (cobimetib)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SRSF2 (H3B-8800)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SF3B1 (H3B-8800)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
U2AF1 (H3B-8800)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ATM (olparib)					✓								✓

Figure 6. Protein expression matrix of targetable proteins. Expression levels (high, dark green; medium to low, light green; white, not detected) of proteins with potential FDA-approved drugs extracted from the OncoKB knowledgebase.

4. Discussion

Here, we built upon our existing body of work to develop a well-described and characterised open-access collection of relevant GBM models for the neuro-oncology research community. Despite significant effort, there is yet to be meaningful increases in survival for GBM sufferers. Our previous Q-Cell resource characterisation study primarily focused on transcriptome-based mRNA expression analysis [11]. mRNA translation is tightly regulated by numerous processes, and gene expression is not always an accurate predictor of protein abundance [36]. We performed MS-based proteomics analysis of the thirteen Q-Cell models, generating a quantitative and deep proteome abundance atlas representing 6167 human proteins. In our analysis, we observed a number of key GBM-associated genes that were not detected at the protein level and may reflect either a lack of translation or limitations of the MS-based approach used. MS-based proteomics has been used effectively by others to measure the proteome, secretome, and glycoproteins in immortalised GBM cell lines [37,38].

Our objective was to further increase the translational significance of the Q-Cell datasets and to provide a platform for researchers to undertake faithful preclinical neuro-oncology studies.

Principal component analysis of Q-Cell replicates grown as 3D neurospheres, showed a high degree of fidelity of protein expression within a given model. All samples in the Q-Cell resource have been generated from IDH1 WT GBM specimens. We found a high degree of common protein expression across the panel, which was expected given the IDH status of the models. Despite this commonality, there were also a number of unique proteins highly expressed in each model.

Elevated expression of common GBM-associated proteins was identified with a trend toward either a broad mesenchymal or neuronal-like signature. An MES-like signature was identified in eight models, while the remaining five models displayed a neuronal-like signature. Given that all models were cultured under the same conditions, this result likely reflects the population of cells captured during the initial cell line generation process. It is worth noting that the majority of Q-Cell models were generated from surgical waste tissue aspirates taken from multiple tumour regions [10]. This was a deliberate attempt to better capture the spectrum of heterogeneous elements within any given tumour, though subclonal elements were likely enriched during the culture generation process.

A major contributing factor to the poor clinical responses observed in GBM patients can be attributed to the significant tumour heterogeneity present within these tumours [8]. Numerous attempts have been made to classify GBM subtypes underlying this heterogeneity [4–6,39]. Suva and colleagues have recently extended current thinking and shown that four functional and dynamic GBM cell-states can exist within a single tumour entity [9]. Given this recent finding, we sought to determine the distribution of these cell-states within the Q-Cell resource. Our initial findings highlighted a good representation of each cell-state across the panel, with a number of models expressing multiple cell–state markers at high levels. These models, termed hybrids, likely contain multiple initiating clones arising from different cell-states or might have arisen from a single clone that has two/more major cell-states. Interestingly, of the 257 cell-state genes described in the Suva study, we detected 153 (57%) at the protein level. MS-based proteomics is not exhaustive, and some key proteins may not have been detected in our analysis. Alternatively, some of the cell-state marker genes may not have been translated to protein. Hence, using gene expression data alone may not be the best approach for accurate biological assessment or translational studies. The NPC-like RKI1 model was derived from a patient with the longest survival from the Q-Cell panel. The NPC-like markers in RKI1 were high, resembling a proneural subtype which is associated with a better prognosis. We also analysed a single pair-matched pre- and post-treatment model (SB2/SB2b). Proteomic results generated from this paired model were intriguing and indicated a number of uniquely overexpressed proteins in the recurrent setting. We also noted a shift to a more mesenchymal phenotype post-treatment, a phenomenon previously reported by others [4,12].

To guide potential future translational studies, we identified a number of unique and common targetable proteins using the OncoKB database. These proteins were targeted with either FDA-approved therapies or other investigational agents and could be used to validate precision oncology or cell-state targeting strategies in the future.

In summary, here we present our detailed proteomics findings of the open-access Q-Cell GBM resource and hope this will accurately guide and inform future studies to improve outcomes for GBM sufferers.

Author Contributions: Conceptualisation, R.C.J.D., T.G.J., A.W.B., and B.W.D.; methodology, R.C.J.D, B.W.S, and B.W.D; formal analysis and investigation, R.C.J.D., C.O.; J.S., U.B., A.K., Y.L., and B.W.D.; resources, B.W.S., B.W.D., H.A., Z.L., P.-L.I., and R.L.J.; data curation, R.C.J.D., C.O., J.S., U.B., A.K., and B.W.D.; writing—original

References

1. Furnari, F.B.; Fenton, T.; Bachoo, R.M.; Mukasa, A.; Stommel, J.M.; Stegh, A.; Hahn, W.C.; Ligon, K.L.; Louis, D.N.; Brennan, C.; et al. Malignant astrocytic glioma: Genetics, biology, and paths to treatment. *Genes Dev.* **2007**, *21*, 2683–2710. [CrossRef] [PubMed]

2. Stupp, R.; Mason, W.P.; van den Bent, M.J.; Weller, M.; Fisher, B.; Taphoorn, M.J.; Belanger, K.; Brandes, A.A.; Marosi, C.; Bogdahn, U.; et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N. Engl. J. Med.* **2005**, *352*, 987–996. [CrossRef] [PubMed]

3. Louis, D.N.; Perry, A.; Reifenberger, G.; von Deimling, A.; Figarella-Branger, D.; Cavenee, W.K.; Ohgaki, H.; Wiestler, O.D.; Kleihues, P.; Ellison, D.W. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: A summary. *Acta Neuropathol.* **2016**, *131*, 803–820. [CrossRef] [PubMed]

4. Phillips, H.S.; Kharbanda, S.; Chen, R.; Forrest, W.F.; Soriano, R.H.; Wu, T.D.; Misra, A.; Nigro, J.M.; Colman, H.; Soroceanu, L.; et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* **2006**, *9*, 157–173. [CrossRef]

5. Verhaak, R.G.; Hoadley, K.A.; Purdom, E.; Wang, V.; Qi, Y.; Wilkerson, M.D.; Miller, C.R.; Ding, L.; Golub, T.; Mesirov, J.P.; et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **2010**, *17*, 98–110. [CrossRef]

6. Wang, Q.; Hu, B.; Hu, X.; Kim, H.; Squatrito, M.; Scarpace, L.; deCarvalho, A.C.; Lyu, S.; Li, P.; Li, Y.; et al. Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer Cell* **2017**, *32*, 42–56.e46. [CrossRef]

7. Akgul, S.; Patch, A.M.; D'Souza, R.C.J.; Mukhopadhyay, P.; Nones, K.; Kempe, S.; Kazakoff, S.H.; Jeffree, R.L.; Stringer, B.W.; Pearson, J.V.; et al. Intratumoural Heterogeneity Underlies Distinct Therapy Responses and Treatment Resistance in Glioblastoma. *Cancers* **2019**, *11*, 190. [CrossRef]

8. Patel, A.P.; Tirosh, I.; Trombetta, J.J.; Shalek, A.K.; Gillespie, S.M.; Wakimoto, H.; Cahill, D.P.; Nahed, B.V.; Curry, W.T.; Martuza, R.L.; et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **2014**, *344*, 1396–1401. [CrossRef]

9. Neftel, C.; Laffy, J.; Filbin, M.G.; Hara, T.; Shore, M.E.; Rahme, G.J.; Richman, A.R.; Silverbush, D.; Shaw, M.L.; Hebert, C.M.; et al. An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **2019**, *178*, 835–849.e821. [CrossRef]

10. Day, B.W.; Stringer, B.W.; Wilson, J.; Jeffree, R.L.; Jamieson, P.J.; Ensbey, K.S.; Bruce, Z.C.; Inglis, P.; Allan, S.; Winter, C.; et al. Glioma Surgical Aspirate: A Viable Source of Tumor Tissue for Experimental Research. *Cancers* **2013**, *5*, 357–371. [CrossRef]

11. Stringer, B.W.; Day, B.W.; D'Souza, R.C.J.; Jamieson, P.R.; Ensbey, K.S.; Bruce, Z.C.; Lim, Y.C.; Goasdoue, K.; Offenhauser, C.; Akgul, S.; et al. A reference collection of patient-derived cell line and xenograft models of proneural, classical and mesenchymal glioblastoma. *Sci. Rep.* **2019**, *9*, 4902. [CrossRef] [PubMed]

12. Bhat, K.P.; Balasubramaniyan, V.; Vaillant, B.; Ezhilarasan, R.; Hummelink, K.; Hollingsworth, F.; Wani, K.; Heathcock, L.; James, J.D.; Goodman, L.D.; et al. Mesenchymal differentiation mediated by NF-kappaB promotes radiation resistance in glioblastoma. *Cancer Cell* **2013**, *24*, 331–346. [CrossRef] [PubMed]

13. Day, B.W.; Stringer, B.W.; Al-Ejeh, F.; Ting, M.J.; Wilson, J.; Ensbey, K.S.; Jamieson, P.R.; Bruce, Z.C.; Lim, Y.C.; Offenhauser, C.; et al. EphA3 maintains tumorigenicity and is a therapeutic target in glioblastoma multiforme. *Cancer Cell* **2013**, *23*, 238–248. [CrossRef]

14. Pollard, S.M.; Yoshikawa, K.; Clarke, I.D.; Danovi, D.; Stricker, S.; Russell, R.; Bayani, J.; Head, R.; Lee, M.; Bernstein, M.; et al. Glioma stem cell lines expanded in adherent culture have tumor-specific phenotypes and are suitable for chemical and genetic screens. *Cell Stem Cell* **2009**, *4*, 568–580. [CrossRef]

15. Reynolds, B.A.; Weiss, S. Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. *Science* **1992**, *255*, 1707–1710. [CrossRef]

16. Wisniewski, J.R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **2009**, *6*, 359–362. [CrossRef]

17. Rappsilber, J.; Ishihama, Y.; Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **2003**, *75*, 663–670. [CrossRef]

18. Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26*, 1367–1372. [CrossRef]

19. Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R.A.; Olsen, J.V.; Mann, M. Andromeda: A peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **2011**, *10*, 1794–1805. [CrossRef]

20. Geiger, T.; Wehner, A.; Schaab, C.; Cox, J.; Mann, M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol. Cell. Proteom.* **2012**, *11*, M111.014050. [CrossRef]

21. Cox, J.; Hein, M.Y.; Luber, C.A.; Paron, I.; Nagaraj, N.; Mann, M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteom.* **2014**, *13*, 2513–2526. [CrossRef] [PubMed]

22. Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M.Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **2016**, *13*, 731–740. [CrossRef] [PubMed]

23. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25–29. [CrossRef] [PubMed]

24. Kanehisa, M.; Goto, S.; Kawashima, S.; Nakaya, A. The KEGG databases at GenomeNet. *Nucleic Acids Res.* **2002**, *30*, 42–46. [CrossRef] [PubMed]

25. Barbie, D.A.; Tamayo, P.; Boehm, J.S.; Kim, S.Y.; Moody, S.E.; Dunn, I.F.; Schinzel, A.C.; Sandy, P.; Meylan, E.; Scholl, C.; et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **2009**, *462*, 108–112. [CrossRef]

26. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef] [PubMed]

27. Bindea, G.; Mlecnik, B.; Hackl, H.; Charoentong, P.; Tosolini, M.; Kirilovsky, A.; Fridman, W.H.; Pages, F.; Trajanoski, Z.; Galon, J. ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **2009**, *25*, 1091–1093. [CrossRef]

28. Rahman, M.; Reyner, K.; Deleyrolle, L.; Millette, S.; Azari, H.; Day, B.W.; Stringer, B.W.; Boyd, A.W.; Johns, T.G.; Blot, V.; et al. Neurosphere and adherent culture conditions are equivalent for malignant glioma stem cell lines. *Anat. Cell Biol.* **2015**, *48*, 25–35. [CrossRef]

29. Yu, Q.; Xue, Y.; Liu, J.; Xi, Z.; Li, Z.; Liu, Y. Fibronectin Promotes the Malignancy of Glioma Stem-Like Cells Via Modulation of Cell Adhesion, Differentiation, Proliferation and Chemoresistance. *Front. Mol. Neurosci.* **2018**, *11*, 130. [CrossRef]

30. Jung, C.S.; Foerch, C.; Schanzer, A.; Heck, A.; Plate, K.H.; Seifert, V.; Steinmetz, H.; Raabe, A.; Sitzer, M. Serum GFAP is a diagnostic marker for glioblastoma multiforme. *Brain* **2007**, *130*, 3336–3341. [CrossRef]

31. Bazzoni, R.; Bentivegna, A. Role of Notch Signaling Pathway in Glioblastoma Pathogenesis. *Cancers* **2019**, *11*, 292. [CrossRef] [PubMed]

32. Wilson, J.J.; Malakhova, M.; Zhang, R.; Joachimiak, A.; Hegde, R.S. Crystal structure of the dachshund homology domain of human SKI. *Structure* **2004**, *12*, 785–792. [CrossRef] [PubMed]

33. Vignais, M.L. Ski and SnoN: Antagonistic proteins of TGFbeta signaling. *Bull. Cancer* **2000**, *87*, 135–137. [PubMed]

34. Behnan, J.; Finocchiaro, G.; Hanna, G. The landscape of the mesenchymal signature in brain tumours. *Brain* **2019**, *142*, 847–866. [CrossRef] [PubMed]

35. Greenall, S.A.; McKenzie, M.; Seminova, E.; Dolezal, O.; Pearce, L.; Bentley, J.; Kuchibhotla, M.; Chen, S.C.; McDonald, K.L.; Kornblum, H.; et al. Most clinical anti-EGFR antibodies do not neutralize both wtEGFR and EGFRvIII activation in glioma. *Neuro Oncol.* **2019**, *21*, 1016–1027. [CrossRef]

36. Baker, K.E.; Coller, J. The many routes to regulating mRNA translation. *Genome Biol.* **2006**, *7*, 332. [CrossRef]

37. Okawa, S.; Gagrica, S.; Blin, C.; Ender, C.; Pollard, S.M.; Krijgsveld, J. Proteome and Secretome Characterization of Glioblastoma-Derived Neural Stem Cells. *Stem Cells* **2017**, *35*, 967–980. [CrossRef]

38. Bock, T.; Moest, H.; Omasits, U.; Dolski, S.; Lundberg, E.; Frei, A.; Hofmann, A.; Bausch-Fluck, D.; Jacobs, A.; Krayenbuehl, N.; et al. Proteomic analysis reveals drug accessible cell surface N-glycoproteins of primary and established glioblastoma cell lines. *J. Proteome Res.* **2012**, *11*, 4885–4893. [CrossRef]

39. Brennan, C.W.; Verhaak, R.G.; McKenna, A.; Campos, B.; Noushmehr, H.; Salama, S.R.; Zheng, S.; Chakravarty, D.; Sanborn, J.Z.; Berman, S.H.; et al. The somatic genomic landscape of glioblastoma. *Cell* **2013**, *155*, 462–477. [CrossRef]