Systematic proteomic profiling and sub-classification of glioblastoma

Background

It is my pleasure to provide my progress report for the Richard Motyka Research Fellowship that I received from the Brain Tumour Foundation of Canada in July, 2017. In my project proposal "Systematic proteomic profiling and sub-classification of glioblastoma" I set out to use state of the art mass spectrometry (MS) proteomic analysis of clinically and molecularly annotated glioma cohorts to define novel molecular predictors of either prognostic or therapeutic relevance. The conception of this project was based on similar experimental approaches I used to spatiotemporally define proteomic landscapes of human fetal brain development[1]. In that study I optimized specialized sample preparation techniques to make use of plentiful archival tissue material stored in formalin-fixed paraffin-embedded (FFPE) blocks. I established that sectioning of this material enables us to by macro-dissect and isolate different cellular layers of the brain to resolve intra-tissue heterogeneity issues for molecular profiling. With the emergence of mass spectrometry approaches to reliably quantitate global proteomic profiles, I decided to translate my developed tools to sub-classification of glioblastomas (GBMs).

GBM is the most common primary brain tumour with a dismal prognosis of <12 months, despite spirited multimodal therapy. Given that GBMs have highly variable clinical outcomes with respect to patient survival and therapeutic responsiveness, we reasoned that proteomic profiling could provide an additional layer of molecular subtyping for more refined clinical patient management. While GBMs with *isocitrate dehydrogenase* mutations (IDH-*mut*) have been shown to exhibit a superior prognosis, this subtype is rare and only found in ~5-10% of GBMs. The vast majority (~90-95%) of GBMs are IDH wild-type (IDH-*wt*) and remain without actionable or prognostic biomarkers. Interestingly, while most IDH-*wt* GBMs follow an aggressive course (baseline survival (BS) <12 months), ~20% of patients survive beyond 3 years (defined as long-term survival, LTS). So far, this biological variability cannot be explained by clinical, treatment or other genomic parameters (e.g. MGMT promoter methylation). To assess the ability of MS-based approaches to define proteomic differences between various brain tumour subtypes, and GBMs in particular, we assembled several clinically-stratified cohorts and tissue culture models (GBM stem cell-like cell lines, GSCs) of GBM development. Parallel profiling of GSCs and primary patient biopsies was designed to establish molecular subtype-specific *in vitro* models for downstream predictive chemical screen experimentation. As a complimentary project with my proteomic profiling, I have been involved in using deep neural networks (DNNs) to develop artificial intelligence (AI) classification algorithms of digitized brain tumour pathology tissue sections. Our prototype DNN is able to accurately discern and highlight tumour lesions within surgically removed material from surrounding necrotic and normal tissues and, thus, provide a workflow for macrodissecting and isolating pure tumour tissues. Ultimately, combinatorial approaches of such computational tools with proteomic profiling of GBM tumours would increase the likelihood of identifying *bona fide* tumour-associated biomarkers for downstream validation.

Progress

MS-based glioma tumour and cell line proteomic profiling and subtyping

I have now translated my abbreviated MS workflow to tissues of common brain tumours. To achieve this goal we have assembled three cohorts containing IDH-*wt* and IDH-*mut* GBMs, low grade oligodendrogliomas and pilocytic astrocytomas, as well as control samples of meningiomas and medulloblastomas. From the three cohorts, our first one was assembled to include the full range of World Health Organization (WHO) grade gliomas (frozen tissues obtained from the Brain Tumour Foundation of Canada, n=15), the second one includes solely higher grade glioma tumours, including IDH-*mt* and *wt* GBMs (FFPE tissues obtained locally from UHN, n=15), and the third one contains only IDH-*wt* GBMs of varying patient survival lengths, classified as either LS or BS (obtained from our collaborators in Hungary, n=32). In addition, I have begun to profile IDH-*mt* and -*wt* GSC *in vitro* models of GBMs, either in their undifferentiated state (in presence of FGF/EGF) or upon growth factor withdrawal (n=18). By growing cells in tissue culture in these differentiation states enables

us to enrich and define proteomic biomarkers of proliferating GSCs that are responsible for therapy resistance commonly observed in GBM patients. My preliminary findings of analyzing these initial cohorts are comprehensively shown in two posters, included in this report, that I presented at the local Advancing Precision Medicine Conference (January, 2018 in Toronto, Ontario) and the Human Proteome Organization conference (Sep, 2017 in Dublin, Ireland). This initial MS-based analysis has had some notable findings. Firstly, our abbreviated MS profiling approach was highly successful in quantitatively detecting upwards of ~2,500 proteins per sample and defining proteomic signatures of gliomas of different WHO grade status, even without performing tumour lesion enrichment through macrodissection. Secondly, we find that GBM microdissection of FFPE tumours increases the likelihood of defining IDH-*mt* and -*wt* GBM tumour proteomic signatures. Thirdly, a subset of differentially abundant proteins in IDH-*wt* and -*mt* GBMs are found in similar levels in *in vitro* GSCs, specifically cultured in undifferentiated conditions. Report of these findings has been prepared into a manuscript and it is now under consideration for publication in Acta Neuropathologica.

Furthermore, I think that in order to define therapy responsiveness and GBM survival-related proteomic signatures in Cohort #3 it may be necessary to perform a more comprehensive proteomic profiling method, using sample fractionation that ensures quantification of even the lower abundance proteins beyond the 5,496 total proteins we have already detected. With our "shallow" proteomic coverage I identify 98 proteins that distinguish LS (>36 month survival) and BS (<13 months survival) GBMs. I expect that applying these "deep" proteome MS approaches will further expand our list of biomarkers of long term survival and therapy sensitivity. In this long-term surivaval-enriched cohort, we have decided to perform comprehensive OMIC analysis to more accurately define molecular events that guide tumour aggressiveness. These profiles were generated in collaboration with OICR and include global DNA methylation analysis, RNAseq and exome sequencing. This approach will allow us to interrogate RNA/DNA and proteomic relationships in different classes of GBM. Furthermore, I have been able to use MS to measure global phosphorylation levels of proteins which will enhance my ability to identify signaling cascades that are perturbed in GBMs. I believe that completion of this high impact project will result in a fantastic manuscript in the near future.

Future Directions:

I plan on completing the bioinformatic analysis of cohort #3, where I will dissect proteogenomic relationships in GBMs of different survival outcomes, within the next 6 months. In parallel I am greatly expanding our proteomic profiling toolbox by performing "deep" proteomic coverage and assessing the "phospho" proteome of our clinically stratified GBM cohorts. Once these tools have been optimized, I will process our greatly expanded GBM cohorts to generate large datasets for inquiry into more refined biomarker identification of GBM survival and therapy response. These profiles will then be further overlayed onto datasets from GSC studies using drug screens to further determine predictive abilities of proteomic profiles to respond to different chemotherapeutic pharmaceutical agents.

Publications

1. Djuric, U., Zadeh, G., Aldape, K. & Diamandis, P. Precision histology: how deep learning is poised to revitalize histomorphology for personalized cancer care. *npj Precis. Oncol.* **1,** (2017). REVIEW published

2. Faust K, Xie Q, Han D, Goyle K, Volynskaya Z, Djuric U, Diamandis P. Visualizing histopathologic deep learning classification and anomaly detection using nonlinear feature space dimensionality reduction. BMC Bioinformatics. **19**, (2018). Manuscript published.

3. Djuric U, Kao J, Batruch I, Jevtic S, Papaioannou M, Diamandis M. Proteomic profiling of diffuse gliomas defines genomically and histopathologically relevant disease subtypes. *Acta Neuropathologica*. Under REVIEW.

References:

1. Djuric, U. *et al.* Spatiotemporal Proteomic Profiling of Human Cerebral Development. *Mol. Cell. Proteomics* **16,** 1548–1562 (2017).
2. Djuric, U., Zadeh, G., Aldape, K. & Diamandis, P. Precision histology: how deep learning is poised to revitalize histomorphology for personalized cancer care. *npj Precis. Oncol.* **1,** (2017).

Presentations

1. Combinatorial Mass Spectrometry (MS) and Artificial Intelligence (AI) Subclassification of Diffuse Glioma. Ugljesa Djuric, Ihor Batruch, Laszlo Bognar, Tibor Hortobagyi, Ken Aldape, Almos Klekner, Phedias Diamandis. **HUPO World Annual Conference** (September, 2017 in Dublin, Ireland).
2. Mass Spectrometry-Based Subclassification of Diffuse Glioma. Ugljesa Djuric, Ihor Batruch, Laszlo Bognar, Tibor Hortobagyi, Ken Aldape, Almos Klekner, Phedias Diamandis. **Advancing Precision Medicine** (January, 2018 in Toronto, Canada).

COMBINATORIAL MASS SPECTROMETRY (MS) AND ARTIFICIAL INTELLIGENCE (AI) SUBCLASSIFICATION OF DIFFUSE GLIOMA

Ugljesa Djuric[1], Ihor Batruch[2], Laszlo Bognar[3], Tibor Hortobagyi[3], Ken Aldape[1], Almos Klekner[3], Phedias Diamandis[1]

1. University Health Network, Princess Margaret Hospital, MacFeeters Hamilton Center for Neuro-Oncology, Toronto, Canada. 2. Mount Sinai Hospital, Toronto, Canada. 3. Departments of Neurosurgery and Neuropathology, University of Debrecen, Debrecen, Hungary.

Summary

BACKGROUND: Diffuse gliomas are the most common primary brain tumors with variable prognosis. While genomic profiling efforts have identified modest genomic predictors of glioma patient survivals in ~8% of cases, to date large-scale proteomic profiles have not been performed. Similarly, little progress has been made to refine histologic classification and risk stratification of diffuse gliomas and would benefit from artificial intelligence (AI)-based image analysis of glioma biopsies using convolutional neural networks (CNNs). We hypothesize that this combinatorial analysis will improve understanding and prognosis of diffuse gliomas.

METHODS: We utilize MS and CNNs to establish (1) protein and (2) morphometry-based ("phenotypic") predictive diffuse glioma clinical subgroups. Towards AIM1, we apply a developed pipeline utilizing Q Exactive high resolution label-free quantifcation (LFQ) MS to characterize proteomic signatures in a cohort of diverse clinically well-annotated brain tumor specimens (n=50). Towards AIM2, we utilize a CNN-based image analysis for automated brain tumor diagnosis. We are, thus, in a position to leverage histologic analyitical outcomes with glioma proteomic profiles.

RESULTS: Our LFQ MS analytical method is well validated with ~2,500 protein quantifications per tumour sample identifying distinct proteomic-based glioma subtypes (ie. oligodendroglioma, astrocytoma and glioblastoma) based on 92 changes in protein abundance (p<0.005). GOterm based pathway analysis demonstrates that glioma-associated molecualr pathways are perturbed in correct tumour types, providing validation that MS-based proteomic measurements are identifying unbiased proteomic signatures of glioma subtypes. Similarly, training our CNN using tumour images produces a tumour identification tool enabling further glioma subtype classifications.

CONCLUSIONS: Our combinatorial approach identifies molecular- and image-based glioma subtypes and, thus, has the potential to provide precise and cost-effective clinical prognosis with faster turn-around times than classical neuropathology workflows

Rationale and Aims

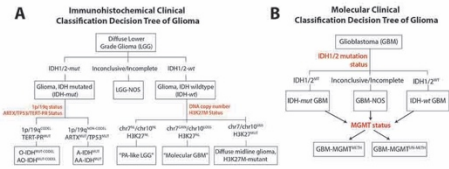
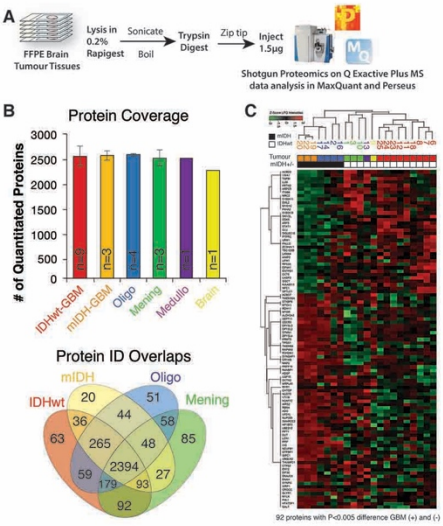


Figure 1 | Integrated histologic and molecular approaches to the classification of diffuse **(A)** lower grade gliomas (LGGs) and **(B)** glioblastoma (GBMs). **A.** Performed by immunohistochemical (IHC) assessment of the most common IDH1-R132H mutation found in ~90% of IDH-*mut* gliomas. If IHC is negative, sequencing for rare non-canonical IDH1/2 mutations is necessary. When IDH1/2 testing is indeterminate or incomplete, the "not otherwise specified" (Glioma-NOS) designation is used. In the setting of IDH-mutant (IDH1/2-*mut*) LGGs, 1p/19q codeletion ($1p/19q^{CODEL}$) status is also assessed to characterize "molecular" WHO grade II & III (anaplastic) oligodendrogliomas ($O\text{-}IDH^{MUT\text{-}CODEL}$ & $AO\text{-}IDH^{MUT\text{-}CODEL}$), $IDH1/2^{MUT}$ LGG without $1p/19q^{CODEL}$ are categorized as astrocytomas WHO grade II & III ($A/AA\text{-}IDH1/2^{MUT}$). The right portion of the figure addresses IDH-wildtype (IDH1/2-*wt*) LGGs. These often represent aggressive tumors and carry gain of chromosome 7 and loss of chromosome 10 ($chr7^{GAIN}$/chr 10^{LOSS}) and are considered "molecular GBMs". In the absence of chromosome 7 and 10 alterations (chr7NL/chr 10NL), a subset of IDH-*wt* LGGs have been proposed as "PA-like LGG", based on molecular characteristics similar to pilocytic astrocytoma (e.g. BRAR/NF1 alterations). Finally, mutations in histone H3 family members ($H3K27M^{MUT}$), can also occur. **B.** Molecular workup of a diffuse glioma with a GBM morphology is often abbreviated and most centers focus only on prognostically significant IDH mutation and MGMT promoter methylation status. **C.** Project objective is to develop robust protein-based biomarkers and deep neural network image-based classifications of glioma with improved survival and risk stratification schemes.

Results



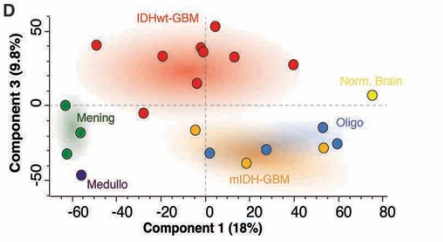


Figure 2 | Figure 2 | Shotgun LC-MS/MS proteomic analysis of a pilot set of CNS Neoplasms. **A.** Shotgun LC-MS/MS of FFPE tumour samples analytical workflow. **B.** Approximately ~2300 proteins are routinely quantified per sample. Venn diagram highlights proportion of similar and unique proteins identified within the different tumor samples. **C.** Hierarchical clustering based on 96 proteins that are significantly different between mIDH-GBMs and IDHwt-GBMs (p<0.005). This preliminary analysis shows a IDH-mutated specific molecular signature that is also shared among Oligodendrogliomas. **D.** Principal component analysis (PCA) demonstrates spatial segregation between the different tumor types, IDH mutated and IDH wild-type gliomas. **E.** PCA loadings of the protein intensity values distinguishing the tumour types and the control tissue. Proteins with the highest resolving power are highlighted.analysis of different brain tumor types. **Abbreviations:** IDH-wildtype glioblastoma (IDHwt-GBM, n=9), IDH-mutated GBM (mIDH-GBM, n=3), IDH-mutated 1p/19q Oligodendrogliomas (Oligo, n=3), Meningioma (Mening, n=3), Medulloblastoma (Medullo, n=1)] and control brain tissue.

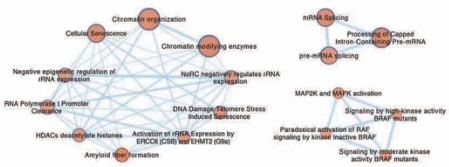


Figure 3 | Pathway analysis in Cytoscape of proteins with abundance changes between mIDH-GBMs and IDHwt-GBMs. The more aggressive wt-GBMs exhibit an increase in epigenetic proteins involved in stem cell replication and telomere maintenance as well as deacetylase activity. mRNA processing proteins are increased in mIDH-GBMs indicating a higher rate of RNA modifications in this less aggressive GBM subtype. MAP2K and BRAF signaling, which mediates cellular responsees to cell growth signals also show protein abundance changes and demonstrates that our shotgun proteomic profiles of FFPE tissues are capturing proteins involved in cell growth.

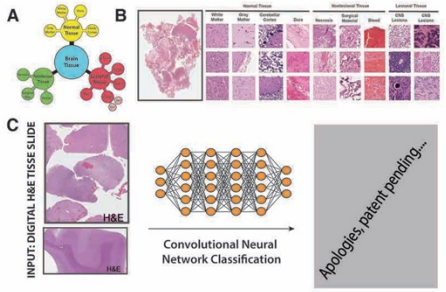


Figure 4 | Inter- and Intra-slide tissue class variability in surgical neuropathology. **A.** Examples of trained convolutional neural network training class es of tissue. **B.** WSI of a glioblastoma containing a heterogenous mixture of tumor, necrosis, normal brain tissue, blood and surgical material. The tumor comprises less than 30% of the surface area on the slide. This diversity, if not accounted for, can result in erroneously classification errors (e.g. mistaking dura for schwannoma or surgical material for calcification). **C.** Inclusion of these classes allow accurate detailed annotation of slide constituents and more accurate delineation of true lesion for future isolation or classification tasks. Example of digital slide classification of our current CNN tissue classifier.

Conclusions and Future

Discovery of genomic changes (i.e. IDH1/2 mutations) in a small subset of diffuse gliomas has revolutionized clinical practice of modern neuro-oncology. However, additional discovery of protein biomarkers in larger, molecularly-undefined, subgroups of GBMs (e.g. IDH-wt) would provide further prognostic significance for risk stratification. Our optimized FFPE-based LC-MS/MS workflow aims to translate this promising technology to clinically stratified cohorts of diffuse gliomas. Shotgun LFQ LC-MS/MS of FFPE tissues achieves sufficient proteome resolving power to discriminate between aggressive (glioma) and benign (meningioma) brain tumour types and, importantly, between IDHwt- and IDHmt-GBMs. Candidate proteins are currently being confirmed in larger clinical cohorts. Convolutional neural networks being developed in our lab successfully stratifiy digital images of H&E tissue slides based on gross tissue morphology.

Future Directions

Proteomics:

- Profiling of a larger clinical cohort and glioma cell lines is being conducted.
- Fractionated peptide LC-MS/MS will be performed to increase coverage and determine whether additional low abundance proteins change in glioma subtypes.
- Post-translational modifications will be assessed in selected cases by global phosphoproteome analysis

CNN:

- Further training with sufficiently large image dataset from online image libraries (TCGA) and the UHN slide digitization service
- Train CNN using well-annotated subgroups of glioma with patient metadata.

Institutions and Funding

This work is supported by the Brain Tumour Foundation of Canada Richard Motyka research fellowship to Ugljesa Djuric.

MOUNT SINAI HOSPITAL Bright Minds. Big Hearts. The Best Medicine. UNIVERSITY OF TORONTO UHN

MASS SPECTROMETRY-BASED PROTEOMIC SUBCLASSIFICATION OF DIFFUSE GLIOMAS

Ugljesa Djuric[1], Ihor Batruch[2], Laszlo Bognar[3], Tibor Hortobagyi[3], Ken Aldape[1], Almos Klekner[3], Phedias Diamandis[1]

1. University Health Network, Princess Margaret Hospital, MacFeeters Hamilton Center for Neuro-Oncology, Toronto, Canada. 2. Mount Sinai Hospital, Toronto, Canada. 3. Departments of Neurosurgery and Neuropathology, University of Debrecen, Debrecen, Hungary.

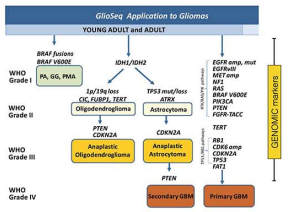
Abstract

BACKGROUND: Diffuse gliomas are the most common primary brain tumors with variable prognosis. While genomic profiling efforts have identified modest genomic predictors of glioma patient survivals in ~8% of cases, to date large-scale proteomic profiles have not been performed. Similarly, little progress has been made to refine histologic classification and risk stratification of diffuse gliomas and would benefit from artificial intelligence (AI)-based image analysis of glioma biopsies using convolutional neural networks (CNNs). We hypothesize that this combinatorial analysis will improve understanding and prognosis of diffuse gliomas.

METHODS: We utilize MS to define glioma subtype through (1) global proteomics and (2) predictive molecular signatures of tumour behaviour. Towards AIM1, we apply a developed pipeline utilizing Q Exactive high resolution label-free quantifcation (LFQ) MS to characterize proteomic signatures in a cohort of diverse clinically well-annotated brain tumor specimens (n=50). Towards AIM2, we utilize a bioinformatic analysis for defining proteins differentially expressed in clinically distinct (ie. patient survival) but genomically uniform tumour subtypes, currently treated as non-unique entities. We are, thus, in a position to leverage histologic analyitical outcomes with glioma proteomic profiles.

RESULTS: Our LFQ MS analytical method is well validated with ~2,500 protein quantifications per tumour sample identifying distinct proteomic-based glioma subtypes (ie. oligodendroglioma, astrocytoma and glioblastoma) based on changes in protein abundance (FDR or p<0.05). GOterm based pathway analysis demonstrates that glioma-associated molecualr pathways are perturbed in correct tumour types, providing validation that MS-based proteomic measurements are identifying unbiased proteomic signatures of glioma subtypes. However, tumour and/or patient heterogeneity seem to confound our abilities to accurately predict patient survival subgroups. Our ongoing training and validation of CNNs for brain tumour subtype classifications are presented elsewhere (See poster of Kevin Faust and Quin Xie).

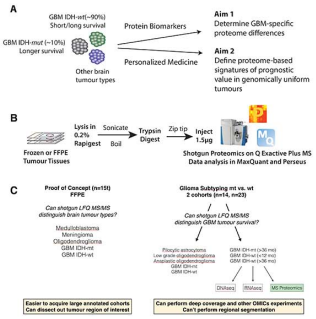
Rationale



Hypothesis:
Global proteome analysis can reveal additional tumour subtypes of clinical relevance

Figure 1 | Integrated histologic and genetic definitions of diffuse lower grade gliomas and glioblastomas. **Problem:** Majority of glioma patients remain without a molecular- based prognosis.

Experimental Design



Results

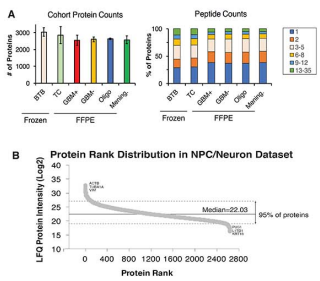


Figure 2 | **A.** Average number of quantified proteins and peptide counts/protein in different brain tumour types used in the study. **B.** Representative dynamic range of LFQ values spans 8.5 orders of magnitude for 95% of the dataset.

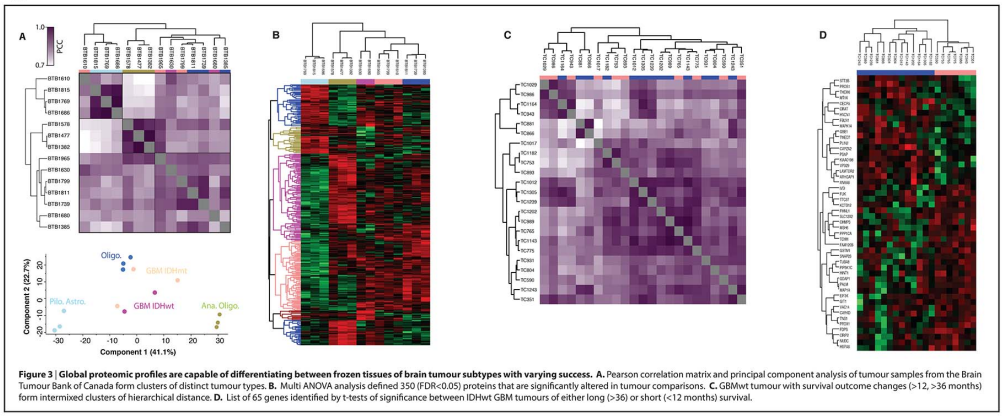


Figure 3 | **Global proteomic profiles are capable of differentiating between frozen tissues of brain tumour subtypes with varying success. A.** Pearson correlation matrix and principal component analysis of tumour samples from the Brain Tumour Bank of Canada form clusters of distinct tumour types. **B.** Multi ANOVA analysis defined 350 (FDR<0.05) proteins that are significantly altered in tumour comparisons. **C.** GBMwt tumour with survival outcome changes (>12, >36 months) form intermixed clusters of hierarchical distance. **D.** List of 65 genes identified by t-tests of significance between IDHwt GBM tumours of either long (>36) or short (<12 months) survival.

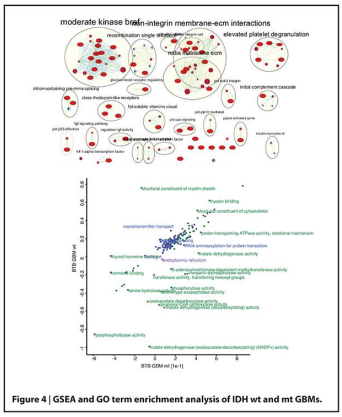


Figure 4 | **GSEA and GO term enrichment analysis of IDH wt and mt GBMs.**

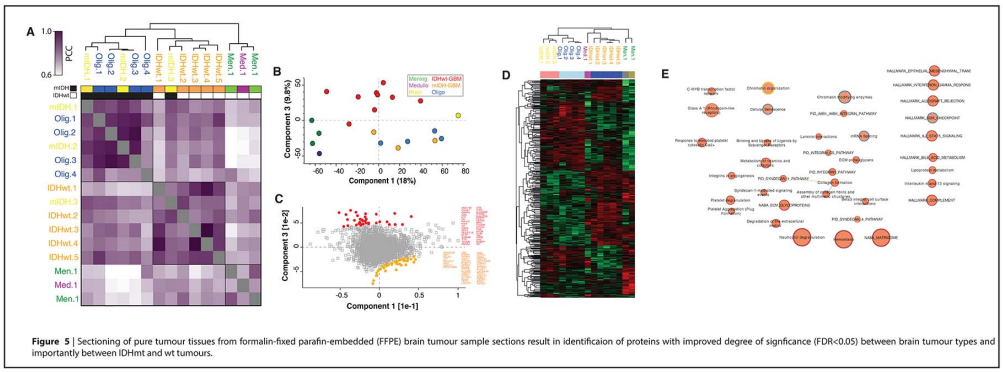
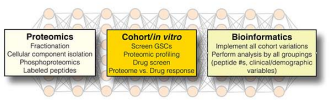


Figure 5 | Sectioning of pure tumour tissues from formalin-fixed parafin-embedded (FFPE) brain tumour sample sections result in identificaion of proteins with improved degree of signficance (FDR<0.05) between brain tumour types and importantly between IDHmt and wt tumours.

Conclusions

Shotgun LFQ mass spectromtry can define distinct protein modules involved in tumour behaviour (ie. Class iii vs iV) but clinical outcomes such as patient survival are affected by high degrees of intragroup sample variabilities. Performing similar analysis on FFPE tumours where samples can be enriched for pure tumour tissue by macrodissection, more reliable and larger lists of candidate proteins can be defined to classify subgroups of glioma tumours. In addiition, biological pathways and functions involved in tumour behaviour can be further explored by performing deep proteome coverage and phosphoproteomic analysis for accurate pinpointing of pathways with optimal drug targets for novel personalized chemotherapeutic regimens.

Future Directions



Institutions and Funding
This work is supported by the Brain Tumour Foundation of Canada Richard Motyka research fellowship to Ugljesa Djuric.

MOUNT SINAI HOSPITAL Bright Minds. Big Hearts. The Best Medicine. TORONTO UHN