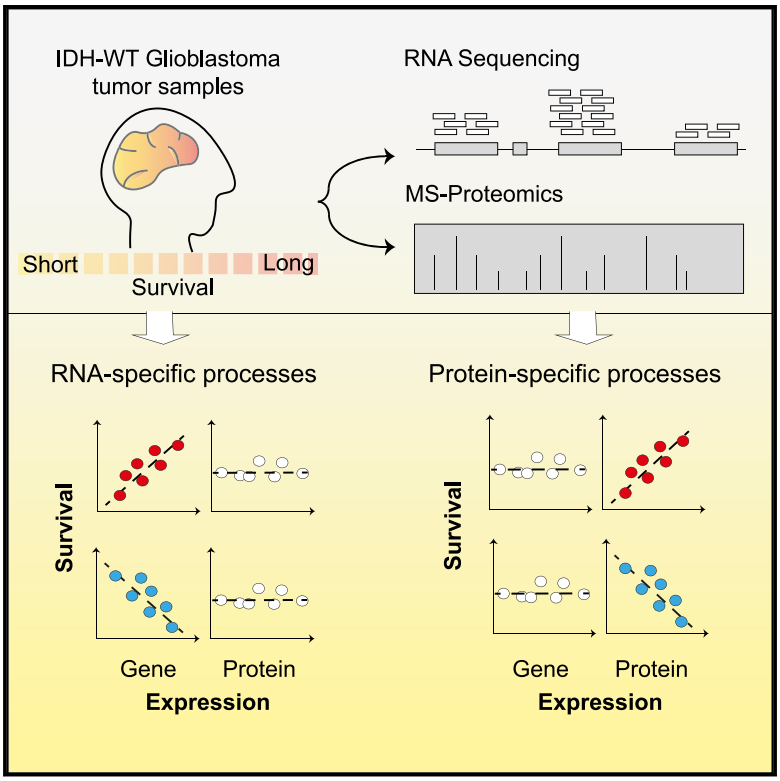
Cell Reports

Resource

Proteogenomics of glioblastoma associates molecular patterns with survival

Graphical Abstract



Highlights

- Integrating proteomics and transcriptomics to study survival in GBM patients
- Proteomics-based classification identifies three subtypes
- Correlation analysis reveals survival patterns unique to the protein level
- Incorporating single-cell published data associates subpopulations with survival

Authors

Gali Yanovich-Arad, Paula Ofek, Eilam Yeini, ..., Rachel Grossman, Ronit Satchi-Fainaro, Tamar Geiger

Correspondence

geiger@tauex.tau.ac.il

In brief

Yanovich-Arad et al. perform a proteogenomic analysis of IDH-WT GBM tumors in which they combine proteomics, transcriptomics, and patient clinical information. Integrative analysis generates proteomic tumor subtypes different from transcriptomic subtypes and identifies biological processes associated with survival that are either common to RNA and protein or are expression-level specific.

Yanovich-Arad et al., 2021, Cell Reports *34*, 108787
March 2, 2021
https://doi.org/10.1016/j.celrep.2021.108787

CellPress

Cell Reports

CellPress
OPEN ACCESS

Resource

Proteogenomics of glioblastoma associates molecular patterns with survival

Gali Yanovich-Arad,[1] Paula Ofek,[2] Eilam Yeini,[2] Mariya Mardamshina,[1] Artem Danilevsky,[3] Noam Shomron,[3] Rachel Grossman,[4] Ronit Satchi-Fainaro,[2] and Tamar Geiger[1,5,*]

[1]Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
[2]Department of Physiology and Pharmacology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
[3]Department of Cell and Developmental Biology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
[4]Department of Neurosurgery, Tel Aviv Medical Center, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
[5]Lead contact
*Correspondence: geiger@tauex.tau.ac.il
https://doi.org/10.1016/j.celrep.2021.108787

SUMMARY

Glioblastoma (GBM) is the most aggressive form of glioma, with poor prognosis exhibited by most patients, and a median survival time of less than 2 years. We assemble a cohort of 87 GBM patients whose survival ranges from less than 3 months and up to 10 years and perform both high-resolution mass spectrometry proteomics and RNA sequencing (RNA-seq). Integrative analysis of protein expression, RNA expression, and patient clinical information enables us to identify specific immune, metabolic, and developmental processes associated with survival as well as determine whether they are shared between expression layers or are layer specific. Our analyses reveal a stronger association between proteomic profiles and survival and identify unique protein-based classification, distinct from the established RNA-based classification. By integrating published single-cell RNA-seq data, we find a connection between subpopulations of GBM tumors and survival. Overall, our findings establish proteomic heterogeneity in GBM as a gateway to understanding poor survival.

INTRODUCTION

Glioblastoma (GBM) is the most common high-grade adult brain tumor. Despite aggressive treatment combining radio- and chemotherapy, as well as gross total resection of the tumor, disease usually progresses and the median survival time of patients is less than 2 years (Ostrom et al., 2015; Patel et al., 2019; Strobel et al., 2019; Stupp et al., 2017). A major challenge in GBM therapy is the tumor heterogeneity, which has been characterized on the genomic and transcriptomic levels.

Several studies conducted by The Cancer Genome Atlas (TCGA) described the genomic landscape of GBM tumors, defined by tumors invariably bearing epidermal growth factor receptor (EGFR) amplification and TP53 and NF1 mutations, in addition to other genetic aberrations (Cancer Genome Atlas Research Network, 2008). GBM classification based on gene expression signatures has established transcriptional heterogeneity (Nutt et al., 2003; Phillips et al., 2006). These analyses identified molecular subtypes associated with prognosis, which were later refined based on RNA sequencing (RNA-seq) (Verhaak et al., 2010) and more uniform sample selection (Wang et al., 2017) to yield three subtypes termed proneural (PN), mesenchymal (MES), and classical (Cla). More recent studies, however, found little-to-no association between subtypes and prognosis (Wang et al., 2017). This is in part due to elimination of mutant (IDH1-mut) tumors from the analyzed cohorts, as they are known to be less aggressive and were previously defined as part of the PN subgroup (Wang et al., 2017).

Beyond bulk tumor analysis, single-cell RNA-seq (scRNA-seq) studies have shown that several transcriptional signatures exist within single tumors, representing different biological processes such as hypoxia and cell cycle (Patel et al., 2014). Furthermore, Neftel et al. (2019) recently showed that transcriptional heterogeneity of GBM tumors converges to four signatures that resemble MES and normal brain lineage stages and represent four tumor cell subpopulations (Neftel et al., 2019).

Omics-based studies have largely contributed to elucidation of these sources of heterogeneity. However, we hypothesized that tumor heterogeneity at the proteomic level should also be comprehensively evaluated, given that mass spectrometry (MS)-based proteomics has become an integral part of cancer research, shedding light on the functional profile of the cancer cell (Coscia et al., 2018; Harel et al., 2019; Mertins et al., 2016; Pozniak et al., 2016; Tyanova et al., 2016a; Vasaikar et al., 2019; Yanovich et al., 2018; Zhang et al., 2016). In GBM, early proteomic studies have mostly utilized matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) technology and applied it to find secreted tumor biomarkers from either cell lines or patient samples (Gautam et al., 2012; Kumar et al., 2010; Thirant et al., 2012). Other studies applied the technology to study the tumor cell proteome of GBM xenograft rat models (Rajcevic et al., 2009) or of cell lines undergoing different

treatments (Puchades et al., 2007). More recent proteomic analyses of gliomas applied higher resolution MS techniques to study clinical samples of grade II–IV gliomas (Buser et al., 2019) and to identify proteomic differences between gliomas of various grades and genomic alterations (Djuric et al., 2019). However, GBM represents a small fraction of the tumors analyzed in these studies.

Integration of multiple omics-based methods further advances the comprehensive view of cancer. Specifically, proteogenomics combines MS-based proteomic data with whole-exome and/or whole-transcriptome sequencing in order to identify the functional outcome of genetic alterations and to evaluate the differences between expression layers. Recently published studies by the Clinical Proteomic Tumor Analysis Consortium (CPTAC) and others include proteogenomic analyses of ovarian, breast, colon, hepatocellular, lung, and gastric cancer (Chen et al., 2020; Gao et al., 2019; Gillette et al., 2020; Mertins et al., 2016; Mun et al., 2019; Zhang et al., 2014, 2016). Among other findings, these studies report the correlation between mRNA and protein levels of the same tumor tissues and show a median correlation that is usually modest, ranging from 0.28 in gastric cancer to 0.54 in hepatocellular carcinoma. This further supports the potential benefit that can be provided by the proteomic layer.

In the current work, we present a proteogenomic dataset of GBM clinical samples. We have assembled a cohort of 87 GBM patients of varying survival rates and performed MS-based proteomics as well as RNA-seq in order to identify the molecular differences associated with survival and examine the contribution of each layer to GBM landscape. We show that the protein layer is more significantly associated with patient survival, but in addition, RNA-protein integration identifies clear patterns of layer-specific and layer-common processes specifically contributing to either short-term or long-term survival periods of patients. We validated our results analytically using published RNA-seq datasets and experimentally using real-time PCR and immunohistochemistry (IHC). Furthermore, we compared our data with published scRNA-seq of GBM tumors and found a link between tumor subpopulations and survival. Altogether, these results highlight the potential of proteogenomics to further stratify heterogeneity in GBM tumors and identify processes contributing to poorer survival.

RESULTS

Unsupervised clustering of samples differs between the proteomic and transcriptomic levels

We collected tumor specimens from 87 patients that were all pathologically defined as GBM, for which we generated either MS-based proteomics data, RNA-seq data, or both (Figure 1A). We performed the analysis with 84 samples out of which 54 had high-resolution proteomic data, 65 had high-quality RNA-seq data, and 32 had both (Figure 1B). Using tandem mass tag (TMT)-10plex chemical labeling, we identified 7,096 proteins in total, out of which 4,567 were used for downstream analyses (see STAR methods). To account for stromal contamination of the transcriptomic samples, we filtered the RNA-seq gene list according to a bona fide glioma (BFG) gene list generated by Wang et al. (2017) and performed the downstream analyses with the resulting 11,459 genes (see STAR methods). To minimize sample variation, we took mostly IDH1-wild type (WT), untreated primary tumors. To initially increase the number of samples with IDH1 mutation annotation, we classified the samples using a published RNA-seq-based signature (Baysan et al., 2012) and found eight samples to be IDH1 mutant, out of which only one had both RNA and protein data (Figure 1C; Figure S1A; Table S1).

In order to globally compare the RNA and protein layers and maximize the statistical power of our analyses, we performed unsupervised classification independently on the proteomic data (53 samples) and transcriptomic data (65 samples) using consensus clustering algorithm (Monti et al., 2003). In addition, we classified the tumors in our cohort according to established RNA-based signatures (Wang et al., 2017). Out of 65 analyzed RNA samples, 19 were defined as the Cla subtype, 16 as MES, and 17 as PN (hypergeometric p value < 0.05; see STAR methods). The transcriptomic consensus clustering resulted in three robust clusters that match the published transcriptional subtypes (group 1 is Cla, group 2 is PN, and group 3 is MES) (Figure 2A; Figure S2A). The proteomic classification also resulted in three groups (Figure 2B; Figure S2B), but with only 25% of the overlapping samples clustering similarly in both layers. In order to try and match between the groups identified in each layer, we looked for differentially expressed genes or proteins in each classification (ANOVA test, false discovery rate [FDR] $<$ 0.01) and examined their enriched functionalities. Hierarchical clustering showed that the differentially expressed features in each layer divide into three clusters. The functional enrichment in the RNA clusters mostly recapitulated the known gene expression signatures for each of the subtypes, e.g., immune signaling in the MES subgroup and Notch signaling in the Cla subgroup (Brennan et al., 2013) (Fisher's exact test, FDR < 0.05; Figure 2C). In each classification, two out of three subgroups were enriched for biological processes that are expression-level specific, such as nuclear factor κB (NF-κB) signaling and epigenetic regulation in RNA clusters, or proteasomal regulation and translation in protein clusters (Figures 2C and 2D). Intriguingly, one cluster was functionally similar in RNA and protein, displaying a neuronal profile and enriched for processes such as synaptic transmission and neuron generation. In addition, protein networks associated with these enrichments include overlapping and inter-connected members, such as guanine nucleotide-binding proteins (GNGs) and the neuronal calcium sensor neurocalcin delta (NCALD) (Figure 2E). These proteins were highly expressed in protein group 3, and RNA group 2, which share only two samples. These results show that neuronal features appear in both layers, but are apparent in a different set of tumors. This discrepancy may partially result from internal tumor heterogeneity. Nevertheless, it highlights proteomics as another layer of tumor heterogeneity in GBM (see Tables S2A–S2D for lists of differentiating features and significantly enriched processes).

Given the profound differences in the clustering analyses, we examined the global RNA-protein agreement of processes, and specifically, the processes observed in the clustering analysis. To that end, we combined the two datasets and calculated the Spearman rank correlation between RNA and protein expression over 4,514 genes quantified in both layers

CellPress
OPEN ACCESS

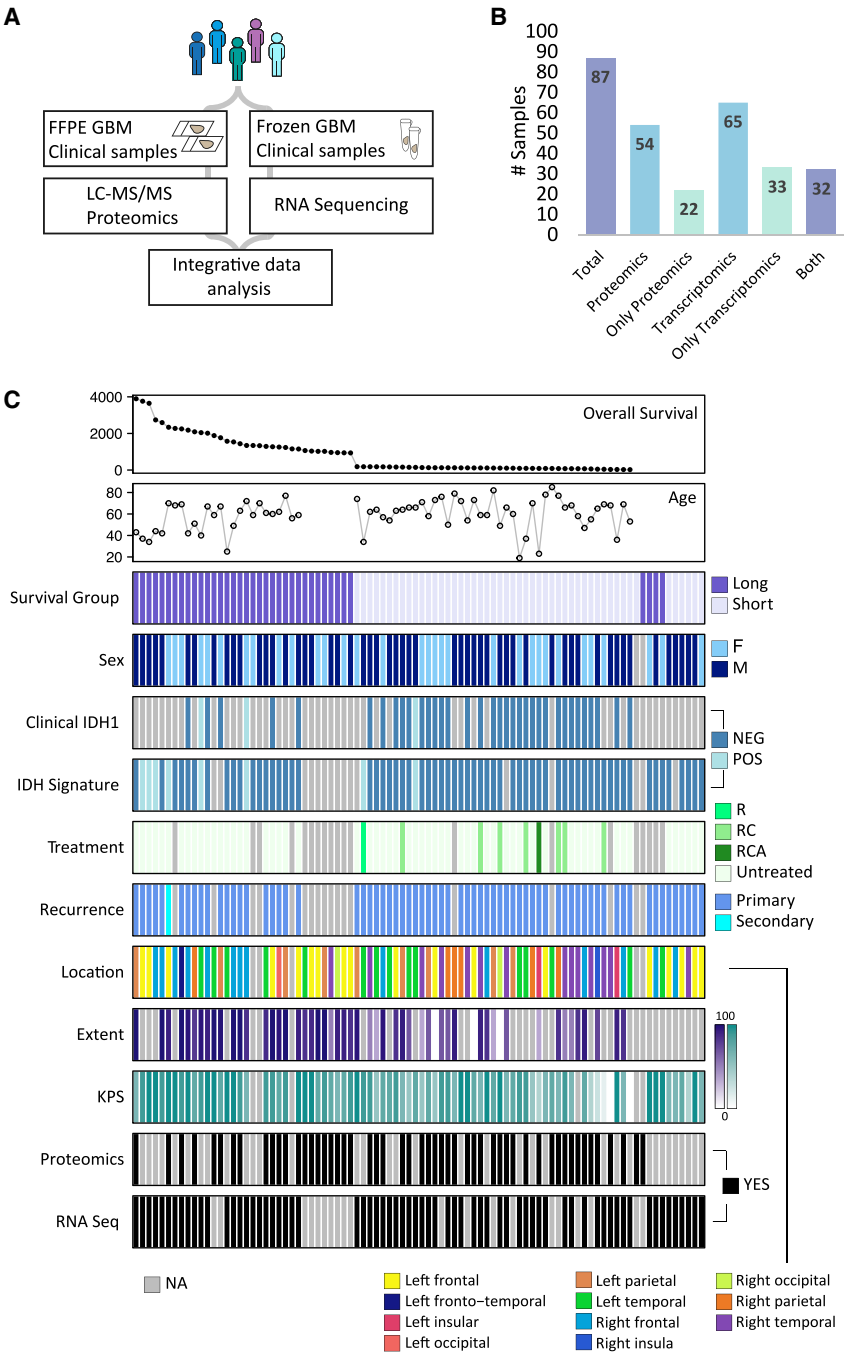


Figure 1. Generation of proteogenomic cohort of glioblastoma patients

(A) For the proteogenomic workflow, samples were collected from resected tumors and were subject to mass spectrometry analysis and RNA-sequencing (RNA-seq).

(B) Number of samples in total and per data type. In each column, different color represents number of samples that have only the data type in the column specified.

(C) Heatmap describing samples collected from 87 GBM patients. Clinical parameters are indicated: R, radiation therapy; RC, radiation and chemotherapy; RCA, radiation and chemotherapy combined with Avastin. Missing values in overall survival panel indicate precise survival days were not available. Extent refers to the extent of surgical resection. KPS stands for Karnofsky performance status.

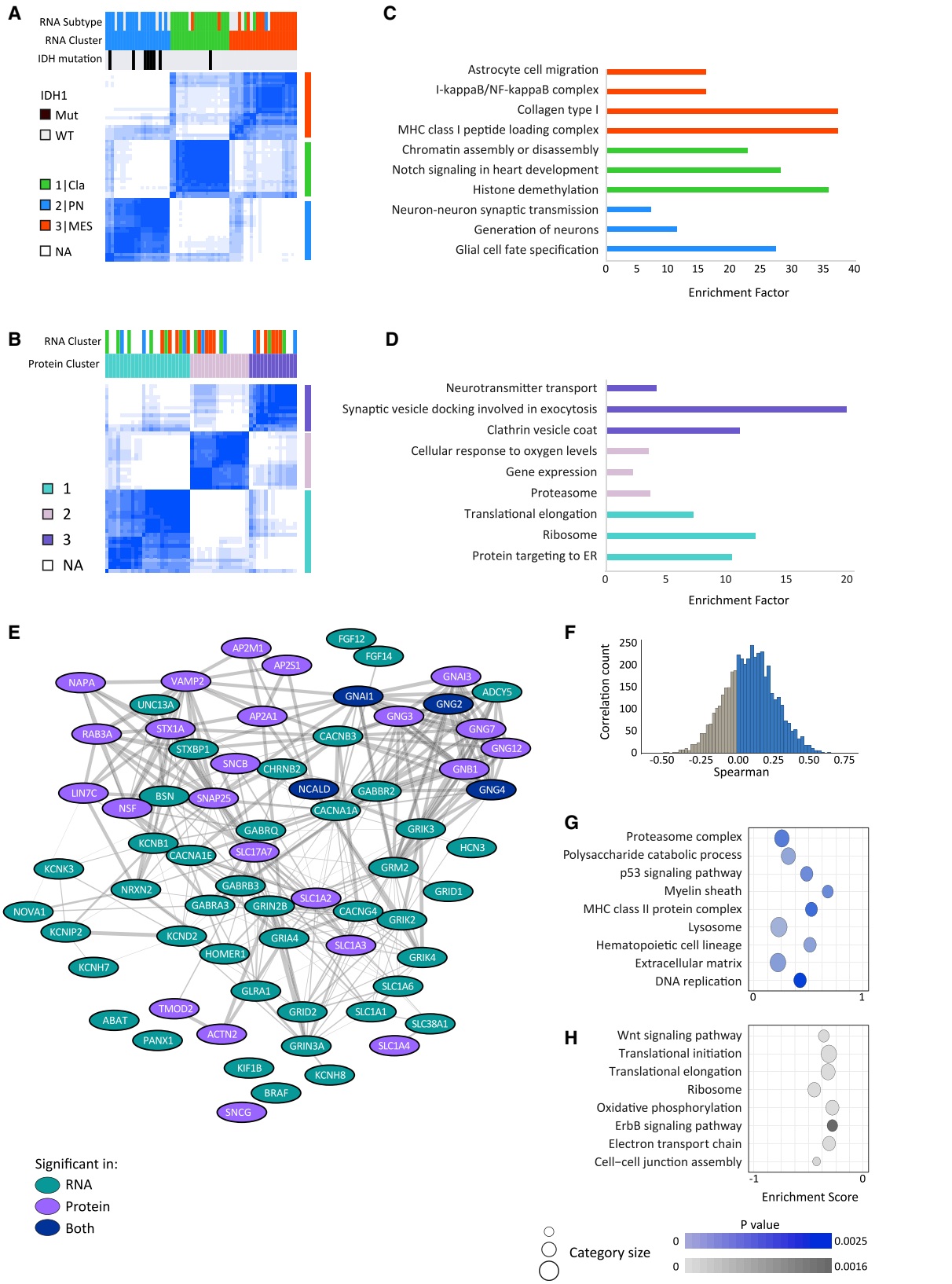
See also Figure S1.

of the 32 shared samples (see STAR methods). The median Spearman correlation was rather modest (r = 0.16 for 67% positive correlations; Figure 2F), compared with published proteogenomic studies, and we assume that the lower correlation results from the different tumor blocks used for each analysis. Nevertheless, the biological processes enriched in each extreme of the RNA-protein agreement axis (1D annotation enrichment test, Benjamini-Hochberg FDR < 0.05) recapitulate known findings regarding shared and layer-specific processes, as shown in Figure S2C (Clark et al., 2019; Gillette et al., 2020; Mertins et al., 2016; Zhang et al., 2014, 2016). For example, ribosome and translation processes are enriched within the negatively correlating genes, which may explain the identification of these processes only in the proteomic classification (Figure 2D); processes such as proteasome and extracellular matrix appear to be enriched within positively correlating genes (Figures 2G and 2H; Table S2E).

Proteogenomic association with GBM clinical parameters

Our main aim was to associate between RNA and protein profiles and the patient clinical parameters: survival, prior treatment, recurrence, sex, and age. We hypothesized these could be identified despite the RNA-protein discrepancies in tumor classification. We excluded IDH1-mut tumors from these analyses, given their different overall survival and clinical features (Yan et al., 2009).

To find functionally related genes and their clinical associations, we performed weighted gene correlation network analysis (WGCNA) independently for transcriptomics and for proteomics of IDH1-WT tumors (n = 52 proteomics, n = 56 transcriptomics). This unsupervised analysis enables a biologically informed dimensionality reduction, by searching for groups of co-expressed genes/proteins termed modules. Each module is then represented by a module eigengene, which is a linear combination of its members' expression. We found 41 and 34 modules in proteomics and transcriptomics data, respectively. We then calculated the correlation between each module's eigengene and clinical traits of interest and retained only significant (p < 0.05) correlations. In both



(legend on next page)

expression layers, we identified modules that correlate to age, sex, and treatment. Processes such as chromatin regulation and cell differentiation were enriched within modules positively correlating with age. Other processes associated with age included NF-κB signaling and were limited to the RNA level. Interestingly, younger age correlated with histone methylation, both known as markers for good prognosis (Batchelor et al., 2004; Zheng et al., 2011). Resemblance between transcriptomics and proteomics was also observed in sex-correlating modules, specifically within female patients wherein commonly upregulated processes included cell-cycle regulation, proteasome, and innate immune response. The major difference was observed in survival-correlating modules: while six proteomic modules correlated with survival (three of them positively and three of them negatively), none of the RNA modules did (Figure 3A). It is worth noting that while some of these survival-associated modules correlated only with survival, the long survival modules also presented a significant and opposite correlation to treatment. However, exclusion of treated samples resulted in similar long-survival-correlating modules (Figure S3A; Tables S3A–S3C) and yielded a single RNA survival module that contains 32 genes and was not enriched for any biological processes. Interestingly, when including IDH1-mut samples in the RNA WGCNA, we do find RNA modules associated with survival. However, each of these modules is also significantly correlated to IDH1 status, indicating that transcriptomic expression profiles do not inform survival beyond IDH1 status (Figure S3B).

To further examine the association with survival on the individual protein level, rather than module level, we performed Kaplan-Meier analysis of all proteins belonging to each of the six survival modules (from Figure 3A). Out of 510 proteins, we found 11 to have a significant association with survival (log-rank test adjusted p value < 0.05; Figures 3B and 3C). Interestingly, one of the long survival modules is represented by the macrophage-secreted protein CD5L (see Discussion). Other than this protein, this module consists of multiple complement system components, implying a potential anti-tumorigenic role (for full details regarding module membership and functional enrichments, see Tables S3A–S3D). Next, we validated the survival analysis using two external RNA-seq datasets (TCGA and Chinese Glioma Genome Atlas [CGGA]). Consistent with our findings, none of the 11 genes was found significant in the RNA level (Figure S3C). It is noteworthy that the expression trend—whether expression level of a certain gene corresponds to better or poorer survival—is similar when comparing our proteomics data with TCGA (64%, 7 out of 11 common genes) and less so when comparing with CGGA (40%, 4 out of 10 common genes). While ethnicity-specific genomic features are well established in lung adenocarcinoma (Chen et al., 2020; Gillette et al., 2020), in GBM there is some evidence regarding differences in molecular classification (Yan et al., 2012); however, there is no difference in overall survival between western and Chinese populations (Ma et al., 2009). Further validation should be performed on other large-scale proteomics datasets of GBM, once these become available.

Integrating scRNA data to link RNA expression patterns with survival

We hypothesized that the lack of significant association between RNA expression levels and survival could be due to high internal tumor heterogeneity and might be unmasked by examining specific tumor subpopulations. Neftel et al. (2019) generated scRNA-seq data and used it to identify four different tumor subpopulations in IDH1-WT GBMs: MES, astrocytic (AC), neural progenitor cell-like (NPC), and oligodendrocytic precursor cell-like (OPC). For each subpopulation, they established a gene signature that enables calculation of the population dominance in bulk tumor. Using these signatures, we defined the dominant subpopulation per tumor both in our RNA-seq data and in TCGA GBM RNA-seq data, as described previously (Neftel et al., 2019) (see STAR methods). In bulk TCGA tumors, the single-cell-based subtypes (defined by the most dominant subpopulation in the tumor) were more evenly distributed (Figure 3D); therefore, we used this dataset to examine whether different subpopulations tend to be highly correlated to survival. We calculated the Pearson correlation between survival and the expression of each signature gene. We then ranked all the genes according to this correlation and examined the enrichment of each of the four signatures in the positively (high rank) or negatively (low rank) correlating genes (1D annotation enrichment test, Benjamini-Hochberg FDR < 0.05). Analysis of the TCGA bulk RNA data showed significant association of NPC and MES signature genes with shorter survival and of OPC and AC signature genes with longer survival (Figure 3E). Only the AC signature was also significantly enriched in our RNA-seq data. Other signatures showed no clear trend, presumably due to their under-representation in the current cohort. Thus, combining bulk tumor correlation analysis with single-cell data provided a potential association between tumor subpopulations and survival.

Integrating RNA and protein to identify layer-specific contribution to survival

The global correlation analysis and the layer-specific classification show that the differences between RNA and protein

Figure 2. Comparing unsupervised clustering of each expression layer

(A and B) Consensus clustering heatmap showing three clusters found in either transcriptomic (A) or proteomic (B) data. Blue color represents sample consensus score. Cluster annotations are indicated as well as IDH1 mutation status in the RNA heatmap.

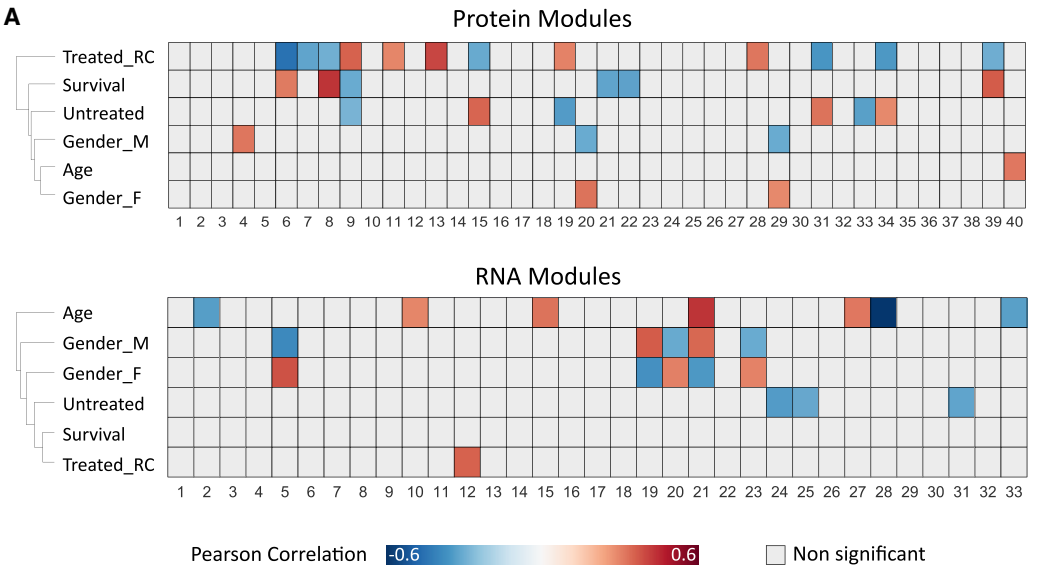
(C and D) Enriched biological processes and pathways in each cluster in each classification (Fisher's exact test, FDR < 0.05).

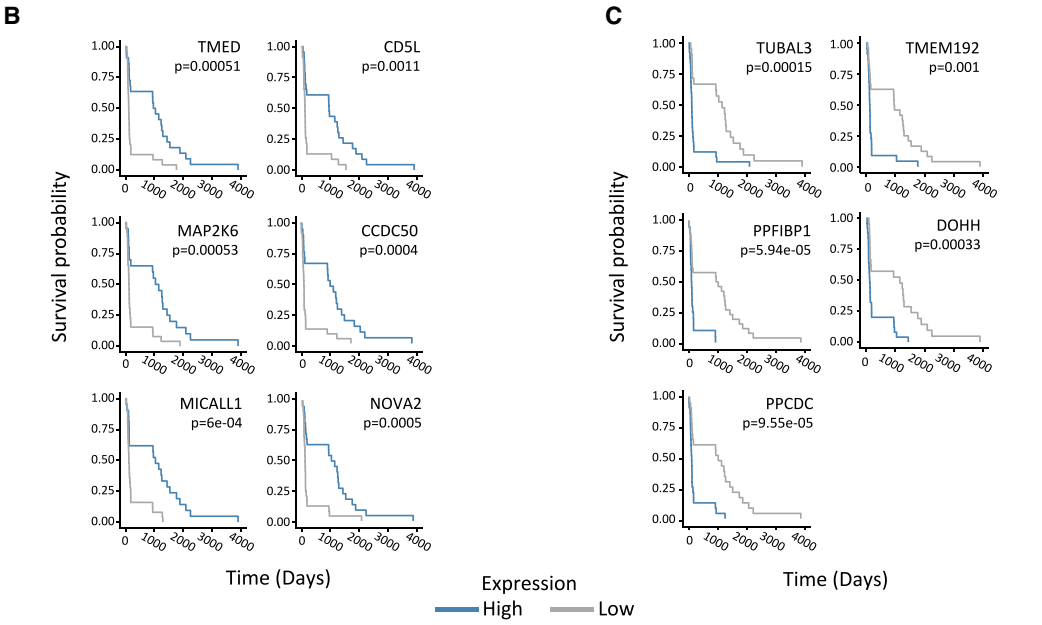
(E) Protein-protein interaction network of the "synaptic transmission" category in protein and RNA. Edge width represents level of confidence for the interaction evidence. Color represents layer: protein, purple; RNA, turquoise; both, dark blue.

(F) Global RNA-protein correlation based on 32 overlapping samples.

(G and H) Enriched biological processes within positively (G) or negatively (H) correlating genes. Circle size stands for GO category size, and p values are indicated by respective color bars.

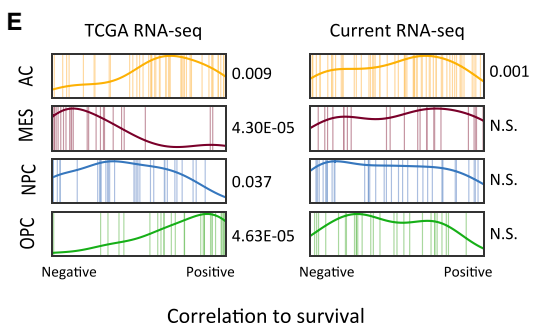
See also Figure S2.





D

	Number of samples	
Subpopulation	**TCGA**	**Current**
AC	**38**	**39**
MES	**46**	**21**
NPC	**26**	**1**
OPC	**31**	**4**



(legend on next page)

potentially represent yet another layer of inter-tumor heterogeneity in GBM, in the gene expression level. Given the contribution of correlation-based analysis to connect single-cell-based signatures and survival, we wanted to extend the analysis to identify other groups of genes and proteins that show a specific survival pattern in the IDH1-WT tumors. To evaluate the contribution of each layer to survival, we filtered the data to retain only genes for which expression was quantified in both layers (n = 3,407; see STAR methods) and then calculated the correlation between patient survival and the gene's protein or RNA expression across all patients. Clustering of the significantly correlating genes/proteins with survival (n = 1,240; permutation-based adjusted p value < 0.1; Figure 4A) separated them to seven clusters, and each of those was tested for its enrichment for protein or RNA correlations with short or long survival (Fisher's exact test, Benjamini-Hochberg FDR < 0.05; Figure 4B). While some genes correlated highly and negatively with survival in both RNA and protein (common-short), some correlated highly and positively in both layers (common-long). Interestingly, the latter is the smallest cluster, indicating that RNA and protein correlate similarly to survival mostly when survival is shorter and suggesting that longer survival is defined by layer-specific processes. Furthermore, several clusters contain genes correlating to survival only in one layer, and one cluster was enriched for genes whose survival-related behavior was altogether opposite in the two layers. A parallel analysis that includes IDH1-mut samples produced similar results except for several common correlations and processes (Figure S3D).

Gene Ontology (GO) enrichment analysis showed that enriched processes in the common-short cluster include lysosomal activity and amino and nucleotide sugar metabolism. We did not find enriched processes in the common-long cluster, presumably due to its small size.

Long-survival processes in the RNA level include layer formation in cerebral cortex and neuron projection, while the protein-long cluster, in addition to gene expression and chromatin regulation processes, is enriched for oxidative phosphorylation and the electron transport chain. Glycolytic metabolism is one of the known biological aspects of GBM tumor aggressiveness (Murat et al., 2009; Wolf et al., 2011). Thus, the processes characterizing the protein-long cluster could be interpreted as a metabolic mirror image of the short-term survival glycolytic metabolism, suggesting an attenuated Warburg effect in less aggressive GBM tumors.

Inflammatory processes are another aspect of GBM aggressiveness (Reynés et al., 2011; Yeung et al., 2013), and indeed we found them to be associated with short-term survival: the RNA-short cluster is enriched for T cell immunity, whereas enriched antigen processing and presentation are observed in the protein-short cluster. To further examine the relationship between survival and immunity, we used the ESTIMATE algorithm (Yoshihara et al., 2013) to calculate immune and stromal infiltration and infer tumor purity. Using linear modeling, we found a significant negative association only between immune score and survival (p value = 0.026), indicating that immune infiltration is higher in short-term survivors (Table S1).

The protein-short cluster is also enriched for cell-cycle signaling and regulation of stem cell maintenance, known to be a marker for poorer prognosis in GBM. In addition, it reveals a metabolic profile that goes beyond glycolysis and includes β-oxidation (see Table S3E for all significantly enriched processes). Despite relying primarily on glucose metabolism, fatty acid oxidation also has a potential role in glioma cell growth (Lin et al., 2017; see Discussion).

Interestingly, we found a cluster of 65 genes that correlate oppositely to survival, out of which 33 have an opposite and significant correlation in both RNA and protein (Figure 4C). Among these genes, we found the NF-κB2-RELA transcription factor complex, calcium/calmodulin-dependent protein kinase type 1 (CAMK1), and TP35BP2. RELA (RELA Proto-oncogene, NF-κB subunit) and TP53BP2 correlate with short survival in the RNA level and long survival in the protein level. NF-κB signaling is tumor promoting in GBM, and its opposite trend in the protein level is yet to be evaluated. TP53BP2 interaction with RELA was shown to inhibit cell death and thus potentially contributes to tumorigenesis (Yang et al., 1999). However, TP53BP2 is a known tumor suppressor (Takahashi et al., 2004), and its upregulation might contribute to better outcome in GBM. CAMK1 shows an opposite pattern: it is correlated with short survival in the protein level and long in the RNA level. Calcium signaling in the brain is associated with neurogenesis in normal brain and with neuro-progenitor/stem-like cells in GBM (Leclerc et al., 2016). In order to validate the opposite patterns of these three markers, we quantified mRNA expression using quantitative real-time PCR and protein expression using IHC, in representative short-term (less than 6 months) and long-term (more than 2 years) samples from our cohort. While only CAMK1 and RELA were statistically significant in RNA and protein, respectively, all three markers showed the expected trend in both expression layers (Figure 4D; Figure S4). Altogether, the integrated proteogenomic analyses show the value of each of the expression

Figure 3. RNA and protein associations to clinical parameters

(A) Heatmap describing modules resulting from independent WGCNA analyses in protein (top) and RNA (bottom). Treated_RC, treated with radiation and chemotherapy. Significant Pearson correlations ($p < 0.05$) to clinical parameters are indicated in red-to-blue color, as indicated in the color bar below. Module 0 (gray), which includes features below the module adjacency cutoff, is not shown.

(B and C) Kaplan-Meier plots of selected proteins from short (B) and long (C) survival modules show significant association with survival (log rank corrected $p < 0.05$).

(D) Table describing the distribution of singe-cell RNA-seq-based dominant subpopulations in TCGA bulk RNA and in the current RNA datasets. AC, astrocytic; MES, mesenchymal; NPC, neural progenitor cell-like; OPC, oligodendrocytic precursor cell-like.

(E) Barcode plot showing the rank of each single-cell subpopulation signature in a scale of correlation to survival in either TCGA (left) or current (right) RNA-seq dataset. High rank indicates positive correlation to survival, while low rank indicates negative correlation. Numbers indicate Fisher enrichment test q-values. N.S., non-significant.

See also Figure S3.

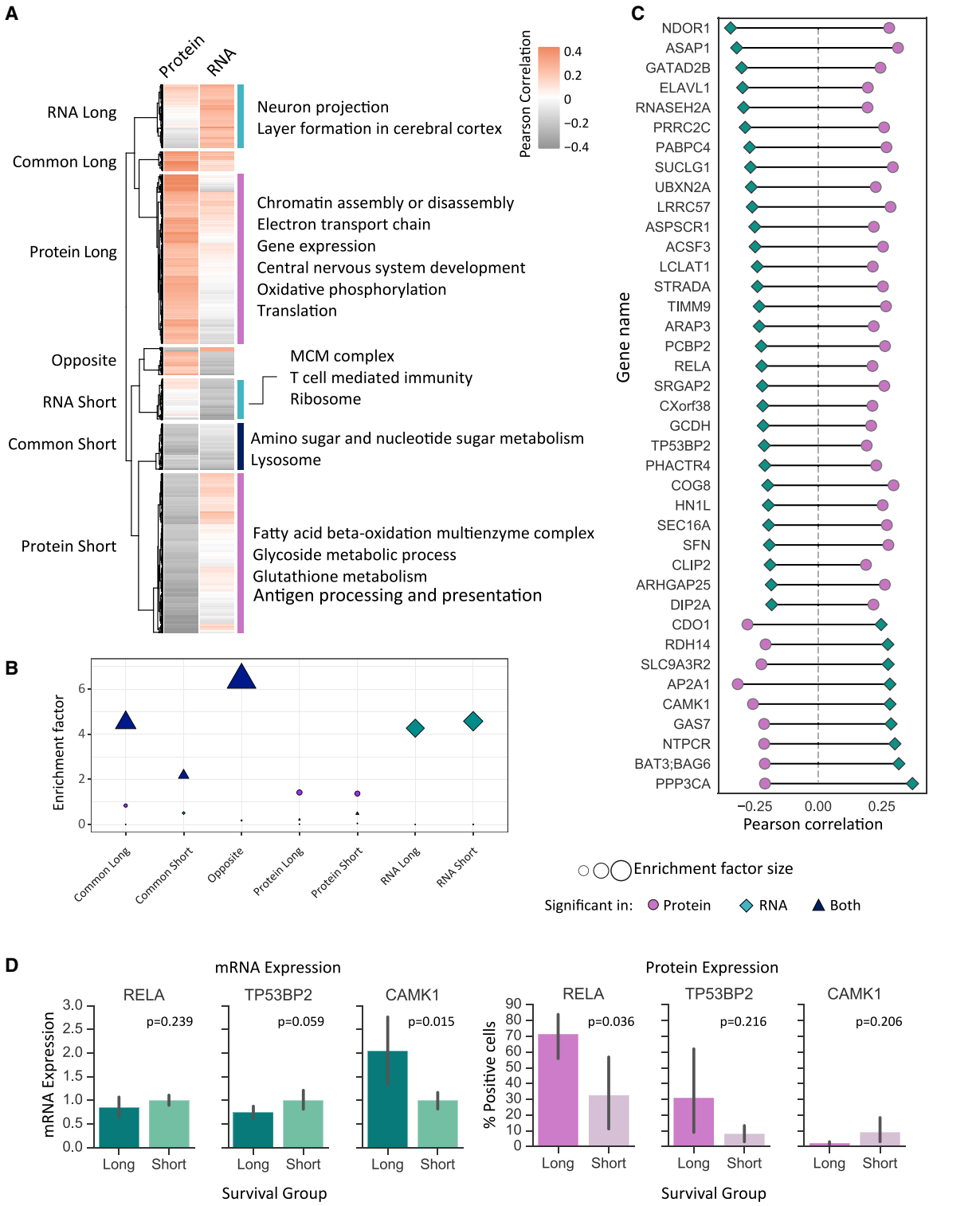


Figure 4. Integrated analysis of survival

(A) Hierarchical clustering of correlation coefficients of 1,240 genes significantly correlating to survival in either RNA or protein (adjusted p value < 0.1) identifies seven clusters.

(B) Each of the seven identified clusters is enriched (BH-FDR < 0.05) for layer-specific significant correlations. Common and opposite clusters are enriched for genes significantly correlating to survival in both RNA and protein. Shape type indicates expression layer, and shape size corresponds to enrichment factor.

(legend continued on next page)

layers and provide the basis for future establishment of prognostic biomarkers of GBM.

DISCUSSION

In this work, we performed a proteogenomic analysis of GBM, focusing on primary IDH1-WT tumors. We compared proteomic findings to established transcriptomic features and examined each layer's association with patient survival. We found that proteomic profiles generate tumor subtypes different from established transcriptomic subtypes and are more robustly associated with survival as shown by the WGCNA analysis. Taking internal tumor heterogeneity into account by using published scRNA-seq data enabled us to link survival with specific tumor subpopulations.

The current cohort posed two initial challenges: samples for proteomics and transcriptomics were not taken from the exact same region, and not all samples had data in both omics. To address the first challenge, we performed most of the analyses in a layer-independent examination and then compared the results. To overcome the partial overlap between RNA and protein samples, we used the 32 samples for which we had data in both layers, compared protein to RNA expression, and verified that enriched processes within either positively or negatively correlating genes recapitulate previous proteogenomics findings.

The discrepancy between RNA and protein was largest in the tumor classification analyses. While some of these differences may be apparent due to the internal tumor heterogeneity, previous classifications of other tumor types (e.g., breast cancer [Mertins et al., 2016; Yanovich et al., 2018]) did not show such major differences, despite analyzing distinct regions/tumors. In contrast, in GBM we found minor concordance between the functionalities of the RNA and protein clusters. Using a different unsupervised approach, in which genes are clustered rather than samples, we did find functional similarities between RNA/protein expression and clinical traits such as sex and age. However, in the case of overall survival, we found proteomic expression profiles that correlate with survival in IDH1-WT tumors, but could not find such profiles in the RNA data. This finding recapitulates other studies that utilized RNA-seq on IDH1-WT tumors and also did not observe a clear association with prognosis (Wang et al., 2017).

Internal tumor heterogeneity is an inherent limitation of omic data generated from bulk tissues. Therefore, we hypothesized that the link between RNA expression and survival might be unmasked by incorporating publically available scRNA-seq data. Applying a correlation-based approach enabled us to significantly link two single-cell-based subpopulation signatures with short survival and two other signatures with long survival. In order to substantiate the connection between RNA levels, internal tumor heterogeneity, and patient outcome, it would be necessary to check whether the ratio between long and short subpopulations within a tumor reflects patient outcome, once single-cell techniques enable the collection of larger patient cohorts. Furthermore, single-cell proteomic analyses are expected to unravel whether the bulk inter-layer differences are also apparent in the individual cell populations.

We used a similar integrative approach to combine protein expression, RNA expression, and survival correlation, which led us to identify survival patterns in both layers, and define whether they are shared between expression layers or layer specific. Since our proteomic cohort included only one IDH1-mut sample, follow-up studies could refine this analysis by comparing the IDH1-WT and mutant proteomic profiles of GBM, as recent evidence suggests a prominent proteomic difference between these groups in lower grade gliomas (Djuric et al., 2019).

Our analyses highlight three main mechanisms to be associated with survival: immune processes, metabolic processes, and developmental processes. We found short survival to be associated with inflammation and glycolytic metabolism, both with established roles in tumor growth and aggressiveness in GBM (Reynés et al., 2011; Waters et al., 2019; Yeung et al., 2013). Interestingly, while NF-κB signaling and T cell immunity negatively correlate with survival and reflect a pro-tumorigenic inflammatory response, supported by a higher immune infiltration score, the WGCNA results also show a positive association between immunity and survival in the form of immunoglobulins and complement system components, suggesting a potential anti-tumorigenic immune response. Furthermore, the proteomic layer provides an extended metabolic context for patient survival: linking increased fatty acid oxidation with short survival time and oxidative phosphorylation (OXPHOS) with longer survival time. These findings are supported by the identification of the role of fatty acid oxidation in tumor growth and oxidative stress mitigation (Duman et al., 2019; Pike et al., 2011) and the role of OXPHOS in tumor suppression and mitochondria-promoted apoptosis in GBM (Michelakis et al., 2010). Additionally, our integrative analysis revealed that stemness-development axis is also evident in the protein level. Traditionally, cancer stem cell populations in gliomas have been shown to possess higher tumorigenic capacity and as such were considered a promising therapeutic target (Berger et al., 2004; Galli et al., 2004; Suvà et al., 2014). However, while the association between stemness and aggressiveness was observed in lower grade gliomas (Tirosh et al., 2016; Venteicher et al., 2017), it is not straightforward in GBM tumors, in which multiple cells invariably express stemness markers (Patel et al., 2014). In addition, the high plasticity of GBM cells enables differentiated tumor cells to undergo de-differentiation, regulated by the tumor microenvironment, epigenomics, and other factors (Dirkse et al., 2019). For example, it was recently shown that differentiation-inducing treatment in GBM cell lines is mediated by elevated mitochondrial metabolism and specifically OXPHOS (Xing et al., 2017).

(C) Thirty-nine genes of the "opposite" cluster that have opposite and significant correlations to survival.
(D) Quantitative real-time PCR and IHC validation of three selected markers showing opposite survival patterns in RNA and protein. mRNA relative expression values normalized to glyceraldehyde 3-phosphate dehydrogenase (GAPDH) are shown in purple. Protein expression represented by percentage of stained cells is shown in turquoise. Vertical lines represent standard error values.
See also Figures S3 and S4.

Our findings reinforce the functional connection between stemness, metabolism, and overall survival in GBM. This suggests that further investigation of the proteomic profiles associated with stemness in shorter and longer survival might benefit stemness-targeted clinical efforts.

In conclusion, we presented here a global, RNA-seq, and MS-based characterization of gene and protein expression in clinical samples of GBM. The identified proteomic and transcriptomic patterns shed light on the intriguing molecular heterogeneity of GBM tumors by revealing functionalities related to patient survival and disentangling the contribution of each expression layer. In addition, this dataset can serve as a resource to further study the proteogenomic landscape of GBM and to evaluate specific genes of interest.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - RNA extraction and sequencing
  - Protein extraction
  - Liquid Chromatography – Mass Spectrometry (LC-MS) Analysis
  - Gene expression analysis (qPCR)
  - Immunohistochemistry staining
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Protein Identification and Quantification
  - RNA-sequencing data acquisition and analysis
  - Proteomic data pre-processing and statistics
  - Data filtration per analysis
  - Protein-RNA correlation
  - Weighted gene correlation network analysis (WGCNA) and unsupervised clustering
  - Integrated Pattern analysis
  - Validation using TCGA and Chinese Glioma datasets
  - Integration of single cell signatures
  - Survival Kaplan-Meier analysis and log rank test
  - Immunohistochemistry quantification

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2021.108787.

ACKNOWLEDGMENTS

The Geiger laboratory has received partial funding from the European Research Council (ERC) starting grant (639534), the Israel Science Foundation (748/16), and the Gail White and Anne and William Cohen Multidisciplinary Brain Cancer Research Program. The Satchi-Fainaro laboratory has received partial funding from the ERC Consolidator Grant agreement no. 617445-PolyDorm and ERC Advanced Grant agreement no. 835227-3DBrainStrom; The Israel Science Foundation (grants 918/14 and 1969/18); Israel Cancer Research Fund; Nancy and Peter Brown friends of The Israel Cancer Association USA (ICA), in memory of Kenny and Michael Adler (20150909); and the Morris Kahn Foundation. E.Y. is a fellow at the Gail White and Anne and William Cohen Multidisciplinary Brain Cancer Research program. P.O. thanks the Naomi Foundation for the Global Research and Training Fellowship in Medical and Life Sciences. We thank Georgina Barnabas for her assistance and Naama Knafo, Michal Harel, and Lir Beck for their critical evaluation of the manuscript and all members of the Geiger laboratory for fruitful discussions. We thank Tom Rabinowitz for technical assistance in cluster-based analyses. RNA-seq was performed by Otogenetics.

AUTHOR CONTRIBUTIONS

Conceptualization, G.Y.-A., T.G., and R.S.-F.; formal analysis and investigation, G.Y.-A., P.O., M.M., and E.Y.; resources, R.G., R.S.-F., P.O., E.Y., G.Y.-A., A.D., and N.S.; writing – original draft, G.Y.-A. and T.G.; writing – review & editing, G.Y.-A., P.O., E.Y., R.G., N.S., and T.G.; funding acquisition, R.S.-F. and T.G.; supervision, T.G.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 13, 2020
Revised: December 8, 2020
Accepted: February 3, 2021
Published: March 2, 2021

REFERENCES

Batchelor, T.T., Betensky, R.A., Esposito, J.M., Pham, L.D., Dorfman, M.V., Piscatelli, N., Jhung, S., Rhee, D., and Louis, D.N. (2004). Age-dependent prognostic effects of genetic alterations in glioblastoma. Clin. Cancer Res. *10*, 228–233.

Baysan, M., Bozdag, S., Cam, M.C., Kotliarova, S., Ahn, S., Walling, J., Killian, J.K., Stevenson, H., Meltzer, P., and Fine, H.A. (2012). G-cimp status prediction of glioblastoma samples using mRNA expression data. PLoS ONE *7*, e47839.

Berger, F., Gay, E., Pelletier, L., Tropel, P., and Wion, D. (2004). Development of gliomas: potential role of asymmetrical cell division of neural stem cells. Lancet Oncol. *5*, 511–514.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

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.; TCGA Research Network (2013). The somatic genomic landscape of glioblastoma. Cell *155*, 462–477.

Buser, D.P., Ritz, M.F., Moes, S., Tostado, C., Frank, S., Spiess, M., Mariani, L., Jenö, P., Boulay, J.L., and Hutter, G. (2019). Quantitative proteomics reveals reduction of endocytic machinery components in gliomas. EBioMedicine *46*, 32–41.

Cancer Genome Atlas Research Network (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature *455*, 1061–1068.

Chen, Y.J., Roumeliotis, T.I., Chang, Y.H., Chen, C.T., Han, C.L., Lin, M.H., Chen, H.W., Chang, G.C., Chang, Y.L., Wu, C.T., et al. (2020). Proteogenomics of Non-smoking Lung Cancer in East Asia Delineates Molecular Signatures of Pathogenesis and Progression. Cell *182*, 226–244.e17.

Clark, D.J., Dhanasekaran, S.M., Petralia, F., Pan, J., Song, X., Hu, Y., da Veiga Leprevost, F., Reva, B., Lih, T.M., Chang, H.Y., et al.; Clinical Proteomic Tumor Analysis Consortium (2019). Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. Cell *179*, 964–983.e31.

Coscia, F., Lengyel, E., Duraiswamy, J., Ashcroft, B., Bassani-Sternberg, M., Wierer, M., Johnson, A., Wroblewski, K., Montag, A., Yamada, S.D., et al.

(2018). Multi-level Proteomics Identifies CT45 as a Chemosensitivity Mediator and Immunotherapy Target in Ovarian Cancer. Cell *175*, 159–170.e16.

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

Cox, J., and Mann, M. (2012). 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. BMC Bioinformatics *13* (*Suppl 16*), S12.

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

Dirkse, A., Golebiewska, A., Buder, T., Nazarov, P.V., Muller, A., Poovathingal, S., Brons, N.H.C., Leite, S., Sauvageot, N., Sarkisjan, D., et al. (2019). Stem cell-associated heterogeneity in Glioblastoma results from intrinsic tumor plasticity shaped by the microenvironment. Nat. Commun. *10*, 1787.

Djuric, U., Lam, K.H.B., Kao, J., Batruch, I., Jevtic, S., Papaioannou, M.D., and Diamandis, P. (2019). Defining Protein Pattern Differences Among Molecular Subtypes of Diffuse Gliomas Using Mass Spectrometry. Mol. Cell. Proteomics *18*, 2029–2043.

Duman, C., Yaqubi, K., Hoffmann, A., Acikgöz, A.A., Korshunov, A., Bendszus, M., Herold-Mende, C., Liu, H.K., and Alfonso, J. (2019). Acyl-CoA-Binding Protein Drives Glioblastoma Tumorigenesis by Sustaining Fatty Acid Oxidation. Cell Metab. *30*, 274–289.e5.

Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. *30*, 207–210.

Galli, R., Binda, E., Orfanelli, U., Cipelletti, B., Gritti, A., De Vitis, S., Fiocco, R., Foroni, C., Dimeco, F., and Vescovi, A. (2004). Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma. Cancer Res. *64*, 7011–7021.

Gao, Q., Zhu, H., Dong, L., Shi, W., Chen, R., Song, Z., Huang, C., Li, J., Dong, X., Zhou, Y., et al. (2019). Integrated Proteogenomic Characterization of HBV-Related Hepatocellular Carcinoma. Cell *179*, 561–577.e22.

Gautam, P., Nair, S.C., Gupta, M.K., Sharma, R., Polisetty, R.V., Uppin, M.S., Sundaram, C., Puligopu, A.K., Ankathi, P., Purohit, A.K., et al. (2012). Proteins with altered levels in plasma from glioblastoma patients as revealed by iTRAQ-based quantitative proteomic analysis. PLoS ONE *7*, e46153.

Gillette, M.A., Satpathy, S., Cao, S., Dhanasekaran, S.M., Vasaikar, S.V., Krug, K., Petralia, F., Li, Y., Liang, W.W., Reva, B., et al.; Clinical Proteomic Tumor Analysis Consortium (2020). Proteogenomic Characterization Reveals Therapeutic Vulnerabilities in Lung Adenocarcinoma. Cell *182*, 200–225.e35.

Harel, M., Ortenberg, R., Varanasi, S.K., Mangalhara, K.C., Mardamshina, M., Markovits, E., Baruch, E.N., Tripple, V., Arama-Chayoth, M., Greenberg, E., et al. (2019). Proteomics of Melanoma Response to Immunotherapy Reveals Mitochondrial Dependence. Cell *179*, 236–250.e18.

Kumar, D.M., Thota, B., Shinde, S.V., Prasanna, K.V., Hegde, A.S., Arivazhagan, A., Chandramouli, B.A., Santosh, V., and Somasundaram, K. (2010). Proteomic identification of haptoglobin α2 as a glioblastoma serum biomarker: implications in cancer cell migration and tumor growth. J. Proteome Res. *9*, 5557–5567.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Leclerc, C., Haeich, J., Aulestia, F.J., Kilhoffer, M.C., Miller, A.L., Néant, I., Webb, S.E., Schaeffer, E., Junier, M.P., Chneiweiss, H., and Moreau, M. (2016). Calcium signaling orchestrates glioblastoma development: Facts and conjunctures. Biochim. Biophys. Acta *1863* (*6 Pt B*), 1447–1459.

Lin, H., Patel, S., Affleck, V.S., Wilson, I., Turnbull, D.M., Joshi, A.R., Maxwell, R., and Stoll, E.A. (2017). Fatty acid oxidation is required for the respiration and proliferation of malignant glioma cells. Neuro-oncol. *19*, 43–54.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Ma, X., Lv, Y., Liu, J., Wang, D., Huang, Q., Wang, X., Li, G., Xu, S., and Li, X. (2009). Survival analysis of 205 patients with glioblastoma multiforme: clinical characteristics, treatment and prognosis in China. J. Clin. Neurosci. *16*, 1595–1598.

Mertins, P., Mani, D.R., Ruggles, K.V., Gillette, M.A., Clauser, K.R., Wang, P., Wang, X., Qiao, J.W., Cao, S., Petralia, F., et al.; NCI CPTAC (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. Nature *534*, 55–62.

Michelakis, E.D., Sutendra, G., Dromparis, P., Webster, L., Haromy, A., Niven, E., Maguire, C., Gammer, T.L., Mackey, J.R., Fulton, D., et al. (2010). Metabolic modulation of glioblastoma with dichloroacetate. Sci. Transl. Med. *2*, 31ra34.

Monti, S., Tamayo, P., Mesirov, J., and Golub, T. (2003). Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. Mach. Learn. *52*, 91–118.

Mun, D.G., Bhin, J., Kim, S., Kim, H., Jung, J.H., Jung, Y., Jang, Y.E., Park, J.M., Kim, H., Jung, Y., et al. (2019). Proteogenomic Characterization of Human Early-Onset Gastric Cancer. Cancer Cell *35*, 111–124.e10.

Murat, A., Migliavacca, E., Hussain, S.F., Heimberger, A.B., Desbaillets, I., Hamou, M.F., Rüegg, C., Stupp, R., Delorenzi, M., and Hegi, M.E. (2009). Modulation of angiogenic and inflammatory response in glioblastoma by hypoxia. PLoS ONE *4*, e5947.

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. (2019). An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. Cell *178*, 835–849.e21.

Nutt, C.L., Mani, D.R., Betensky, R.A., Tamayo, P., Cairncross, J.G., Ladd, C., Pohl, U., Hartmann, C., McLaughlin, M.E., Batchelor, T.T., et al. (2003). Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. Cancer Res. *63*, 1602–1607.

Ostrom, Q.T., Gittleman, H., Fulop, J., Liu, M., Blanda, R., Kromer, C., Wolinsky, Y., Kruchko, C., and Barnholtz-Sloan, J.S. (2015). CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2008-2012. Neuro-oncol. *17* (*Suppl 4*), iv1–iv62.

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. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science *344*, 1396–1401.

Patel, A.P., Fisher, J.L., Nichols, E., Abd-Allah, F., Abdela, J., Abdelalim, A., Abraha, H.N., Agius, D., Alahdab, F., Alam, T., et al.; GBD 2016 Brain and Other CNS Cancer Collaborators (2019). Global, regional, and national burden of brain and other CNS cancer, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol. *18*, 376–393.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419.

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., et al. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res. *47* (*D1*), D442–D450.

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. (2006). Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. Cancer Cell *9*, 157–173.

Pike, L.S., Smift, A.L., Croteau, N.J., Ferrick, D.A., and Wu, M. (2011). Inhibition of fatty acid oxidation by etomoxir impairs NADPH production and increases reactive oxygen species resulting in ATP depletion and cell death in human glioblastoma cells. Biochim. Biophys. Acta *1807*, 726–734.

Pozniak, Y., Balint-Lahat, N., Rudolph, J.D., Lindskog, C., Katzir, R., Avivi, C., Pontén, F., Ruppin, E., Barshack, I., and Geiger, T. (2016). System-wide Clinical Proteomics of Breast Cancer Reveals Global Remodeling of Tissue Homeostasis. Cell Syst. *2*, 172–184.

Puchades, M., Nilsson, C.L., Emmett, M.R., Aldape, K.D., Ji, Y., Lang, F.F., Liu, T.J., and Conrad, C.A. (2007). Proteomic investigation of glioblastoma cell lines treated with wild-type p53 and cytotoxic chemotherapy demonstrates

an association between galectin-1 and p53 expression. J. Proteome Res. *6*, 869–875.

Rajcevic, U., Petersen, K., Knol, J.C., Loos, M., Bougnaud, S., Klychnikov, O., Li, K.W., Pham, T.V., Wang, J., Miletic, H., et al. (2009). iTRAQ-based proteomics profiling reveals increased metabolic activity and cellular cross-talk in angiogenic compared with invasive glioblastoma phenotype. Mol. Cell. Proteomics *8*, 2595–2612.

Reynés, G., Vila, V., Martín, M., Parada, A., Fleitas, T., Reganon, E., and Martínez-Sales, V. (2011). Circulating markers of angiogenesis, inflammation, and coagulation in patients with glioblastoma. J. Neurooncol. *102*, 35–41.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Rudolph, J.D., and Cox, J. (2019). A Network Module for the Perseus Software for Computational Proteomics Facilitates Proteome Interaction Graph Analysis. J. Proteome Res. *18*, 2052–2064.

Strobel, H., Baisch, T., Fitzel, R., Schilberg, K., Siegelin, M.D., Karpel-Massler, G., Debatin, K.M., and Westhoff, M.A. (2019). Temozolomide and Other Alkylating Agents in Glioblastoma Therapy. Biomedicines *7*, 69.

Stupp, R., Taillibert, S., Kanner, A., Read, W., Steinberg, D., Lhermitte, B., Toms, S., Idbaih, A., Ahluwalia, M.S., Fink, K., et al. (2017). Effect of Tumor-Treating Fields Plus Maintenance Temozolomide vs Maintenance Temozolomide Alone on Survival in Patients With Glioblastoma: A Randomized Clinical Trial. JAMA *318*, 2306–2316.

Suvà, M.L., Rheinbay, E., Gillespie, S.M., Patel, A.P., Wakimoto, H., Rabkin, S.D., Riggi, N., Chi, A.S., Cahill, D.P., Nahed, B.V., et al. (2014). Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. Cell *157*, 580–594.

Takahashi, N., Kobayashi, S., Jiang, X., Kitagori, K., Imai, K., Hibi, Y., and Okamoto, T. (2004). Expression of 53BP2 and ASPP2 proteins from TP53BP2 gene by alternative splicing. Biochem. Biophys. Res. Commun. *315*, 434–438.

Thirant, C., Galan-Moya, E.M., Dubois, L.G., Pinte, S., Chafey, P., Broussard, C., Varlet, P., Devaux, B., Soncin, F., Gavard, J., et al. (2012). Differential proteomic analysis of human glioblastoma and neural stem cells reveals HDGF as a novel angiogenic secreted factor. Stem Cells *30*, 845–853.

Tirosh, I., Venteicher, A.S., Hebert, C., Escalante, L.E., Patel, A.P., Yizhak, K., Fisher, J.M., Rodman, C., Mount, C., Filbin, M.G., et al. (2016). Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. Nature *539*, 309–313.

Tyanova, S., Albrechtsen, R., Kronqvist, P., Cox, J., Mann, M., and Geiger, T. (2016a). Proteomic maps of breast cancer subtypes. Nat. Commun. *7*, 10259.

Tyanova, S., Temu, T., and Cox, J. (2016b). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat. Protoc. *11*, 2301–2319.

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

Vasaikar, S., Huang, C., Wang, X., Petyuk, V.A., Savage, S.R., Wen, B., Dou, Y., Zhang, Y., Shi, Z., Arshad, O.A., et al.; Clinical Proteomic Tumor Analysis Consortium (2019). Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. Cell *177*, 1035–1049.e19.

Venteicher, A.S., Tirosh, I., Hebert, C., Yizhak, K., Neftel, C., Filbin, M.G., Hovestadt, V., Escalante, L.E., Shaw, M.L., Rodman, C., et al. (2017). Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. Science *355*, eaai8478.

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.; Cancer Genome Atlas Research Network (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell *17*, 98–110.

Wang, Q., Hu, B., Hu, X., Kim, H., Squatrito, M., Scarpace, L., deCarvalho, A.C., Lyu, S., Li, P., Li, Y., et al. (2017). Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. Cancer Cell *32*, 42–56.e6.

Waters, M.R., Gupta, A.S., Mockenhaupt, K., Brown, L.N., Biswas, D.D., and Kordula, T. (2019). RelB acts as a molecular switch driving chronic inflammation in glioblastoma multiforme. Oncogenesis *8*, 37.

Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics *26*, 1572–1573.

Wolf, A., Agnihotri, S., Micallef, J., Mukherjee, J., Sabha, N., Cairns, R., Hawkins, C., and Guha, A. (2011). Hexokinase 2 is a key mediator of aerobic glycolysis and promotes tumor growth in human glioblastoma multiforme. J. Exp. Med. *208*, 313–326.

Xing, F., Luan, Y., Cai, J., Wu, S., Mai, J., Gu, J., Zhang, H., Li, K., Lin, Y., Xiao, X., et al. (2017). The Anti-Warburg Effect Elicited by the cAMP-PGC1α Pathway Drives Differentiation of Glioblastoma Cells into Astrocytes. Cell Rep. *18*, 468–481.

Yan, H., Parsons, D.W., Jin, G., McLendon, R., Rasheed, B.A., Yuan, W., Kos, I., Batinic-Haberle, I., Jones, S., Riggins, G.J., et al. (2009). IDH1 and IDH2 mutations in gliomas. N. Engl. J. Med. *360*, 765–773.

Yan, W., Zhang, W., You, G., Zhang, J., Han, L., Bao, Z., Wang, Y., Liu, Y., Jiang, C., Kang, C., et al. (2012). Molecular classification of gliomas based on whole genome gene expression: a systematic report of 225 samples from the Chinese Glioma Cooperative Group. Neuro-oncol. *14*, 1432–1440.

Yang, J.P., Hori, M., Takahashi, N., Kawabe, T., Kato, H., and Okamoto, T. (1999). NF-kappaB subunit p65 binds to 53BP2 and inhibits cell death induced by 53BP2. Oncogene *18*, 5177–5186.

Yanovich, G., Agmon, H., Harel, M., Sonnenblick, A., Peretz, T., and Geiger, T. (2018). Clinical Proteomics of Breast Cancer Reveals a Novel Layer of Breast Cancer Classification. Cancer Res. *78*, 6001–6010.

Yeung, Y.T., McDonald, K.L., Grewal, T., and Munoz, L. (2013). Interleukins in glioblastoma pathophysiology: implications for therapy. Br. J. Pharmacol. *168*, 591–606.

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. Nat. Commun. *4*, 2612.

Zhang, B., Wang, J., Wang, X., Zhu, J., Liu, Q., Shi, Z., Chambers, M.C., Zimmerman, L.J., Shaddox, K.F., Kim, S., et al.; NCI CPTAC (2014). Proteogenomic characterization of human colon and rectal cancer. Nature *513*, 382–387.

Zhang, H., Liu, T., Zhang, Z., Payne, S.H., Zhang, B., McDermott, J.E., Zhou, J.Y., Petyuk, V.A., Chen, L., Ray, D., et al.; CPTAC Investigators (2016). Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. Cell *166*, 755–765.

Zheng, S., Houseman, E.A., Morrison, Z., Wrensch, M.R., Patoka, J.S., Ramos, C., Haas-Kogan, D.A., McBride, S., Marsit, C.J., Christensen, B.C., et al. (2011). DNA hypermethylation profiles associated with glioma subtypes and EZH2 and IGFBP2 mRNA expression. Neuro-oncol. *13*, 280–289.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CAMK1 antibody	Atlas Antibodies	Cat#HPA051409; RRID:AB_2681473
TP53BP2 antibody	Sigma-Aldrich	Cat#HPA021603; RRID:AB_1844384
RELA antibody	Cell Signaling Technology	Cat#8242; RRID:AB_10859369
Biological samples		
Formalin-fixed paraffin-embedded (FFPE) glioblastoma samples	Tel Aviv Sourasky Medical Center	N/A
Fresh frozen glioblastoma samples	Tel Aviv Sourasky Medical Center	N/A
Chemicals, peptides, and recombinant proteins		
PerFecTa SYBR Green FastMix ROX	Quanta BioSciences	Cat#95073
Fluorescamine	Sigma-Aldrich	Cat#F9015
Micromount solution	Leica Biosystems	Cat#3801730
Bond Dewax Solution	Leica Biosystems	Cat#LE-AR9222
BondTM Epitope Retrieval 1	Leica Biosystems	Cat#LE-AR9961
BondTM Wash Solution 10X	Leica Biosystems	Cat#LE-AR9590
10plex Tandem mass tags (TMT)	Thermo Scientific	Cat#90110
Critical commercial assays		
EZ RNAII isolation kit	Biological Industries	Cat#20-410-100
DNase RNase-free water	Biological Industries	Cat#01-869-1A
Sequencing Grade Modified Trypsin	Promega	Cat#V5113
LysC Trypsin mix	Promega	Cat#V5071
Pierce High pH Reversed-Phase Peptide Fractionation Kit	Thermo Scientific	Cat# 84868
Bond polymer refine detection kit	Leica Biosystems	Cat#DS9800
FluoroProfile® Protein Quantification Kit	Sigma-Aldrich	Cat# FP0010
Deposited Data		
Proteomics data	This paper	PRIDE: PXD018024
Transcriptomics data	This paper	GEO: GSE149009
Oligonucleotides		
CAMKI F: GCAGCCGCAGGAGCC	Integrated DNA Technologies	https://www.idtdna.com/
CAMKI R: AGGATCACCTCCGAGAAGGC	Integrated DNA Technologies	https://www.idtdna.com/
TP53BP2 F: GAAGTGTGGTGTGGCTCTG	Integrated DNA Technologies	https://www.idtdna.com/
TP53BP2 R: AGATCTTGGTCCACTCACAATGTC	Integrated DNA Technologies	https://www.idtdna.com/
RELA F: GGAATTCCAGTACCTGCCAGA	Integrated DNA Technologies	https://www.idtdna.com/
RELA R: GTCGGTGGGTCCGCTGAA	Integrated DNA Technologies	https://www.idtdna.com/
Software and algorithms		
MaxQuant	Cox and Mann, 2008	https://maxquant.org/; RRID:SCR_014485
Andromeda search engine	Cox et al., 2011	N/A
Perseus	(Tyanova et al., 2016c)	https://maxquant.org/perseus/; RRID:SCR_015753
R	The R Foundation	https://www.r-project.org/
Python	Python Software Foundation	https://www.python.org

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Trimmomatic	Bolger et al., 2014	http://www.usadellab.org/cms/?page=trimmomatic; RRID:SCR_011848
Salmon	(Patro et al., 2017)	https://combine-lab.github.io/salmon/
Other		
Empore Octadecyl C18 47mm Extraction disks	Sigma-Aldrich	Cat#66883-U
50 cm EASY-spray PepMap column	Thermo Scientific	Cat#ES803

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Tamar Geiger (geiger@tauex.tau.ac.il).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The accession number for the mass spectrometry proteomics dataset reported in this paper is PRIDE: PXD018024 (Perez-Riverol et al., 2019). The RNA-seq data have been deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) and are accessible through GEO Series accession number GEO: GSE149009.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Frozen tissue blocks and formalin-fixed paraffin-embedded (FFPE) tissues were obtained from the Neurosurgery and Pathology departments of Tel Aviv Sourasky Medical Center, Tel Aviv, Israel. We acquired tumor samples from 87 patients, all taken from tumor resection surgery; 69 were pathologically defined as primary glioblastoma (GBM), one sample was secondary GBM and for 17 patients recurrence information was not available. Only eight patients were given treatment (radiotherapy and chemotherapy) prior to surgery. For two patients (L18 and L19), we had samples from two different foci. The cohort included tumors from 29 females, 49 males and 9 without sex information, with age range of 19 to 85 (median 62). Additionally, in order to perform a molecular analysis of patient survival, samples were specifically selected to have varying survival rates, ranging from less than three months to over 10 years (median ~5 months). Clinical information regarding tumor location, extent of tumor resection and Karnofsky Performance Status (KPS) are also provided. Regression analysis found no significant association between these traits and survival except for the extent of tumor resection (p value = 0.042, Figure S1B). All samples were obtained upon ethical approval from the IRB committee of the Tel Aviv Sourasky Medical Center. Clinical information of each patient is included in Table S1.

METHOD DETAILS

RNA extraction and sequencing

RNA was extracted from fresh frozen GBM tissues of human patients using EZ RNAII isolation kit (Biological Industries, Bet Haemek, Israel) according to manufacturer's instructions. Samples were homogenized in Denaturing Solution (0.5ml/50-100mg tissue) using GentleMACS homogenizer (Miltenyi Biotech, USA) program RNA-02. Homogenates were stored for 5 minutes at room temperature, then 0.4ml Water-saturated phenol was added followed by 0.09ml 1-Bromo-3-chloropropane (BCP) and vigorous shaking for 15 s. The resulting mixture was stored at room temperature for 10 minutes and then centrifuged at 12,000 g for 15 minutes at 4°C. The aqueous colorless (upper) phase was transferred to a fresh tube followed by addition of 0.5ml isopropanol. Mixture was stored for 30 minutes at −20°C and then centrifuged at 12,000 g for 8 minutes at 4°C. Supernatant was removed and RNA pellet was washed (by vortexing) with 1ml 75% ethanol, then centrifuged at 7,500 g for 5 minutes at 4°C. Ethanol was removed, RNA pellet was air-dried for 20-30 minutes and then dissolved in 100 μL of DNase RNase-free water (Biological Industries, Bet Haemek, Israel) by incubating for 10-15 minutes at 55°C. RNA samples were kept at −80°C until sequencing. RNA Integrity Number (RIN) was determined for each sample using the 2200 TapeStation system (Agilent, CA, USA). RNA libraries were prepared according to Illumina protocols. Paired-end RNA sequencing data (read length 100 base pairs, designated 20 million reads per sample) was generated on the Illumina HiSeq 2500 at Otogenetics Corporation, Atlanta, GA USA.

Protein extraction

54 FFPE blocks were macro-dissected from tissue slices by overlaying H&E staining, in order to enrich for cellular areas and exclude stromal components. Dissected samples were lysed in 50% 2-2-2 trifluoroethanol (TFE) in 25mM ammonium bi-carbonate (ABC), incubated with 5mM Dithiothreitol (DTT) and alkylated with 15mM Iodoacetamide (IAA). Prior to protein digestion, samples were diluted 1:10 with 5mM ABC, and then digested overnight with LysC/Trp mix and Trypsin in an enzyme to protein ratio of 1:100 and 1:50, respectively. Prior to labeling, clean-up of digested peptides was performed using C18 Stage-Tips. We proceeded to tandem-mass-tags (TMT) 10plex labeling according to the manufacturer's instructions (Pierce). The 54 samples were divided into six sets of 10plex-TMT, while the tenth sample in each set consisted of a tumor-mix to be used as a quantification standard between different sets. Following labeling, samples were combined and vacuum-concentrated, and then resuspended in 0.1% trifluoroacetic acid (TFA). Resuspended samples were loaded onto high-pH reverse phase columns (Thermo Fisher Scientific) for sample fractionation. Each 10plex-TMT set was fractionated into eight fractions according to the manufacturer's instructions. Resulting fractions were then vacuum-concentrated and resuspended in MS loading buffer (2% acetonitrile, 0.1% formic acid). Protein concentrations were determined using FluroroProfile Protocol and peptide concentrations prior to TMT labeling were determined using Fluorescamine Protocol, both on NanoDrop 3300 Fluorospectrometer (Thermo Fisher Scientific).

Liquid Chromatography – Mass Spectrometry (LC-MS) Analysis

Peptides were separated in the Easy-nLC 1000 nano-HPLC system (Thermo Fisher Scientific) using reverse phase chromatography on a C18 Easy-Spray column; and loaded to the Q-Exactive HF mass spectrometer (Thermo Fisher Scientific). Each sample ran for a 128-min gradient of water and 80% acetonitrile, with an MS resolution of 120,000 (scan range 350-1400 m/z, ion target value of 3e6 and maximum injection time of 100 ms) and MS/MS resolution of 60,000 (scan range 200-2000 m/z, ion target value of 1e5 and maximum injection time of 60 ms). In every MS scan, the top 15 most abundant peaks were selected for higher-energy collision dissociation (HCD) fragmentation.

Gene expression analysis (qPCR)

RNA from 10 GBM frozen tissues was isolated with EZ-RNA II total RNA isolation kit (Biological Industries Ltd., Israel), according to the manufacturers protocol. Samples were lysed with 0.5 mL Denaturing Solution/10 cm2 culture plate. Water saturated phenol was then added, and the samples were centrifuged. Isopropanol was added to precipitate the RNA and the centrifuged RNA pellet was washed with 75% ethanol, centrifuged, and re-suspended with ultra-pure double distilled water. RNA concentration was evaluated using a NanoDrop® ND-1000 Spectrophotometer according to the manufacturer's V3.5 User's Manual (Nano-Drop Technologies, Wilmington, DE). qScriptTM cDNA synthesis kit for RT-PCR was used to synthesize cDNA, according to the manufacturer's guidelines. Briefly, 1 μg of total RNA sample was mixed with qScript Reverse Transcriptase, dNTPs, and nuclease free water. The reaction tube was then incubated at 42°C for 30 min and heated at 85°C for 5 min to stop the reaction.

Expression level of several target genes was assessed by SYBR green real-time PCR (PerFecTa SYBR Green FastMix ROX (Quanta BioSciences) according to manufacturer's protocol, and normalized to the expression of GAPDH housekeeping gene. Difference between long and short term groups across three replicates for each marker was assessed by two-tailed t test (alpha = 0.05).

Immunohistochemistry staining

FFPE embedded tumor samples from 10 patients (5 long and 5 short) were cut into 5 μm thick sections. Staining was performed using BOND RX autostainer (Leica). Briefly, slides were incubated with anti-human TP53BP2, CAMK1, or RELA primary antibody for 1 h. Following washes, slides were incubated with rabbit\mouse IgG binding polymer for 8 min, washed and incubated with DAB for 7 min followed by incubation with hematoxylin for 16 min. Slides were then washed and treated with micromount solution before being covered with coverslips.

QUANTIFICATION AND STATISTICAL ANALYSIS

All analyses were performed using the Perseus software (Tyanova et al., 2016c), R and Python. Biological annotations were taken from Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG).

Protein Identification and Quantification

MS raw files were analyzed using MaxQuant software version 1.6.2.6 (Cox and Mann, 2008; Tyanova et al., 2016b). Peptide search was performed using the Andromeda search engine (Cox et al., 2011) against the Uniprot human protein database release April 2018, with 1% false discovery rate (FDR) at the PSM and protein levels. MS level mass tolerance was set to 4.5 ppm. Peptides were allowed to have methionine oxidation and N-terminal acetylation as variable modifications and cysteine carbimdomethyl as a fixed modification. Quantification was defined based on MS2 reporter ion intensity with TMT channels correction factors supplied by the manufacturer.

RNA-sequencing data acquisition and analysis

Resulting reads in FASTQ format were trimmed and quality filtered using Trimmomatic (Bolger et al., 2014). We used Salmon quasi-mapping tool (Patro et al., 2017) for expression quantification against the transcriptome compiled from the Ensembl human hg38 genome assembly. We used the R package DESeq2 (Love et al., 2014) to convert transcripts per million (TPM) quantifications to gene level, as well as to perform $\log_2$ transformation and variance stabilizing transformations (VST) normalization of the counts for downstream analysis (Tables S4B–C). Three samples were removed from downstream analysis due to low alignment rates. To account for stromal contamination of the samples, we filtered the RNA-seq gene list according the bona-fide glioma (BFG) gene list generated by Wang et al. (2017). Briefly, the authors created this list by integrating RNA-seq data from glioma single cells and the tumor microenvironment (represented by glioma sphere-forming cell cultures and comparing cellular tumor areas to their matching leading edge). They created the BFG gene list by removing genes whose expression was significantly high in the tumor microenvironment from the initial list of identified genes. This resulted in 11,529 genes out of which 11,459 were present in our dataset. We performed all downstream analyses (except for the global protein-RNA correlation) with the resulting 11,459 genes.

Proteomic data pre-processing and statistics

We identified 7,096 proteins in total in 54 patient samples and 6 control channels (5,422 per sample on average, Table S4A). In order to retain high-quality quantifications, we filtered the data to contain only proteins that were quantified in all six standard channels. This resulted in 4,567 proteins, for which missing value imputation was performed sample-wise by drawing values from a normal distribution with a width of 1.5 and down shift of 0.5 standard deviations of the specific sample. Unless otherwise specified, downstream analyses were then performed on calculated protein ratio between each sample and its corresponding standard. As samples of different survival rates were evenly distributed between TMT sets, we performed linear modeling to eliminate small TMT batch effects using R limma package (Ritchie et al., 2015) (Figure S1C).

Data filtration per analysis

For the global RNA-protein correlation analysis we used the Initial transcriptomic data of 23,011 genes, and only the 32 samples for which we had data in both expression layers. In this analysis, matching between proteomic and transcriptomic data based on gene names resulted in 4,514 genes. For clustering, WGCNA, and integrative correlation analyses we used transcriptomic data from the filtered, BFG-based-list (n = 11,459), and valid proteomic data as described above (n = 4,567). In the integrative correlation analysis, matching between proteomic and transcriptomic data based on gene names resulted in 3,407 genes.

Protein-RNA correlation

Out of a total of 84 samples, 32 had both proteomic and transcriptomic data, while 20 had only proteomic data and 32 had only RNA data. To calculate the protein-RNA correlation we used only the 32 samples for which we had data in both layers. We then matched between proteomic and transcriptomic data based on gene names and calculated the Spearman rank correlation coefficient between gene and protein expression for 4,514 genes. To calculate the biological enrichments of either negatively or positively correlating genes we used 1D annotation analysis in which genes are ranked in ascending order according to their correlation, and genes of each biological category are tested for having significantly high or low ranks, as described (Cox and Mann, 2012).

Weighted gene correlation network analysis (WGCNA) and unsupervised clustering

WGCNA was performed using the WGCNA R package (Langfelder and Horvath, 2008) and WGCNA implementation in Perseus software (Rudolph and Cox, 2019). First, an adjacency network of features (genes/proteins) is created where two features are considered as connected based on their co-expression level. We used a soft-threshold beta power = 16 to create a robust and signed network, where two features are connected only if they are positively correlated. Network was then clustered using the algorithm default parameters to create modules, reducing the data from thousands of features to several dozens of groups of co-expressing features. The first principal component of each module is considered the module eigenegne, which is then used to calculate Pearson correlation with clinical traits. Pearson correlations with p value < 0.05 were considered significant.

As the established RNA classification is based on samples with mixed IDH status, we used all samples (n = 65) for the unsupervised classification of the RNA and the proteomics data. Since only one proteomic sample was identified as IDH-mut we removed it from this analysis (n = 53). Clustering was performed using consensus clustering algorithm implemented in ConsensusClusterPlus R package (Monti et al., 2003; Wilkerson and Hayes, 2010), with subsampling of 80% of the samples over 1000 iterations and 10 as maximum k (number of clusters). Before classification, samples were z-score normalized followed by protein/gene expression z-score normalization. Clustering results were evaluated visually in the resulting consensus matrix (one matrix per each k) as well as quantitatively using the cumulative distribution function of the area under the curve for each k (Figures S2A and S2B).

Integrated Pattern analysis

For the integration of proteomic and transcriptomic datasets we matched the two matrices based on gene names, filtered out genes that were not quantified at all in each one of the two datasets (n = 3407) and merged genes based on gene name (n = 3354). For the correlation analysis, we only kept IDH-WT samples for which we had survival information (n = 49 in protein, n = 49 in RNA) and calculated Pearson correlation twice for each gene: between protein expression and survival, and between gene expression and survival.

We also calculated a permutation based p value for each correlation, by scrambling the expression data within samples and repeating the procedure 1000 times. For downstream analysis we kept genes with a significant correlation to survival (permutation based adjusted p value < 0.1) for either RNA or protein. The resulting 1240 genes, represented by their correlation to survival in each layer, were then hierarchically clustered. Each cluster was then evaluated based on the correlation pattern of its genes. Fisher enrichment test (FDR = 0.05) tested whether each cluster was enriched for significant survival-protein correlations or survival-RNA correlations. Together with the directionality of the correlation, we were able to name each cluster as either "protein," "RNA" or "both"; and as associated with either long-term or short-term survival.

Validation using TCGA and Chinese Glioma datasets

We analyzed published RNA-seq data from two sources to validate the Kaplan-Meier and single cell signature analysis. TCGA RNA-seq data was downloaded from cBioportal for Cancer Genomics (https://www.cbioportal.org/). RNA-seq data of the Chinese cohort was downloaded from the Chinese Glioma Genome Atlas (CGGA, http://www.cgga.org.cn). In both datasets we filtered the samples to contain only IDH-WT GBM samples, which resulted in 141 samples in TCGA and 109 samples in the Chinese cohort.

Integration of single cell signatures

Single cell RNA signatures were downloaded from Neftel et al. (2019). We determined the dominant single cell-based subpopulation as described (Neftel et al., 2019). Briefly, we calculated the average expression of each subpopulation's signature genes. We compared it to background expression, created by randomly drawing 100 genes from the gene's expression bin, for each gene in the signature. This resulted in four scores for each sample, one for each subpopulation. The subpopulation that received the highest score was considered as most dominant in that sample. Before this calculation, the genes in each signature were filtered to adjust for bulk tumor analysis, as described (Neftel et al., 2019). To calculate whether signature gene sets tend to be positively or negatively associated with survival, we calculated Pearson correlation between survival and each gene in the dataset, ranked them in ascending order and performed 1D enrichment test as described above (FDR = 0.05). The same approach was applied for both TCGA and current data.

Survival Kaplan-Meier analysis and log rank test

Kaplan-Meier analysis and log rank test were performed using R's survival (https://cran.r-project.org/web/packages/survival/index.html) and survminer (https://cran.r-project.org/web/packages/survminer/index.html) packages. Log rank test p value < 0.05 was considered significant.

Immunohistochemistry quantification

Obtained slides were scanned using the Leica Aperio VERSA Digital Pathology Scanner platform (Aperio Technologies Inc.). Quantification of the staining was performed using the Aperio eSlide Manager software via the Aperio Cytoplasmic Algorithm (CAMK1, REL-A, and TP53BP2). Areas enriched by cancer cells were defined by a pathologist, subsequently manually annotated and analyzed by optimized cytoplasm algorithm (Leica Biosystems). Percentage of positively stained cells were used for downstream statistical analysis (t test with alpha = 0.05).